



SPIE

# IS&T / SPIE Electronic Imaging

## SCIENCE AND TECHNOLOGY

Technical Summaries

[electronicimaging.org](http://electronicimaging.org)

Conferences/Courses: 23-27 January 2011

Hyatt Regency Hotel  
San Francisco Airport, California, USA



### Contents

7863: Stereoscopic Displays and Applications XXII .....	2	7874: Document Recognition and Retrieval XVIII .....	97
7864A: 3D Imaging Metrology .....	19	7875: Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII .....	105
7864B: 3D Image Processing (3DIP) and Applications II .....	24	7876: Digital Photography VII .....	111
7864C: The Engineering Reality of Virtual Reality 2011 .....	27	7877: Image Processing: Machine Vision Applications IV .....	117
7865: Human Vision and Electronic Imaging XVI .....	29	7878: Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques .....	123
7866: Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications .....	36	7879: Imaging and Printing in a Web 2.0 World II .....	131
7867: Image Quality and System Performance VIII .....	48	7880: Media Watermarking, Security, and Forensics XIII .....	136
7868: Visualization and Data Analysis 2011 .....	56	7881A: Multimedia on Mobile Devices 2011 .....	141
7869: Computer Vision and Image Analysis of Art II .....	62	7881B: Multimedia Content Access: Algorithms and Systems V .....	147
7870: Image Processing: Algorithms and Systems IX .....	68	7882: Visual Information Processing and Communication II .....	150
7871: Real-Time Image and Video Processing 2011 .....	77		
7872: Parallel Processing for Imaging Applications .....	83		
7873: Computational Imaging IX .....	90		

# Conference 7863: Stereoscopic Displays and Applications XXII

Monday-Thursday 24-27 January 2011

Part of Proceedings of SPIE Vol. 7863 Stereoscopic Displays and Applications XXII

## 7863-01, Session 1

### **Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis**

F. Devernay, S. Duchêne, A. Ramos-Peon, INRIA Rhône-Alpes (France)

The 3D shape perceived from viewing a stereoscopic movie depends on the viewing conditions, most notably on the screen size and distance, and depth and size distortions appear because of the differences between the shooting and viewing geometries. When the shooting geometry is constrained, or when the same stereoscopic movie must be displayed with different viewing geometries (e.g. in a movie theater and on a 3DTV), these depth distortions may be reduced by novel view synthesis techniques. They usually involve three steps: computing the stereo disparity, computing a disparity-dependent 2D mapping from the original stereo pair to the synthesized views, and finally composing the synthesized views.

In this paper, we focus on the second and third steps. We examine three different mapping functions: baseline modification, which preserves image content but distorts depth and create divergence, viewpoint modification, which preserves depth but modifies heavily image content, and the newly-introduced hybrid disparity remapping, which preserves both depth and image content.

For the final composition step, we propose an asymmetric view synthesis method, where artifacts are detected and blurred in the synthesized view, while the other view is the same as the original, thus preserving the overall perceived quality of the stereoscopic movie.

## 7863-02, Session 1

### **Visual fatigue monitoring system based on eye-movement and eye-blink detection**

D. Kim, S. Choi, J. Choi, K. Sohn, Yonsei Univ. (Korea, Republic of)

In this paper, we proposed a visual fatigue monitoring system based on eye-movement and eye-blink detection. It analyzes the eye-movement and number of blinks based on the assumption that saccade movement of the eye decreases and the number of eye blink increases when visual fatigue of viewer is accumulated. The proposed system has an infrared single camera and an infrared light source. Then, the pupil of the eye can be detected by applying binary threshold to Purkinje image. The threshold is automatically selected by two constraints which are the angle of eclipse fitting and the size of the pupil. Finally, pattern matching is performed to select the accurate position of the pupil among the candidates and the system estimates the total amount of eye movement and the number of eye blinks. The results were obtained while watching stereoscopic videos after personal calibration procedure. According to subjective evaluation and descriptive self-report, the results show that saccade movement of the eye decreases as the visual fatigue of the viewer is accumulated. However, the number of eye blinks shows large variance along the time axis which implies it is not proper for visual fatigue monitoring system.

## 7863-03, Session 1

### **Factors impacting quality of experience in stereoscopic images**

L. Xing, J. You, Norwegian Univ. of Science and Technology (Norway); T. Ebrahimi, Ecole Polytechnique Fédérale de Lausanne (Switzerland); A. Perkiš, Norwegian Univ. of Science and Technology (Norway)

The stereoscopic 3D industry has fallen short of achieving acceptable Quality of Experience (QoE) because of various technical limitations, such as excessive disparity, accommodation-convergence mismatch. This study investigates the effect on QoE of stereoscopic scene content, camera baseline, screen size and viewing location in a holistic approach. We first designed 240 typical test configurations, in which the disparity constructed from the shooting condition (scene content and camera baseline) is located in different ranges of maximal disparity supported by viewing environment (screen size and viewing location) in order to cover different conditions. Second, extensive subjective tests were conducted using a single stimulus methodology, in which 15 samples for each viewing location were obtained. Finally, a statistical analysis was performed and the results revealed that content, baseline, as well as the interactions between size, content and baseline, have a significant impact on QoE in stereoscopic images, while other combinations, especially viewing location involved, have no significant impact. The resulted Mean Opinion Scores (MOS) and statistical results can be further used to compare and design new stereoscopic quality metrics.

## 7863-04, Session 1

### **Visual discomfort induced by fast salient object motion in stereoscopic video**

S. Lee, Y. J. Jung, H. Sohn, Y. M. Ro, H. W. Park, Korea Advanced Institute of Science and Technology (Korea, Republic of)

In stereoscopic displays, instinct conflicts between accommodation and vergence is well known as one of major factors that may incur visual discomfort. In order to avoid excessive accommodation-vergence conflicts, binocular disparity of 1 degree has been suggested as a guideline for comfortable viewing zone. However, this value is not an absolute threshold since it overlooks effects of space- and time-varying disparities. Also, even within zone of comfortable viewing, there are many factors that may be able to induce visual discomfort, such as individual differences in vision ability and characteristics of stereoscopic contents. To generate a safety of stereoscopic video contents, it is very essential to investigate factors and conditions that induce visual discomfort. In this paper, we focus our scope on the investigation of local motion characteristics that may lead to visual discomfort within the zone of comfortable viewing. Although several researchers have examined the visual discomfort induced by the effect of both disparity and motion, it requires more in-depth studies to understand the effects of motions with space- and time-varying disparities. The contribution of this paper is to investigate the relation between visual discomfort and local motions that vary with different direction, velocity, rotational velocity, and movement artifacts within zone of comfortable viewing.

## 7863-05, Session 1

### 3D video disparity adjustment for preference and prevention of discomfort

H. Pan, C. Yuan, S. J. Daly, Sharp Labs. of America, Inc. (United States)

With the huge success of 3D movies in theaters, 3D TVs and other 3D products are penetrating into the home with ever increasing speed. One of the key issues associated with 3D TVs is the tradeoff between comfort and 3D visual impact. Big disparity is often preferred for strong visual impact but often lead to viewer discomfort based on display size and viewing distances. The goal of the proposed algorithm is to provide viewers a tool for them to adjust disparity according to the environment, contents and their preference in order to have more comfortable and higher quality 3D experiences.

More specifically, given a planar stereoscopic display, the algorithm takes in a stereoscopic image pair that causes viewing discomfort/fatigue, and outputs a modified stereoscopic pair that causes less or no viewing discomfort/fatigue. The algorithm fulfills the functions of disparity estimation, occlusion detection, disparity adjustment and view synthesis. A novel pixel weighting mechanism in regularized-block-matching based disparity estimation helps improve the robustness, accuracy and speed of matching. Occlusion detection uses multiple cues in addition to matching errors to improve the accuracy. An accommodation/vergence mismatch visual model is used in disparity adjustment to predict discomfort/fatigue from the disparity information, the viewing conditions and display characteristics. The hole filling in view synthesis is in the disparity map of the new view instead of the new view itself to reduce the blurriness. The preliminary results are promising.

## 7863-06, Session 2

### Can the depth perception of stereoscopic images be influenced by 3D sound?

A. Turner, N. S. Holliman, Durham Univ. (United Kingdom)

The creation of binocular images for stereoscopic display has benefited from significant research and commercial development in recent years. However, perhaps surprisingly, the effect of adding auditory depth information to stereoscopic images has rarely been studied. Having found few similar studies in the literature we address two preliminary questions. First what is the smallest difference in auditory depth that can be reliably detected using sound alone? Second is it possible that the addition of auditory depth information can enhance the visual perception of depth in a stereoscopic image?

## 7863-07, Session 2

### Evaluating motion and binocular parallax as depth cues for autostereoscopic displays

M. Braun, Fachhochschule für Technik und Wirtschaft Berlin (Germany); U. Leiner, D. Ruschin, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

The perception of space in the real world is based on multifaceted depth cues, most of them monocular, some binocular.

Developing 3D-displays raises the question, which of these depth cues are predominant and should be simulated by computational means in such a panel. Beyond the cues based on image content, such as shadows or patterns, Stereopsis and depth from motion parallax are the most significant mechanisms supporting observers with depth information. We set up a carefully designed test situation, widely excluding undesired other distance hints. Thereafter we conducted a user test to find out, which of these two depth cues is more relevant and whether a combination of both would increase accuracy in a depth estimation task. The trials were conducted utilizing our autostereoscopic "Free2C"-displays, which are capable to detect

the user eye position and steer the image lobes dynamically into that direction. At the same time, eye position was used to update the virtual camera's location and thereby offering motion parallax to the observer. As far as we know, this was the first time that such a test has been conducted using an autostereoscopic display without any assistive technologies. Our results showed, in accordance with prior experiments, that both cues are effective, however Stereopsis is by order of magnitude more relevant. Combining both cues improved the precision of distance estimation by another 30-40%.

## 7863-09, Session 4

### A multi-resolution multi-size windows disparity estimation approach

J. Martinez Bauza, Qualcomm Inc. (United States); M. P. Shiralkar, Clemson Univ. (United States)

This paper describes an algorithm for estimating the disparity between 2 images of a stereo pair. The disparity is related to the depth of the objects in the scene. Being able to obtain the depth of the objects in the scene is useful in many applications such as virtual reality, 3D user interfaces, background-foreground segmentation, or depth-image-based synthesis. This last application has motivated the proposed algorithm as part of a system that estimates disparities from a stereo pair and synthesizes new views. Synthesizing virtual views enables the post-processing of 3D content to adapt to user preferences or viewing conditions, as well as enabling the interface with multi-view auto-stereoscopic displays.

The proposed algorithm has been designed to fulfill the following constraints: (a) low memory requirements, (b) local and parallelizable processing, and (c) adaptability to a sudden reduction in processing resources. Our solution uses a multi-resolution multi-size-windows approach, implemented as a line-independent process, well-suited for GPU implementation. The multi-resolution approach provides adaptability to sudden reduction in processing capabilities, besides computational advantages; the windows-based image processing algorithm guarantees low-memory requirements and local processing.

## 7863-10, Session 4

### Warping error analysis and reduction for depth image based rendering in 3DTV

L. Do, S. Zinger, Technische Univ. Eindhoven (Netherlands); P. H. N. de With, CycloMedia Technology B.V. (Netherlands)

Interactive free-viewpoint selection applied to a 3D multiview video signal is an attractive feature of the rapidly developing 3DTV media. In recent years, significant research has been done on free-viewpoint rendering algorithms which mostly have similar building blocks. In [1], we have analyzed the principal building blocks of most recent rendering algorithms and their contribution to the overall rendering quality. We have discovered that the first step, Warping determines the basic quality level of the complete rendering chain. In this paper, we have analyzed the warping step in more detail since it leads to ways for improvement. We have observed that warping errors consist of mainly three types of errors which are rounding errors when performing pixel-based warping, quantization errors of depth maps and rounding errors at the virtual image. For each error factor, we have proposed a technique that can reduce the errors and thus increase the warping quality. The new techniques are evaluated with two series of experiments using real-life and synthetic data. From the first exploration experiments, we observe that the proposed techniques reduce the warping errors and help to increase the overall rendering quality.

## 7863-11, Session 4

### Novel view synthesis for dynamic scene using moving multi-camera array

T. Yokoi, Nagoya Univ. (Japan); N. Fukushima, Nagoya Institute of Technology (Japan); T. Yendo, M. Panahpour Tehrani, Nagoya Univ. (Japan); T. Fujii, Tokyo Institute of Technology (Japan); M. Tanimoto, Nagoya Univ. (Japan)

We are developing technologies for Free-viewpoint TV (FTV) in which the viewer can freely change the viewpoint. FTV allows us to change the viewpoint freely in a 3D world. Generally, given multi-view images, it is necessary to represent the 3D space to generate an arbitrary viewpoint image.

Ray space method that belongs to image based rendering, can be used for representation of 3-D space and free-viewpoint image generation. The ray space method does not need any 3-D information of the scene for free-viewpoint image generation. A dense ray-space is required to be able to freely change the viewpoint. In order to make a dense ray-space we need to synthesize views between actual images in the ray-space. We focus on free-viewpoint image using multi-view images, using depth maps.

The free-viewpoint image can be generated by using images captured by a static multi-camera system. However, it is hard to render an object that moves widely in the scene. If an object moves widely, the number of the camera should be increased. Thus, the setup and computational costs of such a system are increase. Alternatively, we can reduce the number of camera to reduce the costs by increasing the camera interval. However, in this case, only sparse information can be acquired, so that the free-viewpoint image quality significantly decreases, and cannot be used for photo-realistic applications.

In this paper, we address this problem by proposing moving camera array. Furthermore, for moving camera array, we propose a novel method for synthesizing free-viewpoint images using both spatial and temporal information.

By using moving camera array, not only we can follow an object of interest, but also we can obtain dense ray information of static area of the scene. To make the best use of the dense acquisition of the ray in the static areas, the free-viewpoint image can be generated not only by the images in spatial domain, but also the images in temporal domain. Therefore, in the proposed synthesis method, we use four reference images, which are two views in spatial, another two views in temporal direction. Views in temporal direction belong to different times, before and after, and different views.

In this process we are trying to use closer two views to the location of virtual viewpoint from different frames.

Experiments using a sequence captured by simulated moving multi-camera systems demonstrate objective and subjective improvement of view synthesis quality in comparison with conventional view synthesis scheme.

## 7863-12, Session 4

### Depth-based representations: which coding format for 3D video broadcast applications?

P. Kerbirou, G. Boisson, K. Sidibe, Technicolor R&D France (France); Q. Huynh-Thu, Technicolor S.A. (France)

3D Video (3DV) delivery standardization is currently ongoing in MPEG. Now time is to choose the representation format. Our concern is the final quality for end-users, i.e. synthesized views visual quality. We focus on two major rival depth-based formats, namely Multiview Video plus Depth (MVD) and Layered Depth Video (LDV). MVD can be considered as the basic depth-based format, generated by disparity estimation from multi-view sequences. LDV is more sophisticated, with the compaction of multiview data into colour- and depth-occlusions layers. We compare final views quality using MVD2 and LDV (both containing two color channels plus two depth components) coded with MVC at various compression ratios. Depending on the format, the appropriate synthesis is performed to generate final stereoscopic

pairs. Comparisons are provided in terms of SSIM and PSNR with respect to original views and to synthesized references (obtained without compression). Eventually, LDV outperforms significantly MVD when using state-of-the-art reference synthesis algorithms. Detecting occlusions before encoding reveals beneficial in comparison with handling redundant informations at decoder side. Besides, we observe that depth quantization does not induce much loss on the final view quality until a significant degradation level. Improvements in disparity estimation and view synthesis are therefore still expected for the remaining standardization steps.

## 7863-13, Session 5

### Multiview image compression based on a new basis representation

T. Yamada, T. Fujii, Tokyo Institute of Technology (Japan)

In order to compress or interpolate the multiview image efficiently, we propose a new basis representation by using directional sampling. While the multiview image data has the static correlation related to the camera position, directional sampling can eliminate the correlation efficiently. We apply directional three dimensional discrete cosine transform (directional 3D-DCT) and directional three dimensional discrete wavelet transform (directional 3D-DWT) to three dimensional data that has vertical, horizontal and view direction. In the experimental results, the proposed method showed better quality than previous method, not only in the objective evaluation, but also in the subjective evaluation.

## 7863-14, Session 5

### Design of tuneable anti-aliasing filters for multiview displays

A. R. Boev, R. Bregovic, A. P. Gotchev, Tampere Univ. of Technology (Finland)

In this paper, we address the problem of anti-aliasing filtering of images to be displayed on auto-stereoscopic displays. Multiview displays employ an optical layer, which distributes the light of an underlying TFT-LCD panel in different directions. Certain properties of the layer create specific artifacts, such as ghost images, moiré patterns and masking. These artefacts are especially visible and annoying when 2D imagery, such as graphics and text, is to be displayed on auto-stereoscopic displays. We model the layer as an image processing channel and identify display parameters that are important for the designs of an artifact mitigation filter. The model explains common 3D display artifacts, such as moiré patterns caused by aliasing, and ghost images caused by crosstalk.

It turns out that knowledge of the interleaving pattern and the angular visibility profile of each sub-pixel is not sufficient for predicting the visual output. Due to the optical layer, the visible parts of sub-pixels have a non-rectangular shape and the gaps between them are directionally oriented. The slant of the layer creates a pattern, which interferes with textures with similar orientation in the visualized image and creates masking artifacts similar to the 'imaging' artefacts caused by upsampling in the absence of a post-filter.

Usually, imaging is tackled by an anti-imaging post-filter. As the imaging is created by the physical structure of the display, it is impossible to impose a post-filter. However, the effect can be partially mitigated by a pre-filter. In order to determine the properties of the required 2D filter, and consequently to have the best possible representation of images on the display (minimizing aliasing, imaging and ghosting), it is necessary to determine the performance of the display in the frequency domain; that is, we have to know which frequency components in the image we can keep (ones that will be properly represented on the screen), and which ones we have to attenuate (remove) as potential causes of distortions. Applying such 2D filter will remove moiré artifacts and make masking artifacts less visible. We show that the passband of this filter is optimal, and no further band limitation is necessary.

However, prolonged observation of a multiview display might make the masking artifacts less noticeable, as the human visual system gradually adapts to the constant pattern imposed by the optical layer. Some observers find the 'optimal' 2D filter to create over-smoothed image, and prefer sharper image at the expense of some visible masking artifacts. We propose a tunable filter, with a passband that can be gradually morphed between 'optimal' and 'all-pass' shape. This can be presented to the user as a single "sharpness" control.

In this paper, we describe: 1) measurement methodology for deriving the frequency performance of an arbitrary multiview display; 2) methodology for designing an artifact mitigating filter for that display; 3) algorithm for recalculating the filter that allows gradual change (morphing) of the passband controlled by a single parameter.

## 7863-15, Session 5

### Multiview image compression based on LDV scheme

B. Battin, C. Niquin, P. Vautrot, Univ. de Reims Champagne-Ardenne (France); D. G. Debons, 3DTV Solutions (France); L. Lucas, Univ. de Reims Champagne-Ardenne (France)

In recent years, we have seen several different approaches emerge to deal with multiview compression. First, we can find the H264/MVC extension which generates quite heavy bitstreams when used on n-views autostereoscopic medias and does not allow inter-view reconstruction. Another solution relies in the MVD (MultiView+Depth) scheme which keeps p views ( $n > p > 1$ ) and their associated depth-maps. This method is not suitable for multiview compression since it does not exploit the redundancy between the p views, moreover occlusion areas cannot be accurately filled. In this paper, we present our method based on the LDV (Layered Depth Video) approach which keeps one reference view with its associated depth-map and the  $n - 1$  residual ones required to fill occluded areas. We first perform a global per-pixel matching step (providing a good consistency between each view) in order to generate one unified-color RGB texture (where a unique color is devoted to all pixels corresponding to the same 3D-point, thus avoiding illumination artifacts) and a signed integer disparity texture.

Next, we extract the non-redundant information into two textures (a unified-color one and a disparity one) containing the reference and the  $n-1$  residual views. The RGB texture is compressed with a conventional DCT or DWT-based algorithm and the disparity texture with a lossless dictionary algorithm. Then, we will discuss about the signal deformations generated by our approach.

## 7863-16, Session 5

### Upsampling range camera depth maps using high-resolution vision camera and pixel-level confidence classification

C. Tian, V. A. Vaishampayan, AT&T Labs. Research (United States); Y. Zhang, Texas A&M Univ. (United States)

We consider the problem of upsampling a low-resolution depth map generated by a range camera, by using information from one or more additional high-resolution vision cameras. The goal is to provide an accurate high resolution depth map from the viewpoint of one of the vision cameras. We propose an algorithm that first warps and converts the low resolution depth map into a depth/disparity map in the coordinate frame of and at the same resolution as one vision camera, then classifies the pixels into regions according to whether the range camera depth map is reliable, and finally performs a graph cut optimization on the unreliable regions. To reduce the misalignment caused by using only a single homographic warping, we employ a depth-dependent homographic mapping which has several candidates, which results in more accurate alignment between the camera views. Experimental results show that the proposed method is able to upsample the depth map by a factor of 10-by-10 with very accurate depth details. The improvements are visually perceptible on a 3D auto-stereoscopic display.

## 7863-17, Session 6

### Attack of the s. mutans! A stereoscopic-3D multi-player direct-manipulation behavior-modification serious game for improving oral health in pre-teens

A. Hollander, Firsthand Technology Inc. (United States)

Attack of the S. Mutans! Is a multi-player game designed to harness the immersion and appeal possible with stereoscopic 3D to combat the tooth decay epidemic. Tooth decay is one of the leading causes of school absences. Early problems with teeth can have progressive, systemic health repercussions. The traditional methods of educating the public about how to care for their teeth are failing. In 2008 the authors received a grant from the National Institutes of Health to build a 2000 sqft museum exhibit that included a suite of serious games involving the behaviors and bacteria that cause cavities. One of these is an adventure where five simultaneous players use modified Wii controllers to battle biofilms and bacteria while immersed in environments generated within a 12-foot stereoscopic WUXGA display. The authors describe the system and interface used in this prototype application and some of the ways they attempted to use the power of immersion and the appeal of S3D revolution to change behaviors and lives.

## 7863-18, Session 6

### Stereoscopic multi-perspective capture and display in the performing art

V. Kuchelmeister, The Univ. of New South Wales (Australia)

Multi-perspective image capture constitutes a novel method of documenting the entirety of an event, by recording from multiple points of view. In combination with stereoscopic imaging, it captures the two modalities of three-dimensional representation, perspective and depth. These modalities are then applied for display in a hexagonal stereoscopic multi-screen platform. In this paper the author describes, in form of a case study, the implementation of two installation projects within the performing art context.

## 7863-19, Session 6

### Machine vision and vitrectomy: three-dimensional high definition video for surgical visualization in vitreoretinal surgery

C. D. Riemann, Cincinnati Eye Institute and MedNet Technologies, Inc. (United States)

Machine vision provided by 3 dimensional high definition video systems was successfully used for surgical visualization in 8 vitreoretinal surgeries. Clinical results were excellent and surgeon intraoperative comfort was superb. With continued advancements, this new technology may evolve to profoundly change surgical visualization for vitreoretinal surgery and ophthalmic surgery as a whole.

## 7863-20, Session 7

### High image quality 3D displays with polarizer glasses based on active retarder technology

S. Jung, Y. Lee, H. Park, J. Park, D. Lee, W. Jeong, J. Kim, I. Chung, LG Display (Korea, Republic of)

In this paper, we described the basic concepts of active retarder 3D display which can give full resolution and high luminance in 3D images with simple polarizer glasses. By optimizing the 3D performance of

AR3D with respect to various design parameters, the major problems of AR3D such as the high 3D crosstalk and the 3D image difference were clearly solved. From the experimental results, B/W 3D crosstalk was obtained as 1.0% in both left and right eyes which are similar to that of the conventional PR3D. The luminance is obtained to 75 nit through polarizer glasses under the condition of the surface luminance to be 300 nit without polarizer glasses, leading to 25% light efficiency in 3D modes. The luminance difference was also reduced to be around 0.5%, which can be recognized as same level of left and right luminance. From the optimized results, it was clarified that the AR3D display technology can give superior 3D image quality among the various glasses type 3D technologies with high resolution and luminance.

## 7863-21, Session 7

### High-brightness film projection system for stereoscopic movies

L. Lipton, Oculus3D (United States)

A plano-stereoscopic projection system consisting of a film format and a unique lens for use with the 35mm motion picture infrastructure has been designed and implemented. The system overcomes prior art limitations with regard to ambiguities in threading and assembly of prints which has led to pseudostereoscopic images. The new system is at least three times brighter than most prior single projector film efforts. It is also far brighter than the majority of digital stereoscopic projection products. The images are extremely steady and sharp and capable of being projected onto the largest theatrical screens. The system, by Oculus3D™, costs less than a fifth of that of digital products and fills a need because of the great number of stereoscopic feature films being released given the shortage of available screens.

## 7863-22, Session 7

### New generation of universal active glasses

B. Mendiburu, Volfoni (United States); B. Caillaud, G. Jovene, T. Henkinet, Volfoni (France)

We present a new generation of active glasses that uses new liquid crystal shutters and an electronic capable of recognizing and synchronizing with most part of existing 3D displays. In this paper we focus on the electro-optical characterization of our liquid crystal shutters technology, named ECB, and the performance of our universal electronic demonstrating the advantage of these new glasses.

Our liquid crystal, confined in a very thin cell gap (1,5µm), has homogeneous performance for most important parameters in 3D. Indeed, ECB material has fast switching times, especially a fast relaxation time preventing color banding. Secondly, light transmittance and chromatic dependence have been optimized to improve optical quality. In parallel we have developed an electronic part that can adapt to any display in the market, as it is easily updated by software when new displays appear.

In conclusion, our technology provides 3D glasses with increased image brightness, blocking states, viewing angle and low color dependence. Combined with a programmable electronic they are good candidate for the high quality eyewear needed by the new generation of 3-D displays.

## 7863-23, Session 7

### Continuously adjustable Pulfrich spectacles

K. M. Jacobs, Binghamton Univ. (United States); R. S. Karpf, ADDIS Inc. (United States)

While many Pulfrich 3-D movies have been produced, the standard implementation has inherent drawbacks. The film industry has correctly concluded that the standard Pulfrich 3-D implementation is not a useful commercial 3-D technique.

Continuously Adjustable Pulfrich Spectacles (CAPS) is a new

implementation of the Pulfrich Effect that allows any standard 2-D movie to be optionally viewed in 3-D using inexpensive viewing specs. It works on any viewing device. Without the glasses, the movie will appear as a normal 2-D image.

Recent scientific results in the fields of human perception, optoelectronics, video compression and video format conversion are translated into a new implementation of Pulfrich 3-D. CAPS uses these results to continuously adjust to the movie so that the viewing spectacles always conform to the optical density that optimizes the Pulfrich stereoscopic illusion. This instantly provides 3-D immersion to any moving scene in any 2-D movie.

## 7863-24, Session 8

### Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict

T. Shibata, Univ. of California, Berkeley (United States) and Waseda Univ. (Japan); J. Kim, D. M. Hoffman, M. S. Banks, Univ. of California, Berkeley (United States)

Prolonged use of conventional stereo displays causes viewer discomfort and fatigue because of the vergence-accommodation conflict. We used a novel volumetric display to examine how viewing distance and the sign of the vergence-accommodation conflict affects discomfort and fatigue. In the first experiment, we presented a fixed conflict at short, medium, and long viewing distances. We compared subjects' symptoms in that condition and one in which there was no conflict. We observed more discomfort and fatigue with a given vergence-accommodation conflict at the longer distances. In the second experiment, compared symptoms when the conflict had one sign compared to when it had the opposite sign, and we did so a short, medium, and long distances. We observed greater symptoms with uncrossed disparities at long distances and with crossed disparities at short distances. These findings help define comfortable viewing conditions for stereo displays.

## 7863-66, Session 8

### Effects of 3D display on accommodative and vergent responses and subsequent visual discomfort and motion sickness

S. Yang, Pacific Univ. (United States)

Recent proliferation of 3-D display has been shown to augment viewing experiences. However, 3-D viewing also induces significant visual discomfort and motion sickness. This study investigated whether these symptoms result from the conflict between accommodative demand, determined by display distance, and the vergent demand determined by binocular image disparity. Adult subjects were asked to track the motion of a 3-D target displayed in stepwise or continuous motion. They also watched movies in 2-D and 3-D and reported their visual discomfort and motion sickness symptoms before and after viewing. Subject's accommodation and binocular eye position were continuously measured, as well as their physical symptoms. Results found that subjects made greater accommodative and vergent response in viewing 3-D stimuli in continuous motion, but only when the stimulus was approaching them. Increase in visual symptoms (blurred vision, double vision, and out of focus) and motion sickness (nausea, dizziness, and disorientation) in viewing 3-D movie was correlated to the increase in accommodative and vergent responses in 3-D animation for individual subjects. These findings suggest that reduction in perceived near motion could help maintain 3-D perception while keeping the visual discomfort and motion sickness at an acceptable level.

## 7863-25, Session 9

### Effect of image scaling on stereoscopic movie experience

J. P. Häkkinen, J. Hakala, Aalto Univ. School of Science and Technology (Finland); M. Hannuksela, Nokia Research Ctr. (Finland); P. Oittinen, Aalto Univ. School of Science and Technology (Finland)

As stereoscopic movies have become more popular, the utilization of multiple distribution channels becomes more important. Especially Blu-Ray and mobile devices will be significant channels for distributing consumer content. However, changes in the image size and viewing distance affect the binocular parallax, which might increase the visual load as well as change the viewing experience. In our study we examine the effect of different viewing contexts on the experienced stereoscopic quality. Specifically, we examine the viewing experience of S3D contents on small displays, home theater sized displays and on a cinema screen. In the experiments we change the camera base distance in each condition and measure the effect of change on experienced S3D quality. The same animation contents and base distances are utilized in all three viewing conditions so that we can compare the subjective results with each other. The results will offer us knowledge of the thresholds where the viewing context and scaling begin to be visible and annoying. This knowledge can be utilized to form quality guidelines for stereoscopic repurposing, in which the depth of the contents is changed to achieve more satisfactory viewer experience.

## 7863-26, Session 9

### Relationship between perception of image resolution and peripheral visual field in stereoscopic images

M. Ogawa, K. Shidoji, Kyushu Univ. (Japan)

High-resolution stereoscopic images are effective for virtual reality and teleoperation systems. However, the higher the image resolution, the higher is the cost of computer processing and communication. To reduce the cost, numerous earlier studies have suggested the use of multi-resolution images, which have high resolution in the region of interests (ROI) and low resolution in other areas. However, observers can perceive unpleasant sensations and incorrect depth because they could see the low-resolution area in their field of vision. In this study, we conducted an experiment to research the relationship between the viewing field and the perception of image resolution, and determined the respective thresholds of image-resolution perception for the positions of the viewing field. In the results, participants could not discriminate the high-resolution stimulus from the decreased stimulus, 63 [ppi], at positions more than 8 [deg] outside the gaze point. Moreover, with positions shifted further 14 [deg] from the gaze point, participants could not distinguish between the high-resolution stimulus and the decreased stimuli whose resolution densities were 42 and 25 [ppi]. The results show that we will propose the composition of multi-resolution images in which observers do not perceive the unpleasant sensations and incorrect depth with data reduction (compression).

## 7863-49, Poster Session

### Human perception considerations for 3D content creation

A. Green, Almont Green Studios (United States)

3D content viewed with stereopsis activates regions of the brain that deal with perceiving reality. Where inconsistencies exist, the 3D illusion creates perception conflicts that manifest with a negative reaction.

Observations of people viewing autostereoscopic lenticular photographs show that luminance dynamics matching of imagery to that of human eyes using HDR photographic techniques improves the

subjective response. Matching the size of the photograph to what is perceived in reality also evokes positive response. Smaller sizes create a "Ken and Barbie effect" which refers to the doll like appearance of people depicted in a small size photo.

Using a special 12 camera array to capture multiple perspectives simultaneously with extended dynamics, coupled to long focal length lenses makes it possible to create imagery that closely matches real life perceptions. Lens distortions and interocular spacing problems are mitigated.

The resulting imagery is printed at life size with an inkjet system with HDR inks on paper designed for backlighting. By creating samples with and without extended luminance dynamics and other dynamically adjustable parameters, it is possible to evaluate how aspects of content creation effect perception quality.

Using HDR techniques and precision matching of image size and perspective can be clearly shown to improve the subjective quality of 3D imagery. Binocular rivalry and binocular disparity issues have also been evaluated and observed and the negative aspects of 2D to 3D conversions demonstrated.

## 7863-50, Poster Session

### System crosstalk issues on autostereoscopic displays

P. Wang, S. Hwang, H. Huang, C. Chuang, National Tsing Hua Univ. (Taiwan)

In this study, there were two parts of experiments. The first part of experiment was conducted on a multi-view autostereoscopic display of mirror-type displays. The four experimental pictures randomly appeared in the mirror-type display. This research aimed at investigating the relationship between system crosstalk and perspective image quality. The independent variable was system crosstalk of the experimental pictures. The dependent variables were subjective evaluation, physiological and subjective measures toward asthenopia in the experiments. The second part of experiment investigated the fitting value among 10 levels of system crosstalk on mirror-type display corresponding to the picture shown on the barrier-type display.

In the first experiment, every level of system crosstalk to the viewer crosstalk was significantly different. In addition, the objective and subjective asthenopia in the experiment were not statistically significant. In the second experiment, viewers' evaluations of the system crosstalk on the four pictures were significantly different. There existed interaction effect between the type of picture and the perspective distortion. But the perspective distortion didn't affect distinctly. The type of the picture would affect the tolerance of viewer's crosstalk.

The outcome of the research could provide the display manufacturers a guideline for designing autostereoscopic displays of high quality.

## 7863-51, Poster Session

### Automatic 3D video format detection

T. Zhang, Z. Wang, J. Zhai, Technicolor (United States); D. Doyen, Technicolor S.A. (France)

Many 3D formats exist and will co-exist for a long time since there is no 3D standard that defines a generally accepted 3D format. The support for multiple 3D formats will be important for bringing 3D into home. In this paper, we propose a novel and effective method to detect whether a video is a 3D video or not, and to further identify the exact 3D format. First, we present how to detect those 3D formats that encode a pair of stereo images into a single image. The proposed method detects features and establishes correspondences between features in the left and right view images, and applies the statistics from the distribution of the positional differences between corresponding features to detect the existence of a 3D format and to identify the format. Second, we present how to detect the frame sequential 3D format. In the frame sequential format, the feature points are oscillating from frame to frame. Similarly, the proposed method tracks feature points over consecutive frames,

computes the positional differences between features, and makes a detection decision based on whether the features are oscillating. Experiments show the effectiveness of our method.

## 7863-52, Poster Session

### Low-complexity 2D to 3D video conversion

Y. Chen, R. Zhang, M. Karczewicz, Qualcomm Inc. (United States)

3D film and 3D TV are becoming reality. More facilities and devices are now 3D capable. Compared to capture 3D video content directly, 2D to 3D video conversion is a low-cost, backward compatible alternate. There also exists a tremendous amount of monoscopic 2D video content that are of high interest to be displayed on 3D devices with noticeable immersiveness. 2D to 3D video conversion, therefore, has drawn lots of attention recently. In this paper, a low complexity 2D to 3D conversion algorithm is presented. The conversion generates stereo video pairs by 3D warping based on estimated per-pixel depth maps. The depth maps are estimated jointly by motion and color cues. Subjective tests show that the proposed algorithm achieves 3D perception with acceptable artifact.

## 7863-53, Poster Session

### Development of a modular stereoscopic pre-visualisation and display framework

V. Kuchelmeister, The Univ. of New South Wales (Australia)

The increasing popularity for stereoscopic content in the entertainment industry and computer graphics applications and the availability of affordable capture and display systems is in contrast to the actual knowledge of underlying stereoscopic design principles and fundamental concepts. Content creators and educators unexperienced in stereoscopy require integrated, easy to use and flexible tools which can assist in the process of creating the three-dimensional "look" they are after within the limits of a comfortable viewing experience.

The proposed framework in this paper, a custom stereoscopic export plugin for the popular 3D modelling application Google Sketchup and a flexible stereoscopic display engine, allows for stereoscopic pre-visualisation in near real-time in a format of their choice. The user interface can recommend stereoscopic settings according to the scene, camera and display properties, calculates corresponding values according to manual entries but also leaves unrestricted control over all parameters. The display engine allows for different stereoscopic formats to be shown and saves the result in form of images with metadata for reference. Particular attention is put on usability, accessibility and tight integration.

## 7863-54, Poster Session

### Color appearance in stereoscopy

D. Gadia, A. Rizzi, C. Bonanomi, D. Marini, Univ. degli Studi di Milano (Italy); A. Galmonte, Univ. degli Studi di Verona (Italy); T. Agostini, Univ. degli Studi di Trieste (Italy)

The relationship between color and lightness appearance and the perception of depth has been studied since a while in the field of perceptual psychology and psycho-physiology. It has been found that depth perception affects the final object color and lightness appearance. In the stereoscopy field, many studies have been proposed on human physiological effects, but few has considered color information.

Goal of the paper is realizing some experiments in Virtual Reality in order to determine the effects of depth perception on object color appearance. We consider border effects, luminance gradients, spatial gradients, different parallax values, and we investigate how different choices of these features affect the final perception.

We created a virtual 3D test scene with a simple configuration of

geometric figures over a floating background. We generated different stereoscopic renderings of this scene, changing parameters for color and position of the objects.

We collect the perceptual responses of several users after the observation of the test scene in an immersive Virtual Reality room (the Virtual Theater of the University of Milan).

Users are asked to judge the relative appearance under different versions of the scene varying the relative object depths.

We present an analysis and discussion of these experiments.

## 7863-55, Poster Session

### Coarse integral volumetric imaging with flat screen and wide viewing angle

S. Sawada, A. Nakao, H. Kodaira, H. Kakeya, Univ. of Tsukuba (Japan)

This paper proposes a flat-screen 3D display system with wide viewing angle and little distortion based on coarse integral volumetric imaging (CIVI). CIVI combines multiview and volumetric display solutions and presents undistorted floating 3D image by correcting distortion of volumetric image for each view.

In the conventional CIVI with limited viewing angle, distortions of image plains can be approximated parabolic in the direction of depth, while those in horizontal and vertical directions can be ignored. When the viewing angle becomes wider, however, this approximation cannot realize undistorted image.

To cope with these distortions, each elemental image is approximated with individual second order equations in the method we propose. Also distortions in horizontal and vertical directions are corrected by using texture mapping. To attain precise correction in vertical, horizontal and depth directions, optical paths of light rays between the display panel and each viewpoint are calculated with an optical simulator. Color aberration is corrected by mapping RGB textures separately based on the result of optical simulation.

CIVI prototype system with flat screen and wide viewing angle is produced based on the above distortion correction method. It is confirmed that the proposed system works as expected to correct optical distortions.

## 7863-56, Poster Session

### Coarse integral imaging without pseudo image

T. Kurokawa, H. Kakeya, Univ. of Tsukuba (Japan)

Coarse integral imaging (CII), where each component lens is large enough to cover pixels far more than the number of views, can show clear floating 3D image when distortion is corrected.

The major problem for CII is emergence of pseudo images that appear around the right image to be presented. In this paper we propose two methods to suppress pseudo images.

To realize suppression of pseudo image, we first propose use of large aperture lens with small F number in front of the elemental lenses. When a large aperture lens of small F number is set in front of the elemental lenses so that the distance between them may be the same as the focal distance of large aperture lens, only pseudo images can be erased by using total internal reflection on the outskirts of the large aperture lens.

The second method we propose is use of lens array behind the display panel paired with segmented backlight. When the lens is paired with proper size of backlight, leak of ray out to adjacent elemental lenses and loss of ray into proper elemental lens are both avoided. Since the backlight area is reduced, this method consumes less electric power.



## 7863-57, Poster Session

### Free-viewpoint image generation from a video captured by a handheld camera

K. Takeuchi, Nagoya Univ. (Japan); N. Fukushima, Nagoya Institute of Technology (Japan); T. Yendo, M. Panahpour Tehrani, Nagoya Univ. (Japan); T. Fujii, Tokyo Institute of Technology (Japan); M. Tanimoto, Nagoya Univ. (Japan)

We present a novel system that generates free-viewpoint images using a freely moving handheld camera in static scene.

To generate free-viewpoint images using a captured video by a handheld camera, view frames'pose/position are needed.

Previously, a checkerboard pattern has to be captured in every frame to calculate camera pose/position. This approach obtains frames'pose/position easily, however we need to have a checkerboard pattern with known dimensions that limits the application.

In another method, corresponding feature points in all frame images are used to estimate the view frames' pose/positions, assuming a pseudo perspective projection without checkerboard pattern. However due to the assumption, the handheld camera cannot change angles in capturing the scene.

To address above problems, we propose a method that uses corresponding feature points to calculate camera pose/position using the state-of-art, "Structure from Motion"(SfM). Using this method, we can move a handheld camera freely with any angle.

Moreover, we propose a method that generates a depth map by the nearest viewpoints at the location of the free viewpoint image. Proposed depth estimation scheme uses graph-cuts algorithm for optimization, while reconstructed feature points obtained in SfM are additionally used to enhance the performance.

## 7863-58, Poster Session

### New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone

W. Chen, J. Fournier, France Telecom (France); M. Barkowsky, P. Le Callet, Univ. de Nantes (France)

In this paper, we proposed a new stereoscopic video shooting rule considering two most important issues in 3DTV: stereoscopic distortion and comfortable viewing zone. The results of this study will provide a new method to propose camera parameters based on management of new criteria (depth and shape distortion and depth of focus) in order to produce optimized stereoscopic images and videos.

## 7863-59, Poster Session

### Reduced-view super multi-view display

J. Nakamura, K. Tanaka, Tokyo Univ. of Agriculture and Technology (Japan); C. Tsai, Industrial Technology Research Institute (Taiwan); Y. Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

In order to reduce the number of views of the SMV display, two or more views are generated only around viewer's left and right eyes, with an interval smaller than the pupil diameter. The positions of the views are moved according to the viewer's eye positions to increase viewing freedom. The reduced-view SMV display is implemented using a lenticular 3D display. A cylindrical lens constituting a lenticular lens projects a group of pixels to generate a group of views. The pixel group generating the left view group and that generating the right view group through an identical cylindrical lens are spatially separated to separate the view groups. The left pixel groups and the right pixel groups for different cylindrical lenses are interlaced horizontally on the flat panel display. A prototype reduced-view SMV display was constructed. Each view group consisted of eight views. The interval of the views was 2.6

mm. An LCD panel with a slanted subpixel arrangement was used. The screen size was 2.57 inches and the 3D resolution was 256×192. A USB camera was attached to the display to detect viewer position. The frame rate of the face detection and the image update was 30 Hz.

## 7863-60, Poster Session

### Psycho-physiological effects of visual artifacts by stereoscopic display systems

S. Kim, J. Yoshitake, H. Morikawa, T. Kawai, O. Yamada, A. Iguchi, Waseda Univ. (Japan)

The methods for stereoscopic (3D) displays with glasses can be classified as time-multiplexing and spatial-multiplexing. Each method has its intrinsic visual artifacts. With the time-multiplexing method, an observer perceives three artifacts: flicker, the Mach-Dvorak effect, and phantom array. They only occur under a certain condition: any condition, during smooth pursuit (SPM), and during saccadic eye movements, respectively. With the spatial-multiplexing, temporal-parallax (due to interlaced video signal), binocular rivalry and low-resolution would be induced. These artifacts are considered one of the major problems to safety and comfort of observers while viewing 3D display. In this study, in order to evaluate the implications of the visual artifacts to the safety and comfort, physiological changes were examined through subjective symptoms of visual fatigue and depth sensation. Also, to understand the characteristics of each artifact and the combining effects of the artifacts, four experiment conditions were designed. The results showed that the perception of the visual artifacts differs from visual environments and display methods. Furthermore visual fatigue and depth sensation was influenced by individual characteristics of visual artifacts.

## 7863-62, Poster Session

### 2D viewing experience with fixed 3D displays

M. Salmimaa, T. Jarvenpaa, M. Polonen, Nokia Research Ctr. (Finland)

The main goal of the paper is to present results from subjective studies where participants evaluate 2D content created by using different representation methods and rendered on a small-size autostereoscopic 3D display. The display in question has a lenticular lens as a stereo structure, and the stereo structure as such cannot be switched between 2D and 3D modes. Subjective opinions on the different representations of the 2D content on a stereoscopic 3D display have been studied and these viewing experiences compared.

## 7863-63, Poster Session

### Interestingness of stereoscopic images

J. Hakala, M. Nuutinen, P. Oittinen, Aalto Univ. School of Science and Technology (Finland)

The added value of stereoscopy is an important factor for stereoscopic product development and content production. Previous studies have shown that 'image quality' does not encompass the added value of stereoscopy, and thus the attributes naturalness and viewing experience have been used to evaluate stereoscopic content. The objective of this study was to explore what the added value of stereoscopy may consist of and what are the content properties that contribute to the magnitude of the added value. The hypothesis was that interestingness is a significant component of the added value. A subjective study was conducted where the participants evaluated three attributes of the stimuli in consumer photography domain: viewing experience, naturalness of depth and interestingness. In addition to the no-reference direct scaling method a novel method, the recalled attention map, was used. We conclude from the results that interestingness is a factor of equal importance as naturalness in the

added value of stereoscopy in still images. From the qualitative results and RAMs we found that local differences in distances draw positive attention in stereoscopic images and we propose that a 'local disparity contrast' metric needs to be developed.

## 7863-64, Poster Session

### **Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating**

K. Wang, Acreo AB (Sweden); M. Barkowsky, R. Cousseau, Univ. de Nantes (France); K. E. Brunnström, Acreo AB (Sweden); R. Olsson, Mid Sweden Univ. (Sweden); P. Le Callet, Univ. de Nantes (France); M. Sjöström, Mid Sweden Univ. (Sweden)

In this work a set of processed videos sequences are designed for comparing the coding performance and transmission efficiency of HD S3D stereoscopic sequences in a number of IPTV scenarios. These scenarios include H.264/AVC simulcast coding as well as Multiview Video Coding (MVC). In addition, spatial and temporal subsampling as well as comparison to 2D presentation is considered. A subjective experiment is conducted to investigate the users' experience of stereoscopic video quality by using the Absolute Category Rating method (ACR) method with a training session that uses Double Stimulus Continuous Quality Rating (DSCQR).

A pre-test of subjective experiment result shows the spatial and temporal downsampling may be considered as an alternative to increasing the coding quantization parameter for improving compression efficiency. The influence of the content type on the optimal choice of spatial and temporal downsampling, the appropriate encoder, and its parameters will be analyzed in the final paper.

## 7863-65, Poster Session

### **Improved depth map estimation in stereo vision**

H. Fradi, J. E. Dugelay, EURECOM (France)

Researchers have been giving especial attention to stereo vision systems capable of perceiving accurate depth information. In this article, we propose a new stereo matching algorithm based on correlation and showing progress in handling problem of mismatched points. It is a three steps framework; first, an appropriate cost matching is used to avoid the drawbacks of possible ambiguous matches caused by the violation of the resemblance constraint. Then, a bidirectional matching is applied to detect and to reject mismatches. A matching is valid only if after a return (e.g. right-left-right) the final position is the same as the initial one. Third, the created holes will be filled in by incorporating edges detection to avoid that a window contains more than one object. It is an efficient method for selecting a variable window size with adaptive shape in order to get accurate results at depth discontinuities and in homogeneous areas while keeping a low complexity of the whole system. The resulting disparity map can be converted to depth map using simple triangulation. Experimental results using the Middlebury datasets demonstrate the validity of our presented approach. The main domain of applications for this study is the design of new functionalities within the context of mobile devices.

## 7863-67, Poster Session

### **Is visual fatigue changing the perceived depth accuracy on an autostereoscopic display?**

M. Barkowsky, R. Cousseau, P. Le Callet, Univ. de Nantes (France)

Recently, 3DTV services have been introduced to the public together

with a variety of stereoscopic and auto-stereoscopic displays. However, visual fatigue is a still a serious threat to the wide adoption of binocular presentation of video contents. In this paper, a subjective study is presented which aims to measure the minimum perceivable depth difference on an autostereoscopic display. The developed experimental setup was used to compare the subject's performance before and after 3D excitation on an autostereoscopic display. By comparing the results to a verification session with 2D excitation, the effect of 3D visual fatigue can be examined.

## 7863-68, Poster Session

### **Interlaced MVD format for free viewpoint video**

S. Lee, S. Lee, J. Lee, H. Wey, D. Park, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

A new 3D video format which consists of one full resolution mono video and half resolution left/right videos is proposed. The proposed 3D video format can generate high quality virtual views from small amount of input data while preserving the compatibility for legacy mono and frame compatible stereo video systems. The center view video is the same with normal mono video data, but left/right views are frame compatible stereo video data. This format was tested in terms of compression efficiency, rendering capability, and backward compatibility. Especially we compared view synthesis quality when virtual views are made from full resolution two views or one original view and the other half resolution view. For frame compatible stereo format, experiments were performed on two sampling cases, interlaced and quincunx method. For each case, the proposed format gives BD bit-rate gains up to 15 .

## 7863-69, Poster Session

### **Visual discomfort prediction for stereo contents**

S. He, T. Zhang, Technicolor (United States); D. Doyen, Technicolor S.A. (France)

The current renaissance of 3D movies has drawn more and more attention from the audience. While people are enjoying the 3D movies in the theater, they also hope to bring the 3D experience to home. Three-dimensional television (3DTV) has been expected to be the next advance in television. In 3DTV scenario, much smaller screen sizes and viewing distances in home theater setup put more restrictions on the 3D content fed into the 3DTV, e.g. a smaller depth range. On the other hand, different people have different comfort range of depth in a 3D content. The version of the 3D content sent to home will not stratify all the people in one family. In this paper, we try to solve this problem by providing a prediction of viewing discomfort of certain input content by certain viewer. Our method is based on the Disparity Discomfort Profile (DDP) built through subjective test for each viewer. The input content is analyzed by studying its disparity distribution. The prediction of discomfort is performed by matching the disparity distribution with the viewer's DDP. Then a mechanism to allow the viewers to adjust the depth range according to their visual comfort profile or viewing preference is used to minimize the discomfort. Experiments show promising results of the proposed method.

## 7863-70, Poster Session

### **Three-dimensional holographic display using active shutter for head mounted display application**

H. Kim, J. Park, Chungbuk National Univ. (Korea, Republic of)

Three-dimensional holographic system using active shutters for head mounted display is proposed. Conventional three-dimensional

head mounted display suffers from eye-fatigue since it only provides binocular disparity, not monocular depth cues like accommodation. The proposed method presents two holograms of a 3D scene to corresponding eyes using active shutters. Since a holography delivered to each eye has full three-dimensional information, not only the binocular depth cues but also monocular depth cues are presented, eliminating eye-fatigue. The application to the head mounted display also greatly relaxes the viewing angle requirement that is one of the main issues of the conventional holographic displays. In presentation, the proposed optical system will be explained in detail with experimental results.

## 7863-71, Poster Session

### Pixel-offset position detection using lens array for integral three-dimensional display

H. Sasaki, NHK Science & Technical Research Labs. (Japan); M. Kawakita, National Institute of Information and Communications Technology (Japan) and NHK Science & Technical Research Labs. (Japan); K. Masaoka, J. Arai, M. Okui, F. Okano, NHK Science & Technical Research Labs. (Japan); Y. Haino, M. Yoshimura, M. Sato, JVC KENWOOD Holdings, Inc. (Japan)

In the integral 3DTV system, high-density elemental images are necessary to enhance reconstructed 3D image quality (resolution, viewing zone and angle, and depth directional representability). The dual-green pixel-offset method, which uses two green images (G1 and G2), is well known as a means of achieving ultra high-resolution imagery.

We propose a precise and easy method for detecting the pixel-offset distance when the lens array is mounted in front of the display surface. In this method, pattern luminance distributions based on sine waves are displayed on each G1 and G2 panel. The difference between phases (amount of phase variation) of these patterns is conserved when the patterns are sampled and aliased to the lower frequency by the lens array. This allows the pixel-offset distance of the display panel to be measured in a state of magnification.

In this case, relation between the contrast and the amount of phase variation of the pattern is contradicted in relation to frequency. We therefore devised a way to find the optimal spatial frequency of the pattern by regarding the product of contrast and amount of phase variation of the patterns as an indicator of accuracy.

We measured and adjusted the pixel-offset detection method described above with the developed display system. Results demonstrate that the MTF of elemental images were refined. We expect that the resolution characteristics of the depth direction will be improved by it, thus ensuring higher quality reconstructed 3D images.

## 7863-72, Poster Session

### 3D imaging for glasses free multi-view 3D displays

S. Gurbuz, M. Kawakita, National Institute of Information and Communications Technology (Japan); S. Yano, NHK Science & Technical Research Labs. (Japan); S. Iwasawa, H. Ando, National Institute of Information and Communications Technology (Japan)

In this paper, a multi-view based 3D imaging technique is described for lifelike 3D visualization on multi-view auto-stereoscopic displays. The lifelike multi-view 3D visualization requires regeneration of the light field of a scene for every view. The complete light field of a scene can be reconstructed from the images of a scene ideally taken from infinite viewpoints. However, capturing the images of a scene from infinite viewpoints is not feasible for practical applications. Therefore, in this work, we provide the details of the multi-camera image alignment procedure and virtual camera view image generation that are necessary to achieve good visualization quality. For the task, we utilize an array of hardware-synchronized 30 cameras to capture multiple perspective real-views + 29 virtual views of the scene with proper horizontal parallax and occlusion relationships. Thus, at every time instance, 59

independent perspective (30 real + 29 virtual) views at 1080×1920 resolution are displayed on the multi-view auto-stereoscopic display. The major contributions are computational alignment technique for multi-camera images and novel view video rendering where a new algorithm efficiently renders a novel view from two (left and right camera) video streams.

## 7863-73, Poster Session

### Reduction of image blurring in an autostereoscopic multilayer liquid crystal display

H. Gotoda, National Institute of Informatics (Japan)

A multilayer liquid crystal display (LCD) is a display device constructed by stacking multiple liquid crystal layers on top of a light source. In a previous study, we have already shown that a multilayer LCD can deliver varying images depending on the viewers' eye positions, and can be used for auto-stereoscopic 3D viewing. However, undesirable blurring is sometimes observed in the images that a viewer receives from the display. Such blurring is notable especially around objects in the scene that are far away from the viewer.

To address this problem, we propose to place an optical lens in front of the liquid crystal layers. The lens refracts the beams of light, thus bringing the effects of moving the far objects to nearer positions. Through a simulation-based study, we show that an optimal choice exists for the focal length of the lens, which reduces the local image blurring while not compromising the overall image quality.

A prototype display with 4 layers has been implemented to demonstrate that autostereoscopic 3D viewing is really possible with a multilayer LCD. The implementation also indicates that the proposed method is comparable to other autostereoscopic methods such as the integral imaging using an array of microlenses.

## 7863-74, Poster Session

### A new volumetric 3D display using multi-varifocal lens and high-speed 2D display

T. Sonoda, H. Yamamoto, S. Suyama, Univ. of Tokushima (Japan)

We have developed a new volumetric 3-D display using the multi-varifocal lens and high-speed 2-D display. Floating clear 3-D image in space can be successfully obtained. Our volumetric 3-D image is composed of many 2-D layered images by using multi-varifocal lens. Many 2-D images can be layered by changing their depth position using the discrete focal length change of multi-varifocal lens. The high-speed multi-varifocal lens is composed of several sets of a birefringent lens and a polarization switching device. The total lens power is the sum of the lens powers of these sets. The number of lens sets, N, can yield 2N variations of total focal lengths. In order to re-position many 2-D images within afterimage time, high-speed 2-D display is newly constructed by multi-projectors using LED light-sources. Multi-projector images are projected to the same position of one screen. By switching multi-projectors quickly, 2-D images on the screen can be displayed at high-speed. This high-speed 2-D display can successfully provide bright and clear 2-D layered images by using point light sources of LED.

## 7863-75, Poster Session

### A novel super-multi-view display containing 7.680 perspective views

A. Grasnick, Sunny Ocean Studios Pte. Ltd. (Singapore)

With the increasing availability of 3D movies, a gaining interest in autostereoscopic (AS) 3D displays becomes obvious.

One of the most common applications for glasses free systems is the use in public presentations (out-of-home advertising). But almost every

AS-3D display has a limited viewing area. The resulting "sweet spot" prevents these displays from a universal usage.

The raise of the number of perspective views could extend the viewing area.

A projection of more than one perspective image in the pupil of the observer's eye will reduce the vergence-accommodation-conflict and allow focusing on the virtual images. But an adaptation of the Super-Multi-View (SMV) condition to an actual display will reduce the resolution dramatically.

The 7.680 view display uses a concept of superposition and multiplexing of different perspective views and display pixels.

In this concept the main advantage is an enlargement of the viewing area without the same level of resolution reduction as in a conventional AS-3D display.

SMV using superposition and multiplexing can be used with different AS-3D technologies, i.e. lenticulars, zone plates or parallax barrier.

The purpose of the laboratory setup in this research was to show a functional model using up to 7.680 different views, whereas every view has a very low resolution. With the same number of columns and views, each displayed perspective view has an effective horizontal dimension of only 1 subpixel (1/3 pixel).

The screen image for one frame of the 7.680 view display was made from a stack of perspective images. Based on the results of the tests, it is discussed if and how the SMV-AS-3D technology with many thousand (or more) views can become a trend in 3D television.

## 7863-76, Poster Session

### Use of camera drive in stereoscopic display of learning contents of introductory physics

S. Matsuura, Tokyo Gakugei Univ. (Japan)

Stereogram display of physics simulations for introductory physics learners were created. Since the designs of simulations are simple, and their images have less pictorial clues, the 3D models projected onto the 2D plane are not easy to understand the composition. Then, the stereoscopic representation strongly improves easiness for understanding. Also, the camera-driving means seemed to the learner feel exploratory on the simulations.

## 7863-77, Poster Session

### Producing content for 3D home theater

J. J. Karns, XD Images (United States)

3-dimensional displays for home use can be divided into 3 broad classes: Virtual Reality displays, Stereoscopic displays requiring glasses, and Autostereographic lenticular displays. Virtually all 3D content produced for the home entertainment market today can be viewed using any of the 3 classes of 3D displays with roughly the same level of 3D quality. Full appreciation of autostereographic lenticular displays requires multiple-perspective 3D content created using an adjustable-parallax, multi-view system. Most notably, autostereographic lenticular displays provide motion parallax and changes in occlusion within a single temporal frame that are not realized in content derived from 2-view stereo content.

We have been experimenting with a newly invented, adjustable-parallax, multi-perspective camera system to identify consumer preferences in multi-perspective, 3D, lenticular images (parallax panoramagrams). We have found that much of the conventional wisdom and research related to 2-view stereograms must be significantly modified as it relates to parallax panoramagrams. Our objective is to identify and codify factors and methods that maximize consumer acceptance and enjoyment of multi-perspective parallax panoramagrams.

## 7863-78, Poster Session

### DWT-based stereoscopic image watermarking

M. P. Mitrea, A. Chammem, F. J. Prêteux, TELECOM & Management SudParis (France)

The present paper deals with stereoscopic image protection by means of watermarking techniques. First, a study on the optimal insertion domain is carried out. Secondly, in order to reach the trade-off between transparency and robustness, principles from spread spectrum and side information are combined. Finally, the experiments were performed on the stereoscopic image database organised at Ecole Polytechnique Fédérale de Lausanne par T. Ebrahimi and L. Goldman (<http://mmspl.epfl.ch/page38841.html>) and on a stereoscopic medical image database built at the ARTEMIS Department (Dr. Catalin Fetita, [www.it-sudparis.eu/artemis](http://www.it-sudparis.eu/artemis)).

The mark is represented by 64 bits to be inserted in the (9,7) DWT (Discrete Wavelet Transport) corresponding to the right view of the stereoscopic image. In order to benefit from both high robustness and large data payload, the insertion technique combines spread-spectrum and informed-embedding principles (according to a basic idea presented into a previous patent the authors filled-in). The marked stereoscopic image is obtained by combining the watermarking right view and the disparity information computed between the watermarked right view and the original left view.

The experiments exhibited strong robustness against linear (Gaussian) and non-linear (sharpening, median) filtering and geometric attacks. The transparency was established by both objective and subjective criteria.

## 7863-79, Poster Session

### Development of a new HD multi-view camera and processing system

C. Park, J. Lee, J. Kang, Korean Broadcasting System (Korea, Republic of); K. Lee, Korean Broadcasting System (United States)

We have developed a new multi-view camera system which consist 9 pieces of HD camcorders, a mechanical apparatus and a control PC. By using this system, some test multi-view videos were made and successfully displayed in the lenticular type multi-view display.

The 9 pieces of HD camcorders are located on the linear type stand which is supported by three camera tripods. Each view of 9 cameras is recorded into each camcorder, at the same time, two view videos are transported to the control PC during picture-taking for monitoring. We also developed other capturing method which captures all views simultaneously into PC via IEEE-1394.

The distance between each camera can be calibrated manually from the minimum contacted position and the maximum 25cm, and the optical axis of each camera can be rotated by servo-motor. Therefore we can configure not only the toed-in camera system but also the parallel axis camera system.

Some test videos were captured by this system and converted to multi-view format video of the multi-view display. Then, we successfully watched the multi-view video via the lenticular display without wearing glasses.

In the future, we will research on the best capturing condition of this multi-view camera system.

## 7863-80, Poster Session

### Multi-view video codec based on KTA techniques

J. Seo, K. Sohn, Yonsei Univ. (Korea, Republic of)

Multi-view video coding (MVC) is video coding standard for multi-

view video data by ISO/IEC and ITU-T. It showed average PSNR gain of 1.5dB compared with view-independent coding by H.264/AVC. However, because resolution of multi-view and stereoscopic video is getting higher for 3D effect and reality, high performance video codec is required.

The MVC adopted hierarchical B-picture structure and inter-view prediction for coding efficiency. Hierarchical B-picture structure removes redundancy in time axis regardless of the characteristics of multi-view video, and inter-view prediction reduces inter-view redundancy by prediction from reconstructed neighbor views. In standardization process of the MVC, Other techniques were proposed, such as illumination compensation between views, motion information skip mode and view synthesis mode. However, they were not adopted for the MVC, because they did not show sufficient coding gain. Thus, we propose enhanced video codec for multi-view video by Key Technology Area (KTA) techniques. The KTA is a new video codec by Video Coding Expert Group (VCEG), and it is being carried out for coding efficiency and lower computational complexity. The KTA software showed better coding gain than H.264/AVC by using additional coding techniques. The techniques were proposed for 2D video, but we applied them for multi-view video.

## 7863-81, Poster Session

### On-screen-display (OSD) menu detection for proper stereo content reproduction for 3D TV

E. V. Tolstaya, V. V. Bucha, M. Rychagov, Samsung Electronics Co., Ltd. (Russian Federation)

Modern consumer 3D TV sets are equipped with special devices allowing switching between mono and stereo modes. Besides, some outside devices such as DVD or Blue-ray players, are able to display some additional information by imposing an overlay picture on video content, an On-Screen-Display (OSD) menu. In this case TV set must recognize the type of OSD, whether it is mono or stereo, and visualize it correctly by either switching off stereo mode, or continue demonstration of stereo content.

We propose a new stable method for detection of mono OSD on stereo content. Mono OSD is set to be a rectangular area with letters and pictograms. OSD menu can be of different transparency levels and colors. The main problem in detecting OSD is to distinguish whether the color difference is due to OSD presence, or due to stereo parallax. We applied special techniques to find reliable image difference and additionally used a cue that usually OSD has very implicit geometrical features: straight parallel lines. The developed algorithm was tested on our video sequences database, with several types of OSD with different colors and transparency levels overlaid upon video content. Detection quality exceeded 99% of true answers.

## 7863-82, Poster Session

### Real-time ray-space transfer of cylindrical objective space

T. Yendo, Nagoya Univ. (Japan); T. Fujii, Tokyo Institute of Technology (Japan); M. Panahpour Tehrani, M. Tanimoto, Nagoya Univ. (Japan)

We propose a real-time ray-space transfer system that captures rays around a cylindrical objective space from 360 degree horizontally and reconstructs it as 3D image which can be seen from 360 degree. The system consists of a ray capturing unit, a display unit, and a data converter.

The ray capturing unit is the same as we have already proposed; it consists of a scanning optics system and a high-speed camera. The scanning optics system is composed of a double-parabolic mirror shell and a rotating flat mirror tilted at 45 degrees to the horizontal plane. The mirror shell produces a real image of an object that is placed at the bottom of the shell and the rotating mirror reflect the image to the

camera, that enables capturing images from different angles as if the camera is moving.

According to the property of the scanning optics system, captured image includes distortion and is rotated. The data converter processes compensation of the distortion and the rotation in real-time. Due to extremely high frame rate, we developed a special hardware using FPGA to achieve real-time image processing. The data converter also converts image data into suitable format for transmitting to the display.

The display unit is also known as the Seelinder, now we have developed new version of the display. It can receives image data of over 300 views and display it in real-time with 24 bit color depth.

## 7863-83, Poster Session

### Analysis of scene distortions in stereoscopic images due to the variation of the ideal viewing conditions

A. Viale, D. Villa, D. Marini, Univ. degli Studi di Milano (Italy)

Recently stereoscopy has increased a lot its popularity and various technologies are spreading in theaters and homes allowing observation of stereoscopic images and movies, becoming affordable even for home users. However there are some "golden rules" that users should follow to ensure a better enjoyment of stereoscopic images, first of all the viewing condition should not be too different from the ideal ones, which were assumed during the production process.

To perceive stereo depth instead of a flat image, two different views of the same scene are shown to the subject, one is seen just through his left eye and the other just through the right one; the vision process is making the work of merging the two images in a virtual three-dimensional scene, giving to the user the perception of depth. The problem is that viewing the same two images from a different position will result in "merging" a different virtual three-dimensional scene.

With the set of instruments we have developed, we can analyze different viewing conditions of the stereoscopic scene in order to configure a viewing environment, either a movie or a home theater, to allow a correct vision as independent as possible from the ideal viewing conditions.

## 7863-84, Poster Session

### Analysis of resolution limitation of glasses-free 3D tabletop display

D. Moldovan, S. Yoshida, M. Kawakita, H. Ando, National Institute of Information and Communications Technology (Japan)

In this work we propose a method to compute the maximum displayable resolution of a tabletop display using conic-shaped optical device. Our approach employs computation of the Nyquist frequency of the 3-D image in both vertical and horizontal planes. By being able to extract a formula that combines projector's pitch and optical features of the screen for computing the maximum resolution, future applications will benefit by knowing how to design the tabletop display in order to obtain a 3-D image of a certain resolution.

## 7863-85, Poster Session

### Image quality of up-converted 2D video from frame-compatible 3D video

F. Speranza, W. J. Tam, C. A. Vázquez, A. Vincent, R. Renaud, R. Klepko, Communications Research Ctr. Canada (Canada)

How to provide both 2D and 3D video with high picture quality is a key concern for broadcasters and content providers. Video format plays a role. Although Service-compatible formats provide 3D video capabilities with the ability to deliver regular 2D video services, they require more bandwidth for transmission. With Frame-compatible

formats, the separate left and right views are reduced in resolution and packed to fit within the same video frame as a conventional 2D signal. Thus, they do not require more bandwidth than current 2D systems. However, with Frame-compatible formats the quality of the 2D video is an issue because a 3D service should also be able to deliver a conventional 2D signal. Such as, when the viewer elects to watch a program in 2D or when a 2D signal needs to be intermixed with 3D programming in advertising. We report in two experiments the loss in video quality of 2D video material up-converted from side-by-side frame compatible 3D video at different bit-rates and under both 3D and 2D viewing conditions. The results confirmed the loss of video quality of the 2D video up-converted material. However, the loss was rather small, was not affected by bit rate and varied marginally for the 2D and 3D presentation modes.

## 7863-86, Poster Session

### **System crosstalk measurement of a time-sequential 3D display using ideal shutter glasses**

F. H. Chen, K. Huang, L. D. Lin, C. Wu, K. Lee, Industrial Technology Research Institute (Taiwan)

the market of stereoscopic 3D TV grows up fast recently; however, for 3D TV really taking off, the interoperability of SG to view different TV sets must be solved, so developing a measurement method to separate time-sequential stereoscopic displays and shutter glasses (SG) is necessary. For measuring the 3D performance of time-sequential stereoscopic 3D displays, the optical characterization of time-sequential displays and SG should be separated. The advantages are that the sources of optical characterization are distinguished, and the interoperability of SG is broadened. Hence, this paper proposed an "ideal shutter glasses" (ISG), whose non-ideal properties are eliminated, as a platform to evaluate the optical characterization purely from the display. In the ISG method, the illuminance of the display was measured in time domain to analyze the optical characterization of the display. In this experiment, the ISG method was used to measure system crosstalk (SCT) with a high-speed-response illuminance meter. From the time-resolved illuminance signals, the slow time response of liquid crystal leading to SCT is visualized clearly. Furthermore, an intriguing phenomenon that SCT measured through SG increases with shortening viewing distance was observed, and it may arise from LC leakage of the display and shutter leakage at large viewing angle. Thus, we measured how LC and shutter leakage depending on viewing angle and verified our argument. Therefore, the ISG method offers a platform to evaluate SCT from time-sequential 3D displays, and the resources of SCT are distinguished.

## 7863-87, Poster Session

### **Guidance for horizontal image translation (HIT) on high definition stereoscopic video production**

D. K. Broberg, Cable Television Labs., Inc. (United States)

Horizontal image translation (HIT) is an electronic process for shifting the left-eye and right-eye images horizontally as a way to alter the stereoscopic characteristics and alignment of 3D content after signals have been captured by the stereoscopic cameras. HIT can be a valuable tool in the post production process as a means to modify stereoscopic content for more comfortable viewing on television screens. Most commonly it can be used to alter the zero parallax setting (ZPS), to compensate for stereo window violations or to compensate for excessive positive or negative parallax in the source material.

HIT must be used cautiously and with full awareness of the impact on other interrelated aspects of the stereography. As more and more cinematic 3D content migrates to television distribution channels the use of this tool will likely expand. With appropriate consideration, HIT can be used effectively to adjust such content for more comfortable

viewing on television. However, disregard of certain guidelines can actually harm the 3d viewing experience of such content by television audiences. This paper provides guidance on its most effective use and describes some of the interrelationships and trade-offs. The paper recommends the adoption of the cinematic 2K video format as a 3D source master format for high definition television distribution of stereoscopic 3D video programming.

## 7863-27, Session 10

### **Implementation of autostereoscopic HD projection display with dense horizontal parallax**

S. Iwasawa, M. Kawakita, National Institute of Information and Communications Technology (Japan); S. Yano, NHK Science & Technical Research Labs. (Japan); H. Ando, National Institute of Information and Communications Technology (Japan)

Our final goal is to develop advanced autostereoscopy in case of not compelling viewers to wear 3D glasses on, or not to be shown under insufficient resolution even if it frees 3D glasses. We believe that larger screen size, higher image quality, natural image appearance such motion parallax and multiple viewer capable are priority targets for professional 3D display applications. By combining the proprietary screen and the developed projector array, we've designed and implemented a kind of autostereoscopic projection display. Enough number of pixels to render true high definition is assigned for every viewpoint. The initial implementation has more than 100 million overall pixels. An actual observed horizontal motion parallax is quite smooth and reduced flipping. Through out this feasibility study, we've learned especially following two practices; a strong requirement of "array friendly feature" ready projector, and existence of some imagery glitches. Appearances of moirés and ghost images are most significant visual fatigue on the implementation. Some of these problems were tackled and already suppressed. Projectors for the array should be prepared with a management of color space and brightness, geometric image compensation, and accurate frame synchronization, and so forth. To extract and examine practical problems along with autostereoscopic projection display was a first step as a feasibility study.

## 7863-28, Session 10

### **Full-parallax 360 degrees horizontal viewing integral imaging using anamorphic optics**

M. Erdenebat, G. Baasantseren, J. Park, N. Kim, K. Kwon, Chungbuk National Univ. (Korea, Republic of)

We proposed full-parallax integral imaging display with 360 degrees horizontal viewing angle. The elemental images are projected by the high-speed DMD projector and integrated into three-dimensional image by the lens array. The cylindrical lens system tailors the horizontal and vertical viewing angle of the integrated 3D images in order to obtain high angular ray density in horizontal direction and large viewing angle in vertical direction. Finally, the mirror screen that rotates in synchronization with the DMD projector presents the integrated three-dimensional images to desired direction accordingly. By this method, full-parallax 360 degree horizontal viewing angle three-dimensional images with both of monocular and binocular depth cues can be achieved.

## 7863-29, Session 10

### **Optical characterization of autostereoscopic 3D displays**

M. J. Sykora, 3M Co. (United States)

Recently, there have been many exciting announcements for

autostereoscopic 3D (AS3D) displays; particularly for mobile applications. This paper reviews the measurement of AS3D displays using different optical techniques. The quality and usability of an AS3D display is highly dependant on its optical properties. Some of the properties measured are the viewing distance, viewing offset, bias angle, and crosstalk. The analysis techniques used to glean the critical information about these optical properties will be described. A comparison is made between various types of AS3D displays, including a time sequential 3D display. The measurement results from the metrology methods are compared where ever possible and examples will be shown.

## 7863-30, Session 11

### Depth cube display using depth map

B. Song, S. Min, J. Jung, Kyung Hee Univ. (Korea, Republic of)

We propose a depth cube display (DCD) method using depth map. The structure of the proposed method consists of two parts: A projection part composed of projector for generating image and a twisted nematic liquid crystal display (TN-LCD) as polarization modulating device for adjusting proper depth and a display part composed of an air-spaced stack of selective scattering polarizers which make the incident light to scatter selectively as the polarization of the light rays. An image from projector whose depth is determined as passing through the TN-LCD displaying depth map progresses into the stack of selective scattering polarizers and then three-dimensional image is generated. At that time, the polarization of each polarizer is set 0°, 45° and 90° sequentially, and then the incident light rays are scattered by different polarizer as the polarization of these rays. If the light ray has the polarization between those of polarizers, this light ray is scattered by multi polarizers and the image of this ray is generated on air between polarizers. The proposed method is more simple structure and implemented easily than previous DCD method. We explain and verify the proposed method.

## 7863-31, Session 11

### Surface representation of 3D objects for aerial 3D display

H. Ishikawa, H. Watanabe, S. Aoki, H. Saito, Keio Univ. (Japan); S. Shimada, M. Kakehata, Y. Tsukada, National Institute of Advanced Industrial Science and Technology (Japan); H. Kimura, Aerial Systems Inc. (Japan) and Burton Inc. (Japan)

A new type 3D display, which is small and high-speed desktop aerial 3D display (desktop system), has been developed by Burton Inc. and AIST. Even if the display area is small, the desktop system can create a dot of light at 50 kHz.

In this presentation, we propose a novel method for drawing the complex surface of 3D objects by vector scanning, which is suitable to the desktop system. The proposed method represents the surface with cross sections of an object against the vertical direction. This means that the object is represented by a set of contours on the cross sections. The drawing route in each section is determined by the connection of polygonal patches taking into account a burden of 3D scanner. As the result, point sequence data is created from polygonal models.

From the experiments of drawing, 3D objects, for example such as a hand, can successfully be drawn by the proposed method in the desktop system. Based on these experiments, we confirm that it is appropriate to represent 3D objects by section contours, when the display system which generates dots of light at high frequency and which scans at high speed vector scanning is used.

## 7863-33, Session 13

### How are crosstalk and ghosting defined in the stereoscopic literature?

A. J. Woods, Curtin Univ. of Technology (Australia)

Crosstalk is a critical factor determining the image quality of stereoscopic displays. Also known as ghosting or leakage, high levels of crosstalk can make stereoscopic images hard to fuse and lack fidelity; hence it is important to achieve low levels of crosstalk in the development of high-quality stereoscopic displays. In the wider academic literature, the terms crosstalk, ghosting and leakage are often used interchangeably and unfortunately very few publications actually provide a descriptive or mathematical definition of these terms. Unfortunately when definitions are provided they are sometimes contradictory.

This paper reviews how the terms crosstalk, ghosting and associated terms (system crosstalk, viewer crosstalk, gray-to-gray crosstalk, leakage, extinction and extinction ratio, and 3D contrast) are defined and used in the stereoscopic literature. Both descriptive definitions and mathematical definitions are considered.

The paper will also briefly discuss literature on the perception of crosstalk in stereoscopic displays.

## 7863-34, Session 13

### A simple method for measuring crosstalk in stereoscopic displays

M. A. Weissman, TrueVision Systems (United States); A. J. Woods, Curtin Univ. of Technology (Australia)

Maintaining low crosstalk in a stereoscopic display system - that is, reducing the amount of "wrong" image in each eye - is critically important for comfortable and high-quality 3D viewing. A moderate amount can cause eyestrain; a large amount will prevent fusing the 3D scene. However, because of

- The lack of measurements,
  - The complexity of making a measurement, or
  - The reluctance of manufacturers to release measurement data,
- it is often difficult for the user to know how much crosstalk is in any particular display.

We will propose here a simple method of measuring crosstalk (also known as "ghosting"), one that relies on only viewing test patterns on the display. No electronic or optical instruments are needed. Our hope is that this tool can be distributed widely and will lead to the collection of consistent information about 3D displays, and therefore, to the production of the best stereoscopic displays possible.

We will also present the results of optical measurements that confirm that the simple method gives good crosstalk measurements

## 7863-35, Session 13

### Ergonomic evaluation of crosstalk in stereoscopy through heart activity and forehead blood flow

S. Toyosawa, H. Morikawa, K. Nakano, T. Kawai, Waseda Univ. (Japan); C. Chen, H. Chang, J. Yang, Industrial Technology Research Institute (Taiwan)

Crosstalk is a phenomena in stereoscopy where objects become blurry due to leakage of the left image into the right eye and vice versa, and considered one of the most serious problems in stereoscopy. The current study aims at examining mental activity under a various level of crosstalk through heart activity and forehead blood flow. In the experiment that presented three still images and one video with a various crosstalk ratios, heart rate showed tri-phasic pattern: decelerative - accelerative - decelerative - accelerative for all the image

types. The pattern suggests the change in mental state in accordance to the crosstalk level: i.e. orientation response under no-crosstalk, mental elaboration upon noticing the presence of crosstalk, reduced level of elaboration as crosstalk progressed, and stressed state when the crosstalk exceeds tolerance limit. However, the patterns in the ratio of the low and high frequency component of the heart rate variability (LF/HF) and forehead blood flow showed slight deviation from the heart rate pattern in some stimulus types. This suggests that the mental states under crosstalked image viewing could be more complex than simple combination of orientation response and mental elaboration.

## 7863-36, Session 13

### Optical characterization of shutter glasses stereoscopic 3D displays

P. M. Boher, T. R. Leroux, V. Collomb-Patton, ELDIM (France)

A shutter glasses stereoscopic 3D display is a combination of one display working at high frequency and liquid crystal shutter glasses. Both components have their own imperfections that must be taken into account simultaneously to measure the performances of such systems. In the proposed paper a 3D ready Samsung SyncMaster 2233RZ 120Hz LCD display coupled with a NVIDIA 3D vision system is measured. Transmittance and response time of the shutter glasses are measured using a static light source. Grey to grey level response times and luminance targets of the LCD are also measured. Finally, the temporal behavior of the complete system is modeled and grey to grey luminance across shutter glasses are deduced. Visual impact is checked using grey level test patterns and imaging colorimeter. Grey level variations due to the crosstalk between the two eyes and the temporal synchronization are the main source of imperfection for this type of display.

## 7863-37, Session 13

### The effect of crosstalk on depth magnitude in thin structures

I. Tsirlin, R. S. Allison, L. M. Wilcox, York Univ. (Canada)

Stereoscopic displays must present separate images to the viewer's left and right eyes. Crosstalk is the unwanted contamination of one eye's image from the image of the other eye. It has been shown to cause distortions, reduce image quality and visual comfort and increase perceived workload when performing visual tasks. Crosstalk also affects one's ability to perceive stereoscopic depth although little consideration has been given to the perception of depth magnitude in the presence of crosstalk. In this paper we extend a previous study (Tsirlin, Allison & Wilcox, 2010, submitted) on the perception of depth magnitude in stereoscopic occluding and non-occluding surfaces to the special case of crosstalk in thin structures. Crosstalk in thin structures differs qualitatively from that in larger objects due to the separation of the ghost and real images and thus theoretically could have distinct perceptual consequences. To address this question we used a psychophysical paradigm, where observers estimated the perceived depth difference between two thin vertical bars using a measurement scale. Our data show that crosstalk degrades perceived depth. As crosstalk levels increased the magnitude of perceived depth decreased, especially for stimuli with larger relative disparities. In contrast to the effect of crosstalk on depth magnitude in larger objects, in thin structures, a significant detrimental effect was found at all disparities. Our findings, when considered with the other perceptual consequences of crosstalk, suggest that its presence in S3D media even in modest amounts will reduce observers' satisfaction.

## 7863-38, Session 14

### Effects of stereoscopic presentation on visually induced motion sickness

H. Ujike, H. Watanabe, National Institute of Advanced Industrial

Science and Technology (Japan)

The present study investigates whether VIMS, which can be induced in 2D images, is affected by stereoscopic presentation. To do this, we conducted an experiment to measure the effects psychologically and physiologically. Thirty-five adults, aged 21-77 years, participated in the experiment. Visual stimulus was computer graphics that simulates traveling along streets for 10 minutes. The visual rotations, (+/- 30deg, 0.167Hz), along pitch and roll axes were alternatively added at intervals. The stimulus were created as either the stereoscopic, "3D", or "2D" images, and they were presented on 3D LC displays. Each observer watched both 2D and 3D images with one-hour rest between them. We measured Simulator Sickness Questionnaire (SSQ) before and after each trial, and subjective comfort level every one minute during each trial. Moreover, we measured electrocardiogram, plethysmograph, respiration as indices of autonomic nervous activity. The results showed that higher SSQ scores and lower comfortable level for the 3D image than for the 2D image. Moreover, %RR50, which is the index of parasympathetic nerve activity, clearly decreased more for the 3D image than for the 2D image. We conclude that stereoscopic presentation enhances biomedical effects of VIMS. We speculate that stereoscopic images can be efficient reference of spatial orientation.

## 7863-39, Session 14

### Vergence and accommodation to multiple-image-plane stereoscopic displays: 'Real world' responses with practical image-plane separations?

K. J. MacKenzie, R. Dickson, S. J. Watt, Bangor Univ. (United Kingdom)

Conventional stereoscopic displays present images on a single focal plane. The resulting 'conflict' between the stimuli to the eyes' focusing response (accommodation) and to convergence causes fatigue and poor stereo performance. One promising solution is to distribute image intensity across a number of relatively widely spaced image planes—a technique referred to as depth filtering. Previously, we found this elicits accurate, continuous monocular accommodation responses with image-plane separations up to ~1.1 Diopters (MacKenzie et al., 2010), suggesting that a relatively small (i.e. practical) number of image planes is sufficient to eliminate vergence-accommodation conflicts over a large range of simulated distances. However, accommodation responses overshoot systematically when the same stimuli are viewed binocularly, due to convergence-driven accommodation (MacKenzie & Watt, 2010). Here, we examined the minimum image-plane spacing required for accurate accommodation to binocular depth-filtered images. We compared accommodation (and vergence) responses to step changes in depth (0.3-1.2 D) for depth-filtered stimuli, using image-plane separations of 0.6-1.2 D, and equivalent real stimuli. Accommodation responses were accurate for image-plane separations of ~0.6-0.9 D. Thus, depth filtering can be used to precisely match accommodation and vergence demand in a practical stereoscopic display, using a relatively small number of image planes.

## 7863-40, Session 14

### Both efficiency measures and perceived workload sensitive for manipulations in binocular disparity

M. van Beurden, W. Ijsselstein, Technische Univ. Eindhoven (Netherlands)

Stereoscopic displays are known to offer a number of key advantages in visualizing complex 3D structures or datasets. The large majority of studies that focus on evaluating stereoscopic displays for professional applications use completion time and/or percentage correct answers to measure potential performance advantages. However, both completion time and accuracy might not fully reflect all the benefits of stereoscopic



displays. In this paper, we argue that perceived workload is an additional valuable indicator reflecting the extent to which users can benefit from using stereoscopic displays. Overall, the results showed that the performance (completion time and accuracy) was optimal around 25 min of arc, and significantly decreased for a disparity level of 50 min arc. Further perceived mental workload decreased with increasing disparity. When the disparity level becomes 50 min of arc, perceived workload significantly increased. Perceived discomfort gradually increased with increasing disparity levels. The results further suggest that perceived workload was shown to be sensitive for variations in disparity and therefore introduces a promising theoretical concept as well as a useful measurement tool to aid human factors research on stereoscopic displays.

## 7863-41, Session 14

### Comparison of relative (mouse-like) and absolute (tablet-like) interaction with a large stereoscopic work-space

M. Averkiou, N. A. Dodgson, Univ. of Cambridge (United Kingdom)

We compare two different modes of interaction with a large stereoscopic display. In absolute mode, the physical pointer's position exactly maps to position in the display volume, analogous to a 2D graphics table and 2D screen. In relative mode, the connection is between the physical pointer's motion and the motion of the pointer in the volume is analogous to that obtained with a 2D mouse and 2D screen.

Both statistical analysis and participants' feedback indicated a strong preference for absolute mode over relative mode. This is in contrast to 2D displays where relative mode (mouse) is far more prevalent than absolute mode (tablet). We also compared head-tracking against no head-tracking. There was no statistically-significant advantage to using head-tracking, however almost all participants strongly favoured head-tracking.

## 7863-42, Session 15

### Optimal design and critical analysis of a high resolution video plenoptic demonstrator

V. Drazic, J. Sacré, J. Bertrand, A. Schubert, Technicolor (France)

A plenoptic camera is a natural multi-view acquisition device also capable of measuring distances by correlating a set of images acquired under different parallaxes. Its single lens and single sensor architecture have two downsides : limited resolution and depth sensitivity. In a very first step and in order to circumvent those shortcomings, we have investigated how the basic design parameters of a plenoptic camera optimize both the resolution of each view and also its depth measuring capability. In a second step we have built a prototype based on a very high resolution Redone movie camera with an external plenoptic adapter and a relay lens. The prototype delivers 5 video views of 820x410. The main limitation in our prototype is view cross talk due to optical aberrations which reduce the depth accuracy performance. We have simulated some limiting optical aberrations and predicted the impact on the performances of the camera. We have developed adjustment protocols based on a simple pattern and analyzing programs which investigate the view mapping and amount of parallax crosstalk on the sensor on a pixel basis. The results of these developments enable us to adjust the lenslet array with a sub micrometer precision and to mark the pixels of the sensor where the views do not register properly.

## 7863-43, Session 15

### Geometric and subjective analysis of stereoscopic I3A cluster images

M. Kytö, J. Hakala, P. Oittinen, Aalto Univ. School of Science and Technology (Finland)

It can be expected that stereoscopic photography will be incorporated in mobile phones in near future. The typical scenes in mobile phone photos are divided into clusters in International Imaging Industry Association's (I3A) Camera Phone Image Quality Initiative. This paper presents a geometric and subjective analysis for stereoscopic versions of four I3A clusters.

The geometry of the stereoscopic pipeline from scene to viewer's eye is a very relevant issue in stereoscopic media. One important factor is the camera separation, because it can be used to control the perceived depth of stereoscopic images. The computational camera separations were compared to subjectively preferred camera separations.

Participants evaluated the strength and naturalness of depth sensation and overall viewing experience from the still images with single-stimulus method. Results showed that participants were able to perceive the change of depth scale even though the images were shown in random order without a reference depth scale.

The mild depth sensation was preferred over strong depth sensation. The computational camera separation differed from the subjectively preferred camera separation when the depth range of the scene was narrow. This result indicates that scenes with narrow depth should not be imaged with a long camera separation just to fill the depth budget of the display.

## 7863-44, Session 16

### The Dynamic Floating Window: a new creative tool for 3D movies

B. R. Gardner, Independent 3D Consultant (United States)

Unlike the real world, stereoscopic cinemas have a bordered frame which can unnaturally cut off the image, creating conflicting visual cues. This can diminish the 3-D effect and cause visual fatigue.

OBJECTIVE: A solution is sought which meets four key criteria:

- (1) remove "Window Violations" (visual conflict)
- (2) offer controls to minimize "Retinal Rivalry Zones"
- (3) "invisible" - it's use is neither apparent nor distracting to audiences
- (4) controllable - should support filmmaker's storytelling

METHOD: In 1952, Spottiswoode applied a static Float Window technique to the short film, "The Black Swan".

In this paper, the concept of the static Floating Window is greatly expanded to match the dynamic nature of movies.

By positioning a dynamic mask at the borders of the 3-D movie, then animating the Left and Right eye stereoscopic border masks differentially, a stereoscopic parallax is created. Thus the screen borders appear to "float" off of the device screen in three dimensions. By varying the border widths and angles, this Window can even be re-oriented to be non-parallel to the display screen. This decouples the perceived Window (screen border) from the screen, making it a continuously controllable element by the 3-D filmmaker.

The Dynamic Floating Window has been successfully used to achieve all four of the stated Objectives, and they've been applied to over a dozen 3-D feature films.

## 7863-45, Session 16

### Stereo video inpainting

F. Raimbault, A. Kokaram, Trinity College Dublin (Ireland)

As the production of stereoscopic content increases, so does the need

for post-production tools for that content. Video inpainting has become an important tool for rig removal but there has been little consideration of the problem in stereo. An algorithm for stereo video inpainting that builds on existing exemplar based work and also considers the issues of view consistency will be presented.

Given user selected patches in the sequence which may be in the same location in several frames and in both views, the objective is to fill in this patch using all the available picture information. Existing algorithms lack temporal consistency, causing flickering and other artefacts. The use of long term picture information across many frames in order to achieve temporal consistency at the same time as exploiting inter-view dependencies will be discussed.

The core of the process is built on finding matching patches in surrounding picture information by extending an exemplar-based framework. A sample area is constructed from frames in time, views (using inter view disparity vectors) and in the current frame. Matching patches in those frames are used to fill the missing hole recursively pixel by pixel.

### 7863-46, Session 16

#### **A modified non-local mean inpainting technique for occlusion filling in depth-image based rendering**

L. Azzari, F. Battisti, Univ. degli Studi di Roma Tre (Italy); A. P. Gotchev, Tampere Univ. of Technology (Finland); M. Carli, Univ. degli Studi di Roma Tre (Italy); K. Egiastian, Tampere Univ. of Technology (Finland)

A technique for filling disocclusion holes arising from depth-image based rendering is proposed. It adapts a state-of-the-art exemplar-based inpainting algorithm to the specifics of depth-based view synthesis in 3D video display. Modifications are suggested in the so-called priority map and also in the way patches are searched for similarity and to be used in forming non-local mean estimates.

Objective and subjective tests have been conducted to evaluate the performance of the proposed technique against state-of-the-art occlusion-filling approaches and the results show an improved performance along with an improved efficiency compared with the original method.

### 7863-47, Session 16

#### **A study on the stereoscopic codecs for non-real time 3DTV services**

B. Lee, Electronics and Telecommunications Research Institute (Korea, Republic of)

This paper presents a study on the stereoscopic 3D codec for the non-real time 3DTV services. For the DTV (Digital Television) system where its bandwidth is limited to accommodate the full 3D HD quality video, a complementary enabler is the non-real time delivery scheme which 3D video component is downloaded in advance. From the codec perspective, if stereo videos are coded in independently, the coding/decoding can be simply managed compared to the case where stereo videos are interview prediction coded. In DTV environment where it is built on ATSC (Advanced Television System Committee) standard, making a choice of an optimal codec is regarded as one of the key issues for 3DTV services. When we view the NRT service, the choice of codec is also a primary concern because the combination of codec and the use of interview prediction have an effect on the performance of NRT services. So in this paper the combinations of available codecs are evaluated and also the performance comparisons are conducted to find the optimal codec for NRT services. Amongst various codec combinations, we evaluated MPEG2+MPEG2, MPEG2+H.264 Simulcast, MPEG2+H/264 Interview Prediction and MVC. Based on this evaluation, this paper also addresses the optimal condition of NRT delivery. It covers the NRT scenarios, 3D object size, object segmentation and aggregation, scheduling, and delivery scheme.

### 7863-48, Session 16

#### **A modular cross-platform GPU-based approach for flexible 3D video playback**

R. Olsson, H. Andersson, M. Sjöström, Mid Sweden Univ. (Sweden)

Different compression formats for stereo and multiview based 3D video is being standardized and software players capable of decoding and presenting these formats onto different display types is a vital part in the commercialization and evolution of 3D video.

This paper describes the design and implementation of a GPU-based real-time 3D video playback solution, built on top of cross-platform, open source libraries for video decoding and hardware accelerated graphics. A software architecture is proposed that efficiently process and presents high definition 3D video in real-time and in a flexible manner support both current 3D video formats and emerging standards. The presented solution is a stand-alone 3D video player application built on top of FFmpeg, using libavformat for media container demultiplexing and libavcodec for video decoding. To increase modularity and flexibility the 3D video player functionality has been divided into two separate components: a video player and a 3D video filter.

The proposed prototype 3D video playback solution shows that it is possible to build a 3D video player application relying completely upon open-source and cross-platform libraries. Compressed tiled and multi-view video formats up to resolutions of 1080p have successfully been verified to be displayed at their intended native frame rate.

# Conference 7864A: 3D Imaging Metrology

Monday-Tuesday 24-25 January 2011

Part of Proceedings of SPIE Vol. 7864A 3D Imaging Metrology

## 7864A-01, Session 1

### Traceable hierarchical procedures for dimensional metrology

D. K. MacKinnon, J. A. Beraldin, L. Cournoyer, B. Carrier, National Research Council Canada (Canada)

We present a series of dimensional metrology procedures that have been either designed or modified from existing procedures to ensure traceability of each metric from the certified reference surface to the certifying laboratory. These metrics are divided into surface form precision, surface fit trueness, and surface response. The procedures for generating these metrics would form the basis of a volumetric analysis of the characteristic profile of a 3D imaging system. We use a hierarchical approach in which each metric builds on either certified reference values or previously-generated characteristic values. Starting from simple planar and spherical surfaces using fitting procedures recommended by NIST, we demonstrate how metrics for plane form spread (flatness), sphere form spread, sphere form error, sphere-spacing error, plane-spacing error, planar uncertainty resolution, intensity resolution, and spatial resolution are built upon each other. Both simulated and real data are used to demonstrate how these procedures are used as part of a process for characterizing the performance of a 3D imaging system.

## 7864A-02, Session 1

### Harmonic distortion free distance estimation in ToF camera

B. Kang, S. Kim, K. Lee, J. D. K. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

A Time-of-Flight (ToF) depth camera captures the distance from the camera to an object using a near infrared (NIR). The distance can be calculated from the phase shift between the emitted and reflected NIR. ToF depth cameras usually modulate the NIR with a square wave rather than a sinusoidal wave due to its difficulty in hardware implementation. Previous method uses a simple trigonometric function to estimate the phase shift using the difference of electrons generated by the reflected square wave so that the estimated phase shift can include a harmonic distortion. This is because the phase shift should be linearly proportional to the difference of electrons along the distance to an object. The trigonometric function, however, nonlinearly estimates the phase shift. In this paper, we propose a new estimation method based on the sign of the difference of electrons to reduce the distortion of the phase shift. For quantitative evaluation, the previous and proposed methods are tested on our prototype ToF depth camera. Experimental results show that the distance calculated from the proposed method is more accurate than that from the previous one.

## 7864A-03, Session 1

### Separating true range measurements from multi-path and scattering interference in commercial range cameras

A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, L. V. Streeter, The Univ. of Waikato (New Zealand)

Time-of-flight range cameras acquire a three-dimensional image of a scene simultaneously for all pixels from a single viewing location. Attempts to use range cameras for metrology applications have been hampered by the multi-path problem, which causes range distortions when stray light interferes with the range measurement in a given pixel. Correcting multi-path distortions by post-processing the three-dimensional measurement data has been investigated, but enjoys

limited success because the interference is highly scene dependent. An alternative approach based on separating the strongest and weaker sources of light returned to each pixel, prior to range decoding, is more successful, but has only been demonstrated on custom built range cameras, and has not been suitable for general metrology applications. In this paper we demonstrate an algorithm applied to both the Mesa Imaging SR4000 and Canesta Inc. XZ422 Demonstrator unmodified off-the-shelf range cameras. Additional raw images are acquired and processed using an optimization approach, rather than relying on the processing provided by the manufacturer, to determine the individual component returns in each pixel. Substantial improvements in accuracy are observed, especially in the darker regions of the scene.

## 7864A-04, Session 1

### 3D imaging studies of rigid-fiber sedimentation

D. W. Vahey, T. Scott, U.S. Forest Service (United States); E. J. Tozzi, Univ. of California, Davis (United States); D. J. Klingenberg, Univ. of Wisconsin-Madison (United States)

Fibers are industrially important particles that experience coupling between rotational and translational motion during sedimentation. This leads to helical trajectories that have been poorly understood from both theoretical and experimental perspectives.

Sedimentation experiments and hydrodynamic analysis were performed on eleven copper "fibers" of average length 10.5 mm and diameter 0.20 mm. Each fiber contained three linear but non-coplanar segments. The fibers were characterized by their 2D projections on orthogonal planes.

The fibers were sequentially released into silicone oil contained in a transparent cylinder of square cross section. Identical, synchronized cameras were mounted to a moveable platform and imaged the cylinder from orthogonal directions. The cameras were fixed in position during the time that a fiber remained in the field of view. Subsequently, the cameras were controllably moved to the next, lowered field of view. The trajectory of descending fibers was followed over a range of 250 mm.

Custom software was written to decouple fiber orientation and trajectory from the 3D images. Fibers with similar terminal velocity were found to have highly variable angular velocities. Both terminal and angular velocities were well-predicted by theory. Helical radius was hard to predict when angular velocity was small, probably reflecting errors in measuring fiber shape.

## 7864A-05, Session 1

### Depth upsampling method using the confidence map for a fusion of a high resolution color sensor and low resolution time-of-flight depth sensor

K. Bae, K. Kyung, T. Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

This paper proposes a depth upsampling method using the confidence map for a fusion of a high resolution color sensor and low resolution time-of-flight depth sensor. The confidence map represents the accuracy of depth depending on the reflectance of a measured object and is estimated with amplitude, offset, and reconstructed error of a received signal. The proposed method suppresses the depth artifacts that are caused by difference between low and high reflective materials on an object at a distance. Although the surface of an object is located at the same distance, the reflectance of small regions within the surface depends on constituent materials. Weighted filter

generated by confidence map is added to modified noise-aware filter for depth upsampling (NAFDU) that is proposed by Derek et al., and is adaptively selected. The proposed method consists of followings; the normalization, the reconstruction, the confidence map estimation, and the modified noise-aware filtering. In the normalization, amplitudes and offsets of received signals are calculated and received signals are normalized by those. The normalized signals are denoised, and then the phase shifts are measured between transmitted and received signals. In the reconstruction, received signals are reconstructed using only the values of phase shifts and the reconstruction errors are calculated. The confidence map is estimated with amplitudes, offsets, and reconstruction errors. The coefficients of a modified noise-aware filter are adaptively selected by referring to the confidence map. The proposed method shows the enhanced results of removing depth artifacts in the experiments.

## 7864A-06, Session 2

### Instrument for 3D characterization of autostereoscopic displays

J. PrevotEAU, Univ. de Reims Champagne-Ardenne (France) and 3DTV Solutions (France); S. Chalencon-Piotin, Univ. de Reims Champagne-Ardenne (France); D. G. Debons, 3DTV Solutions (France); L. Lucas, Y. Remion, Univ. de Reims Champagne-Ardenne (France) and 3DTV Solutions (France)

We now have numerous autostereoscopic displays, and it is mandatory to characterize them because it will allow to optimize their performances and to make efficient comparison between them. Therefore we need standards so we have to be able to quantify the quality of the viewer's perception. The purpose of the present paper is twofold; we first present a new instrument of characterization of the 3D perception on a given autostereoscopic displays; then we propose a new way to realize an experimental protocol allowing to get a full characterization. This instrument will allow us to compare efficiently the different autostereoscopic displays but it will also validate practically the adequacy between the shooting and rendering geometries. In this aim, we are going to match a perceived scene with the virtual scene. It is hardly possible to determinate the scene perceived by a viewer placed in front of an autostereoscopic display. Indeed if it may be executable on the pop-out, it is impossible on the depth effect because the depth of the virtual scene is set behind the screen. Therefore, we will have to use an optical illusion based on the deflection of light by a mirror to know the position which the viewer perceives some points of the virtual scene on an autostereoscopic display.

## 7864A-07, Session 2

### Accurate stereo matching based on multiband imaging

M. Doi, A. Minami, Osaka Electro-Communication Univ. (Japan); S. Tominaga, Chiba Univ. (Japan)

The accuracy of stereo matching depends on precise detection of corresponding points in a pair of stereo images by template matching. A multiband imaging system captures more than three channels in a visible range. The multiband imaging technique is useful for improving the accuracy of the stereo matching. In this paper, we propose an imaging system and an algorithm for stereo matching based on multiband images. The imaging system is composed of a liquid-crystal tunable filter and a high sensitive monochrome camera. In our modified SSDA algorithm, the similarity is calculated for each band in the descending order of the variance of band image intensity in template. Since the temporary similarity at non-candidate points exceeds threshold quickly in band with large variance, the processing time is shortened by quick interruption of the calculation at the point. Experimental results show that multiband stereo matching is accurate compared with RGB stereo matching. A sheet of color texture patches with small color differences was used as a measurement target. The rate of detection of correct corresponding points was 98.4% by multiband stereo matching, while the rate was 34.7% by RGB stereo

matching. Moreover, use of the modified SSDA reduced 17% of the CPU time.

## 7864A-24, Session 2

### Flash trajectory imaging of target 3D motion

X. Wang, Y. Zhou, S. Fan, J. He, Y. Liu, Institute of Semiconductors (China)

Determination of 3D motion parameters and imaging of moving targets have been research hotspots in recent years. In astronomy, remote sensing, tracking targets and estimating their motion parameters are very important. Traditional trajectory prediction technique by image sequences and corresponding imaging processing are complicated, especially for targets in complex background. In order to solve the above problems, we present a flash trajectory imaging (FTI) technique for target 3D motion. This technique uses a pulsed laser to illuminate targets and a camera with a microchannel plate to take images. The microchannel plate acts as both an amplifier and a shutter. In the FTI, time delay integration and time slicing are used. For moving targets, the mode of time delay integration increases information of one single frame image so that one can directly gain the moving trajectory. Time slicing gives the range of targets and realizes silhouette detection which can directly extract targets from complex background. Therefore, the complexity of image processing decreases. By two successive frames including target trajectory, the motion parameters and flight attitude can be given. Since the minimum gating rate per frame is one and the maximum can reach several millions, the FTI can effectively image low or high speed moving targets and also give their 3D motion parameters. In addition, the manner of time slicing makes the FTI has properties such as high suppression of backscatter from fog and other obscurants, high signal-to-noise ratio, and long detection range. For the FTI, we have researched it in experiments and also studied the algorithm about it. Our research demonstrates that the FTI is an effective approach to determine the motion parameters of 3D motion targets and image moving targets.

## 7864A-10, Session 3

### The ASTM E57 file format for 3D imaging data exchange

D. Huber, Carnegie Mellon Univ. (United States)

There is currently no general-purpose, open standard for storing data produced by three dimensional (3D) imaging systems, such as laser scanners. As a result, producers and consumers of such data rely on proprietary or ad-hoc formats to store and exchange data. There is a critical need in the 3D imaging industry for open standards that promote data interoperability among 3D imaging hardware and software systems. For the past two years, a group of volunteers has been working within the ASTM E57 Committee on 3D Imaging Systems to develop an open standard for 3D imaging system data exchange to meet this need. The E57 File Format for 3D Imaging Data Exchange (E57 format hereafter) is capable of storing point cloud data from laser scanners and other 3D imaging systems, as well as associated 2D imagery and core meta-data. This paper will describe the motivation, requirements, design, and implementation of the E57 format, and will highlight the technical concepts developed for the standard. We will also compare the format with other proprietary or special purpose 3D imaging formats, such as the LAS format, and we will discuss and analyze the open source library implementation designed to read, write, and validate E57 files.

## 7864A-11, Session 3

## The impact of different alignment strategies on the overall performance of a white light scanner according to the uncertainty especial according to sphere spacing error specified in VDI 2634

E. Klaas, Breuckmann GmbH (Germany)

This paper is about accuracy of optical white light or so called "topometric" scanners. In almost any application of such scanners it is necessary to put together scans from different directions: from a couple of scans to a couple of hundred scans. Accuracy can be usually well described for a single scan, whereas accuracy for those assembled data sets is harder to estimate and specify: it depends on much more parameters as well as on the alignment strategy being used. This paper will describe different alignment strategies including using robots and tracking systems, targets as well as using best fitting methods. The impact of these methods on the resulting overall accuracy is described and demonstrated using real live examples. Also different methods of achieving these accuracy numbers are being presented including guidelines such as VDI 2634. It will briefly touch on the basic principle of white light scanning to understand the potential but also limitations of these technique.

This should be a useful guideline for engineers or quality managers who want to establish or learn about new scanning technologies with special attention to accuracy issues.

## 7864A-12, Session 3

## Simulation-based determination of local probing uncertainty for fringe projection measurements

J. Weickmann, A. A. Weckenmann, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany)

Fringe projection sensors gain in importance in manufacturing quality control due to their multiple advantages. In order to adapt the measurement strategy to a specific inspection task, both a suitable sensor and the necessary measurements have to be chosen, so that the complete workpiece shape is recorded with a tolerance-compatible measurement uncertainty, according to DIN EN ISO 14253-1. Thus a reliable forecast of the measurement uncertainty is crucial for an effective inspection-planning procedure. There are multiple influences, whose impacts on the measurement result vary dependent on the position of the measured point. So the local measurement uncertainty at each measured point - that means the 'probing uncertainty' - is individual. Today, the probing uncertainty cannot be forecasted locally. Thus, the expected uncertainty cannot be taken into account, when an inspection is planned. This paper shows a simulation-based approach to eliminate this shortfall. Firstly, a definition for probing uncertainty is given. Then the model for the simulation of planned fringe projection measurements - including a GUM-conformant forecast for the local probing uncertainty - is described. This simulation is then implemented into a prototype of an assistance system that supports the inspection planner when setting up the measurement strategy. Finally a method for the experimental verification of the local probing uncertainty is introduced and the simulation results are verified.

## 7864A-13, Session 3

## Low cost characterization of TOF range sensors resolution

G. Guidi, M. Russo, G. Magrassi, M. Bordegoni, Politecnico di Milano (Italy)

The purpose of this paper is to define a low cost approach for estimating the resolution of a TOF laser scanner, porting to the 3D

environment concepts already accepted and implemented in 2D ISO standards for electronic imaging systems such as ISO 12233 and 16067.

The methodology presented is based on the production of images similar to those generate by 2D imaging of the cited ISO targets, from the laser scanning of specifically developed 3D targets, made by windows of varying size on the horizontal plane (xy) and abrupt jumps on the laser beam direction (z). Each range map is transformed in a B/W image by coding z in grey levels. On each image the cited ISO methods can be applied. These can be divided in two categories: one based on the estimation of resolution from the direct evaluation of the image generated by geometric features progressively closer each other, and the other based on the frequency domain analysis of slanted edges the evaluation of the associated Modulation Transfer Function (MTF). A few different TOF range cameras based on light pulses and on phase detection of continuous modulated light have been considered. The experimental results are finally presented and discussed.

## 7864A-14, Session 3

## Introducing the depth transfer curve for 3D capture system characterization

K. Atanassov, V. Ramachandra, S. R. Goma, Qualcomm Inc. (United States)

Evaluating depth characteristics subjectively is a very difficult task since the intended or side effects from image fusion (depth interpretation) by the brain are not immediately perceived by the observer. Objective evaluation of 3D camera depth characteristics is a useful tool that can be used for "black box" characterization of 3D cameras. In this paper we propose a methodology to evaluate 3D cameras depth capture capabilities using a special test chart - with 3D features- that is practical and contains necessary structure to extract important 3D depth statistics and we present a processing algorithm that extracts the relevant depth information.

## 7864A-16, Session 4

## Assessment of the quality of as-is building information models generated from point clouds using deviation analysis

E. B. Anil, P. Tang, B. Akinci, D. Huber, Carnegie Mellon Univ. (United States)

Three dimensional (3D) imaging sensors, such as laser scanners, are being used to create Building Information Models (BIMs) of the as-is conditions of facilities. Quality assurance (QA) needs to be conducted to ensure that the models accurately depict the as-is conditions. We propose a new approach for QA that analyzes patterns in the raw 3D data and compares the 3D data with the as-is BIM geometry to identify potential errors in the model. This "deviation analysis" approach to QA enables users to analyze the regions with significant differences between the 3D data and the reconstructed model or between the 3D data of individual laser scans. This method can help identify the source of the error and does not require additional physical access to the facility. To show the approach's potential effectiveness, we conducted case studies of several professionally conducted as-is BIM projects. We compared the deviation analysis method to an alternative method - the physical measurement approach - in terms of coverage of the environment, types of errors detected, and time requirements. We also conducted a survey and evaluation of commercial software with relevant capabilities and identified technology gaps that need to be addressed to fully exploit the deviation analysis approach.

## 7864A-17, Session 4

**Content-based depth estimation in focused plenoptic camera**

S. R. Goma, K. Atanassov, Qualcomm Inc. (United States); T. G. Georgiev, Adobe Systems Inc. (United States)

The plenoptic camera has been used for generating novel views and refocusing. However the original purpose of this, as well as the improved focused plenoptic camera, has been measuring depth. Depth estimation in focused plenoptic camera is a critical step for most applications of this technology and poses interesting challenges, as this estimation is content based. We present an iterative algorithm, content adaptive, that exploits the redundancy found in images captured with focused plenoptic camera. Our algorithm determines for each point its depth along with a measure of reliability allowing subsequent enhancements of spatial resolution of the depth map. Also, our capture is corrected for the distortion between images of central and non-central microlenses. We remark that the spatial resolution of the recovered depth corresponds to discrete values of depth in the captured scene to which we refer as slices. More, each slice has a different depth and will allow extraction of different spatial resolutions of depth, depending on the scene content being present in that slice along with occluding areas. Interestingly, as focused plenoptic camera is not theoretically limited in spatial resolution, we show that the recovered spatial resolution is depth related, and as such, rendering of a focused plenoptic image is content dependant.

## 7864A-18, Session 5

**Measurement of micro gears: comparison of optical, tactile-optical, and CT-measurements**

U. Neuschaefer-Rube, M. Bartscher, F. Härtig, M. Neukamm, Physikalisch-Technische Bundesanstalt (Germany); J. Goebbels, Bundesanstalt für Materialforschung und -prüfung (Germany)

Micro gears are applied in high quantity in a lot of applications. Therefore, precise measurements are of growing importance to ensure their quality. The presentation describes the measurement of gears of a micro planetary gear set with an optical sensor, a tactile-optical probe and computed tomography (CT).

An imaging sensor based on focus variation is used for the optical measurements. The tactile-optical measurements were carried out with a so called fiber probe. This sensor applies image processing to determine the position of the tactile probing element. For the measurements single point probing was used. Due to limited accessibility at some gears not all regions can be measured by the optical sensor and the tactile-optical probe. In contrast to this, with CT always the whole part can be measured with high point density.

All the used sensors deliver measurement data in Cartesian coordinates. A challenge is to transfer the data in coordinates in which gear parameters are defined. For this, special attention must be paid to the center point of the gear and to the orientation of the teeth.

The comparison between data of different measurements was carried out successfully. The deviations between tactile-optical and the CT data are only a few micrometers.

## 7864A-19, Session 5

**Method for measuring stereo camera depth accuracy based on stereoscopic vision**

M. Kytö, M. Nuutinen, P. T. Oittinen, Aalto Univ. School of Science and Technology (Finland)

We present a method to evaluate stereo camera depth accuracy in human centered applications. It enables the comparison between stereo camera depth resolution and human depth resolution.

Our method uses a multilevel test target which can be easily assembled and used in various studies. Binocular disparity enables humans to perceive relative depths accurately, making a multilevel test target applicable for evaluating the stereo camera depth accuracy when the accuracy requirements come from stereoscopic vision.

The method was validated with a stereo camera built of two SLRs. The depth resolution of the SLRs was better than normal stereo acuity at all measured distances ranging from 0.7 m to 5.8 m. The method was used to evaluate the accuracy of a lower quality stereo camera. Two parameters, focal length and baseline, were varied. Focal length had a larger effect on stereo camera's depth accuracy than baseline. The tests showed that normal stereo acuity was achieved only using a tele lens.

However, a user's depth resolution in a video see-through system differs from naked eye viewing. The same test target can be used to evaluate this by mixing the levels randomly and asking users to sort the levels according to their depth. The comparison can be done by calculating correlations between the measured order, perceived order and the actual order of the levels.

## 7864A-20, Session 5

**Best practices for the 3D documentation of the Grotta Dei Cervi of Porto Badisco, Italy**

J. A. Beraldin, M. Picard, National Research Council Canada (Canada); V. Valzano, A. Bandiera, Univ. del Salento (Italy); F. Negro, CASPUR (Italy)

final/partial results:

- Creation of a 3D polygonal model of the central 300-m corridor.
- Texture mapping of most of the cave especially where some pictographs show deterioration.
- Creation of transversal and longitudinal cross-sections of the cave with a 1-mm resolution.
- Creation of high resolution realistic renderings (up to 17.000 x 8.000) using colour and shape with a 1-mm resolution.
- Video animation from 3D digital polygonal model (1080p).
- Creation of a CD-ROM for public consumption.
- Creation of a web site.
- Creation of a 3D polygonal model for stereoscopic displays.
- Extraction of information from model in order to assist restoration/preservation work.

## 7864A-21, Session 6

**NPL freeform artefact for verification of non-contact measuring systems**

M. McCarthy, National Physical Lab. (United Kingdom)

New machining techniques make it possible to manufacture ranges of advanced freeform components and with appropriate metrology, higher precisions (sub-micrometre) can potentially be achieved. However, such rapid developments are limited due to a lack of measurement confidence and suitable measurement traceability. Tactile co-ordinate measuring machines (CMMs) are able to scan complex surfaces, but this process is often very time consuming. In contrast, many portable optical-based non-contact co-ordinate measuring system are now commercially available which can capture vast quantities of point-cloud data in a fraction of the time. Verification of these non-contact systems is complex and useful guides such as the VDI/VDE 2634 series demonstrate capability while employs test artefacts such as spheres and planes. The guide does not extend to fully address performance verification when freeform surfaces are to be measured.

To assist industry, NPL has developed a range of freeform-based reference artefacts, allowing the performance of non-contact systems, such as fringe projector, laser scanners and other similar optically based systems to be verified against a set of known surface conditions.

The purpose of the artefacts is to demonstrate the capability of selected non-contact technologies and certain systems, to measure specific forms and surface conditions, rather than to be a universal standard.

The 'NPL Freeform artefacts' have been designed, manufactured and then calibrated at NPL using an ultra high accuracy tactile based CMM. A chosen artefact, having a nominally foot print of 120 mm x 120 mm has subsequently been used to verify the performance of a number of commercial freeform optical measuring systems. This paper will describe the generic design of the 'NPL FreeForm artefact', the calibration procedure adopted and discuss a measurement inter-comparison involving a number of non-contact industrial measuring systems.

## 7864A-22, Session 6

### **Proposed NRC-CNRC portable target case for short-range triangulation-based 3D imaging systems characterization**

B. Carrier, D. K. MacKinnon, L. Cournoyer, J. A. Beraldin,  
National Research Council Canada (Canada)

The National Research Council of Canada (NRCC) is currently evaluating and designing artifacts and methods to completely characterize 3-D imaging systems. We have gathered a set of artifacts to form a low-cost portable case and provide a clearly-defined set of procedures for generating characteristic values using these artifacts. In its current version, this case is specifically designed for the characterization of short-range (standoff distance of 1 centimeter to 3 meters) triangulation-based 3-D imaging systems. The case is known as the "NRC-CNRC Portable Target Case for Short-Range Triangulation-based 3-D Imaging Systems" (NRC-PTC). The artifacts in the case have been carefully chosen for their geometric, thermal, and optical properties. A set of characterization procedures are provided with these artifacts based on procedures either already in use or are based on knowledge acquired from various tests carried out by the NRCC. Geometric dimensioning and tolerancing (GD&T), a well-known language in the industrial field, was used to define the set of tests. The following parameters of a system are characterized: dimensional properties, form properties, orientation properties, localization properties, profile properties, repeatability, intermediate precision, and reproducibility. A number of tests were performed in a special dimensional metrology laboratory to validate the capability of the NRC-PTC. The NRC-PTC will soon be subjected to reproducibility testing using an intercomparison evaluation to validate its use in different laboratories.

# Conference 7864B: 3D Image Processing (3DIP) and Applications II

Wednesday-Thursday 26-27 January 2011

Part of Proceedings of SPIE Vol. 7864B 3D Image Processing (3DIP) and Applications II

## 7864B-25, Session 8

### 3D shape descriptors for face segmentation and fiducial points detection: an anatomical-based analysis

A. E. Salazar Jiménez, Univ. Nacional de Colombia Sede Medellín (Colombia); A. Cerón, Univ. Militar Nueva Granada (Colombia); F. A. Prieto Ortiz, Univ. Nacional de Colombia (Colombia)

The behavior of nine 3D shape descriptors which were computed on the surface of 3D face models, is studied. The set of descriptors includes six curvature-based ones, SPIN images, Folded SPIN Images, and Finger prints. Instead of defining clusters of vertices based on the value of a given primitive surface feature, a face template composed by 28 anatomical regions, is used to segment the models and to extract the location of different landmarks and fiducial points. Vertices are grouped by: region, region boundaries, and subsampled versions of them. The aim of this study is to analyze the discriminant capacity of each descriptor to characterize regions and to identify key points on the facial surface. The experiment includes testing with data from neutral faces and faces showing expressions. Also, in order to see the usefulness of the bending-invariant canonical form (BICF) to handle variations due to facial expressions, the descriptors are computed directly from the surface and also from its BICF. In the results: the values, distributions, and relevance indexes of each set of vertices, were analyzed.

## 7864B-28, Session 9

### Deformable shape retrieval using bag-of-feature techniques

H. Tabia, M. Daoudi, J. Vandeborre, TELECOM Lille 1 (France); O. Colot, Univ. des Sciences et Technologies de Lille (France)

We present a novel method for 3D-shape matching using Bag-of-Feature techniques (BoF). The method starts by selecting and then describing a set of points from the 3D-object. Such descriptors have the advantage of being invariant to different transformations that a shape can undergo. Based on vector quantization, we cluster those descriptors to form a shape vocabulary. Then, each point selected in the object is associated to a cluster (word) in that vocabulary. Finally, a BoF histogram counting the occurrences of every word is computed. These results clearly demonstrate that the method is robust to non-rigid and deformable shapes, in which the class of transformations may be very wide due to the capability of such shapes to bend and assume different forms.

## 7864B-32, Session 10

### Automatic generation of 3D building models from orthogonal building footprint

K. Sugihara, Gifu Keizai Univ. (Japan); X. Zhou, Nagoya Bunri Univ. (Japan); T. Murase, Chukyo Gakuin Univ. (Japan)

Based on building polygons or building footprints on digital maps, we propose a GIS and CG integrated system that automatically generates 3-D building models. A 3-D urban model is an important information infrastructure that can be utilized in several fields. However, enormous time and labor has to be consumed to create these 3-D models, using 3D modeling software such as SketchUp. In order to automate the laborious steps, we proposed the GIS and CG integrated system that automatically generates 3-D building models from building polygons

on a digital map. Most building polygons' edges meet at right angles (orthogonal polygon). A complicated orthogonal polygon can be partitioned into a set of rectangles. The integrated system partitions orthogonal building polygons into a set of rectangles and places rectangular roofs and box-shaped building bodies on these rectangles. In order to partition an orthogonal polygon, we proposed a useful polygon expression (RL expression) and a partitioning scheme that is used in deciding from which vertex a dividing line (DL) is drawn. In this paper, we propose a new scheme for partitioning building polygons and for creating a complicated shape of building models based on orthogonal building polygons.

## 7864B-35, Session 11

### Feature vertices for 3D synchronization using Euclidean minimum spanning tree

N. Tournier, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France) and Stratégies S.A. (France); W. Puech, G. Subsol, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); J. Pedeboy, Stratégies S.A. (France)

Synchronization in 3D data hiding is one of the main problems. We need to know where we can embed information, and be able to find this space after the insertion of the message. Various algorithms propose their synchronization techniques by triangle or vertex path in a 3D mesh.

In this paper, we proposed a new synchronization technique based on Euclidean minimum spanning tree computing (EMST) and the analysis of the displacement of the vertices without moving the connections in the tree. Based on the analysis of the vertices, we select the most mobile and synchronize these areas by computing a new EMST on them, that we call "robust EMST".

Then, we analyze the robustness of the technique, i.e. the stability of the most mobile vertices selection; and demonstrate the consistence of the criterion selection with the vertex displacement.

## 7864B-36, Session 11

### Probability distributions from Riemannian geometry, generalized hybrid Monte Carlo sampling, and path integrals

E. Paquet, National Research Council Canada (Canada); H. L. Viktor, Univ. of Ottawa (Canada)

When considering probabilistic pattern recognition methods, especially methods based on Bayesian analysis, the probabilistic distribution is of the utmost importance. However, despite the fact that the geometry associated with the probability distribution constitutes essential background information, it is often not ascertained. This paper discusses how the standard Euclidean geometry should be generalized to the Riemannian geometry when a curvature is observed in the distribution. To this end, the probability distribution is defined for curved geometry. In order to calculate the probability distribution, a Lagrangian and a Hamiltonian constructed from curvature invariants are associated with the Riemannian geometry and a generalized hybrid Monte Carlo sampling is introduced. Finally, we consider the calculation of the probability distribution and the expectation in Riemannian space with path integrals, which allows a direct extension of the concept of probability to curved space.



### 7864B-37, Session 11

#### **Compression method by using the motion estimation of residual image transformed from elemental image array in three-dimensional integral imaging**

C. H. Yoo, J. Lee, H. Kang, E. Kim, Kwangwoon Univ. (Korea, Republic of)

In this paper, we proposed a highly enhanced compression scheme of Integral Imaging (InIm) by use of sub-images (SIs) removing the Motion Vector (MV) of residual image array transformed from Sub-Image Array (SIA). In the pickup process, SIA is generated from EIA after the object through the virtual pinhole array is recorded as Elemental Image Array (EIA). It provides enhanced compression efficiency by improving the similarity among SIs. In the proposed method, a segmented area, which is a macroblock, in the reference SI is matched on current SIs applying to MSE. MVs occurred among SIs might result in an additional increase for data compression. Accordingly, the computed motion estimation from the block-matching is saved as MV and all objects in each current SI are shifted to the object position of the reference SI to compensate their MV based on the motion estimation. We can enhance the similarity of SIs removing MV, so that an improvement of compression efficiency of the SIA could be obtained. In addition, the video compression scheme such as MPEG-4 can be applied to data reduction of the consecutive frames. The proposed algorithm outperforms the baseline JPEG and the conventional EIA compression scheme applied to InIm and is compared with simulations produced using these schemes.

### 7864B-38, Session 11

#### **Detection of the aortic intimal tears by using 3D digital topology**

C. Lohou, Instituts Universitaires de Technologie (France); B. Miguel, CHU Clermont-Ferrand (France)

Several works about aorta segmentation have already been performed: most of them concern aortic outer walls and are mainly proposed in the case of abdominal aortic aneurysms and are usually based on deformable techniques.

Aortic dissections is a real problem of Public Health, and may quickly lead to death. Aortic dissections are due to presence of tears inside lumens (or holes in the intimal tissue). These tears are difficult to detect because they do not correspond to a filled organ to segment; these tears are usually visually retrieved by radiologists by examining gray level variation on successive planes, which is a very difficult and error-prone task.

Our purpose is to detect these intimal tears to help cardiac surgeons to establish a diagnosis: the visualization of intimal tears could lead to choose a size of endoprosthesis, and the add of this data could also help surgeons during the intervention.

At this aim, we use Aktouf et al.'s holes filling algorithms proposed in the field of digital topology. This algorithm permits the filling of holes on a 3D object by using topological notions — holes are intimal tears in our images.

We also think that this approach would gained to be known to specialists of other diseases.

### 7864B-40, Session 12

#### **Data processing path from multimodal 3D measurement to realistic virtual model**

R. Sitnik, J. F. Krzeslowski, G. Maczkowski, Warsaw Univ. of Technology (Poland)

A set of calculation methods has been developed and tested to provide means of creating virtual copies of three dimensional (3D) historical objects with minimal user input. We present a step by step data processing path along with algorithm description required to reconstruct a realistic 3D model of a culturally significant object. The important feature for archiving historical objects is the ability to include both information about its shape and texture, allowing visualization using arbitrary conditions of illumination. Data samples used as input for the processing method chain were collected using an integrated device consisting of shape, multispectral color and simplified BRDF measurements. To confirm the usability of presented methods, it has been tested by example of real life object - statue of an ancient Greek goddess Kybele. Additional visualization methods have also been examined to render a realistic virtual representation satisfying intrinsic surface properties of the investigated specimen.

### 7864B-41, Session 12

#### **Preliminary study of statistical pattern recognition-based coin counterfeit detection by means of high resolution 3D scanners**

M. Leich, S. Kiltz, C. Kraetzer, J. Dittmann, C. Vielhauer, Otto-von-Guericke-Univ. Magdeburg (Germany)

According to the European Commission around 200,000 counterfeit Euro coins are removed from circulation every year. While there exist approaches to automatically detect these coins, satisfying error rates are usually only reached for low quality forgeries, so-called "local classes". Minted forgeries of very high quality ("common classes") pose a problem for these methods as well as for trained humans.

This paper presents a first approach for statistical analysis of coins based on high resolution 3D data acquired with a chromatic white light sensor. The goal of this analysis is to determine whether two coins are of common origin. The test set for these first and new investigations will consist of about 50 coins from not more than five different sources to assess the overall potential. The analysis is based on the assumption that, apart from markings caused by wear such as scratches and residue consisting of grease and dust, coins from equal origin have a more similar height field than coins from different mints. The influence of wear markings is discussed and an approach for eliminating this influence is outlined.

### 7864B-42, Session 12

#### **3D digitization of metallic specular surfaces using scanning from heating approach**

A. Bajard, O. Aubreton, Univ. de Bourgogne (France)

Because of the difficulty of dealing with specularity of several surfaces, few methods have been proposed to measure three-dimensional shapes of specular metallic objects. In this paper we present an application on this kind of material of an approach called "Scanning From Heating". This approach has been developed initially for 3D reconstruction of transparent objects. This article presents an application of the working principle of SFH method on material with high thermal conductivity.

7864B-43, Session 12

**3D image processing architecture for  
camera phones**

K. Atanassov, V. Ramachandra, M. Aleksic, S. R. Goma,  
Qualcomm Inc. (United States)

Recent developments in 3D display technology have created a demand for consumer generated 3D content. Camera phones are ubiquitous and as such, one of the first choices for user generated content. We present a solution to a full 3D datapath implementation for camera phones, providing practical arguments to 3D technology issues such as camera positioning, disparity control rationale, and screen geometry dependency. Implementing successfully 3D capture functionality on phone cameras requires algorithms that fit within the processing capabilities of the device. Various constraints like sensor position tolerances, module to module variation, post-processing, 3D video resolution and frame rate, should be carefully considered for their influence on 3D experience. Migrating user functionality from the 2D usage model, such as zoom and pan (on capture and display) requires additional consideration. It is also important that the user interaction with both the capture and the display device is both intuitive and efficient. Finally, both the processing power of the device and the practicality of the concept need to be taken into account in the calibration and processing methodology.

7864B-44, Session 12

**Return detection for outdoor active  
triangulation**

D. M. Ilstrup, R. Manduchi, Univ. of California, Santa Cruz  
(United States)

No abstract available.

# Conference 7864C: The Engineering Reality of Virtual Reality 2011

Tuesday 25 January 2011

Part of Proceedings of SPIE Vol. 7864C The Engineering Reality of Virtual Reality 2011

## 7864C-46, Session 13

### Acquisition of stereo panoramas for display in VR environments

R. A. Ainsworth, Ainsworth & Partners, Inc. (United States); D. J. Sandin, Univ. of California, San Diego (United States) and Univ. of Illinois at Chicago (United States); J. P. Schulze, A. Prudhomme, T. A. DeFanti, Univ. of California, San Diego (United States)

Virtual reality systems are an excellent environment for stereo panorama displays. The acquisition and display methods described here combine high-resolution photography with surround vision and full stereo view in an immersive environment. This combination provides photographic stereo-panoramas for a variety of VR displays, including the StarCAVE and NextCAVE.

The zero parallax point used in conventional panorama photography is also the center of horizontal and vertical rotation when creating photographs for stereo panoramas. The two photographically created images are displayed on a cylinder or a sphere. The radius from the viewer to the image is set at approximately 20 feet, or at the object of major interest.

A full stereo view is presented in all directions. The two spherical images are displaced horizontally by the interocular distance, as seen from the viewer's perspective. This presents correct stereo separation in whatever direction the viewer is looking, even up and down. Objects at infinity will move with the viewer, contributing to an immersive experience.

Stereo panoramas created with this acquisition and display technique can be applied without modification to a large array of VR devices having different screen arrangements and different VR libraries.

## 7864C-47, Session 13

### Low cost heads-up virtual reality (HUVR) with optical tracking and haptic feedback

T. Margolis, T. A. DeFanti, G. Dawe, A. Prudhomme, J. P. Schulze, Univ. of California, San Diego (United States)

Researchers at the University of California, San Diego, have created a new, relatively low-cost augmented reality system that enables users to touch the visual environment they are immersed in.

The Heads-Up Virtual Reality device (HUVR) couples a consumer 3D HD flatscreen TV with a half-silvered mirror to project any graphic image onto the user's hands and into the space surrounding them. With his or her head position optically tracked to generate the correct perspective view, the user maneuvers a force-feedback (haptic) device to interact with the generated image, literally 'touching' the object's angles and contours as if it was a tangible physical object.

HUVR can be used for training and education in structural and mechanical engineering, archaeology and medicine as well as other tasks that require hand-eye coordination. One of the most unique characteristics of HUVR is that a user can place their hands inside of the virtual environment without occluding the 3D image. Built using open-source software and consumer level hardware, HUVR offers users a tactile experience in an immersive environment that is both functional and affordable.

## 7864C-48, Session 13

### An integrated pipeline to create and experience compelling scenarios in virtual reality

C. Cruz-Neira, D. Reiners, J. Springer, Univ. of Louisiana at Lafayette (United States)

We present in this paper our research on designing a software pipeline that enables us to create compelling scenarios with a fair degree of visual and interaction complexity in a semi-automated way. Specifically, we are targeting "drivable" urban scenarios, ranging from large cities to sparsely populated rural villages that incorporate both static components, such as houses, trees, telephone poles, etc. and dynamic components such as people, vehicles, as well as events, such as explosions, sudden noises, etc.

Our pipeline has four basic components: 1) An environment designer, where users sketch the overall layout of the scenario and an automated method constructs the 3D environment from the information in the sketch; 2) A scenario editor used for authoring the complete scenario, incorporate the dynamic elements and events, tweak the automatically generated environment, define the execution conditions of the scenario (type of device, interactions, data inputs, etc), and set any data gathering that may be necessary during the execution of the scenario; 3) A run-time environment for the different virtual reality systems providing users with the interactive experience designed through the designer and the editor; 4) A bi-directional monitoring system that allows us to capture and modify information from the virtual environment.

The main use of this pipeline is for the rapid development of scenarios for human factors studies, however, it can be applied in a much more general context.

## 7864C-49, Session 14

### Whose point-of-view is it anyway?

G. P. Garvey, Quinnipiac Univ. (United States)

Shared virtual worlds such as Second Life privilege a single point-of-view, namely that of the user. When logged into Second Life a user sees the virtual world from a default viewpoint, which is from slightly above and behind the user's avatar (the user's alter ego 'in-world.'). This point-of-view is as if the user were viewing his or her avatar using a camera floating a few feet behind it. In fact it is possible to set the view to as if you were seeing the world through the eyes of your avatar or you can even move the camera completely independent of your avatar.

A change in point-of-view means more than just a different camera point-of-view. The practice of using multiple avatars requires a transformation of identity and personality. When a user 'enacts' the identity of a particular avatar, their 'real' personality is masked by the assumed personality.

The technology of virtual worlds permits both a change of point-of-view and also facilitates a change in identity. Does this cause any psychological distress? Or is the ability to be someone else and see a world (a game, a virtual world) through a different set of eyes somehow liberating and even beneficial?

## 7864C-50, Session 14

### **Biocybrid systems and the reengineering of life**

D. Domingues, Univ. de Brasília (Brazil) and LART-Lab. Research in Art and TechnoScience (Brazil); A. Rocha, C. Hamdan, L. Augusto, Univ. de Brasília (Brazil)

Our collaborative researches in Art and TechnoScience are based on biocybrid systems (bio+cybrid+hybrid) engineering immersive multimodal technologies for Caves, and wearable devices to expand perceptual experiences and the sense of presence in Virtual Reality and Augmented Reality, locative, mobile and transparent interfaces in urban mixed life. We explore the artists' creativity close to the inventivity of scientists and mutual capacity for the generation of biocybrid systems which implies in the human existence, being co-located in the continuum zone between of body and flesh - cyberspace and data - and the hybrid properties of physical world. Anthropological issues and the sense of presence being adequated by the technological apparatus for human life in Software Art practices require the interface design for intertwined relationships body/environment/nets reaffirming the ubiquitous and mobile condition in physical world and cyberspace. Enactive interfaces and the changes and challenges of the reengineering of life are discussed in three artworks. Ecological Biocybrid Emotions in Paracosmos offer the immersion in our Cave to receiving signals from remote landscapes from the distant South Mato Grosso, Pantanal zone, Brazil, by mixing the life of biological community of frogs and snakes natural behavior to biological signals of a tracked body and its physiology transduced into a sophisticated system of biofeedback and mutations in a virtual datalandscape. Mutual behaviors and the systematic structuring of the relations between sensory patterns ultimately derived from manufactured sensor technologies and those given by locomotion, heat, heart beats and breathing, and electrical potentials of electrooculogram, by taking the internal paradigms and the parameters should result in those contingencies leading to the externalization of a percept. Opened Body Connection (performance) is a biocybrid system in augmented reality expanded to mixed reality in a bodily continuum of flesh - cyberdata and physical space. During a ritual, in a public event, a tattoo artist inscribed on the back of the performer the computation language of the code a wing shape. The tattoo became a mixed reality system, only read in computer vision and during two hours, other printed tags on the body provided three-dimensional animations traveled through the body transmitted on line. Animated wings respond to the artist dream of flying. Expanded interactions and mixed landscape generated the 14 Bis, biocybrid system, by exploring mobile and locative interface in mixed reality which allow the computer vision in the sky of the Brazilian capital of the little plane invented by the pioneer Santos Dumont. The creative practice in mixed reality placing the tag code in the geolocated Global Positioning System result of the modelled historical plane in real scale apparition in the city sky. The absence/presence state is homologated through the computer vision of a mobile cellcam. In a post biological extrusion of human vision, the act of seeing is shared with the satellite eye in the sky and the handled eye in the mobile device, by expanding the neuropsychophysiology of human perception. Synthetic senses and the reengineering of life offer a different scenario for human narrative in the theater of life.

## 7864C-51, Session 14

### **Twisting the sense of space in immersive trees**

M. Song, Simon Fraser Univ. (Canada); S. J. Barnes, Univ. of British Columbia (Canada); D. Gromala, T. Fox, Simon Fraser Univ. (Canada); D. Barnes, Independent Artist (Canada)

What consequences do we pay for sitting in bed and mindlessly surfing the web to read about what people ate for breakfast, buying yet again another pair of jeans or watching the latest TV shows? How are our online lives affecting our real-world environments and the ecological systems that we live in but forget as we are so engrossed in our digital lives? We seek to explore these relationships and the effects of our actions through an information-rich, inhabitable exhibition that paradoxically blurs multiple aspects of the virtual and the physical.

A 16-foot high Banyan tree - a metaphor for ancient and contemporary connectedness - is constructed from multiple, inter-nested layers of a projection material that is alternately translucent and opaque. Through exploration and habitation of a penetrable Banyan tree, interactors experience the ecological and environmental costs of internet consumption through stereoscopic projections that vary in depth-of-field and twist around the core of the tree. The interactors are the source of pleasure and the consumed energy that moves through the inter-connected branches and roots of the Banyan tree - passing through xylem, phloem and pith.

## 7864C-52, Session 14

### **Productive confusions: learning from simulations of pandemic virus outbreaks in Second Life**

M. Gardenas, Univ. of California, San Diego (United States); L. S. Greci, Univ. of California, San Diego (United States) and Veterans Administration San Diego (United States); S. Hurst, K. Garman, H. Hoffman, R. Huang, M. Gates, K. Kho, E. Mehrmand, T. Porteous, Univ. of California, San Diego (United States); A. Calvitti, Univ. of California, San Diego (United States) and Veterans Administration San Diego (United States); E. Higginbotha, Veterans Administration San Diego (United States); Z. Agha, Univ. of California, San Diego (United States) and Veterans Administration San Diego (United States)

Users of immersive virtual reality environments often report a side-effect of feeling like their real experiences after immersive sessions resemble elements of the virtual world. Yet perhaps this side-effect can be turned around to explore the possibilities for immersion in virtual world group training simulations. This paper will describe observations from my time working as an artist/researcher with the UCSD Medical School and VA Healthcare System to develop trainings for nurses, doctors and Hospital Incident Command staff which simulate pandemic virus outbreaks. By examining moments of slippage between realities, both into and out of the virtual environment, moments of the confusion of boundaries, we can better understand methods for creating immersion. I will use the mixing of realities as a transversal line of inquiry, borrowing from science studies, game studies, and performance studies to reveal the social implications of technology. Focusing on drills conducted in Second Life, I will examine both moments within the drill, interviews after the drill and the feedback with actual hospital procedures.

# Conference 7865: Human Vision and Electronic Imaging XVI

Monday-Thursday 24-27 January 2011

Part of Proceedings of SPIE Vol. 7865 Human Vision and Electronic Imaging XVI

## 7865-05, Session 2

### What makes good image composition?

R. Banner, Hewlett-Packard Labs. Israel Ltd. (Israel)

Composition is one of the most important features for the visual representations of ideas and messages. In particular, the ability to evaluate compositions and separate between “good” and “bad” compositions is of major importance for graphic design applications that facilitate the creation of compelling digital jobs. Given a compositional arrangement of objects in the pictorial space, this work proposes a measure that evaluates how balanced the given composition is. To that end, we first review several important perceptual concepts that artists intuitively obey to produce balanced compositions. Based on these concepts we suggest a novel composition measure that makes a use of the electrostatic model. Finally, we demonstrate the validity of this measure in theory and through simulations.

## 7865-06, Session 2

### A comparison of perceived lighting characteristics in simulations versus real-life setup

B. A. Salters, P. J. H. Seuntjens, Philips Research Nederland B.V. (Netherlands)

The advance of LED technology enables a whole new range of luminaire designs, which previously were not possible. The optical performance of a design in terms of e.g. brightness distributions can be simulated quite well by means of raytracing software. These simulations are not very suitable to rate a certain design on perceptual aspects such as cosiness or diffuseness, and visibility of artifacts such as non-uniformities. For this kind of perceptual research, it is still required to have an actual prototype of a certain design which is costly and time consuming.

Therefore, it is extremely useful to have an understanding if perception questions could be answered by means of photorealistic renderings. For this we have built a room where several luminaires can be tested. The same room has been simulated in Lighttools generating photorealistic renderings. Finally, several different perceptual questions have been investigated statistically, both in the experimental room, as well as using a picture shown on a normal LCD monitor, and printed on paper. We will discuss similarities and differences between the results of both types of tests. In this way, relations can be established between all aspects in designing luminaires: design and simulation, prototyping, and perception studies.

## 7865-07, Session 2

### Investigating two features of aesthetic perception in consumer photographic images: clutter and center

C. D. Cerosaletti, A. C. P. Loui, A. C. Gallagher, Eastman Kodak Co. (United States)

No abstract available.

## 7865-08, Session 3

### Analyzing near-infrared images in utility assessment

N. Salamati, Z. Sadeghipoor, S. E. Süssstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

Visual cognition is of significant importance in imaging applications, such as security and surveillance, where it is crucial to determine the maximum distortion level that can be applied to the images while still insuring that enough information is conveyed to recognize the scene. Since reaction in the NIR part of the spectrum is material dependent, an object made of a specific material is more probable to have uniform response in the NIR images. Consequently, edges in the NIR images likely correspond to the physical boundaries of the objects rather than changes in color within the object.

In this paper, we evaluate the usefulness of NIR images for cognition tasks. We compare the maximum distortion level of visible and NIR images where scenes are still correctly recognized. We performed a subjective study on six scenes, each one represented by its visible and NIR image. The images were compressed to different Q factors using JPEG compression. We found that in this preliminary test, four out of the six images were more easily recognized based on the NIR images.

## 7865-09, Session 3

### Appearance-based human gesture recognition using multimodal features for human computer interaction

D. Luo, Waseda Univ. (Japan); H. Gao, H. K. Ekenel, Karlsruher Institut für Technologie (Germany); J. Ohya, Waseda Univ. (Japan)

The use of gesture as a natural interface plays an utmost important role for achieving intelligent Human Computer Interaction (HCI). Human gestures include different components of visual actions such as motion of hands, face, and torso, to convey meaning. So far, in the field of gesture recognition, most previous works have focused on the manual component of gestures. In this paper, we present an appearance-based multimodal gesture recognition framework, which combines the different groups of features such as head motion, facial expression and hand motions which have been extracted from the image frames captured directly by a web camera. We refer 12 classes of human gestures with facial expression including neutral (e.g. a sign “feel”), negative (e.g. “angry”) and positive (e.g. “excited”) meanings from American Sign Languages. We combine the features in two levels by employing two fusion strategies. At the feature level, an early feature combination can be performed by concatenating and weighting different feature groups, and LDA is used to choose the most discriminate elements by projecting the feature on a discriminative expression space. The second strategy is applied on decision level. Weighted decisions from single modality are fused in a late stage. A Condensation-based algorithm is adopted for classification. We collected a data set with three recording sessions and conducted experiments with the combination techniques. Experimental results showed that the combined recognition outperforms the single modality significantly.

### 7865-10, Session 3

#### **Adaptive user interfaces for relating high-level concepts to low-level photographic parameters**

E. Scott, P. A. Madhawa Silva, B. Pardo, T. N. Pappas, Northwestern Univ. (United States)

Common controls for photographic editing can be difficult to use and have a significant learning curve. Often, a user does not know a direct mapping from a high-level concept (such as “soft”) to the available parameters or controls. In addition, many concepts are subjective in nature, and the appropriate mapping may vary from user to user. To overcome these problems, we propose a system that can quickly learn a mapping from a high-level subjective concept onto low-level image controls using machine learning techniques. To learn such a concept, the system shows the user a series of training images that are generated by modifying a seed image along different dimensions (e.g., color, sharpness), and collects the user ratings of how well each training image matches the concept. Since it is known precisely how each modified example is different from the original, the system can determine the correlation between the user ratings and the image parameters to generate a controller tailored to the concept for the given user. The end result — a personalized image controller — is applicable to a variety of concepts. We have demonstrated the utility of this approach to relate low-level parameters, such as color balance and sharpness, to simple concepts, such as “lightness” and “crispness,” as well as more complex and subjective concepts, such as “pleasantness.” We have also applied the proposed approach to relate subband statistics (variance) to perceived roughness of visual textures (from the CURET database).

### 7865-11, Session 3

#### **Parametric quality assessment of synthesized textures**

D. Siddalinga Swamy, D. M. Chandler, K. J. Butler, Oklahoma State Univ. (United States); S. S. Hemami, Cornell Univ. (United States)

No abstract available.

### 7865-12, Session 3

#### **On the perception of band-limited phase distortion in natural scenes**

K. P. Vilankar, L. Vasu, D. M. Chandler, Oklahoma State Univ. (United States)

This paper presents the results of a psychophysical experiment performed to find the sensitivity of the human visual system to bandlimited phase distortions in spatial frequency bands ranging from approximately 0.6-16 cycles/degree. A complex wavelet transform was used to generate phase-distorted images by adding Gaussian noise to the phase component of the complex wavelet coefficients within individual subbands. Phase-distortion sensitivity was measured using a spatial two-alternative forced-choice procedure. Three natural scenes and five vertical sine-wave gratings were used in the experiment. The results revealed that for individual sine-wave gratings, phase-distortion sensitivity was found to increase with decreasing contrast, which suggests the effect of contrast masking. For a compound sine-wave grating (containing four spatial frequencies 15.8, 5.05, 1.65 and 0.57 cycles/degree), and for natural scenes, phase-distortion sensitivity demonstrated an image-specific trend, suggesting the occurrence of both masking and cueing during phase-distortion detection. Based on these results, an algorithm is developed which attempts to predict the visual quality of phase-distorted images. We demonstrate that our algorithm performs well in quality estimation of phase-distorted images compared to other modern image quality assessment algorithms.

### 7865-13, Session 4

#### **Complex bioinformatics data: insights from data visualization and perception**

T. Munzner, The Univ. of British Columbia (Canada)

No abstract available

### 7865-15, Session 4

#### **Perceptual issues in the recovery and visualisation of integrated systems biology data**

T. P. Pridmore, The Univ. of Nottingham (United Kingdom)

The systems approach to biological research emphasises understanding of complete biological systems, rather than a reductionist focus on tightly defined component parts. Systems biology is naturally interdisciplinary; research groups active in this area typically contain experimental and theoretical biologists, mathematicians, statisticians, computer scientists and engineers. A wide range of tools are used to generate a variety of data types which must be integrated, presented to and analysed by researchers from any and all of the contributing disciplines. The goal here is to create predictive models of the system of interest; the models produced must also be analysed, and in the context of the data from which they were generated. Effective, integrated data and model visualisation methods are crucial if scientifically-appropriate judgments are to be made.

The Nottingham Centre for Plant Integrative Biology (CPiB) takes a systems approach to the study of the root of the model plant *Arabidopsis thaliana*. A rich mixture of data types, many extracted via automatic analysis of individual and time-ordered sequences of standard CCD and confocal laser microscope images, is used to create models of different aspects of the growth of the *Arabidopsis* root. In this paper we briefly review the data sets and modelling formalisms employed within CPiB, and discuss issues raised by the need to interpret images of the *Arabidopsis* root and integrate and present the resulting data and models to an interdisciplinary audience.

### 7865-16, Session 4

#### **Using cellular network diagrams to interpret large-scale datasets: past progress and future challenges**

P. D. Karp, M. Latendresse, S. Paley, SRI International (United States)

No abstract available

### 7865-17, Session 4

#### **Visualizing large high-throughput datasets based on the cognitive representation of biological pathways**

A. Nagel, Max-Planck-Institut für Molekulare Pflanzenphysiologie (Germany); O. Thimm, BASF Plant Science Company GmbH (Germany); M. Stitt, B. Usadel, Max-Planck-Institut für Molekulare Pflanzenphysiologie (Germany)

No abstract available

## 7865-18, Session 4

### **Metadata Mapper: a user interface web service for mapping data between independent visual analysis components, guided by perceptual rules**

B. E. Rogowitz, Visual Perspectives Consulting (United States);  
N. Matasci, The Univ. of Arizona (United States)

No abstract available

## 7865-47, Session 4

### **Hypergraph visualization and enrichment statistics: how the EGAN paradigm facilitates organic discovery from Big Data**

J. Paquette, Life Technologies Corp. (United States)

No abstract available

## 7865-20, Session 6

### **Examination of 3D visual attention in stereoscopic video content**

Q. Huynh-Thu, L. Schiatti, Technicolor (France)

Studies have indicated that viewers tend to focus their attention on specific areas of interest in a video and two mechanisms of visual attention have been identified: bottom-up and top-down mechanisms. Bottom-up attention corresponds to involuntary and unconscious eye movements, which are driven mostly by the signal content. On the other hand, top-down attention is mostly driven by task, context, semantics and experience. Very few studies have investigated visual attention and eye movement patterns on 3D stereoscopic moving sequences. In this paper, we investigate 3D visual attention and differences in visual attention between 2D and 3D content. For that purpose, we conducted a subjective experiment using an eye-tracking apparatus and a 3D stereoscopic display. Content was viewed using passive polarized glasses technology in a free-viewing scenario. Observers were asked to view the 2D and 3D versions of the same video content. We discuss our observations in terms of the attentive behavior and compare this aspect between the 2D and 3D scenarios. Finally, we also discuss issues related to the set-up of a 3D eye tracking experiment and the measurement of the human gaze in three dimensions.

## 7865-21, Session 6

### **Quantifying how the combination of blur and disparity affects the perceived depth**

J. Wang, M. Barkowsky, V. Ricordel, P. Le Callet, Univ. de Nantes (France)

The influence of a monocular depth cue, blur, on the apparent depth of stereoscopic scenes will be studied in this paper. When 3D images are shown on a planar stereoscopic display, binocular disparity becomes a pre-eminent depth cue. But it induces simultaneously the conflict between accommodation and vergence, which is often considered as a main reason for visual discomfort.

We propose to decrease the (binocular) disparity of 3D presentations, and to reinforce (monocular) cues to compensate the loss of perceived depth and keep an unaltered apparent depth. The limitation of depth-of-field of human eyes causes blur in the retinal image which is known as an important monocular depth cue.

We conduct a subjective experiment with a background plane and a single object in the foreground. A Siemens Star is used as the

foreground object, since it contains regions with both low frequency and high frequency. In the subjective experiment, two sources of perceived depth are used: disparity and blur. The perceived depth from disparity stems from the difference of disparity between the foreground object and the background. The perceived depth from blur stems from the amount of blur introduced to the background by convolution with a Gaussian kernel. A detailed statistical analysis is performed revealing the interaction of the two depth cues at different depth levels on a 3D flat screen display. The results are used to develop a computational model of perceived depth depending on blur and disparity.

## 7865-40, Poster Session

### **Depth perception enhancement based on chromostereopsis**

J. Hong, H. Lee, D. Park, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

The goal of this study is to enhance the cubic effect in images with reduced depth or cubic effect, using chromostereopsis among the characteristics of human visual perception. Chromostereopsis generally refers to a phenomenon in which Red, a long wavelength, seems to be more projected than Blue, a short wavelength. With this chromostereopsis, in ordinary times, colors with a long wavelength seem closer than those with a short wavelength due to the sense of distance; but as the lightness of the background layer is closer to White, a reversal effect takes place so that the sense of depth may differ depending on the lightness of the background layer. With this, the information on the lightness of the background layer becomes another depth cue. Based on this property, a psychophysical experiment was carried out to examine the impact of chromostereopsis and the neighbor color between layers on the sense of depth based on the lightness of the background layer.

## 7865-41, Poster Session

### **An evaluation of perceived color break-up on field-sequential color displays**

M. Kobayashi, A. Yoshida, Y. Yoshida, Sharp Corp. (Japan)

Field sequential color (FSC) displays have a major problem: color break-up (CBU). Moreover, it is difficult to quantify the CBU in saccadic eye movements, because the phenomenon occurs as quickly as a flash in saccadic eye movements, and there are individual variabilities for perceiving the CBU. Some previous study have presented assessments of saccadic CBU, but not indicated the detection and allowance thresholds of the target size in horizontal saccadic eye movements. Then, we conducted psychophysical experiments based on an FSC display driving with sub-frame frequency of 240Hz-1440Hz (each frame consist of red, green, and blue sub-frames). We employed a simple stimulus for our experiment, a static white bar with variable width. We tasked ten subjects a fixed saccade length of 58.4 visual degrees in horizontal eye movements, and a fixed target luminance of 15.25 cd/m<sup>2</sup>. We examined PEST method to find detection and allowance thresholds of white bar width for saccadic CBU.

This paper provides correlations between target sizes and sub-frame frequencies of an FSC display device, and proposes an easily evaluation method of perceiving saccadic CBU on FSC displays.

## 7865-42, Poster Session

### **Text detection: effect of size and eccentricity**

C. H. Kao, C. Chen, National Taiwan Univ. (Taiwan)

We studied the detectability of displayed texts as a function text size and retinal eccentricity. We measured the contrast detection threshold for various types of text-like stimuli (real-, non-, oracle bone-, and scrambled-characters) at different size and presented at different eccentricities. The contrast threshold was measured with a spatial two-

alternative forced-choice paradigm and PSI adaptive method. When the text size is small, the detection threshold of a character decreased with the increase of its size with a slope  $-1/2$  on log-log coordinates up to a critical size at all eccentricities and for all stimulus types. Beyond this critical size, there was little, if any, improvement of text detectability as text size further increased. The critical size depended on the stimulus type and eccentricity. The sensitivity for all types of stimuli decreased as the eccentricity increased. The estimated receptive field size of the text detectors is greater than that of the line detectors. In addition, the logarithm of receptive field size increases with eccentricity with a slope 0.22. With this information, we can construct a model to estimate the ability of a human observer to detect a character by the size and eccentricity of that character.

### 7865-43, Poster Session

#### **Image enhancement of high digital magnification for patient with central vision loss**

Z. Li, G. Luo, E. Peli, Schepens Eye Research Institute (United States)

Central vision loss is a leading vision impairment affecting millions of people worldwide. We are developing a head mounted camera and display device for mobile use, which can provide a wide range of magnification levels by means of digital zooming (scaling). The exposure level may not be appropriate for the sub-image to be magnified, and therefore often results in too dark, too bright, or low contrast sub-images. This problem is prevalent especially at high magnification levels.

For small sub-images, conventional histogram stretching enhancement often causes over- or under-enhancement, and flickering video images. According to human observers' selection, our new enhancement methods work better, and therefore are suitable to be implemented in the digital magnification device.

### 7865-44, Poster Session

#### **Quality versus intelligibility: studying human preferences for American Sign Language video**

F. M. Ciaramello, S. S. Hemami, Cornell Univ. (United States)

Real-time video conferencing using cellular devices provides natural communication to the Deaf community. For this application, compressed American Sign Language (ASL) video must be evaluated in terms of the intelligibility of the conversation and not in terms of the overall aesthetic quality of the video. This work reports an experiment to determine the subjective preferences of ASL users in terms of the trade-off between intelligibility and aesthetic quality when varying the proportion of the bitrate allocated explicitly to the regions of the video containing the signer. Test videos are encoded using a rate-distortion optimization technique that jointly optimizes for quality (measured using PSNR) and intelligibility (measured using a computational model of intelligibility) according to a single, user-specified encoding parameter. Experimental results suggest that at high bitrates, users prefer videos coded with the quality criteria, in which the non-signer regions in the video are encoded with some nominal rate. As the total encoding bitrate decreases, users prefer video coded with the intelligibility criteria, in which a greater proportion of the rate is allocated to the signer.

### 7865-22, Session 7

#### **Preferences for the balance between true image detail and noise**

S. G. Deshpande, S. J. Daly, Sharp Labs. of America, Inc. (United States)

No abstract available.

### 7865-23, Session 7

#### **Measurement of compression-induced temporal artifacts in subjective and objective video quality assessment**

C. Mantel, Gipsa-lab (France) and STMicroelectronics (France); P. Ladret, Gipsa-lab (France); T. Kunlin, STMicroelectronics (France)

Temporal pooling and temporal defects are the two differences between image and video quality assessment. Whereas temporal pooling has been the object of two recent studies, this paper focuses on the rarely addressed topic of compression-induced temporal artifacts, such as mosquito noise. To study temporal aspects in subjective quality assessment, we compared the perceived quality of two versions of a mosquito noise corrector: one purely spatial and the other spatio-temporal. We set up a paired-comparison experiment and choose videos whose compression mainly creates temporal artifacts. Results proved the existence of a purely temporal aspect in video quality perception. We investigate the correlation between subjective results from the experiment and three video metrics (VQM, MOVIE, VQEM), as well as two temporally-pooled image metrics (SSIM and PSNR). SSIM and PSNR metrics find the corrected sequences of better quality than the compressed ones but do not distinguish spatial and spatio-temporal processings. The confrontation of those results with the VQM and Movie objective metrics show that they do not account for this type of defects. A detailed study highlights that either they do not detect them or the response of their temporal component is masked by the one of their spatial components.

### 7865-24, Session 7

#### **Perceived contrast of electronically magnified video**

A. M. Haun, R. L. Woods, E. Peli, Schepens Eye Research Institute (United States)

It is frequently observed that electronic magnification of imagery results in a decrease in the apparent contrast of the magnified image relative to the original. Magnification is typically used in one of two contexts: either the entire image is enlarged to fill a larger display area, or a portion of an image is enlarged to fill the same display area. The decrease in perceived contrast might be due to a combination of image blur and of sub-sampling the larger range of contrasts in the original image. In a series of experiments, we measured the effect on apparent contrast of magnification in both contexts; both as a function of magnification power and of viewing distance (visibility of blur induced by magnification). We found a significant difference in the apparent contrast of magnified versus unmagnified video sequences. The effect on apparent contrast was found to increase with increasing magnification, and to decrease with increasing viewing distance (or with decreasing apparent size). There was significant variation between observers in the magnitude of the perceived difference, particularly between experienced and naive observers, but across observers and conditions the reduction in perceived contrast was reliably on the order of 0.1 to 0.2 log units (80% to 63%). These effects are consistent with expectations based on both the contrast statistics of natural images and the contrast sensitivity of the human visual system. It is demonstrated that 1) local areas within larger images or videos will usually have lower physical contrast than the whole; and 2) visibility of



'missing content' (e.g. blur) in an image is generally interpreted as a decrease in contrast, and this visibility declines with viewing distance. The perceptual magnitudes of both effects are thereby predictable by a model of human contrast sensitivity and perceived contrast.

Supported by NIH grants EY05957 and EY19100.

## 7865-25, Session 7

### Estimating the impact of single and multiple freeze occurrences on video quality

T. Xiao, K. E. Brunnström, Acreo AB (Sweden)

Hybrid models need modules for analyzing the bitstream as well as analyzing the decoded video. For the decoded video interesting to estimate the impact of packet loss after concealment. One common concealment strategy is to simply freeze the video to the last good frame.

This work is based on an algorithm developed by Wolf (2009) that can detect dropped or repeated frames, which may be perceived as freeze or pause by the viewer. A mapping from the detected freezing time and the number of freezing occurrences to video quality will be presented. As a starting point we consider the relationship that Pastrana - Vidal et al. (2004) suggested between probability of detection and the duration of the dropped frames. They also found that it is important to consider not only the duration of the freeze but also the number of freeze occurrences. Two different relationships between the total duration of freeze and the number of occurrences have been derived, based on extension of models for a single freeze occurrence. One was based on Pastrana - Vidal et al. (2004) and the other was based on Van Kester et al (2011). A subjective test was designed to evaluate the performance of the models. Good performance was found on this data i.e. more than 0.9 correlation, which was similar for both models. The RMSE was better for one of the models.

## 7865-26, Session 7

### The effects of scene characteristics, resolution, and compression on the ability to recognize objects in video

J. Dumke, C. G. Ford, I. W. Stange, Institute for Telecommunication Sciences (United States)

The quality of video used in public safety applications must be evaluated in terms of its usability for specific tasks performed by the end user, and this task is often object recognition. As video applications in public safety become more widespread, guidance to end users regarding how to identify the level of video quality necessary for their application becomes necessary.

The Public Safety Communication Research (PSCR) project performed a subjective test as one of the first in a series to explore visual intelligibility in video - the ability of the user to recognize an object in a video stream given various conditions. The test sought to measure the effects on visual intelligibility of three scene parameters (target size, scene complexity, scene lighting), several compression rates and two resolutions (VGA (640x480) and CIF (352x288)). Seven similarly sized objects were used as targets in nine sets of near-identical source scenes, where each set was created using a different permutation of the parameters under study. Viewers were asked to identify the objects via multiple choice questions. Objective measurements were performed on each of the scenes, and the predictive ability of the measurement on visual intelligibility is presented.

## 7865-27, Session 7

### Supplemental subjective testing to evaluate the performance of image and video quality estimators

A. R. Reibman, AT&T Labs. Research (United States); F. M. Ciaramello, Cornell Univ. (United States)

The subjective tests used to evaluate image and video quality estimators (QEs) are expensive and time consuming. More problematic, the majority of subjective testing is not designed to find systematic weaknesses in the evaluated QEs.

As a result, a motivated attacker can take advantage of these systematic weaknesses to gain unfair monetary advantage. In this paper, we draw on some lessons of software testing to propose additional testing procedures that target a specific QE under test. These procedures supplement, but do not replace, the traditional subjective testing procedures that are currently used. The goal is to motivate the design of objective QEs which are better able to accurately characterize human quality assessment.

## 7865-28, Session 7

### On evaluation of video quality metrics

M. Cadik, T. O. Aydin, K. Myszkowski, H. P. Seidel, Max-Planck-Institut für Informatik (Germany)

We propose and make publicly available a new dataset for evaluation of image/video quality metrics with emphasis on applications in computer graphics. Several aspects were influential while designing the dataset: (i) in addition to the assessment of the quality of LDR videos, the assessment of high-dynamic range (HDR) videos, as well as comparing HDR videos with LDR videos and vice versa, and (ii) the outcome of the subjective experiment in the form of distortion maps that show quality prediction as a function of spatial position which is especially important for applications in computer graphics. Furthermore, we show an example evaluation of recent image and video quality metrics that were successfully applied in the field of computer graphics. The goal of this evaluation was to examine the correlation between the objective quality predictions computed by the video quality metrics, and the subjective responses obtained by the experimental procedure. To that end the proposed dataset and the subjective study have the following unique features over previous studies on video quality assessment: 1) The test set includes LDR-LDR, HDR-HDR, and HDR-LDR reference-test video pairs with various types of distortions. 2) A BrightSide DR37-P HDR display (max. luminance  $\sim 30000$  cd/m<sup>2</sup>) was used for displaying the videos. 3) The subjects were not asked to assess only an overall quality of the video, but to mark the regions where they saw differences between test and reference videos, resulting in distortion maps similar to the metric outcome.

## 7865-29, Session 8

### Interactions of visual attention and quality perception

J. A. Redi, EURECOM (France); H. Liu, Technische Univ. Delft (Netherlands); R. Zunino, Univ. degli Studi di Genova (Italy); I. Heynderickx, Philips Research Nederland B.V. (Netherlands)

Several attempts have been made in literature to integrate quality metrics computation with visual saliency information, with contrasting results. A careful design of the integration strategy should reflect the mechanisms underlying the interaction between image quality assessment and visual attention. A subjective study, based on tracking eye-movements during quality assessment was performed to better understand them. The aim of the experiment was to analyze the effect of quality evaluation on the attention deviation from Natural Scene Saliency (NSS), and, in particular, whether, and if so how, this deviation depended on the distortion kind and/or amount. Deviated saliency

maps were derived from the scoring of distorted images, and then compared to the corresponding NSS, derived from the free-looking of high quality images. The study revealed some differences between the Deviated and the Natural Scene saliency maps, related to the quality level of the images. The higher the quality, the more the deviated attention was spread in the background regions. No evident role for the kind of distortion in the saliency changes was found instead. Rather, the quality assessment task seemed to prevail on the natural attention, forcing it to deviate in order to better evaluate the impact of artifacts.

## 7865-30, Session 8

### **Task dependence of visual attention on compressed videos: point of gaze statistics and analysis**

A. Mittal, A. K. Moorthy, W. S. Geisler, A. C. Bovik, The Univ. of Texas at Austin (United States)

Previously, we presented some preliminary results on how task influences visual attention when a subject views compressed videos at varying levels of compression. Our analysis suggested that the task does indeed influence fixations and that the points of gaze were definitely influenced by the amount of distortion conditioned on the task. However, this analysis did not provide further insight into low level statistical features that attract visual attention. Here, we analyze the eye movement data by computing statistics at point of gaze for each of the two tasks - quality assessment and summarization. We compute luminance, root-mean-squared (RMS) contrast, motion and quality at points of gaze and analyze how these statistics behave across tasks as well as within tasks across varying levels of compression/distortion.

## 7865-31, Session 8

### **Measuring contour degradation in natural image utility assessment: methods and analysis**

G. O. Pinto, D. M. Rouse, S. S. Hemami, Cornell Univ. (United States)

Utility estimation algorithms predict the usefulness of an image for a particular task to be performed by a human. They differ from quality assessment algorithms in that they should provide accurate estimates even when images are extremely visibly distorted relative to the original, yet are still sufficient for the task. Our group has previously proposed the Natural Image Contour Evaluation (NICE) algorithm for utility estimation. This work investigates novel ways to measure the degradation of image contours and applies the results to improve the performance of NICE. NICE estimates perceived utility by comparing the intensity edge maps of both the reference and distorted images using a Hamming distance. This work considers both gradient orientations and quality assessment algorithms spatially limited to edges instead of comparing edge maps, and replaces the Hamming distance with distance transforms. The performance of these alternative implementations was evaluated on the CU-Nantes database, which provides perceived utility scores for a collection of distorted images. By applying VIF solely on the contour regions we obtained a statistically significant improvement with respect to the Pearson correlation metric. This result suggests that we can recast the image utility assessment problem as the problem of image quality assessment on the image contours. The other approaches yielded statistically equivalent performance to the standard implementation of NICE, suggesting that utility estimation is robust to the exact implementation of the edge degradation measure and the distance measure.

## 7865-32, Session 9

### **Evolution of attention mechanisms for early visual processing**

T. Müller, A. Knoll, Technische Univ. München (Germany)

Being inspired by the hierarchical architecture of the human visual processing apparatus, an proposed approach introduces an early processing system based on saliency features for attentional filtering. Areas containing a high saliency values are then composed into regions of interest for further object recognition and tracking. The areas are determined from both bottom-up computation, e.g. in dynamic environments connected blobs of textures moving in a uniform way within the visual field are considered salient; and top-down computation, where saliency effects are either unconscious through inherent mechanisms like inhibition-of-return, or volitional through cognitive feedback, e.g. when an object moves consistently in the visual field but unexpectedly disappears, attention is directed to this region in order to investigate this unexpected behavior.

In this paper furthermore a multi-threaded extension to this saliency mechanism applying evolutionary processes is proposed. Here, multiple saliency units are used to produce the regions of attention. All of these units have different parameter-sets. Now, a population of saliency units creates regions of attention first, then the results are evaluated by cognitive/top-down feedback and finally the genetic mechanism is applied to the population units: mutation and cloning of the best performers and extinction of the worst performers. The fitness function is defined by measurement of the probability for the regions being task-relevant, i.e. containing relevant objects for the human-robot-interaction task.

## 7865-33, Session 9

### **Learned saliency transformations for gaze guidance**

E. Vig, Univ. zu Lübeck (Germany); M. Dorr, Schepens Eye Research Institute (United States) and Univ. zu Lübeck (Germany); E. Barth, Univ. zu Lübeck (Germany)

The saliency of an image or video region indicates how likely it is that the viewer of the image or video fixates that region due to its conspicuity. An intriguing question is how we can change the video region to make it more or less salient. Here, we address this problem by using a machine learning framework to learn from a large set of eye movements collected on real-world dynamic scenes how to alter the saliency level of the video locally. We derive saliency transformation rules by performing spatio-temporal contrast manipulations (on a spatio-temporal Laplacian pyramid) on the particular video region. Our goal is to improve visual communication by designing gaze-contingent interactive displays that change, in real time, the saliency distribution of the scene.

## 7865-34, Session 10

### **On the relationship between selective visual attention and visual consciousness**

N. Tsuchiya, C. Koch, California Institute of Technology (United States)

The relationship between attention and consciousness is a close one, leading many scholars to conflate the two. I distinguish between exogenous, saliency-driven, task-independent attention and top-down, endogenous and voluntary attention. In the first part I will summarize powerful computational models of saliency-driven attention that capture a large fraction of eye movements in normal subjects inspecting natural scenes. In the second part, I will summarize psychophysical evidence arguing that top-down attention and consciousness are distinct phenomena that need not occur together and that can be manipulated using distinct paradigms. Subjects can become conscious of an isolated object, or the gist of the scene in

the near absence of top-down attention. Conversely, subjects can attend to perceptually invisible objects. In particular, I will describe a full factorial study of the influences of attention and consciousness on afterimage formation. These data provide clear evidence for distinctive influence of attention and consciousness on perception.

### 7865-35, Session 10

#### **A gaze-contingent display to study contrast sensitivity under natural viewing conditions**

M. Dorr, P. J. Bex, Schepens Eye Research Institute (United States)

Contrast sensitivity has been extensively studied over the last decades and there are well-established models of early vision that were derived by presenting the visual system with synthetic stimuli such as sine-wave gratings near threshold levels. Natural scenes, however, contain a much wider distribution of orientations, spatial frequency content, and both luminance and contrast values. Furthermore, humans typically move their eyes two to three times per second under natural viewing conditions, but most laboratory experiments require subjects to maintain central fixation.

We here describe a gaze-contingent display capable of performing real-time contrast modulations of video in retinal coordinates, thus allowing us to study contrast sensitivity when dynamically viewing dynamic scenes.

Our system is based on a Laplacian pyramid for each frame that efficiently represents individual frequency bands. Each output pixel is then computed as a locally weighted sum of pyramid levels to introduce local contrast changes as a function of gaze.

Our GPU implementation achieves real-time performance with more than 100 fps on high-resolution video (1920 by 1080 pixels) and a synthesis latency of only 1.5 ms.

Psychophysical data show that contrast sensitivity is greatly decreased in natural videos and under dynamic viewing conditions. Synthetic stimuli therefore only poorly represent natural vision.

### 7865-36, Session 10

#### **Analyzing complex gaze behavior in the natural world**

J. B. Pelz, T. Kinsman, K. M. Evans, Rochester Institute of Technology (United States)

The history of eye-movement research extends back at least to 1794, when Erasmus Darwin (Charles' grandfather) published *Zoonomia*, including descriptions of eye movements due to self motion. But research on eye movements was restricted to the laboratory for 200 years, until Land built the first wearable eyetracker at the University of Sussex and published the seminal paper "Where we look when we steer" [Land & Lee 1994]. In the intervening centuries, we learned a tremendous amount about the mechanics of the oculomotor system and how it responds to isolated stimuli, but virtually nothing about how we actually use our eyes to explore, gather information, navigate, and communicate in the real world.

Inspired by Land's work, we have been working to extend knowledge in these areas by developing hardware, algorithms, and software that have allowed researchers to ask questions about how we actually use vision in the real world. Central to that effort are new methods for analyzing the volumes of data that come from the experiments made possible by the new systems.

We will describe a number of recent experiments and demonstrate SemantiCode, a new program that supports assisted coding of eye-movement data collected in unrestricted environments.

### 7865-37, Session 10

#### **What your visual system sees where you are not looking: implications for imaging applications**

R. E. Rosenholtz, Massachusetts Institute of Technology (United States)

What is the representation in early vision? Considerable research has demonstrated that the representation is not equally faithful throughout the visual field; representation appears to be coarser in peripheral and unattended vision, perhaps as a strategy for dealing with an information bottleneck in visual processing. In the last few years, a convergence of evidence has suggested that in peripheral and unattended regions, the information available consists of local summary statistics. Given a rich set of these statistics, many attributes of a pattern may be perceived, yet precise location and configuration information is lost in favor of the statistical summary. This representation impacts a wide range of visual tasks, including visual search as well as visual cognition of complex displays. This talk discusses the implications for both perception, and for imaging applications such as information visualization.

### 7865-38, Session 10

#### **Attention as a Bayesian inference process**

S. Chikkerur, Massachusetts Institute of Technology (United States)

David Marr famously defined vision as "knowing what is where by seeing". In the framework described here, attention is the inference process that solves the visual recognition problem of what is where. The theory proposes a computational role for attention and leads to a model that performs well in recognition tasks and that predicts some of the main properties of attention at the level of psychophysics and physiology. We propose an algorithmic implementation - a Bayesian network that can be mapped into the basic functional anatomy of attention involving the ventral stream and the dorsal stream. This description integrates bottom-up, feature-based as well as spatial (context based) attentional mechanisms. Attentional phenomena such as pop-out, multiplicative modulation and change in contrast response, which have been described in the recent literature as fundamentally different and in some cases as conflicting findings, are all directly predicted by the same model. We also show that the Bayesian model predicts well human eye fixations (considered as a proxy for shifts of attention) in natural scenes, and can improve accuracy in object recognition tasks involving cluttered real world images. In both cases, we found that the proposed model can predict human performance better than existing bottom-up and top-down computational models.

### 7865-39, Session 10

#### **Statistical modeling of surprise with applications to predicting attention and gaze**

L. Itti, The Univ. of Southern California (United States)

No abstract available

# Conference 7866: Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications

Monday-Thursday 24-27 January 2011

Part of Proceedings of SPIE Vol. 7866 Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications

## 7866-01, Session 1

### **Image reconstruction on multi-primary displays**

C. Brown Elliott, Nouvoyance, Inc. (United States)

No abstract available.

## 7866-02, Session 1

### **Adaptive color visualization for dichromats using a customized hierarchical palette**

C. E. Rodriguez Pardo, G. Sharma, Univ. of Rochester (United States)

Color display systems have been growing rapidly, and the emergence of new technologies, materials and applications, have enabled display designers to improve the observer perception as well as physical constraints in the display performance. In particular, multi-primary color techniques has been recently explored, showing advantages in expanding the color gamut, widening the viewing angle and power saving, when comparing with the conventional display systems with only three primaries. These works assume a standard trichromat observer for which the display primaries and other parameters are optimized. However, it is known that nearly 10% of the male population has color deficiencies. Although some work has been made at the stages of image processing and computer vision in order to improve the perception of these observers, few research has been developed in the display part. The aim of this work is to improve the perception of color for dichromat observers, by using a multi-primary system, that allows to choose the best combination of light sources in order to maintain proportional the differences in color perception between normal observers and color deficiency individuals.

## 7866-03, Session 1

### **R/G/B color crosstalk characterization and calibration for LCD displays**

R. Safaee-Rad, Qualcomm Inc. (Canada); M. Aleksic, Qualcomm Inc. (United States)

LCD displays exhibit significant color crosstalks between their red, green and blue channels (more or less depending on the type of LCD technology). This problem, if it is not addressed properly, leads to (a) a significant color errors in the rendered images on LCD displays, (b) significant shifts in red, green and blue primaries and (c) a significant gray tracking problem. The traditional method for addressing this problem has been using a 3x3 color correction matrix in the display processing pipeline. Experimental data clearly shows that this linear model for color correction is not sufficient to address the above problems. Herein, it is proposed to use higher order polynomials for color correction in the display processing pipeline. This paper presents detailed experimental results and comparative analysis on using polynomial models with different orders for color correction.

## 7866-04, Session 1

### **Color gamut boundary optimization of wide gamut display devices**

F. Lebowsky, STMicroelectronics (France)

High-end monitors based on LCD technology increasingly address wide color gamut implementations featuring precise color calibration within a variety of different color spaces such as extended sRGB or AdobeRGB. Combining a Look-Up-Table method with linear interpolation in RGB component space using 3x3 matrix multiplication provides optimized means of tone curve adjustments as well as independent adjustment of device primaries. The proposed calibration method completes within several seconds compared to traditional color calibration procedures easily taking several minutes. In addition, the user can be given subjective control over color gamut boundary settings based on dynamic gamut boundary visualization. The proposed component architecture not only provides independent control over 8 color vertices but also enables adjustment in quantities of 10-4 of full amplitude range. User defined color patches can be adjusted manually while simultaneously tracking color gamut boundaries and visualizing gamut boundary violation in real time. All this provides a convenient approach to fine tuning tone curves and matching particular constraints with regard to user preferences, for example specific ambient lighting conditions, across different devices such as monitors and printers.

## 7866-05, Session 2

### **Color correction for projected image on colored-screen based on a camera**

D. Kim, T. Lee, Kyungpook National Univ. (Korea, Republic of); M. Choi, Daegu Polytechnic College Univ. (Korea, Republic of); Y. Ha, Kyungpook National Univ. (Korea, Republic of)

This paper proposes a color correction method for images projected on colored surfaces. To achieve our object without characterization process, we replace measurement device for characterization with a digital camera. We estimate a color correction matrix by linear regression using input digital values which produce the same color on both white and colored screen. Differently from previous methods, the use of general still camera allows to measure regardless of places. In addition, two captured images on white and colored screen with ramp patches form the color shift for 9 steps of each channel, enabling accurate construction of the transform matrix. Nonlinearity of camera characteristics is also considered by using regression method to construct a transform matrix. In the experimental results, corrected image using the proposed method on colored screen have shown better performance than previous methods in both objective and subjective evaluations.

## 7866-06, Session 2

### **Modeling LCD displays with local backlight dimming for image quality assessment**

J. Korhonen, N. Burini, S. Forchhammer, Technical Univ. of Denmark (Denmark); J. M. Pedersen, Bang & Olufsen A/S (Denmark)

Traditionally, objective image and video quality assessment methods operate with the numerical presentation of the signal, and they do not take the characteristics of the actual output device into account. This is a reasonable approach, when quality assessment is needed for evaluating the signal quality distortion related directly to digital signal processing, such as compression. However, the physical characteristics of the display device also pose a significant impact on the overall perception. In order to facilitate image quality assessment on modern liquid crystal displays (LCD) using light emitting diode (LED) backlight with local dimming, we present the essential considerations and guidelines for modeling the characteristics of displays with high

dynamic range (HDR) and locally adjustable backlight segments. The representation of the image generated by the model can be assessed using the traditional objective metrics, and therefore the proposed approach is useful for assessing the performance of different backlight dimming algorithms in terms of resulting quality and power consumption in a simulated environment. We have implemented the proposed model in Matlab environment and compared the visual results produced by the model against respective images displayed on a real display with locally controlled backlight units.

## 7866-07, Session 2

### Content dependent selection of image enhancement parameters for mobile displays

Y. Lee, C. Kim, Y. Kang, G. Kim, H. Kim, Inha Univ. (Korea, Republic of)

Mobile devices such as cellular phones and portable multimedia player with capability of playing terrestrial digital multimedia broadcasting (T-DMB) contents have been introduced into consumer markets. As size and resolution of mobile displays increase, market demands for improved image quality are gradually increasing. Various image enhancement techniques are applied to meet the needs for improved image quality on mobile displays. They include noise reduction, contrast and sharpness enhancement and color correction, etc. Unlike TV application, there are strict restrictions on memory resources and computational complexity for mobile applications. In this paper, parameters for image enhancement techniques are adaptively determined based on image contents. Minimization of flicker due to sudden scene change is also proposed. Experimental results indicate that dynamic selection of image control parameters exhibits better performance than traditional imaging chain with fixed parameters.

## 7866-08, Session 2

### Saliency-driven black point compensation

A. J. Lindner, Ecole Polytechnique Fédérale de Lausanne (Switzerland); N. Bonnier, Océ Print Logic Technologies (France); S. E. Süssstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

We present a novel framework for automatically determining whether or not to apply black point compensation (BPC) in image reproduction. Visually salient objects have a larger influence on determining image quality than the number of dark pixels in an image, and thus should drive the use of BPC. We propose a simple and efficient algorithmic implementation to determine when to apply BPC based on low-level saliency estimation. We evaluate our algorithm with a psychophysical experiment on an image dataset printed with or without BPC on a Canon printer. We find that our algorithm is correctly able to predict the observers' preferences in all cases where the saliency maps are unambiguous and accurate.

## 7866-09, Session 2

### DIN 6164 for gamut mapping?

U. Caluori, D. Küpper, K. Simon, EMPA (Switzerland)

Color as a perceptual phenomenon is, to date, an only approximately understood research topic. In general, color spaces address specific applications or situations, such that they may not be adequate in other contexts. In this work, we investigate four device-independent color spaces (color order systems) with regard to their suitability for a specific gamut mapping concept.

We implemented the transformation of colors to and from CIELAB and IPT by symbolic computation, and for Munsell and DIN 6164 by building an efficient three-dimensional interpolation structure. We encountered several difficulties with DIN 6164 and Munsell,

like defining a good distance formula, artifacts that occur when compressing colors, discontinuity, boundary handling, extrapolation and sample intersection. We describe how we tackled these problems and finally verified and evaluated our implementation by conducting a psycho-visual experiment on images processed with one of our latest gamut mapping algorithms.

Preliminary results show that our modified DIN 6164 space surprisingly performs best in terms of image quality, which motivates further investigation of DIN 6164 as a serious alternative in this context.

## 7866-10, Session 3

### Estimation of low dynamic range images from single Bayer image using exposure look-up table for high dynamic range image

T. Lee, Y. Ha, Kyungpook National Univ. (Korea, Republic of); C. Lee, Andong National Univ. (Korea, Republic of)

HDR(high dynamic range) imaging techniques guarantees wider dynamic range for images than those captured from general still camera. Usually, these techniques emerge several LDR(low dynamic range) images with different exposure. It means that they need additional process for capturing LDR images, as follows they cause ghost effect for moving objects in HDR images. Accordingly, this paper suggests a method to estimate arbitrary LDR images from a single Bayer image using exposure LUT(look-up table), considering channel dependency. Bayer image provides sufficient luminance information with 14bit data for each channel than LDR images with 8bit. We first construct three LUTs for each RGB channel using Bayer image. It has the relationship between input scene and output average luminance. Then, from input image captured by auto mode, corresponding input digital value is estimated by using current exposure and LUTs. Next, target exposures which are corresponded to target average luminance from user choice are estimated using estimated digital value and LUTs. After that, final LDR images with target exposures are estimated using input image and LUT. To improve the accuracy of estimation, saturated area is estimated by considering channel dependency. In experiments, high PSNR values are obtained between estimated and captured images.

## 7866-11, Session 3

### Flicker reduction in tone mapped HDR videos

B. Guthier, S. Kopf, M. Eble, W. Effelsberg, Univ. Mannheim (Germany)

In order to display a HDR video on a regular low dynamic range (LDR) screen, it needs to be tone mapped (TM). A great number of tone mapping operators exist - all designed to tone map one image at a time. Using them on each frame of a HDR video individually leads to flickering in the full image.

In our work, we analyze three tone mapping operators with respect to flickering. We propose a criterion for the automatic detection of image flicker by analyzing the average pixel brightness of the tone mapped frame. Flicker is detected if the difference between the averages of two consecutive frames is larger than a threshold derived from Stevens' power law. Fine-tuning of the threshold is done in a subjective study.

Additionally, we propose a generic method to reduce flickering. It is applicable to all parameter driven tone mapping operators. We begin by tone mapping a frame with the parameters set to default values. If the flicker detection reports a visible variation in the frame's brightness, the parameters are adjusted and the frame is tone mapped again. As a result, the brightness variation is smoothed out over several frames, becoming less disturbing.

## 7866-12, Session 4

### Applying AR technology with a projector-camera system in a history museum

K. Miyata, National Museum of Japanese History (Japan); R. Shiroishi, Ochanomizu Univ. (Japan); Y. Inoue, Bunkyo Univ. (Japan)

In this research, an AR (augmented reality) technology with projector-camera system is applied in a history museum to provide user-friendly interface and pseudo hands-on exhibition. The proposed system is a desktop system and designed for old Japanese coins to enhance the visitors' interest and motivation to investigate them. The surface of the old coins has fine structures on both sides, so it is meaningful to show the reverse side to the visitors for enhancing their interest and motivation. The detection of the AR markers and rendering of the processed images were performed using ARToolKit, and the appearances of the old coins were calculated based on the photometric stereo algorithm. The users could observe the images of the old coins with changing appearance followed by the movements of the AR markers.

The proposed system contributes to develop an exhibition method based on the combinations of the real artifacts and the AR technology, and demonstrated the flexibility and capability to offer background information relating to the old Japanese coins. However, the accuracy of the detection and tracking of the markers and more detailed visitor evaluation survey are required to improve the effectiveness of the system.

## 7866-13, Session 4

### Memory preservation made prestigious but easy

R. Fageth, C. Debus, CeWe Color AG & Co. OHG (Germany); P. Sandhaus, Carl von Ossietzky Univ. Oldenburg (Germany)

Preserving memories combined with story telling using either photo books for multiple images in order to tell stories and using high quality products such as images printed on canvas or images behind acryl in order to use it as prestigious wall decoration are substituting more and more classical 4\*6 prints and classical silver halide posters. Digital printing via electro photography and ink jet is substituting more and more classical silver halide technology as dominant production technology for these kinds of products.

Images used in order to generate tangible output are stored in several locations (desktop and/or online in galleries or social networks) or originated at least from two and more image taking devices. This makes the generation of a compelling output more complicated as well as the confirmation process before generating tangible output of the people involved in the design process is becoming a challenge.

This paper describes a unique approach of combining desktop based software to initiate a compelling project but additionally use online capabilities in order to finalize and optimize that project online in a community process. A comparison of the consumer behavior between online and desktop based solutions for generating photo books will be presented.

The paper also analyzes the user behavior of generating pages; how many pages are containing text or cliparts and pre-defined styles versus own designed products. How many images are geo tagged or/and originated by mobile phones, point and shoot cameras or DSLRs.

Additionally we describe an own developed process to keep the quality standard the same over different factories and digital printers (electro photography and ink jet) supplied by different manufacturers. Customer satisfaction while expectations similar quality of an image to be viewed on a wall or seen in a photo book is the driver for the verified parameters during our production processes and the test set of images used to derive these parameters.

## 7866-14, Session 4

### A method to estimate the UV content of illumination sources

P. Green, Y. Chang,

The need to standardise the UV content of incident illumination has led to both ISO 3664 and ISO 13655 adopting precise requirements for the spectral power of illumination sources in both visible and UV regions of a daylight simulator.

By varying the power of UV and daylight-simulating sources, and measuring the incident irradiance and the reflected radiance from a selection of fluorescent and non-fluorescent substrates, it was found that for a given paper the fluorescent radiance increases linearly with incident UV irradiance. It was possible to model this relationship so that the UV energy (and hence the UV content of the source) could be predicted with reasonable accuracy from measurements of reflectance factor or reflected radiance.

Two particular applications where this is of interest are:

- estimating the UV content of an illumination source used in viewing colour samples, from a reflected radiance measurement of a sample
- estimating the UV content of an illumination source used in a spectrophotometer or other colour measurement instrument with an internal lamp.

We also describe how to estimate a UV Metamerism Index (MI-UV) value for common lamp types, using some assumptions about the typical spectral power distribution of such lamps in the UV region, and compare these estimated MI-UV values against those computed directly from the lamp spectral irradiance.

## 7866-15, Session 5

### Knowledge exchange in the CREATE project: Colour Research for European Advanced Technology Employment

C. E. Parraman, Univ. of the West of England (United Kingdom); A. Rizzi, Univ. degli Studi di Milano (Italy)

The presentation will review a 4 year European funded project CREATE (Colour Research for European Advanced Technology Employment), which was established in 2006. The group came together to promote and exchange research and knowledge through a series of conferences and training courses to researchers working in Europe who were in the early stages of their career. The long-term objective was to address a broad range of themes in colour and to develop with artists, designers, technologists and scientists a cross disciplinary approach to improving colour communication and education and to provide a forum for dialogue between different fields. Now at the end of its funding, this paper will highlight some of the key milestones of the project. Moreover, having completed a supplementary workshop event in October 2010, researchers considered new themes for a future Re-CREATE.

## 7866-16, Session 5

### Is it turquoise + fuchsia = purple or is it turquoise + fuchsia = blue?

G. B. Beretta, N. Moroney, Hewlett-Packard Labs. (United States)

The first step in communicating color is to name it. The second step is color semiotics. The third step is introducing structure in the set of colors. In color education at all levels, this structure often takes the form of formulas, like red + green = yellow, or turquoise + red = black. In recent times, Johannes Itten's color theory and its associated color wheel has been very influential, mostly through its impact on Bauhaus, although a number of color order systems and circles have been introduced over the centuries.

Students get confused when they are trying to formulate the color name arithmetic using the structure of color order systems and concepts like complementary colors and opponent colors. Suddenly turquoise + fuchsia = purple instead of blue; purple and violet become blurred, and finally the student's head explodes under epistemological pressures of Itten, Albers, Goethe, Newton, da Vinci, and all the other monsters of color structure.

In this contribution we propose a systematic presentation of structure in color, from color theories to color naming. We start from the concept of color perception introduced by da Vinci and work ourselves through color measurement, color formation, and color naming, to develop the basis for a robust system based on table lookup and interpolation.

One cause of confusion is that color naming has been quite loose in color theory, where for example red can be used interchangeably with fuchsia and blue with turquoise. Furthermore, common color terms are intermingled with technical colorant terms, for example cyan and turquoise or fuchsia and magenta. We present the evolution of a few color terms, some of which have experienced a radical transition over the centuries.

## 7866-17, Session 5

### Human vision based color edge detection

A. Kim, H. S. Kim, S. Park, Daejin Univ. (Korea, Republic of)

Edge detection can be of great importance to sharpness enhancement in various digital imaging applications such as digital television and camera. Therefore, extracting more accurate edge properties is significantly demanded for achieving a better image quality. In vector gradient edge detection, absolute difference of RGB values between a center pixel value and its neighborhood values are usually used, although such a device-dependent color space does not account for human visual characteristics well. The goal of this study is to test a variety of color difference equations and propose the most effective model that can be used for the purpose of color edge detection. A set of 7 color difference equations is selected and implemented in this study. Those include delta RGB, CIELAB delta E, CMC delta E, Leeds color difference, CIEDE94, CIEDE2000 (dE00) and CIECAM02-UCS delta E (dECAM-UCS). Consequently, there were not significant performance variations observed between those 7 color difference equations for the purpose of edge detection. However, dE00 and dECAM-UCS showed slightly higher z scores than the others. Observer accuracy was less than 20% value of CV so the agreement between the all human observers participated in this experiment can be thought of as reasonable.

## 7866-18, Session 5

### Color universal design: analysis of color category dependency on color vision type (2)

N. Kojima, M. G. Kamachi, Y. G. Ichihara, Kogakuin Univ. (Japan); K. Ito, The Univ. of Tokyo (Japan)

The present study investigates the tendency of individuals to categorize colors. Humans recognize colors by categorizing them using specific color names, such as red, blue, and yellow. When an individual with a certain type of color vision observes an object, they categorize its color using a particular color name and assume that other people will perceive the color in an identical manner. However, there are many variations in human color vision as a result of differences in photoreceptors in the eye, including red and green confusion. Thus, another person with a different type of color vision may categorize a color using a completely different name. To address this issue, we attempted to determine the differences in the ranges of color that people with different types of color vision observe. This is an important step towards achieving Color Universal Design, a visual communication method that is viewer-friendly irrespective of color vision type. Herein, we report on a systematic comparison among individuals with trichromacy (C-type), protan (P-type) and deuteran (D-type) color vision. This paper is a follow-up to SPIE-IS & T / Vol. 7528 752805-1.

## 7866-19, Session 6

### Object classification by color normalization or calibration?

W. Hans, D. W. Paulus, Univ. Koblenz-Landau (Germany)

Model-based approaches to object recognition rely on shape and contours while appearance-based approaches use information provided by the object intensity or color. Color histograms as an object characteristics are commonly used to solve this task and described in detail by Swain and Ballard. RGB color values formed by a camera depend heavily on the image formation process - especially the illumination involved. Mainly for this reason color normalization algorithms are applied to estimate the impact of position and color of the illumination and eliminate or at least minimize their influence to the image appearance.

The KOPID dataset used in our research consist of several items on carton boxes with many colors. Since the images in this set contains a GretagMacbeth ColorChecker c , another color normalization is applicable: color calibration. We compare several color normalization procedures to a calibration method proposed by Raymond L. Lee, Jr. By estimating the spectral reflectances of object surfaces one obtain a colorimetrically correct image representation.

Our experiments perform several histogram distance measures used for histogram based object classification. Additionally we vary the number of bins used, the order of some processing steps, and the dimensionality of color histograms to determine a most suitable parameter setting for object classification.

## 7866-20, Session 6

### Contrast preserving color fusion

J. Kamenicky, B. Zitova, Institute of Information Theory and Automation (Czech Republic)

We will address the problem of edge preserving image fusion for visualization and for printing of two intensity images from different modalities. The most important point is that instead of degrading the contained information by intensity output only, we will compute a color image in the way that gives us better possibility to control edge preservation.

The most common approach in these situations is the use of an alpha-blending, leading in some cases to worse visibility of edges present in one of the input images. The proposed method is meant to solve this issue by preserving intensity changes in the input images equally, independently of the other modality.

The main idea of the method is based on the perceptual color difference. A 2D rectangular color mapping scheme is created in such a way, that color differences as perceived by the human eye in all points are nearly the same. Then, this mapping scheme is applied to generate the output color.

The proposed method can help to distinguish even slight differences in any input modality without risk of losing details. Modifications of the proposed method can be derived for special cases where certain intensity intervals are more important than others.

## 7866-21, Session 6

### Evaluating the smoothness of color transformations

A. Aristova, Z. Wang, J. Y. Hardeberg, Gjøvik Univ. College (Norway)

Color image quality is an important factor in various media such as digital cameras, displays and printing systems. The employment of different color imaging media leads to a constant problem that each device produces color differently. It makes various manufacturers focus on the technology to achieve successful cross-media color reproduction. In this case the image reproduction quality depends

on processes of device characterization. The device characterization and profiling are central processes which allow predicting a result of device reproduction according to the known input, and provide communication of devices. Color look-up tables (LUTs) are the most common empirical approach for device characterization, and are the basis for ICC profiles. Correct 3DLUT-based color conversion in device characterization is an important factor for achieving high quality of the reproduced color image. Such factors as LUTs size, interpolation methods and unavoidable noise in color measurement process and unstable printing process influence on smoothness of 3DLUT-based color transforms, and may result in the appearance of artifacts in the final reproduced images. It is quite common to evaluate the quality of color transforms in terms of colorimetric accuracy, but smoothness is often neglected, even though its importance is now generally agreed on (Olson, 1999).

So an important problem in this case is to find a way of quantifying to which extent different transforms produce smooth or not smooth output images. The evaluation of smoothness of LUTs-based color transforms will allow avoiding undesirable results in image color reproduction systems such as artifacts and distortions of image content and improving device characterization process for achieving smooth color transforms. There are some scientific studies dedicated to evaluating smoothness of color transforms but the proposed algorithms were applied and tested only on well-designed 3D-LUTs, device characterization process and experimental data. So these metrics still require testing and evaluation using complex images and on profiles obtained during different measurements and in different environments.

A new method of evaluating smoothness of color transformations was proposed based on extension of second derivative method suggested earlier (Green, 2008). The algorithm is based on considering 3DLUTs of ICC profiles (AToB# table) as set of color planes in the printer color space and corresponding to them values in CIELAB color space. Each color plane in CIELAB color space presents set of color transitions in horizontal and vertical directions. The second derivative of  $\Delta L^*$ ,  $\Delta a^*$  and  $\Delta b^*$  between adjusted points of each vertical and horizontal transitions it was found. Statistical estimations were used for deriving general result among color planes for profiles.

In this research we have also proposed approaches using image difference metrics for evaluating smoothness of color transformations. For these goals 45 ICC profiles were generated from measurements with different repeatability and from measurements of consecutive printed charts on substrates of the same paper type for providing 3DLUTs with various noise characteristics. The process of profiling was designed to be close to a real practical case. Four test images containing smooth transitions of colors were converted using the profiles for obtaining images with varying smoothness. A psychophysical experiment involving 20 observers was conducted for evaluating perceived difference in smoothness between softcopy of original image and softcopies of images' reproductions using a category scale from perfect match to worst match in smoothness.

The proposed method have shown better performance in predicting smoothness of color transformation by particular profiles - compared with previous metrics by Green (2008), and Kim et al. (2010). Full-reference image quality metrics SSIM, GSSIM, pixel wise CIELAB Delta E, sCIELAB, Adaptive bilateral filter, Edge similarity, and Structural content - were compared for evaluating difference in smoothness between original and reproduced images. GSSIM and Structural content have shown higher Pearson correlation with visual judgments and representation than the other metrics.

## 7866-22, Session 6

### High capacity image barcodes using color separability

O. Bulan, G. Sharma, Univ. of Rochester (United States); B. Oztan, Rensselaer Polytechnic Institute (United States)

Two-dimensional barcodes are widely used for encoding data in printed documents. In a number of applications, the visual appearance of the barcode constitutes a fundamental restriction. In this paper, we propose high capacity color image barcodes that encode data in an

image while preserving its basic appearance. Our method aims at high embedding rates and sacrifices image fidelity in favor of embedding robustness in regions where these two goals conflict with each other. The method operates by utilizing cyan, magenta, and yellow printing channels with elongated dots whose orientations are modulated in order to encode the data. At the receiver, by using the complementary sensor channels to estimate the colorant channels, data is extracted in each individual colorant channel. In order to recover errors introduced in the channel, error correction coding is employed. Our results indicate that the proposed method can achieve embedding rates of traditional two-dimensional barcodes while preserving the appearance of the base image.

## 7866-23, Session 7

### The color side of dark

R. Bala, Xerox Corp. (United States)

No abstract available

## 7866-24, Session 7

### What a bad signal from this strange device

A. Rizzi, Univ. degli Studi di Milano (Italy)

No abstract available

## 7866-25, Session 7

### HDR imaging and color constancy: Two sides of the same coin?

J. J. McCann, McCann Imaging (United States)

Dark Side of Color

At first, we think that High Dynamic Range (HDR) imaging is a technique for improved recordings of scene radiances. Many of us think that human color constancy is a variation of a camera's automatic white balance algorithm. However, on closer inspection, glare limits the range of light we can detect in cameras and retinas. All scene regions below middle gray are influenced, more or less, by the glare from the bright scene segments. Instead of accurate radiance reproduction, HDR imaging works well because it preserves the details in the scene's spatial contrast. Similarly, human color constancy, also on closer inspection, depends on spatial comparisons that synthesize appearances from all the scene segments. Can spatial image processing play similar principle roles in both HDR imaging and color constancy?

## 7866-26, Session 7

### ICC profiles: are we better off without them?

G. B. Beretta, G. J. Dispoto, E. Hoarau, I. Lin, J. Zeng, Hewlett-Packard Labs. (United States)

Before ICC profiles, a device-independent page description document would encode all color in a device independent CIE space like CIELAB. When the document was to be printed, the press person would measure a target and create a color transformation from the CIE coordinates to device coordinates. For office and consumer color printers, the color transformation for as standard paper would be hardwired in the printer driver or the printer firmware.

This procedure had two disadvantages: the color transformations required deep expertise to produce and where hard to manage (the latter making them hard to share), and the image data was transformed twice (from input device to colorimetric and then to output device coordinates) introducing discretization errors twice. The first problem was solved with the ICC profile standard, and the last problem was



solved by storing original the device dependent coordinates in the document, together with an input ICC profile, so the color management system could first collapse the two profiles and then perform a single color transformation.

Unfortunately, there is a wide variety in the quality of ICC profiles (see figure at <http://www.mostlycolor.ch/2010/04/color-errors-in-printing.html>). Even worse, the real nightmare is that quite frequently the incorrect ICC profiles are embedded in page description documents or the color management systems apply the wrong profiles (see paper described in <http://www.mostlycolor.ch/2010/04/vanity-publishing.html>).

For consumer and office printers, the solution is to forgo ICC profiles and reduce everything to the single sRGB color space, so only one profile is required. However, the sRGB quality is insufficient for print solution providers. How can a modern print workflow solve the ICC profile nightmare?

## 7866-27, Session 7

### Green halftoning: Can less be more?

J. P. Allebach, Purdue Univ. (United States)

No abstract available

## 7866-28, Session 7

### Can displays go wild?

G. G. Marcu, Apple Inc. (United States)

No abstract available

## 7866-55, Poster Session

### Spectral reflection and transmission prediction model of halftone image on fluorescent supports

G. Shi, Y. Zhang, J. Chen, W. Ni, Jiangnan Univ. (China)

A new model which predicts the reflectance of halftone image on fluorescent supports is established in this article. The reflected light from fluorescent supports is divided into two parts: the primary streams which consist of originally incident light and the fluorescent streams which are created by absorption of the UV lights. By analyzing the different transmission paths of the primary lights flow and the fluorescence flow in ink-layer and paper, the total reflectance formula to predict the colors printed on fluorescent supports is derived. Besides the attenuation of the primary streams and the fluorescent streams in ink layer /paper bulk are considered, the ink-spreading model with which we can derive the effective dot coverage is accounted for. It can improve the accuracy of the prediction model after using the effective dot coverage in the total reflectance formula. With such a calibrated prediction model, we can predict the total reflectance of any halftone image on fluorescent supports. Then we use the Matlab software to make data simulation to verify the accuracy of the model we derived. At last, two reflectance curves were generated and it's proved that the new model significantly improve the accuracy of prediction, compared with the previous model.

## 7866-56, Poster Session

### The transmission of light affect the color reproduction of plastic print

J. Chen, Y. Zhang, G. Shi, Z. Xu, Jiangnan Univ. (China)

We analyze that the diffuse light transmission of the plastic printing affects the color reproduction of the product. The paper uses Kubelka-Munk model and radiative transfer theory, and considers the impact

of the plastic-printed film and the ink-color factors to study the color reproduction of plastic print. In this paper, we do not consider the white ink that was printed on the bottom of the plastic substrate firstly (then the plastic substrate was coated with color inks), and inside printing of the plastics. We only consider the transmission of light in the printing process was used in the hierarchical thinking, and consider the different optical properties of the ink layer and the plastic substrate. The paper has important significance on the security of printed plastic.

## 7866-57, Poster Session

### Reflectance model of plastic substrate halftone image based on Markov chain

W. Ni, Y. Zhang, G. Shi, J. Huang, Jiangnan Univ. (China)

Markov chain expresses the condition probability of a random sequence. It is relate to the latest state, is not relate to the previous state. The use of Markov chains provides two main advantages compared with the classical method of expressing multiple reflection-transmission processes by geometric series. The distance of the light scattering landscape orientation in lower biface is much larger than the size of dot while screening high frequency. So we just considered the situation that multilayer specimens composed of layers having distinct refractive indices and expanded the mathematic model. The transfer matrix of halftone homochromous image is the addition of the transfer matrix of every color and blank substrate. We obtain the reflectivity model of halftone plastic color presswork. This mathematic model is important to the digital workflow and autocontrol detection of presswork.

## 7866-58, Poster Session

### Color image segmentation on region growing and multiscale clustering

W. Wang, Henan Polytechnic Univ. (China)

This paper presents a color image segmentation method by combining region growing and color clustering algorithms. This method considers the both color and location information in a transformed color space. After multi-scale clustering (MSC), it does a spatial processing region growing. MSC can perform better in conquering the over-segment problem than equal distance clustering. Compared with the previous methods only depended MSC, the region growing can enhance the ability of noise suppression. This method inherits the idea that operates clustering first and then carries out spatial processing. Both clustering algorithm and spatial processing method are improved, so this method can get more satisfied results.

This paper presents a color image segmentation method by combining region growing and color clustering algorithms. This method considers the both of color information and location information in a transformed color space. After the MSC, it does a spatial processing region growing. MSC can do better in conquering the over-segment problem than equal distance clustering. Compared with the previous method only depended MSC, the region growing can enhance the ability of noise suppression. The whole algorithm does not need priori data and can find appropriate number of clusters adaptively.

Of course, there are many deficiencies in this algorithm need to be improved. For example: (1) when fix on the core pixel, the size of region D is not only associated with the size of the image, but also involved with attention to detail neglect. If we need to neglect more details, the size of region D will be bigger; and 2) in the MSC, one should join some special processing before and after the clustering to reduce the amount of calculation.

## 7866-59, Poster Session

### Advanced spectral reflectance prediction model for color prints

D. Tian, Q. Wang, Y. Zhang, Jiangnan Univ. (China)

We presented a model that enable the characterisation of a press in a environment with frequently changing inks. In the study, paper substrate is assumed to be a perfect diffuser, i.e., the intensity scattered by the paper follows Lambert law, the colorant layer is low scattering. The spectral reflectance prediction model is based on a description of the multiple reflections of light between the paper substrate and the print-air interfaces. In our experiment, we tested our model on an offset printing machine and obtained an accuracy in terms of root mean square error of 0.58% and of 1.21 dE. Good agreement was found between the simulations and both the analytical and measured reflectances for the printed patches. The model only needs a single print of a target and a spectral measurement of the new ink set for the future print job. The performance of the system lies at about 1-2 dE from the measurement in the root mean square error. The model could therefore be used in printing systems to print machine characterize. It could also be used to predict the color of images as a function of illumination and viewing angles.

## 7866-60, Poster Session

### Color reproduction performance of halftone color image printed on paper substrate

W. Ni, Y. Zhang, G. Shi, J. Huang, Jiangnan Univ. (China)

We can obtain the light intensity and reflectivity which send up from screen dots or the paper among screen dots after analyzing ink layer probability transfer matrix when the source and detector are on the top, ink layer probability transfer inverse matrix when the source is in the bottom and detector is on the top, paper probability transfer matrix when the source and detector are on the top and paper probability transfer inverse matrix when the source is in the bottom and detector is on the top synthetically. By the relationship between  $P_i(a)$  and  $P_p(a)$ , we can get the reflectivity of presswork.

In the new model, the effort of ink spreading (physics dot gain), ink penetration and optics dot gain are considered. From our new model, the reflectivity of homochromy halftone screen dot can be obtained. When multilayer ink combined, three kinds of ink can form screen dots of 8 kinds of colors called Neugebauer Primaries. If overprint screen dot area satisfies statistically independent, we can gain spectrum reflectivity of color halftone presswork. This mathematic model is important to the digital workflow and autocontrol detection of presswork.

## 7866-61, Poster Session

### Color image segmentation on region growing and multiscale clustering

J. Sun, W. Wang, Z. Jia, Henan Polytechnic Univ. (China)

This paper presents a color image segmentation method by combining region growing and color clustering algorithms. This method considers the both color and location information in a transformed color space. After multi-scale clustering (MSC), it does a spatial processing region growing. MSC can perform better in conquering the over-segment problem than equal distance clustering. Compared with the previous methods only depended MSC, the region growing can enhance the ability of noise suppression. This method inherits the idea that operates clustering first and then carries out spatial processing. Both clustering algorithm and spatial processing method are improved, so this method can get more satisfied results.

This paper presents a color image segmentation method by combining region growing and color clustering algorithms. This method considers the both of color information and location information in a transformed color space. After the MSC, it does a spatial processing region growing. MSC can do better in conquering the over-segment problem than equal distance clustering. Compared with the previous method only depended MSC, the region growing can enhance the ability of noise suppression. The whole algorithm does not need priori data and can find appropriate number of clusters adaptively.

Of course, there are many deficiencies in this algorithm need to be

improved. For example: (1) when fix on the core pixel, the size of region D is not only associated with the size of the image, but also involved with attention to detail neglect. If we need to neglect more details, the size of region D will be bigger; and 2) in the MSC, one should join some special processing before and after the clustering to reduce the amount of calculation.

## 7866-62, Poster Session

### Regression based characterization of color measurement instruments in printing applications

P. Nussbaum, J. Y. Hardeberg, Gjøvik Univ. College (Norway)

In the context of print quality and process control colorimetric parameters and tolerance values are clearly defined. Although, calibration procedures are well defined for color measurement instruments, in printing workflow using more than one color measurement instrument measuring the same color wedge can produce obviously different results. In certain situations where one instrument gives values which are just inside the given tolerances and the second measurement instruments produces values which exceeds the predefined tolerance parameters the question arises whether the print or proof is approved or not according to standard parameters.

The aim of this paper was to find an appropriate model to characterize color measurement instruments for printing applications to reduce color differences due to instrument uncertainties. The method used is derived from color measurement instrument characterization methods which have been applied by performing the polynomial regression with least square technique. Six color measurement instruments were used measuring color patches of a control color wedge on three different types of substrates. The characterization functions were derived using polynomial regression and the non linear optimization routine, based on the training set of 14 color reference patches and the corresponding colorimetric measurements obtained by the measurement instruments. The derived functions are then used to predict the colorimetric values of 46 color wedge patches independent of the training set. The estimated colorimetric values from one instrument were then compared to the estimated colorimetric values from a different measurement instrument. Comparing the color differences of the raw measurement data obtained by two instruments, the applied characterization model was reducing the color differences significant dependent on the instrument combination (product families).

## 7866-63, Poster Session

### Printing anaglyph maps optimized for display

H. Zeng, Hewlett-Packard Co. (United States); R. Zeng, Liming Univ. (China)

Although anaglyphs have a big advantage that they can be presented using traditional single channel media such as print, film, display, etc., a media type must be determined as a pair of views is combined into a single image to minimize retinal rivalry and stereo crosstalk. Most of anaglyph maps and map tools are optimized for display and assumed using red-cyan filtered glasses for viewing. Due to the large difference between a display gamut and a printer gamut, red and cyan colors that are used to separate the left view and the right view are changed considerably as they are mapped from a display color space to a printer color space for printing and results in serious retinal rivalry. A solution using a special gamut mapping method to preserve the relative relationship of cyanish and reddish colors was developed to gamut map colors from display to printer. And the color characterization to balance neutral colors for specific red/cyan glasses is applied to further improve the color appearance.

## 7866-64, Poster Session

### **A restoration method for book scan images**

H. Ohk, S. H. Kim, D. Choi, Samsung Electronics Co., Ltd.  
(Korea, Republic of)

When a bound document such as a book is scanned or copied with a flat-bed scanner, there are two kinds of defects in the scanned image; the geometric and photometric distortion. The root cause of these two defects is the imperfect contact between the book to be scanned and the scanner glass plate. The long gap between the book center and the glass plate causes the optical path from the surface of the book and the imaging unit (CCD/CIS) to be different from the optimal condition. The main distortion of photometric one is "Shadow Distortion". Near the spine, reflectance of light is reduced because of the distance between book surface and document glass and the curvature of book surface. And it is the main reason of "shadow distortion". This shadow distortion can make spine region entirely black in copy process. And Geometric distortion occurs because of structure of lens and CCD. In scanner, there is lens at the middle of CCD or CIS. If a scan object is laid on document glass, its length at CCD is measured correctly. But if a scan object is not contacted with document glass, its length is shorter than original one.

We suggest a method for restoring bound document scan images without any additional information or sensor. We correct the bound document images based on the estimation of the boundary feature and background profile. Boundary Feature is obtained after calculating and analyzing the Minimum Boundary Rectangle which encloses the whole foreground contents with minimum size. From the boundary information we can estimate some information - folding point position, left and right page information, skew angle of each page and so on and the extracted feature is used for correcting geometric distortion de-skew, warping, and page separation. Background profile is estimated from the gradient map and it is utilized to correct photometric distortion; exposure problem. Experimental results show effectiveness of our proposed method.

In this paper, we propose a solution for copying or scanning thick, bound document. This solution contains "Page separating", "Shadow Distortion Correcting", "Skew Correction" and "Perspective Distortion Correction". Page separation cuts each page of book automatically and "Shadow Distortion Correction" corrects shadowed area near the spin region and "Perspective Distortion Correction" warps input image with effect.

## 7866-29, Session 8

### **Soft proofing of printed colours on substrates with optical brightening agents**

N. S. Parab, P. J. Green, London College of Communication  
(United Kingdom)

The appearance of colours printed on substrates with optical brightening agents has been studied with help of a colour matching experiment where the observers matched a colour patch displayed on a LCD monitor, by adjusting its  $L^*a^*b^*$  values, to another color patch printed out on paper viewed under varying amount of UV content in lighting condition in the viewing booth. A customised viewing booth was built for this purpose and substrates with varying amount of optical brighteners were considered for the study.

A model based on CIECAM02 and a scaling technique has been developed to predict the perceived colour match on a LCD display, of colours printed on substrates with optical brighteners and viewed under the viewing booth with varying amount of UV content in the viewing illumination. According to the obtained results, the appearance of the colours printed on substrates containing optical brighteners varied with variation in the amount of UV content in the viewing illumination. The developed model gave good prediction of the XYZ tristimulus values for the perceived match on the LCD display from the XYZ tristimulus values of the printed colours on the substrate with acceptable  $\Delta E_{ab}$ . This shows that CIECAM02 can be effectively used for soft proofing.

## 7866-30, Session 8

### **Ghostscript color management**

M. J. Vrhel, R. P. Johnston, Artifex Software Inc. (United States)

Ghostscript is a well known open source document conversion engine introduced in 1986 by L. Peter Deutsch. Initially, Ghostscript was designed to render PostScript documents but has grown to handle conversions amongst several high level vector languages including PostScript (PS), PDF, XPS, SVG and PCL. Today, Ghostscript is found on every Linux system and is still freely available for users and commercial development under GPL licensing.

The initial design of Ghostscript was well before the definition of the ICC format and likely even before there was much thought about digital color management. When PS color management (PCM) was introduced by Adobe, it was incorporated into Ghostscript and became the primary method to achieve managed color within Ghostscript through the use of color rendering dictionaries (CRDs) and color space arrays. ICC support to Ghostscript was added with the support of PDF1.3 but the solution was restricted in a performance sense due to the architecture relying upon PS color management.

This paper reviews a significant update to the color architecture within Ghostscript. The ICC-based design operates efficiently in Ghostscript's multi-threaded rendering environment, allows easy interfacing of external CMMs and includes conversion of PS and PDF color objects to ICC objects.

## 7866-31, Session 8

### **Color control of a lighting system using RGBW LEDs**

M. Tanaka, T. Horiuchi, S. Tominaga, Chiba Univ. (Japan)

A lighting system is proposed to render objects under a variety of colored illuminations. The system is constructed with a LED unit, white diffusion filters, dimmers, and a personal computer as a controller. The LED unit is composed of four kinds of color LED lamps which are 12 red (R), 14 green (G), 12 blue (B) and 10 white (W) colors. We determine the digital control signals of the RGBW lights for generating the colored light with the target XYZ tristimulus values. Then we have a mapping problem from the 3D color space of the tristimulus values to the 4D control space of the RGBW digital values. We develop an effective algorithm by decomposing the target XYZ values into (XYZ) rgb for RGB lights and (XYZ)w values for W lights. When a colored light with the target tristimulus values can be generated without the white LED lights, the conventional XYZ-RGB color coordinate conversion is used for obtaining input RGB values only. For a high luminance range in the XYZ space, the XYZ-RGBW coordinate conversion is used for determining input RGBW digital values. The accuracy of the colored light is examined with regard to the CIE color difference.

## 7866-32, Session 9

### **Brightness contrast under high surround luminance levels: psychophysical data vs CIECAM02**

Y. S. Baek, H. S. Kim, S. Park, Daejin Univ. (Korea, Republic of)

This study aims to evaluate brightness contrast calculated by CIECAM02 and compare the performance with the psychophysical. The surround conditions were changed 7 levels from Dark to Over Bright (2087 cd/m<sup>2</sup>). For this study, six neutral colors used as test images uniformly filled the entire display screen. Psychophysical experiments were carried out to investigate the relation between perceived brightness and luminance using magnitude estimation method. The reference brightness was set white test image under Average 3 (200 cd/m<sup>2</sup>). To investigate perceived image contrast, the Michelson contrast was calculated using perceived brightness of white and black test image. As a result, we found out that the perceived image contrast is highest at Average 3 surround condition. Consequently, brightness

contrast increases as surround luminance increases but it decreases from average surround condition to over bright surround condition. Furthermore, Michelson contrast was computed using brightness of CIECAM02. Because CIECAM02 considers only three surround conditions, it has same brightness contrast regardless of the variation in the surround luminance. Consequently, this shows different result with our result. Higher surround luminance levels should be taken into account.

## 7866-33, Session 9

### LabRGB: optimization of bit allocation

F. Nakaya, Fuji Xerox Co., Ltd. (Japan)

Spectral distribution can be written as a linear combination of eigenvectors and the eigenvectors method gives the least estimation error, but eigenvectors depend on a sample selection of population and encoding values have no physical meaning. Recently reported LabPQR is to convey physical values, but still is dependent on a sample selection of population. Thus, LabRGB, was proposed in 2007. LabRGB is to provide "sample selection of population" free spectral encoding/decoding methods, which consists of six unique trigonometric base functions and physically meaningful encoding values. LabRGB was applied to the real multi spectral images and showed a good performance in spectral as well as colorimetric estimation. In this paper, the allocation of a bit depth to the weighting factors is examined in terms of spectral and colorimetric distance of nearest neighbors. The optimum way to minimize the unusable combination of weighting factors is obtained by using the correlation of the weighting factors. The optimum way to minimize the spectral and colorimetric distance of nearest neighbors is also obtained by using the non linear mapping method. The two methods thus obtained give a good clue for explicitly defining the bit depths of respective scores for the future applications and standardization.

## 7866-34, Session 9

### Spatio-temporal colour correction of strongly degraded movies

A. B. M. T. Islam, I. Farup, Gjøvik Univ. College (Norway)

We present a method for digital colour restoration of strongly degraded movie material. The method is based upon the existing STRESS algorithm. In order to cope with the problem of highly correlated colour channels, we implemented a preprocessing step in which saturation enhancement is performed in a PCA space. Spatial colour algorithms tends to emphasise all details in the images, including dust and scratches. Surprisingly, we found that the presence of these defects does not affect the behaviour of the colour correction algorithm. Although the STRESS algorithm is already in itself more efficient than traditional spatial colour algorithms, it is still computationally expensive. To speed it up further, we went beyond the spatial domain of the frames and extended the algorithm to the temporal domain. This way, we were able to achieve an 80 percent reduction of the computational time compared to processing every single frame individually. We performed a user experiment, and found that our method produces significantly better results than the existing methods. Thus, our method outperforms the existing ones in terms of both visual quality and computational efficiency.

## 7866-35, Session 9

### Color correction optimization with hue regularization

H. Zhang, H. Liu, Oregon State Univ. (United States); S. Quan, Broadcom Corp. (United States)

Previous work has suggested that observers are capable of judging the quality of an image without any knowledge of the original scene.

When reference is not available, observers extract the apparent objects in an image and compare them with typical colors of similar objects recalled from their memories. Some generally agreed research results indicate that although perfect colorimetric rendering is not conspicuous and color errors can be well tolerated, observers do perceive some memory colors such as skin, grass, and sky fairly consistently and remember them with slightly different hues and higher saturation than their originals, with specific preferences. Appropriate rendition of these colors is necessary and contributes heavily to the overall perceived image quality.

A color correction matrix is the transformation converting the image data from a device dependent color space to a target color space. Color correction matrix can be obtained through linear regression between the two color spaces, minimizing the mean square Euclidean error in certain colorimetric coordinates. Unfortunately, this method could result in objectionable distortions if the color errors biased memory colors undesirably.

We propose a color correction optimization method with preferred color reproduction in mind through hue regularization. Preferred color reproduction will be reviewed, and an optimization method will be proposed using recursive linear regression with additional constraints for hue regularization.

## 7866-36, Session 10

### Spectral model of an electro-photographic printing system

M. A. Kriss, MAK Consulting (United States)

At EI 2007 in San Jose, California a detailed physical models for monochrome and color electro-photographic printers was presented. These models were based on computer simulations of toner-dot formation for a variety of halftone structures. The optical interactions between the toner-dots and the paper substrate were incorporated by means of an optical scattering function, which allowed for the calculation of optical dot-gain (and physical dot-gain) as function of the halftone structure. The color model used simple red-green-blue channels to measure the effect of the absorption and scattering properties of the cyan, magenta, yellow and black toners on the final half-tone image. The new spectral model uses the full absorption and scattering spectrum of the image toners in calculating the final color image in terms of CIE XYZ values for well-defined color and gray patches. The new spectral model will be used to show the impact of halftone structure and toner-layer-order on conventional dot-on-dot and rotated dot color halftone systems and how to minimize the impact of image toner scattering. The model has been expanded to use the Neugebauer equations to approximate the amount of cyan, magenta, and yellow toners required to give a "good" neutral in the rotated dot halftone and fine tuning is achieved by adjusting the development threshold level for each layer to hold a good neutral over the full tonal range. Once a "good" neutral is obtained the impact on dot gain, color reproduction and optimum layer order can studied with an emphasis on how the full spectral model differs from the simpler three-channel model. The model is used to explore the different approaches required in dot-on-dot and rotated dot screens to achieve good results. In the future the model will be applied to stochastic screens.

## 7866-37, Session 10

### Optimized selection of image tiles for ink spreading calibration

T. Bugnon, R. D. Hersch, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The Yule-Nielsen modified spectral Neugebauer model (YNSN) enables predicting reflectance spectra from surface coverages. In order to provide an improved prediction accuracy, this model is enhanced with an ink spreading model accounting for ink spreading in all superposition conditions (IS-YNSN). As any spectral reflection prediction model, the IS-YNSN model is designed to predict the reflection spectra of uniform patches. Instead of uniform patches,

we investigate if tiles located within color images can be accurately predicted and how they can be used to facilitate the calibration of the ink spreading model. In the present contribution, we first detail an algorithm to automatically select image tiles based uniquely on the CMY or CMYK pixel values of these color images and show that such image tiles can be accurately predicted by the IS-YNSN model provided that they are uniform enough. This selection algorithm incorporates additional constraints and is verified on 6 different color images. We finally demonstrate that the ink spreading model can be calibrated with as few as 5 to 10 image tiles provided that the image tiles are chosen by applying the proposed additional constraints.

## 7866-38, Session 10

### **A preferred skin color enhancement method for photographic color reproduction**

H. Zeng, Hewlett-Packard Co. (United States); M. R. Luo, Univ. of Leeds (United Kingdom)

Skin tones are the most important colors among the memory color category. Reproducing skin colors pleasingly is an important factor in photographic color reproduction. Moving skin colors toward their preferred skin color center improves the color preference of skin color reproduction. Several methods to morph skin colors to a smaller preferred skin color region has been reported in the past. In this paper, a new approach is proposed to further improve the result of skin color enhancement. An ellipsoid skin color model is applied to compute skin color probabilities for skin color detection and to determine a weight for skin color adjustment. Preferred skin color centers determined through psychophysical experiments were used for color adjustment. Preferred skin color centers for dark, medium, and light skin color categories are applied to adjust skin colors differently. Skin colors are morphed toward their preferred color center in both chroma and hue angle. A special processing for highlight skin colors is applied to avoid contrast loss in highlight. A 3-D interpolation method is applied to fix a potential contouring problem and to increase color processing efficiency. An initial psychophysical experiment validates that the method of preferred skin color enhancement effectively identifies skin colors, improves the skin color preference, and does not objectionably affect preferred skin colors in original images.

## 7866-39, Session 10

### **Kubelka-Munk theory for efficient spectral printer modeling**

M. A. Abebe, J. Gerhardt, J. Y. Hardeberg, Gjøvik Univ. College (Norway)

In spectral color reproduction, we reproduce a color based on its spectral reflectance rather than on its colorimetric values. This has the potential of increasing the colour fidelity quality of the reproduction under different illumination conditions and gives higher general color accuracy.

In our work we particularly focus on the quality of spectral color image reproduction by multi-channel inkjet printing; a key element of this process is to accurately model the colorimetric and spectral behavior of the printer. For instance the Yule-Nielsen modified spectral Neugebauer (YNSN) model is a very much used spectral printer model. In this model the spectral reflectance of the different colorant combinations is estimated as a convex combination of the reflectances of the so-called Neugebauer Primaries (NP), which are the primaries, secondaries, tertiaries, etc. In order to set up such a model for a printing system with a high number of colorant channels, many color patches must be measured; in fact the necessary number of patches increases exponentially with the number of channels, this is a costly and time-consuming task. Another problem with this approach is that for NPs of more than 3 colorants, limitations of paper total ink coverage start to play an important role.

We propose to use Kubelka-Munk theory to estimate the spectral reflectances of the Neugebauer Primaries instead of printing and

measuring them, and subsequently to use these estimated NPs as the basis of our YNSN printer modeling. We have evaluated this approach experimentally on several different paper types and on the HP Deskjet 1220C CMYK inkjet printer and the Xerox Phaser 7760 CMYK laser printer, using both the conventional spectral Neugebauer (SN) model and the YNSN model.

Using this approach we find that we achieve not only less time consuming model establishment, but also, somewhat unexpectedly, improved model precision over the models using the real measurements of the NPs. We have also investigated a hybrid model with mixed NPs, half measured and half estimated.

Our results show us the reasonability of the Kubelka-Munk theory for spectral printer modeling. The results differ from printer to printer and from paper to paper. The spectral estimations of both SN and YNSN models in laser printers perform much better than for inkjet printers. We also see that using simple and very cheap copy paper will give us even better performances than using some expensive photo papers. The paper property which seems to be the most important factor for the quality of the model is the opacity, the higher the opacity the higher the performance of the KM theory will be.

## 7866-40, Session 11

### **A simple color prediction model based on multiple dot gain curves**

Y. Y. Qu, S. Gooran, Linköping Univ. (Sweden)

Summary: Most of the color prediction models use a single dot gain curve for each ink, few mention dot gain changes in different ink superposition situations, but still they figure only one dot gain curve for each possible combination situation.

In this paper we present a simple color prediction model. In this model we utilize three different dot gain curves for each primary ink obtained by CIEX, Y and Z, which approximately stand for three different wavelength bands.

In addition, we noticed that the dot gain curves for single ink printed on paper are not representing the dot gain for the same ink when printed on another ink. Therefore, dot gain curves for different ink overlapping situations are optimized by matching calculation of special training patches to the corresponding measured tristimulus values. Regarding certain ink combination, for each ink we finally figure out three dot gain curves, each of which is a combination of different dot gain curves that are weighted based on their probability.

Our work presents a feasible color prediction model considering both ink superposition and light wavelength influence.

## 7866-41, Session 11

### **Subsampled optimal noise management method for a robust separation based calibration of color printing systems**

M. Qiao, J. M. Sanchez, Y. Chen, I. Case, G. Lin, Xerox Corp. (United States)

For many color printing systems, printer calibration is often utilized to return the printer to a known state to ensure consistent color output. In particular, the key visual response of color balance is often controlled by the calibration state return. Input color signal noise, generated from the printing system natural variation when printing the calibration target, affects the accuracy and robustness of the calibration outcome. Noise management techniques for managing input color signal noise prior to system calibration are often absent or rely on ad hoc analysis and are usually not based on the return of a well developed printer response that has been extracted from measured signal using advanced noise management methods. This paper describes Part II of an overall method for developing a robust noise management system for printer calibration. In Part I, an 8-bit full resolution calibration target is described and an iterative filtering noise management metric and method are defined and developed. In this Part II, the specific

development of a low resolution calibration target and corresponding noise free representation of the printer system state, as defined by quantitative metrics relative to the printer response derived from high resolution signal in Part I is defined and developed. This subsampled calibration target using the proposed noise management method can increase the productivity and reduce operator error in print shop workflow with minimal loss of accuracy.

## 7866-42, Session 11

### Investigating the wavelength dependency of dot gain in color print

M. Namedanian, S. Gooran, D. Nyström, Linköping Univ. (Sweden)

Fully characterizing physical and optical dot gain is useful for system calibration and quality control of the color reproduction. The main purpose of this paper is to study the optical dot gain behavior of color prints. We present an approach to separate optical and physical dot gain by using microscopic images. A high resolution camera (2  $\mu\text{m}/\text{pixel}$ ) equipped with a set of seven broadband interference filters illuminating light in 400 nm to 700 nm wavelength band, is used. The camera system can capture both reflected and transmitted images. This phenomenon has been used to separate the optical and physical dot gain. By using the seven interference filters the wavelength dependency of optical dot gain has been studied. As black ink has the absorbing wavelength band of 400 nm to 700 nm, its optical dot gain in different wavelengths has been investigated and the results show that light scattering of the paper is wavelength independent. This means that Point Spread Function (PSF) is independent of the ink. By using the PSF and the physical dot shapes of a color halftoned image, it is possible to predict the resulting color including the effect of optical dot gain.

## 7866-43, Session 11

### Fast approach for toner saving

I. V. Safonov, I. V. Kurilin, M. N. Rychagov, Samsung Electronics Co., Ltd. (Russian Federation); H. K. Lee, S. H. Kim, D. Choi, Samsung Electronics Co., Ltd. (Korea, Republic of)

Saving of toner consumption is an important task in modern printing devices and has a significant ecological impact. Existing toner saving approaches have two main drawbacks: appearance of hardcopy in toner saving mode is worse in comparison with normal mode; processing of whole rendered page bitmap requires significant computational costs.

We propose to add small holes of various shapes and sizes to random places inside a character bitmap during font rendering before storing character bitmap to font cache. This scheme is based on processing pipeline in RIP of standard printing languages Postscript and PCL. Processing of text characters only, and moreover, processing of each character for given font and size alone, is an extremely fast procedure. The approach does not deteriorate halftoned bitmap and business graphics and provide toner saving for typical office documents up to 15-20%. Rate of toner saving is adjustable.

Appearance of characters is almost indistinguishable in comparison with solid black text due to random placement of small holes inside the character regions. The suggested method automatically produces no toner saving on small fonts, since preserves quality of small fonts. Readability of text processed by proposed method is fine. OCR programs process that scanned hardcopy successfully too.

## 7866-44, Session 11

### A virtual printer and reference printing conditions

P. Green, London College of Communication (United Kingdom);

C. Revie, FFEI UK (United Kingdom); D. McDowell, Eastman Kodak Co. (United States)

In a late binding workflow, data is commonly prepared in an output-referred state based on a reference intermediate RGB colour encoding. Such encodings may have a larger gamut than the target printing condition, and so there is some ambiguity over how to preview the data before it has been converted to the target printing condition.

Here we propose an additional intermediate encoding, referred to as a 'virtual printer' which bridges the gap between three-component reference RGB or PCS encodings, and reference CMYK printing conditions.

The virtual printer has a large colour gamut which represents a superset of most available print gamuts. It is defined here in terms of the reflectance and colorimetric coordinates of the virtual colorants, and associated colour mixing model.

When used in a colour reproduction workflow, documents can be initially rendered to the printer-like gamut of the virtual printer, and channel preferences (such as black generation) can be defined. Re-rendering to a reference printing condition and associated colour gamut is deferred, thus supporting re-purposing of the document.

By transforming a colour document to virtual printer CMY or CMYK, it is possible to perform editing, preview and channel specification operations prior to re-rendering to a reference printing condition. In conjunction with the reference printer, whose colour gamut is limited to a specific printing process, the virtual printer provides robust support for late binding workflows in the graphic arts.

## 7866-45, Session 12

### Cost function analysis for stochastic clustered-dot halftoning based on direct binary search

P. Goyal, M. Gupta, Purdue Univ. (United States); C. Staelin, M. Fischer, Hewlett-Packard Labs. Israel Ltd. (Israel); O. Shacham, Hewlett-Packard Indigo Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

No abstract available

## 7866-46, Session 12

### Stochastic clustered-dot screen design for improved smoothness

M. Gupta, P. Goyal, Purdue Univ. (United States); M. Fischer, C. Staelin, Hewlett-Packard Labs. Israel Ltd. (Israel); T. Kashti, O. Shacham, Hewlett-Packard Indigo Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

No abstract available

## 7866-47, Session 12

### Design of color screen tile vector sets

J. Kim, Y. Chen, Purdue Univ. (United States); M. Fischer, Hewlett-Packard Labs. Israel Ltd. (Israel); O. Shacham, Hewlett-Packard Indigo Ltd. (Israel); C. Staelin, Hewlett-Packard Labs. Israel Ltd. (Israel); K. Bengtson, Hewlett-Packard Co. (United States); J. P. Allebach, Purdue Univ. (United States)

No abstract available

7866-48, Session 12

### UV Fluorescent Encoded Image Using Two Successive Filling Halftone Algorithms

Y. Zhao, S. Wang, Xerox Corp. (United States)

Methods are provided for creating a fluorescent watermark within an image on a substrate, such as paper. The method involves creating a halftone image using two different halftone strategies. The halftone method is combined with a binary watermark mask to form two color patterns (e.g., one in a background region of the image and one in a watermark region of the image) and two successive-filling halftone algorithms, such that the ink droplets deposited by one color pattern cover more of the substrate than the ink droplets deposited by the other color pattern, with the two color patterns having approximately the same reflectance under normal light. However, under UV illumination, a visible difference (e.g., the watermark) is seen in the two patterns.

7866-49, Session 12

### Moiré-free color halftoning using hexagonal geometry

R. P. Loce, S. Wang, Xerox Corp. (United States)

No abstract available

7866-50, Session 13

### A hybrid adaptive thresholding method for text with halftone pattern in scanned document images

S. Yu, W. Ming, Konica Minolta Systems Labs., Inc. (United States)

In this paper, a hybrid adaptive thresholding method for scanned document images containing text with halftone pattern is presented. The method is based on the topological feature and gray level statistics of halftone text. Histogram only based thresholding methods often miss some halftone text after binarization, especially with close to background gray level halftone text. The proposed method divides the document image into non overlap windows and extracts text characters as connect component in each window. The Euler number of each text character is then calculated and used as topological features to identify halftone text. After most the halftone text are identified in each window, the pixel value statistics of the halftone text are estimated. Halftone text are first segmented out by using the threshold determined by their pixel value statistics. Then a global threshold is calculated for the remaining pixels in the image to segment out dark text. The final binarization result of the document image is obtained by combining the binarization results of the halftone text and dark text. Comparing to methods based only on histogram, satisfied binarization results are obtained when testing the proposed method on scanned document images containing halftone text.

7866-51, Session 13

### Window-based spectral analysis of periodic color halftone screens

A. H. Eid, B. E. Cooper, E. E. Rippetoe, Lexmark International, Inc. (United States)

Improper design of color halftone screens may create visually objectionable moiré patterns in the final prints due to the interaction between the halftone screens of the color primaries. The prediction of such interactions from the screens' bitmaps helps to identify and avoid problematic patterns, reducing the time required to design effective

color halftone screens.

In this paper, we detect the moiré patterns by examining the spatial frequency spectrum of the mixed screens. We study different windowing techniques including Hamming, Hanning, and Blackman, to better estimate the moiré strength, frequency and orientation. The window-based spectral estimation has the advantage of reducing the effect of spectral leakage associated with the non-windowed discrete signals.

Two methods are used to verify the detected moiré from the bitmaps. First, we analyze scans of the printed halftones, using the same technique that we applied to the bitmaps. Second, we independently inspect the printed halftones visually. Our experiments show promising results by detecting the moiré patterns from both the bitmap images as well as the scans of the actual prints verified by visual inspection.

7866-52, Session 13

### Descreening of color halftone images in the frequency domain

C. J. Stanger, T. Tran, E. H. Barney Smith, Boise State Univ. (United States)

Color printers use three colors, cyan, magenta, and yellow, and sometimes black when printing color images. A wide range of colors is created by varying the size of the dots and the relative contribution of the color primaries. The perceived color depends on the size of the dots and the different combinations of the three colors at different angles. This type of printing is called halftoning. Scanning a halftone image introduces halftone artifacts, known as Moiré patterns, which significantly degrade the image quality. Printers that use amplitude modulation (AM) screening for halftone printing position dots in a periodic pattern. Therefore, frequencies relating to halftoning are easily identifiable in the frequency domain. This paper proposes a method for descreening scanned color halftone images using a custom band reject filter designed to isolate and remove only frequencies related to halftoning while leaving image edges sharp without image segmentation or edge detection. To enable hardware acceleration, the image is processed in small overlapped windows. The windows are filtered individually in the frequency domain, then pieced back together to show the entire filtered image without noticeable blocking artifacts.

7866-53, Session 13

### Analog image backup with steganographic halftones

R. A. Ulichney, I. Tastl, E. Hoarau, Hewlett-Packard Labs. (United States)

Hard copy (analog) backup of photographs is an important alternative to digital storage. It offers a means to visually enjoy the "storage format" decoupled from a digital storage media which can have a shorter archival life than hard copy, along with shorter lifetime of hardware support. The paper describes a means to eliminate the need to include unsightly text that is part of earlier solutions by embedding all required metadata in a small steganographic halftone with the print. The only hardware requirement is an image scanner, which we can safely assume will be available far into the future when readers of today's digital storage media will be long gone. Examples of the resulting archival compositions and metadata-embedded halftones will be included.

# Conference 7867: Image Quality and System Performance VIII

Monday-Wednesday 24-26 January 2011

Part of Proceedings of SPIE Vol. 7867 Image Quality and System Performance VIII

## 7867-01, Session 1

### Image quality metrics for the evaluation of print quality

M. Pedersen, J. Y. Hardeberg, Gjøvik Univ. College (Norway); N. Bonnier, Océ Print Logic Technologies (France); F. Albregtsen, Univ. of Oslo (Norway)

Image quality metrics have become more and more popular in the image processing community. Many different image quality metrics have been proposed, often with the goal of being able to predict perceived image quality.

However, so far, no one has been able to define an image quality metric well correlated with the percept for overall image quality.

In our research we have presented a set of quality attributes built on existing attributes from the literature. The six proposed quality attributes are: sharpness, color, lightness, artifacts, contrast, and physical. An experiment validates the quality attributes as suitable for the evaluation of image quality.

We have then proposed to use image quality metrics for each quality attributes in order to predict perceived image quality. A selection of suitable of image quality metrics for the different quality attributes has been carried out.

Each of the quality attributes has been investigated, and an experimental analysis has been carried out to find the most suitable image quality metrics for each of the given quality attributes

The process of applying image quality metrics to printed images is not straightforward as image quality metrics require a digital input. The printed images need to be transformed from physical copies to digital copies in order to apply metrics. A framework has been developed for this process, which includes the transformation to a digital format, image registration, and the application of image quality metrics.

The results indicate that image quality metrics can correlate with the percept for certain quality attributes, but they are not correlated with overall perceived image quality. Therefore the use of quality attributes together with image quality metrics is interesting, and very promising.

## 7867-02, Session 1

### Hyper error map based document stitching

L. C. Cui, Lexmark International, Inc. (United States)

Document stitching can generate unpleasant stitching artifacts. Here we examine one type of stitching artifacts, the misaligned features and edges, and propose a method to minimize that based on one performance characteristics of the human vision, the hyper acuity.

## 7867-03, Session 1

### Quantification of perceived macro-uniformity

K. Lee, Y. Bang, H. Choh, Samsung Electronics Co., Ltd. (Korea, Republic of)

INCITS W1.1 team defined macro-uniformity, categorized it into five types of sub-attributes, and attempted to evaluate it based on quality ruler method. Quality ruler method is easily used for the end user to judge the level of print defect. However, the precise quantification method without ruler images is more helpful to developer to analyze printing system components affecting print uniformity and to communicate between developers.

In this paper, we propose a method to quantify perceived macro-

uniformity for a given test print. We suppose that macro-uniformity is perceived by blending four kinds of sub-attributes: banding, streaking, 2D noise, and gradient. Each attribute is independently measured by the developed method. The measured values are converted to the same visual scale using by the subjective result. The score of macro-uniformity is determined by the weighted sum of each measured value. Though the subjective test, we make sure the performance of the specific method for measuring sub-attributes of macro-uniformity. Correlations between the specific methods (banding/streaking, 2D noise, and Gradient) and subjective score are 0.92, 0.97, and 0.86, respectively. We obtain the correlation between the proposed method to quantify overall uniformity and subjective score is 0.94.

## 7867-04, Session 1

### Current practices in art image reproduction: image quality experimentation

S. P. Farnand, Rochester Institute of Technology (United States)

A project, supported by the Andrew W. Mellon Foundation, is currently underway to evaluate current practices in fine art image reproduction, determine the image quality generally achievable, and establish a suggested framework for art image interchange. To determine the image quality currently being achieved, experimentation has been conducted in which a set of objective targets and pieces of artwork in various media were imaged by participating museums and other cultural heritage institutions. Print files and guide prints, if these are used in the institution's typical workflow, were delivered to the Rochester Institute of Technology where prints were made on the same Heidelberg Speedmaster press run by the same operator and using the same inks and paper throughout. The resulting prints were used as stimuli in psychometric testing to generate scales of image quality. In this testing, twenty-four observers were asked to rank the prints relative to the original artwork. They are asked specifically to determine which prints were the best reproductions of the original artwork, as opposed to the most personally pleasing images. The results indicated that certain of the workflows provided more consistently accurate reproductions while certain other workflows consistently fell short. The experimental results will be among the inputs used to construct a conceptual framework of the various types of imaging taking place in cultural institutions at present. Based on this framework, an image processing tool that incorporates appearance models that are adequate for the various working environments in cultural heritage institutions will be developed.

## 7867-05, Session 1

### Using metrics to assess the ICC perceptual rendering intent

K. R. Falkenstern, N. Bonnier, Océ Print Logic Technologies (France); H. Brettel, Telecom ParisTech (France); F. Viénot, Muséum national d'Histoire naturelle (France)

Increased interest in color management has resulted in more options for the user to choose between for their color management needs. We propose an evaluation process that uses metrics to assess the quality of ICC profiles, specifically for the perceptual rendering intent. The primary objective of the perceptual rendering intent, unlike the media-relative intent, is a preferred reproduction rather than an exact match. Profile vendors commonly quote a CIE  $\Delta E^*_{ab}$  color difference to define the quality of a profile.

With the perceptual rendering intent, this may or may not correlate to the preferred reproduction.

For this work we compiled a comprehensive list of quality aspects,



used to evaluate the perceptual rendering intent of an ICC printer profile. The aspects are used as tools to individually judge the different qualities that define the overall strength of profiles. The proposed workflow uses metrics to assess each aspect and delivers a relative comparison between different printer profile options. The aim of the research is to improve the current methods used to evaluate a printer profile, while reducing the amount of time required.

## 7867-06, Session 2

### Development of perceptually calibrated objective metrics of noise

E. W. Jin, B. W. Keelan, S. F. Prokushkin, Aptina Imaging Corp. (United States)

This study aims at developing a noise metric with a transform to just noticeable differences (JNDs) of quality in pictorial scenes. Such a perceptually calibrated noise metric is particularly valuable for comparing the impact of noise with that of other attributes and for computing overall image quality. A system simulation model was used to create scene-dependent noise masks that reflect the performance of today's mobile capture devices. Samples with different overall magnitudes of noise and with varying mixtures of red, green, blue, luminance, and chrominance noises were included in the study. Eleven treatments in each of ten pictorial scenes were evaluated by twenty observers using a softcopy ruler method. The most general and best-performing metric tested involved integrating the system noise power spectra over a visual frequency weighting function, and combining the covariances obtained with empirical coefficients. In CIELAB space, inclusion of a normally negative  $L^*a^*$  covariance in addition to  $L^*$  and  $a^*$  variances improved the predictiveness significantly ( $b^*$  variance was found to contribute little). Test targets in linear sRGB and rendered  $L^*a^*b^*$  spaces for each treatment will be made available to enable other researchers to test metrics of their own design and calibrate them to JNDs of quality without performing additional perceptual experiments.

## 7867-07, Session 2

### Perceptually relevant evaluation of noise power spectra in adaptive pictorial systems

B. W. Keelan, R. B. Jenkin, Aptina Imaging Corp. (United States)

Noise Power Spectra (NPS) are traditionally measured using uniform areas of tone. Adaptive algorithms, such as noise reduction, demosaicing, and sharpening, can modify their behavior based on underlying image structure. In particular, noise reduction algorithms may suppress noise more strongly in perfectly uniform areas than they would in areas with modest variations, as found in actual pictorial images, and so yield unrepresentative NPS. This phenomenon would be similar in nature to the susceptibility of high-contrast-edges to adaptive sharpening and the subsequent over-estimation of effective pictorial modulation transfer function. Experimentation is described that examines the effect of modern adaptive noise reduction algorithms on the NPS of images containing varying gradients. Gradients are chosen based on a survey of consumer images from areas where noise is typically noticeable, such as blue sky, walls and faces. Although slight loss in performance of adaptive noise reduction is observed as gradients increase, the effect is perceptually small at gradients of relevance in pictorial imaging. The significant additional complexity of measuring gradient-based NPS at a number of mean levels does not appear to be justified; measuring NPS from uniform areas of tone should suffice for most perceptual work.

## 7867-08, Session 2

### A novel perceptual image quality measure for block based image compression

T. Shoham, D. Gill, S. Carmel, ICVT Ltd. (Israel)

Reliable, low complexity, automatic perceptual evaluation of image quality still remains an open challenge. Specifically, evaluation of image quality that undergoes recompression using a block-based scheme, such as JPEG, is an important enabler for automatic, perceptually lossless, image recompression.

We propose a novel image quality measure that answers this need. The proposed quality measure combines 3 metrics to obtain a score in the range 0-1, where 1 corresponds to identical images. The first component evaluates the average, per-pixel, distortion. The second component measures the extent of blockiness added by the compression process. The third component measures the texture distortion in each 4x4 pixel block. These metrics are then pooled into a single score using a weighted geometric average. The image is divided into tiles, whose size depends on image resolution. The proposed score is calculated for each image tile, and may be used for evaluation of local quality. The tiles scores can also be pooled into a single image quality score, by averaging the lowest tile score with the average score over all tiles with, thus mimicking human perception.

The proposed quality score has been successfully integrated into a real-time, automatic, perceptually lossless, JPEG recompression system.

## 7867-09, Session 2

### A metric for predicting preferred coring level to reduce toner scatter in electrophotographic printing

H. J. Park, Samsung Electronics Co., Ltd. (Korea, Republic of); J. P. Allebach, Purdue Univ. (United States)

As modern technology develops very rapidly, we are swamped with many new cutting-edge products including imaging applications. To survive in a fiercely competitive environment, image quality plays an important role in imaging products. Sharpness that describes the clarity of details and edge transition of an image is one of concerns in the imaging product since it affects overall perception on the image. In spite of convenience of physical assessment, it should be noticed that the perceived data from a human viewer cannot be replaced with the physical measured data, in terms of acceptability to the end-user of the imaging product. Therefore, the measured data by physical assessment should be correlated with the perceived data by psychophysical assessment.

In this paper, we exploit the previously published psychophysical data based on perceived toner scatter. For physical data, we apply the Edge Transition Width (ETW) between T90 and T10 boundaries as a line metric to measure sharpness and the Weighted Differential Toner Scatter (WDTS) within the region between T60 and T05 boundaries to measure the spread of differential toner scatter. Utilizing both perceived and measured data, we predict preferred coring level to reduce toner scatter in electrophotographic (EP) printing.

## 7867-11, Session 3

### A universal and reference-free blurriness measure

C. Chen, W. Chen, J. A. Bloom, Dialogic Media Labs (United States)

Blurriness is among the artifacts that can be introduced into still images and video sequences by processing. Measurements of blurriness can be included as part of an assessment of perceptual quality. In this paper, a novel universal and reference-free blurriness measurement approach is presented. The gradient image generated from the given image is modeled as a Markov chain, specified using a one-step transition probability matrix. The transition probabilities for selected pairs of gradient values are computed and combined to formulate the blurriness measure for a given image. This is the first time that transition probabilities are applied to perceptual quality assessment. Transition probabilities can exploit the relationship between adjacent elements in the gradient image and thus give very promising blurriness

measure. Experimental studies are conducted to compare the proposed method to the state-of-the-art reference-free blurriness measurement algorithms. The results show that the proposed method outperforms the commonly used measures. In this paper we also discuss the computational complexity of the proposed approach and conclude that, for a number of applications, it can be used for real-time video analysis.

### 7867-12, Session 3

#### Issues in the design of a no-reference metric for perceived blur

H. Liu, Technische Univ. Delft (Netherlands); I. Heynderickx, Philips Research Nederland B.V. (Netherlands)

Designing a no-reference (NR) blur metric that can reliably predict what humans perceive remains an academic challenge. In this paper, we address some significant issues relevant to the development of a NR blur metric. Based on state-of-the-art metrics and the data of psychovisual experiments available in the literature, we explain current concerns and difficulties in the metric design: (1) the classification of blur metrics depending on their targeted applications, (2) the effect of edge detection method on the performance of a metric, (3) the sensitivity in performance of a metric in terms of content independency, (4) the added value of including visual attention in the design of a blur metric. These issues are discussed in either quantitative or qualitative terms, which is beneficial for the future research in designing a more reliable NR blur metric.

### 7867-13, Session 3

#### Evaluating super resolution algorithms

Y. Kim, J. H. Park, G. Shin, H. Lee, D. Kim, S. H. Park, J. Kim, Samsung Electronics Co., Ltd. (Korea, Republic of)

This study intends to establish a sound testing and evaluation methodology based upon the human visual characteristics for appreciating the image restoration accuracy; in addition to proposing a color difference equation (CDE) based objective evaluation method. In total, 5 different super resolution (SR) algorithms - such as iterative back-projection (IBP), maximum a posteriori (MAP), robust SR, projections onto convex sets (POCS), and a non-uniform interpolation - were selected. The performance comparison between the SR algorithms in terms of their restoration accuracy was carried out through both subjectively and objectively. The former methodology relies upon the triplet comparison method recommended by ISO 20462-2. For the latter, the two most widely used CIE standard CDEs, i.e. CIELAB and CIEDE2000, were adopted for evaluating the restoration accuracy of those SR algorithms. Consequently, POCS and a non-uniform interpolation outperformed the others for an ideal situation, while no huge algorithm dependency could be observed in a real world case where any prior information about the blur kernel is remained unknown. However, IBP and MAP resulted in higher sharpness if a blur kernel can be accurately estimated. Objective data analysis with a larger number of test stimuli can verify our results and will be discussed in the final manuscript.

### 7867-14, Session 3

#### Image quality assessment based on distortion identification

A. Chetouani, A. Beghdadi, Univ. Paris-Nord (France)

A New Global Full-Reference Image Quality System based on classification and fusion scheme is proposed. It consists of many steps. The first step is devoted to the identification of the type of degradation contained in a given image based a Linear Discriminant Analysis (LDA) classifier using some common Image Quality Metric (IQM) as feature inputs. An IQM per degradation (IQM-D) is then used

to estimate the quality of the image. For a given degradation type, the appropriate IQM-D is derived by combining the top three best IQMs using an Artificial Neural Network model. The performance of the proposed scheme is evaluated first in terms of good degradation identification. Then, for each distortion type the image quality estimation is evaluated in terms of good correlation with the subjective judgments using the TID 2008 image database.

### 7867-15, Session 4

#### Image quality evaluation of light field photography

Q. Fu, Xi'an Institute of Optics and Precision Mechanics (China); Z. Zhou, Univ. of Science and Technology of China (China); Y. Yuan, BeiHang Univ. (China); B. Xiangli, The Academy of Opto-Electronics (China)

Light field photography captures both 2D spatial and 2D angular information of a scene. Digital refocusing and digital correction of aberrations could be done after the photograph is taken. However, capturing 4D light field is costly and tradeoffs between different image quality metrics should be made and evaluated.

This paper explores the effects of light field photography on image quality by quantitatively evaluating some basic criteria for an imaging system. A simulation approach was first developed by ray-tracing the light rays of a designed optical system. A standard testing chart followed by ISO 12233 was provided as the input scene and an image was acquired by light field rendering methods. As a comparison, the same measures were taken for the same main lens system as the results of conventional photography. Image quality metrics were calculated at several different depths. An experimental light field system was built up and its performance was tested.

This work helps better understanding the pros and cons of light field photography in contrast with conventional imaging methods and perceiving the way to optimize the digital-optical design of the system.

### 7867-16, Session 4

#### Feature-based automatic color calibration for networked camera system

S. Yamamoto, Tokyo Metropolitan College of Industrial Technology (Japan); K. Taki, N. Tsumura, T. Nakaguchi, Y. Miyake, Chiba Univ. (Japan)

In this paper, we propose an automatic color calibration technique among the networked cameras. Each camera is assumed to have the overlapping area in their captured area, and to detect one or more common objects at least. Our algorithm firstly searches the same object in each scene. MSER method is applied to detect the area of appropriate object. After that, the feature-based SIFT detection can calculate the shape information on the candidate area as 128 vectors. Next, similarity comparison is performed among the candidate area by using 128 vectors of shape feature. If the similarity of the closest pair is remarkable compared with other pairs, its RGB outputs which are averaged pixel value in this area are used as elements of color calibration. Finally, the comparison of RGB output is performed between most similar objects, and RGB outputs are used to make the color calibration matrix when judged the same object from information of shape and the color. Experimentally we found that RGB output between cameras is gradually corresponding by continuous detection.

### 7867-17, Session 4

#### Analysis of estimation error in image quality measurements

P. D. Burns, Carestream Health, Inc. (United States)

Error propagation analysis is often used to predict the transformation

of variation or error when a signal undergoes a transformation. For example in image capture, when red, green, blue color-signals are transformed by a color-matrix or a colorimetric transformation. Less-often considered, however, is the measurement error inherent in the calculation of several derived measurements, such as modulation transfer function (MTF), contrast-to-noise ratio or noise-power spectrum. Each of these measurements is actually an estimate based on observed image characteristics, such as mean digital signal level, sample standard deviation.

Consider the signal variation that result from exposure and detector characteristics. When signals are combined, either by matrix or spatial operations, so are the variations, and this can be modeled as forming a function of random variables. After describing the technical basis for analysis of the estimation error in terms of component parameter variations, we analyze the origin and magnitude of measurement variation for particular measurements of interest: the spatial frequency response (SFR), as described in the ISO standards, summary measures of image sharpness, such as CMT (visually-weighted) acutance. Based on this approach, is possible to model the propagation of the first- and second-order error statistics in terms of expected bias and variation error when applying such measurements in specific performance monitoring or production tasks.

### 7867-18, Session 5

#### **LCD displays performance comparison by MTF measurement using the white noise stimulus method**

C. Mitjà, J. Escofet, Univ. Politècnica de Catalunya (Spain)

The amount of images produced to be viewed as soft copies on output displays are significantly increasing. This growing occurs at the expense of the images targeted to hard copy versions on paper or any other physical support. Even in the case of high quality hard copy production, people working in professional imaging uses different displays in selecting, editing, processing and showing images, from laptop screen to specialized high end displays. Then, the quality performance of these devices is crucial in the chain of decisions to be taken in image production. Metrics of this quality performance can help in the equipment acquisition. Different metrics and methods have been described to determine the quality performance of CRT and LCD computer displays in clinical area. One of most important metrics in this field is the device spatial frequency response obtained measuring the modulation transfer function (MTF). This work presents a comparison between the MTF of several LCD displays measured by the white noise stimulus method, over vertical and horizontal directions. Additionally, different displays show particular pixels structure pattern. In order to identify this pixel structure, a set of high magnification images is taken from each display to be related with the respective vertical and horizontal MTF.

### 7867-19, Session 5

#### **Improving the quality of H.264 by using a new rate control model**

M. Hrtati, A. Saadane, M. Larabi, XLIM-SIC (France)

To improve the quality results of H.264, we propose in this paper, a new Rate-Quantization (R-Q) model resulting from extensive experiments. This latter is divided into two parts. The first part is an Intra R-Q model used to determine an optimal initial QP for I-Frames and derived from extensive experiments. The QP determination is based on both target bit-rate and I-Frame complexity. The I-frame target bit-rate is obtained from the global target bit-rate by using a new non-linear model. The second part is an Inter R-Q model used to calculate QP. From this, we demonstrate a logarithmic relationship between the QP and the target bit budget used to encode a given Frame or Macroblock. The Inter R-Q model coefficients are updated using the statistics of the previous coded units. The inter R-Q model does not need any complexity measure (such a MAD) and replaces both linear and quadratic models used in H.264 rate controller.

### 7867-20, Session 5

#### **A noble method on no-reference video quality assessment using block modes and quantization parameters of H.264/AVC**

I. Park, T. Na, M. Kim, Korea Advanced Institute of Science and Technology (Korea, Republic of)

Video quality assessment is an important tool of guaranteeing video services in a required level of quality. Although subjective quality assessment is more reliable due to the reflection of Human Visual System (HVS) than objective quality assessment, it is a time-consuming and very expensive approach, and is not appropriate for real-time applications. Therefore, much research has been made for objective video quality assessment instead of subjective video quality assessment.

Among three kinds of objective assessment approaches which are full-reference, reduced-reference and no-reference methods, no-reference method has drawn much attention because it does not require any reference. The encoding parameters are good features to use for no-reference model because the encoded bitstreams carry plenty of information about the video contents and it is easy to extract some coding parameters to assess visual quality.

In this paper, we propose a no-reference quality metric using two kinds of coding parameters in H.264/AVC: quantization and block mode parameters. These parameters are extracted and computed from H.264/AVC bitstreams, without relying on pixel domain processing. We design a linear quality metric composed of these two parameters. The weight values of the parameters are estimated using linear regression with the results of subjective quality assessment which are obtained based on the DSIS (Double Stimulus Impairment Scale) method of ITU-R BT.500-11.

### 7867-21, Session 5

#### **Prioritization of AL-FEC information for improving IP television services QoS**

E. Mammi, Univ. degli Studi di Roma Tre (Italy); G. Russo, P. Talone, Fondazione Ugo Bordoni (Italy)

In the digital television world, an important transformation is represented by the television over IP service. One of the key factors enabling the spreading of television over IP is represented by the quality. Furthermore, packet loss is probably the main service degradation source for that services.

The proposed approach combines the use of AL-FEC with the set-up of a transport quality mechanism based on FEC packets prioritization. To FEC packets is assigned a transfer priority higher than that of the media packets transferred under the best effort paradigm, thus reducing in congested routers the amount of FEC packet losses. In this way the error correction capability is improved. Furthermore, as the FEC stream is usually a percentage of the media one, the choice of applying the prioritization to the FEC stream and not to the whole media allows reducing the impact of prioritization of television service traffic on other types of traffic, concurrent on the same link. The tests have been performed on a simulated network and on a real IP test-bed. The results show the effectiveness of the proposed approach with respect the un-prioritized one, allowing to obtain higher video quality at the same packet loss rate.

### 7867-22, Session 6

#### **Reference image method for measuring quality of photographs produced by digital cameras**

M. Nuutinen, Aalto Univ. School of Science and Technology (Finland); O. Orenius, T. S. Säämänen, Univ. of Helsinki (Finland); P. T. Oittinen, Aalto Univ. School of Science and Technology

(Finland)

Computational image quality metrics can be divided three groups: full-reference (FR), reduced-reference (RR) and no-reference (NR). FR metrics cannot be applied to compute image quality produced by different digital cameras because pixel-wise reference images are missing. NR metrics are applicable only when the distortion type is known and the distortion space is low-dimensional. RR metrics provide a tradeoff between NR and FR metrics. A RR metric does not require pixel-wise reference image, it only needs a set of extracted features. With the aid of RR features it is possible to try to avoid the problems related to the NR metrics. In this study we used RR metrics for measuring image quality of natural images produced by digital cameras. We propose a method where reference images are produced using a reference camera. The reference images are expected to be natural reproductions of the views under study. We tested our method using three RR metrics proposed in the literature. The results suggest that the proposed method is promising when the problem is to measure the quality of natural images produced by digital cameras for the purpose of camera benchmarking.

### 7867-23, Session 6

#### **RAW camera DPCM compression performance analysis**

K. Bouman, V. Ramachandra, K. Atanassov, M. Aleksic, S. R. Goma, Qualcomm Inc. (United States)

The MIPI standard has adopted DPCM compression for RAW data images streamed from mobile cameras. This DPCM is line based and uses either a simple 1 or 2 pixel predictor. In this paper, we analyze the DPCM compression performance as MTF degradation. To test this scheme's performance, we generated Siemens star images and binarized them to 2-level images. These two intensity values were chosen such that their intensity difference corresponds to those pixel differences which result in largest relative errors in the DPCM compressor. (E.g. a pixel transition from 0 to 4095 corresponds to an error of 6 between the DPCM compressed value and the original pixel value). The DPCM scheme introduces different amounts of error based on the pixel difference. We passed these modified Siemens star chart images to this compressor and compared the compressed images with the original images using IT3 MTF response plots for slanted edges. Further, we discuss the PSF influence on DPCM error and its propagation through the image processing pipe.

### 7867-24, Session 7

#### **Brightness, lightness, and specifying color in high-dynamic-range scenes and images**

M. D. Fairchild, P. Chen, Rochester Institute of Technology (United States)

Traditional color spaces have been widely used in a variety of applications including digital color imaging, color image quality, and color management. These spaces, however, were designed for the domain of color stimuli typically encountered with reflecting objects and image displays of such objects. This means the domain of stimuli with luminance levels from slightly above zero to that of a perfect diffuse white (or display white point). This limits the applicability of both of these spaces to color problems in HDR imaging. This is caused by their hard intercepts at zero luminance/lightness and by their uncertain applicability for colors brighter than diffuse white. To address HDR applications, two new color spaces were recently proposed, *hdr-CIELAB* and *hdr-IPT*. They are based on replacing the power-function nonlinearities in *CIELAB* and *IPT* with a more physiologically plausible hyperbolic function optimized to most closely simulate the original color spaces in the diffuse reflecting color domain. This paper will present the formulation of the new models, evaluations using Munsell data in comparison with *CIELAB*, *IPT*, and *CIECAM02*, two sets of lightness-scaling data above diffuse white and various formulations of *hdr-CIELAB* and *hdr-IPT* to predict the visual results.

### 7867-25, Session 7

#### **Evaluating HDR photos using Web 2.0 technology**

G. Qiu, Y. Mei, The Univ. of Nottingham (United Kingdom)

In this work, we exploit Web 2.0 technology to evaluate HDR photographs. We have constructed a website for this purpose. The URL of the website is <http://www.hdri.cs.nott.ac.uk/> which is currently live and has been collecting viewer inputs.

At any one time, two versions of the same HDR photos are rendered by two different tone mapping operators are shown to the viewers. The viewer is asked to click on the one that she/he prefers or click on "cannot tell the difference" button if the user cannot decide which one is better. Such pair comparison results are then put into an SQL database server. These pair comparison results are then processed by a ranking algorithm called the Condorcet method [6]. This method will rank all versions of the same HDR photo according to the perceived visual qualities by the website's visitors. In the website, this is display as the "top 10" with the following URL <http://www.hdri.cs.nott.ac.uk/top10.php>

The website also enables users to up load their own tone mapping operators' results for comparison with results of other operators. Our visions is that as we accumulate more and more operators (especially with users contributing new operators results), the website will eventually enable new tone mapping operators to compare with existing ones thus making it much easier to compare the relative merits of a new tone mapping operator.

### 7867-36, Poster Session

#### **Potential of face area data for predicting sharpness of natural images**

M. Nuutinen, Aalto Univ. School of Science and Technology (Finland); O. Orenius, T. S. Säämänen, Univ. of Helsinki (Finland); P. T. Oittinen, Aalto Univ. School of Science and Technology (Finland)

Face detection techniques have been used for many different applications. For example face detection is a basic component in many consumer still and video cameras. In this study we compare the performance of face area data and freely selected local area data for predicting the subjective sharpness of photographs. The local values were collected in a systematic way from image and for the analyses we selected only the ones with the highest performance. The objective sharpness metric was based on the statistic of the wavelet coefficients of the selected area. We used three image contents whose subjective sharpness values had been measured. The image contents were captured by 13 cameras and the images were evaluated by 25 subjects. The quality of the cameras was between low-end mobile phone cameras to low-end compact cameras. The image contents simulated typical photos that consumers take with their mobile phones. Image size was scaled to 2 Mpix in all cases. The face areas on the images were approximately 7, 20 and 74 kpix. Based on the results the face area data is valuable for measuring the sharpness of photographs if face size is large enough. When the face area size was 20 kpix or 74 kpix the performance of the measured sharpness value equals or is better than the sharpness value measured from the best local areas. If the face area was too small (7 kpix) the performance was low compared to the best local areas.

### 7867-37, Poster Session

#### **A video quality assessment model based on the MPEG-7**

M. Sato, Y. Horita, Univ. of Toyama (Japan)

In this paper we propose a new video quality assessment model based on the MPEG-7 Descriptor. Previously, we have examined the

estimated accuracy in still image by using the MPEG-7 descriptor. We use descriptor ("Color Layout", "Scalable Color", "Color Structure", "Edge Histogram" and "Homogeneous Texture") from the MPEG-7 Descriptor. As a result, we was able to presume the still image by high accuracy.

In this paper, we use the presumption accuracy based on result of previous model (estimate of frame quality) when we propose video quality assessment model.

We estimated Video Quality based on Frame Quality and we encoded by using 10-kinds of video material (bit-rate: 448/1024/4096 kbps, frame-rate: 30/10/5 fps, coding schemes: Windows Media Video 9 and h.264).

The results of the estimation accuracy for WMV9/h.264 are: correlation 0.98/0.98, average error 0.13/0.15 and maximum error 0.48/0.46. The results of the estimation accuracy for both encoding methods (WMV and h.264) are: correlation 0.93, average error 0.22 and maximum error 0.72.

### 7867-38, Poster Session

#### Image quality: a tool for no-reference assessment methods

S. Corchs, F. Gasparini, F. Marini, R. Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

In this work we propose an image quality assessment tool. The tool is composed of different modules that implement several No Reference (NR) metrics (i.e. where the original or ideal image is not available). Different types of image quality attributes can be taken into account by the NR methods, like blurriness, graininess, blockiness, lack of contrast and lack of saturation or colorfulness among others. An extra module permits the user to give a subjective rating of the image quality. Our tool aims to give a structured view of a collection of objective metrics that are available for the different artifacts/attributes within an integrated framework that also offers the possibility of a subjective evaluation. As each metric corresponds to a single module, our tool can be easily extended to include new metrics or to substitute some of them. The software permits to apply the metrics not only globally but also locally to different regions of interest of the image. In this way, if necessary, the user can interact with the tool and eventually choose another NR method or he can decide to apply a certain metric locally (the region of interest can be manually selected) because the global one is not in correspondence with his subjective judgment. This computer-aid process can be iteratively applied for each of the images.

### 7867-39, Poster Session

#### Extending video quality metrics to the temporal dimension with 2D-PCR

C. Keimel, M. Rothbucher, K. Diepold, Technische Univ. München (Germany)

The aim of any video quality metric is to deliver a quality prediction similar to the video quality perceived by human observers. One way to design such a model of human perception is by data analysis. In this contribution we intend to extend this approach to the temporal dimension. Even though video obviously consists of spatial and temporal dimensions, the temporal aspect is often not considered well enough. Instead of including this third dimension in the model itself, the metrics are usually only applied on a frame-by-frame basis and then temporally pooled, commonly by averaging. Therefore we propose to skip the temporal pooling step and use the additional temporal dimension in the model building step of the video quality metric. We propose to use the two dimensional extension of the PCR, the 2D-PCR, in order to obtain an improved model. We conducted extensive subjective tests with different HDTV video sequences at 1920x1080 and 25 frames per seconds. For verification, we performed a cross validation to get a measure for the real-life performance of the acquired model. We will show that the direct inclusion of the temporal dimension of video into the model building improves the overall prediction accuracy of the visual quality significantly

### 7867-41, Poster Session

#### Image quality metric benchmarking on compressed image databases

M. Nauge, M. Larabi, Univ. de Poitiers (France)

In recent literature hundreds of paper has proposed objective quality metrics dedicated to several image and video applications.

Our aim is to simplify the choice of metrics according to the application. We can also verify the benefit of your new image algorithms in term of the reduction of visual degradation.

To successfully complete this task we largely used standard methodology for the evaluation of objective model performance in respect of the report from the Video Quality Experts Group. This benchmarking of metrics uses four image databases from different countries, comprising of a large range of image contents, image distortion types, subjective scores from various observers and varying equipment and display configurations.

In this experiment we test twenty-seven metrics, from simple mathematical measures (like PSNR) to complex modelisation of Human Visual System (like HDR\_VDP).

### 7867-42, Poster Session

#### Optimal front light design for reflective displays in different ambient illumination

S. Wang, T. Chang, C. Li, Y. Bai, K. Hu, Industrial Technology Research Institute (Taiwan)

In this study, a luminance and color-tunable front light device for e-paper display was built. The psychophysical experiment results played an important role in visual perception since it could link the human psychological responses to the physical stimuli. A color calibration procedure was applied on the device to present correct 256 level luminance and 13 different color temperatures at fixed illumination of 200cd/m<sup>2</sup>. When the human visual system was first exposed to a new stimulus, it must be adapted to 13 different color temperatures. Human subjective testing experiments were completed to reveal the human preference for the luminance and color temperature of the front light device in different ambient illumination.

After analyzing the experimental results, it could be concluded that when the ambient illumination was dimmer, the human observers preferred brighter front light. However, in dark surrounding, the highest luminance of front light device was 208 cd/m<sup>2</sup>, and the most preferred color temperature lied between 11000K and 13000K. Also, this study revealed that human observers prefer higher luminance and color temperature while reading text. In the other words, carefully controlled subjective psychophysical experiment will be required to obtain more human visual perception data, and the data can be used to design the front light for reflective displays.

### 7867-43, Poster Session

#### Comparison of HDTV formats in a consumer environment

C. Keimel, A. Redl, K. Diepold, Technische Univ. München (Germany)

High definition television (HDTV) has become quite common in many homes. Still, there are two different formats used currently in commercial broadcasting: one interlaced format, 1080i50/60, and one progressive format, 720p50/60. There have already been quite a few contributions comparing the visual quality of these formats subjectively under common standard conditions. These conditions, however, don't necessarily represent the viewing conditions in the real-life consumer environment. In this contribution we therefore decided to do a comparison under conditions more representative of the consumer environment with respect to display and viewing conditions. Furthermore we decided to select not specially prepared test

sequences, but real-life content and coding conditions. As we were not interested in the influence of the transmission errors, we captured the sequences directly in the play-out centre of a cable network provider in both 1080i50 and 720p50. Also we captured for comparison the same content in digital PAL-SDTV. We conducted extensive subjective tests with overall 25 test subjects and a modified SSIS method. The results show that both HDTV formats outperform SDTV significantly. Although 720p50 is perceived to have a better quality than 1080i50, this difference is not significant in a statistical sense. This supports the validity of previous contribution's results, gained in standard conditions, also for the real-life consumer environment.

## 7867-26, Session 8

### **Just noticeable difference vs. visual difference: hypotheses and how to check whether they are true or not**

S. N. Bezryadin, P. Burov, KWE International, Inc. (United States)

The problem of accurate color reproduction is a hot topic, which is closely linked to the problem of accurate measurement of human visual threshold, or "Just Noticeable Difference" (JND).

The formulas Mac Adam used for his Color Matching Ellipsoids were created with an assumption that human ability to distinguish close stimuli may be described with the Gaussian distribution, a standard distribution for stochastic processes (we name it hypothesis #1). However, this assumption hasn't yet been proved.

Since most imaging scientists believe that JND experiments are too complicated and costly, Visual Difference (dV) experiments have gained high popularity. In connection to this there are also several common beliefs, such as:

- Visual Difference experimental results can be extrapolated on pairs with Color difference less than 1 JND (let us name it hypothesis # 2)
- 1 JND is equal to 1 dV (hypothesis # 3)
- JND is proportional to dV (hypothesis # 4).

However, there have been no experiments devoted to determine the correlation between JND and dV.

This paper presents two experimental techniques that can determine stimuli pairs which p % of observers perceive as different, while the other (100 - p) % cannot distinguish. It also discusses the experiment's requirements such as accuracy and stability of equipment.

The presented technique will help to resolve the following problems:

1. Determine with high accuracy stimuli pairs of 1 JND difference in various Gamut areas.
2. Check whether the assumption that human ability to distinguish similar stimuli may be described with Gaussian distribution is correct.
3. Check if JND and dV are really proportional and whether Visual Difference experimental results may be extrapolated on pairs with Color difference less than 1 JND.

## 7867-27, Session 8

### **Device dependent scene dependent quality predictions using effective pictorial information capacity**

K. H. Oh, S. Triantaphillidou, R. E. Jacobson, Univ. of Westminster (United Kingdom)

This study aims to introduce improvements in the predictions of device-dependent image quality metrics (IQMs). A validation experiment was first carried out to test the success of such a metric, the Effective Pictorial Information Capacity (EPIC), using results from subjective tests involving 32 test scenes replicated with various degrees of sharpness and noisiness. The metric was found to be a good predictor when tested against average ratings but, as expected by device-dependent metrics, it predicted less successfully the perceived quality of individual, non-standard scenes with atypical spatial and structural

content.

Improvement in predictions was attempted by using a modular image quality framework and its implementation with the EPIC metric. It involves modelling a complicated set of conditions, including classifying scenes into a small number of groups. The scene classification employed for the purpose uses objective scene descriptors which correlate with subjective criteria on scene susceptibility to sharpness and noisiness. The implementation thus allows automatic grouping of scenes and calculation of the metric values. Results indicate that metric predictions were improved. Most importantly, they were shown to correlate equally well with subjective quality scales of standard and non-standard scenes. The findings indicate that a device-dependent, scene-dependent IQM can be achieved.

## 7867-28, Session 9

### **Social image quality**

G. Qiu, A. Kheiri, The Univ. of Nottingham (United Kingdom)

No abstract available

## 7867-29, Session 10

### **Utility studies for security encoded office documents: experimental design challenges**

C. A. Deller, G. J. Woolfe, Canon Information Systems Research Australia Pty. Ltd. (Australia)

We have developed methodologies to study the usability of documents. This methodology has been applied to the study of the impact of visible security patterns in the usability of typical office documents. Two specific information retrieval tasks have been examined: the retrieval of text-based information from a written report and the retrieval of numerical information from tables and graphs. The methodology we have developed aims to minimize sources of uncontrolled variability in the measurements while simultaneously avoiding a systematic bias from learning effects and maintaining task equivalence across all documents. We believe the methodologies developed in this work may prove useful in future studies of document usability.

## 7867-30, Session 10

### **Printed fingerprints: a framework and first results toward detection of artificially printed latent fingerprints for forensics**

S. Kiltz, M. Hildebrandt, J. Dittmann, C. Vielhauer, C. Krätzer, Otto-von-Guericke-Univ. Magdeburg (Germany)

In the publication of Lothar Schwarz an amino acid model for printing latent fingerprints to porous surfaces is introduced, motivated by the need for reproducibility tests of different development techniques for forensic investigations. This can be used legitimately for quality assurance. However, this technique also enables the fabrication of artificial traces constituting a possible security threat, motivating a need for research of appropriate fabrication detection techniques.

It is important to detect and distinguish between real, natural latent fingerprint patterns from humans and artificially printed latent fingerprints under the consideration of the Schwarz amino acid model and ink-jet printing techniques. The discrimination should work before and after traditional development/enhancing technologies (e.g. carbon-black) are applied. The main contribution of our work is a first proposal for an extensible framework for the examination of fingerprints as a process-chain composed of fingerprint printing, processing of the physical forensic trace sample, digital acquisition, and a subjective assessment of the digital sample. We evaluate first results and tendencies of the influence of properties on the recognition of artificial traces. Those include properties of the fingerprint, the sample material

and age, the printer, the ink, the contact-less acquisition sensor and the properties related the dactyloscopic expert.

Furthermore, in this paper we contribute as well to a faster saliency map extraction using H.264 compressed stream information.

## 7867-31, Session 10

### Monitoring image quality for security applications

M. Larabi, D. Nicholson, Univ. de Poitiers (France)

Nowadays, security applications are of a big interest for personal and public security. Governments and big cities are putting a lot of effort and money on it. Video surveillance is the most common way to monitor and enhance the security of strategic places. There is a large diversity of systems, video codecs, sensors, and software in this field, which makes difficult the assessment of even the certification of video-surveillance systems and sub-systems. As installed systems have to operate for a while, interoperability and extensibility are very important for the durability of a video surveillance system.

## 7867-32, Session 10

### Video quality and interpretability study using SAMVIQ and Video-NIIRS

D. L. Young, Raytheon Intelligence & Information Systems (United States); J. Ruzczyk, General Dynamics Advanced Information Systems (United States); T. Bakir, Harris (United States)

The results of a video quality and interpretability study are described. The study varied scene content, compression ratio, and encoder implementation specifics. Loss of video quality as measured the Subjective Assessment Methodology Video Image Quality (SAMVIQ) is compared to loss of interpretability as measured by the Video National Intelligence Interpretability Rating Scale (Video-NIIRS). Quality rating results are compared to predictive indicators of quality such as the Visual Information Fidelity (VIF). Interpretability subjective rating results are compared to predictive objective indicators of interpretability using the Motion Imagery Quality Equation (MIQE).

This material is based upon work supported by a DoD contract. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Government.

## 7867-33, Session 11

### Weighted-MSE based on saliency map for assessing video quality of H.264 broadcasted video streams

H. Boujut, Bordeaux Univ. (France); O. Hadar, Ben-Gurion Univ. of the Negev (Israel); J. Benois-Pineau, T. Ahmed, Bordeaux Univ. (France); P. Bonnet, Audemat WorldCast Systems Group (France)

The paper contributes to objective video quality assessment of broadcasted video over DVB and IP networks. We introduce an objective video quality metric based on saliency map to assess packet and signal loss influence on broadcasted video streams. This new metric, we called it Weighted-MSE (WMSE), requires the full-reference video like MSE and SSIM metrics. Unlike MSE which does not consider the Human Visual System (HVS), WMSE uses spatio-temporal saliency maps to increase the contribution of salient regions. We note that there are similar ideas in the literature, but they work on spatial saliency map and use magnitude-error weighting scheme. Despite the fact that SSIM metric and the approach proposed in the literature already take into account the HVS, the WMSE also considers the temporal side of visual perception. We use adequate fusion strategies to combine both spatial and temporal saliency in a single map. The WMSE metrics we propose is designed as a basis of comparison for quality of experience tests.

## 7867-34, Session 12

### Metrics for regression testing and optimization of visual attention (saliency) models

R. J. Moore, B. Stankiewicz, 3M Co. (United States)

When developing a predictive visual tool clear metrics are required to evaluate the model's performance. One area of rich research is in the area of Visual Attention Modeling. In this field of research one typically compares eye tracking data collected from human observers to the predictions made by the model. To evaluate the performance of these predictions research in visual attention modeling typically uses signal detection (Receiver Operating Characteristic (ROC)) to measure the predictive power of the system. Researchers typically compare the model's Saliency map output to eye tracking data to generate ROC curves and values for each saliency map. The average over a set of test images provides a final measure of the system's performance. In releasing a commercial visual attention system, we have spent considerable effort in developing metrics and testing methods that allow for regression testing, and are useful for optimizing the visual attention model. It was determined that ROC alone was not a satisfactory measure of system performance. In this paper we present the methods used to test and measure the performance of our visual attention model, and how they allow us to build regression tests that can be used to optimize the model's parameters.

## 7867-35, Session 12

### Naturalness and interestingness of test images for visual quality evaluation

R. Halonen, S. Westman, P. T. Oittinen, Aalto Univ. School of Science and Technology (Finland)

Balanced and representative test images are needed to study perceived visual quality in various application domains. This study investigates naturalness and interestingness as quality attributes in the context of test images. Taking a top-down approach we aim to find the dimensions which constitute naturalness and interestingness in test images and the relationship of these high-level quality attributes. We compare existing collections of test images (e.g. Sony sRGB images, ISO 12640 images, Kodak images, Nokia images and test images developed within our group) in an experiment involving quality sorting and structured interviews. Based on the data gathered we analyze the viewer-supplied criteria for naturalness and interestingness across image types, quality levels and judges. This study advances our understanding of the subjective criteria used when judging image quality as well as enables the validation of current test images and furthers their development.

### 7868-01, Session 1

#### **Data repository mapping for influenza protein sequence analysis**

D. A. Pellegrino, Jr., C. Chen, Drexel Univ. (United States)

This paper introduces a new method for creating an interactive sequence similarity map of all known influenza virus protein sequences and integrating the map with existing general purpose analytical tools. The NCBI data model was designed to provide a high degree of interconnectedness amongst data objects. Substantial and continuous increase in data volume has led to a large and highly connected information space. Researchers seeking to explore this space are challenged to identify a starting point. They often choose data that is popular in the literature. References in the literature follow a power law distribution and popular data points may bias explorers toward paths that lead only to dead-ends of what is already known. To help discover the unexpected we developed an interactive visual analytics system to map the information space of influenza protein sequence data. The design is motivated by the needs of eScience researchers.

### 7868-02, Session 1

#### **GPU-accelerated visualization of protein dynamics in ribbon mode**

M. Wahle, S. Birmanns, The Univ. of Texas Health Science Ctr. at Houston (United States)

Proteins are biomolecules present in living organisms and essential for carrying out vital functions. Inherent to their functioning is folding into different spatial conformations, and to understand these processes, it is crucial to visually explore the structural changes. In recent years, significant advancements in experimental techniques and novel algorithms for post-processing of protein data have routinely revealed static and dynamic structures of increasing sizes. In turn, interactive visualization of the systems and their transitions became more challenging.

Therefore, much research for the efficient display of protein dynamics has been done, with the focus being space filling models, but for the important class of abstract ribbon or cartoon representations, there exist only few methods for an efficient rendering. Yet, these models are of high interest to scientists, as they provide a compact and concise description of the structure elements along the protein main chain.

In this work, a method was developed to speed up ribbon and cartoon visualizations. Separating two phases in the calculation of geometry allows to offload computational work from the CPU to the GPU. The first phase consists of computing a smooth curve along the protein's main chain on the CPU. In the second phase, conducted independently by the GPU, vertices along that curve are moved to set up the final geometrical representation of the molecule.

### 7868-03, Session 1

#### **OpenOrd: an open-source toolbox for large graph layout**

S. Martin, W. M. Brown, Sandia National Labs. (United States); R. Klavans, K. Boyack, SciTech Strategies, Inc. (United States)

We document an open-source toolbox for drawing large-scale undirected graphs. This toolbox is based on a previously implemented closed-source algorithm known as VxOrd. Our toolbox, which we call OpenOrd, extends the capabilities of VxOrd to large graph layout by incorporating edge-cutting, a multi-level approach, average-link clustering, and a parallel implementation. At each level, vertices are grouped using force-directed layout and average-link clustering. The

clustered vertices are then re-drawn and the process is repeated. When a suitable drawing of the coarsened graph is obtained, the algorithm is reversed to obtain a drawing of the original graph. This approach results in layouts of large graphs which incorporate both local and global structure. A detailed description of the algorithm is provided in this paper. Examples using datasets with over 600K nodes are given. Code is available at [www.cs.sandia.gov/~smartin](http://www.cs.sandia.gov/~smartin).

### 7868-04, Session 1

#### **A pseudo-haptic knot diagram interface**

H. Zhang, Indiana Univ.-Purdue Univ. Indianapolis (United States); J. Weng, Zhejiang Univ. (China); A. Hanson, Indiana Univ. (United States)

To make progress in understanding knot theory, we will need to interact with the projected representations of mathematical knots which are of course continuous in 3D but significantly interrupted in the projective images. One way to achieve such a goal would be to design an interactive system that allows us to sketch 2D knot diagrams by taking advantage of a collision-sensing controller and explore their underlying smooth structures through a continuous motion. Recent advances of interaction techniques have been made that allow progress to be made in this direction. Pseudo-haptics that simulates haptic effects using pure visual feedback can be used to develop such an interactive system. This paper outlines one such pseudo-haptic knot diagram interface. Our interface derives from the familiar pencil-and-paper process of drawing 2D knot diagrams and provides haptic-like sensations to facilitate the creation and exploration of knot diagrams. A centerpiece of the interaction model simulates a "physically" reactive mouse cursor, which is exploited to resolve the apparent conflict between the continuous structure of the actual smooth knot and the visual discontinuities in the knot diagram representation. Another value in exploiting pseudo-haptics is that an acceleration (or deceleration) of the mouse cursor (or surface locator) can be used to indicate the slope of the curve (or surface) of whom the projective image is being explored. By exploiting this additional visual cues, we proceed to a full-featured extension to a pseudo-haptic 4D visualization system that simulates the continuous navigation on 4D objects and allows us to sense the bumps and holes in the fourth dimension. Preliminary tests of the software show that main features of the interface overcome some expected perceptual limitations in our interaction with 2D knot diagrams of 3D knots and 3D projective images of 4D mathematical objects.

### 7868-05, Session 2

#### **Interactive isosurfaces with quadratic C<sup>1</sup> splines on truncated octahedral partitions**

T. Kalbe, A. Marinc, Technische Univ. Darmstadt (Germany); M. Rhein, Univ. Mannheim (Germany); M. Goesele, Technische Univ. Darmstadt (Germany)

The reconstruction of a continuous function from discrete data is a basic task in many applications such as the visualization of 3D volumetric data sets.

Here, we use a local approximation method for quadratic C<sup>1</sup>-splines on uniform tetrahedral partitions to achieve a globally smooth function. The spline is based on a truncated octahedral partition of the volumetric domain, where each truncated octahedron is further split into a fixed number of disjoint tetrahedra. The Bernstein-Bézier coefficients of the piecewise polynomials are thereby directly determined by appropriate combinations of the data values in a local neighborhood. As previously shown, the splines and their derivatives provide an approximation order two for smooth functions as well as their derivatives. We present the first visualizations using these splines and show that they are well-suited for GPU-based, interactive high-quality visualization of discrete data.



## 7868-06, Session 2

**Indirect multi-touch interaction for brushing in parallel coordinates**

R. Kosara, The Univ. of North Carolina at Charlotte (United States)

Interaction in visualization is often complicated and tedious. Brushing data in a visualization such as parallel coordinates is a central part of the data analysis process, and sets visualization apart from static charts. Modifying a brush, or combining it with another one, usually requires a lot of effort and mode switches, though, slowing down interaction and even discouraging more complex questions.

We propose the use of multi-touch interaction to provide fast and convenient interaction with parallel coordinates. By using a multi-touch trackpad rather than the screen directly, the user's hands do not obscure the visualization during interaction. Using one, two, three, or four fingers, the user can easily and quickly perform complex selections. Being able to change the selections rapidly, the user can explore the data set more easily and effectively, and can focus on the data rather than the interaction.

## 7868-07, Session 3

**The science of visual analysis at extreme scale**

L. T. Nowell, U.S. Dept. of Energy (United States)

No abstract available

## 7868-08, Session 4

**A randomized framework for discovery of heterogeneous mixtures**

M. A. Livingston, A. M. Palepu, J. Decker, M. Dermer, U.S. Naval Research Lab. (United States)

"Mixture models" is the term given to models that consist of a combination of independent functions creating the distribution of points within a set. We present a framework for automatically discovering and evaluating candidate models for unstructured data. Our abstraction of models enables us to seamlessly consider different types of functions in different numbers of dimensions as equally possible candidates. Our framework does not require an estimate of the number of underlying models or training on sample data, allows points to be probabilistically classified into multiple models or identified as outliers, and includes a few parameters that an analyst (not typically an expert in statistical methods) may use to adjust the output of the algorithm. We give results from our framework with synthetic data and classic data.

## 7868-09, Session 4

**Exploring height fields: interactive visualization and applications**

M. Allili, A. Villares, Bishop's Univ. (Canada); D. Corriveau, Univ. de Sherbrooke (United States)

Height fields are an important modeling and visualization tool in many applications and their exploration requires their display at interactive frame rates. This is hard to achieve even with high performance graphics computers due to their inherent geometric complexity. Typical solutions consist of using polygonal approximations of the height field to reduce the number of geometric primitives that need to be rendered. Starting from a rough approximation, a refinement process is operated until a desired level of detail is reached. In this work, we present a novel efficient algorithm that starts with an approximation that carries enough information about the height field so that only few refinement steps are

needed to achieve any desired level of detail. Our initial approximation is a simple triangulation whose nodes are the critical points of the height field, that is the peaks, pits, and passes of the surface which give its overall shape.

The extraction of critical points of the surface, which is a discrete structure, is done using a newly designed algorithm based on discrete Morse theory.

## 7868-10, Session 4

**An Evaluation of Methods for Encoding Multiple, 2D Spatial Data**

M. A. Livingston, J. Decker, Z. Ai, U.S. Naval Research Lab. (United States)

Datasets over a spatial domain are common in a number of fields, often with multiple layers (or variables) within data that must be understood together via spatial locality. Thus one area of long-standing interest is increasing the number of variables encoded by properties of the visualization. A number of properties have been demonstrated and/or proven successful with specific tasks or data, but there has been relatively little work comparing the utility of diverse techniques for multi-layer visualization. As part of our efforts to evaluate the applicability of such visualizations, we implemented five techniques which represent a broad range of existing research (Color Blending, Oriented Slivers, Data-Driven Spots, Brush Strokes, and Stick Figures). Then we conducted a user study wherein subjects were presented with composites of three, four, and five layers (variables) using one of these methods and asked to perform a task common to our intended end users (GIS analysts). We found that the Oriented Slivers and Data-Driven Spots performed the best, with Stick Figures yielding the lowest accuracy. Through analyzing our data, we hope to gain insight into which techniques merit further exploration and offer promise for visualization of data sets with ever-increasing size.

## 7868-11, Session 5

**Multivariate visualization of chromatographic systems**

T. Urness, T. Marrinan, A. R. Johnson, M. F. Vitha, Drake Univ. (United States)

Chemists are often faced with the challenge of separating and quantifying the molecules in a complex mixture. Pharmaceutical companies, for example, must test formulations for impurities and to make certain the drug contains the correct amounts of active ingredients. There are hundreds of chromatographic systems to choose from when developing analytical methods. Therefore, to avoid wasting time and money, it is vital that method development be guided by chemical principles rather than by trial and error.

This manuscript is broken into four main sections. In the first, we describe chromatography and the practical problem we are trying to solve. In the second, we discuss several methods for visualizing multivariate data, including glyphs, triangle plots, and parallel coordinates. We then describe a 3D visualization tool that we have created in order to analyze large sets of chromatographic data. As detailed in that section, our approach combines scatter plots, parallel coordinates, and specialized glyphs to assist in the analysis of the data. In the final section, we demonstrate the utility of the visualization tool by applying it to two chromatographic data sets.

## 7868-12, Session 5

**Visualization of dynamic adaptive resolution scientific data**

A. Foulks, R. D. Bergeron, S. H. Vohr, The Univ. of New Hampshire (United States)

Interactive visualization of very large data sets remains a challenging

problem to the visualization community. One promising solution involves using adaptive resolution representations of the data. In this model, important regions of data are identified using reconstructive error analysis and are shown in higher detail. During the visualization, regions with higher error are rendered with high resolution data, while areas of low error are rendered at a lower resolution. We have developed a new dynamic adaptive resolution rendering algorithm along with software support libraries. These libraries are designed to extend the VisIT visualization environment by adding support for adaptive resolution data. VisIT supports domain decomposition of data, which we use to define our AR representation. We show that with this model, we achieve performance gains while maintaining error tolerances specified by the scientist.

## 7868-13, Session 5

### A flexible low-complexity device adaptation approach for data presentation

R. U. Rosenbaum, Sr., A. Gimenez, Univ. of California, Davis (United States); H. Schumann, Univ. Rostock (Germany); B. Hamann, Univ. of California, Davis (United States)

Visual data presentations require adaptation for appropriate display on a viewing device that is limited in resources such as computing power, screen estate, and/or bandwidth. Due to the complexity of suitable adaptation, the few proposed solutions available are either too resource-intensive or inflexible to be applied broadly. Effective use and acceptance of data visualization on constrained viewing devices require adaptation approaches that are tailored to the requirements of the user and the capabilities of the viewing device.

We propose a dynamic device adaptation approach that takes advantage of progressive data refinement. The approach relies on hierarchical data structures that are created once and used multiple times. By incrementally reconstructing the visual presentation on the client with increasing levels of detail and resource utilization, we can determine when to truncate the refinement of detail so as to use the resources of the device to their full capacities. To determine when to finish the refinement for a particular device, we introduce a profile-based strategy which also considers user preferences. We discuss the whole adaptation process from the storage of the data into a scalable structure to the presentation on the respective viewing device. This particular implementation is shown for two common data visualization methods, and empirical results we obtained from our experiments are presented and discussed.

## 7868-14, Session 6

### EdgeMaps: visualizing explicit and implicit relations

M. Doerk, S. Carpendale, C. Williamson, Univ. of Calgary (Canada)

With this work we introduce EdgeMaps: a new method for integrating the visualization of implicit and explicit data relations. Explicit relations are specific connections between entities already present in a given dataset, while implicit relations are derived from multidimensional data based on shared properties and similarity measures. Many datasets include both explicit and implicit relations that are often not accounted for together in information visualizations. Node-link diagrams typically focus on explicit data connections, while not incorporating implicit similarities between entities. Multi-dimensional scaling considers similarities between items, however, explicit links between nodes are not displayed. In contrast, EdgeMaps visualize both implicit and explicit relations by combining and complementing spatialization and graph drawing techniques. As a case study for this approach we chose a dataset of philosophers, their interests, influences, and birthdates. By introducing the limitation of activating only one node at a time, interesting visual patterns emerge that resemble the aesthetics of fireworks and waves. We argue that the interactive exploration of these patterns enables the viewer to grasp the structure of a graph better than complex graph visualizations.

## 7868-15, Session 6

### Visualizing node attribute uncertainty in graphs

N. Cesario, A. Pang, Univ. of California, Santa Cruz (United States); L. Singh, Northwestern Univ. (United States)

Visualization is frequently misrepresentative of actual data as it is an absolute representation when uncertainty often exists in the data. While various techniques and tools exist for visualizing uncertainty in scientific visualizations, these do not exist for visualizing information such as graph/network data. Specifically, to our knowledge, no tool exists that allows a user to view a graph or network with uncertainty inherent within the attributes of nodes and edges. With the recent prevalence in data which can be represented as a graph (e.g. social networks), graphs are no longer simple, bi-modal datasets with only nodes and edges. Our task is often to work with multi-modal graphs that contain multiple types of nodes and edges where each node/edge can have many-perhaps hundreds-of attributes, and these attributes routinely have some uncertainty attached to them. Moreover, it is often useful to compare multiple graphs of this type as well as the ego networks of nodes in these graphs. In this paper we present various techniques and a prototype tool that can be used to visualize multi-modal graph data with uncertainty attached to each attribute and compare multiple such graphs with one another.

## 7868-16, Session 6

### Interactive visualization of scattered moment tensor data

H. Obermaier, Univ. of Kaiserslautern (Germany); M. I. Billen, Univ. of California, Davis (United States); H. Hagen, Univ. of Kaiserslautern (Germany); M. Hering-Bertram, Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (Germany)

In this paper we present a number of novel extraction and visualization techniques for interactive analysis of scattered moment tensor fields.

Symmetric second-order moment tensors derived from seismic measurements during earthquakes are related to stress tensors and contain important geological information about surface displacement in the earth's mantle. For a better understanding of earthquake sources, types and properties, analysis of this type of data sets is crucial. The methods introduced in this work facilitate interactive visualization of scattered moment tensor data to support earthquake, source, and displacement analysis. To this goal, we combine visualizations of three-dimensional spatial location and orientation information derived from moment tensor decompositions and present interaction techniques to provide semantic links between both viewpoints. We develop new tensor glyphs highlighting the indefinite character of moment tensors, while conveying important geological information such as wave propagation and fault orientations, showing significant improvements over classic beachball-glyph based visualization techniques. Additionally, we propose novel tensor clustering and averaging techniques based on a selection of moment tensor similarity measures along with accompanying visualization methods to overcome visual clutter, remove redundant information and help gain an insight into scattered moment tensor data sets for earthquake analysis.

## 7868-17, Session 7

### Visualizing frequent patterns in large multivariate time series

M. C. Hao, M. Marwah, Hewlett-Packard Labs. (United States); H. Janetzko, Univ. Konstanz (Germany); R. K. Sharma, Hewlett-Packard Labs. (United States); D. A. Keim, Univ. Konstanz (Germany); U. Dayal, Hewlett-Packard Labs. (United States); D. Patnaik, N. Ramakrishnan, Virginia Polytechnic Institute and State Univ. (United States)

The detection of previously unknown, frequently occurring patterns in time series, often called motifs, has been recognized as an important task. However, it is difficult to discover and visualize these motifs as their numbers increase, especially in large multivariate time series. To find frequent motifs, we use several temporal data mining and event encoding techniques to cluster and convert a multivariate time series to a sequence of events. Then we quantify the efficiency of the discovered motifs by linking them with a performance metric. To visualize frequent patterns in a large time series with potentially hundreds of nested motifs on a single display, we introduce three novel visual analytics methods: (1) motif layout, using colored rectangles for visualizing the occurrences and hierarchical relationships of motifs in a multivariate time series, (2) motif distortion, for enlarging or shrinking motifs as appropriate for easy analysis and (3) motif merging, to combine a number of identical adjacent motif instances without cluttering the display. Analysts can interactively optimize the degree of distortion and merging to get the best possible view. A specific motif (e.g., the most efficient or least efficient motif) can be quickly detected from a large time series for further investigation. We have applied these methods to two real-world data sets: data center cooling and oil well production. The results provide important new insights into the recurring patterns.

## 7868-18, Session 7

### Visual pattern discovery in timed event data

M. Schaefer, F. Wanner, F. Mansmann, C. Scheible, V. Stennett, A. T. Hasselrot, D. A. Keim, Univ. Konstanz (Germany)

Business processes have tremendously changed the way large companies conduct their business: The integration of information systems into the workflows of their employees ensures a high service level and thus high customer satisfaction. One core aspect of business process engineering are events that steer the workflows and trigger internal processes. Strict requirements on interval-scaled temporal patterns, which are common in time series, are thereby released through the ordinal character of such events. It is this additional degree of freedom that opens unexplored possibilities for visualizing event data.

In this paper, we present a flexible and novel system to find significant events, event clusters and event patterns. Each event is represented as a small rectangle, which is colored according to categorical, ordinal or interval-scaled metadata. Depending on the analysis task, different layout functions are used to highlight either the ordinal character of the data or temporal correlations. The system has built-in features for ordering customers or event groups according to the similarity of their event sequences, temporal gap alignment and stacking of co-occurring events. Two characteristically different case studies dealing with business process events and news articles demonstrate the capabilities of our system to explore event data.

## 7868-19, Session 7

### Enhancing visualization with real-time frequency-based transfer functions

E. Vucini, Technische Univ. Wien (Austria); D. Patel, Christian Michelsen Research AS (Norway); E. Groeller, Technische Univ. Wien (Austria)

Transfer functions have a crucial role in the understanding and visualization of 3D data. While exhaustive research has scrutinized the possible uses of one and multi-dimensional transfer functions in the spatial domain, to our knowledge, no attempt has been done to explore transfer functions in the frequency domain. In this work we propose transfer functions for the purpose of frequency analysis and visualization of 3D data. Frequency-based transfer functions offer the possibility to discriminate signals composed from different frequencies, to analyze problems related to signal processing, and to help understanding the link between the modulation of specific frequencies and their impact on the spatial domain.

We demonstrate the strength of the frequency-based transfer function by applying it to medical CT, ultrasound and MRI data, physics data

as well as synthetic seismic data. The interactivity of the proposed framework in building complex filters and the usage for structure or feature enhancement can be a useful addition to conventional classification techniques.

## 7868-34, Session 8

### Scientific visualization for data analysis

D. L. Kao, NASA Ames Research Ctr. (United States)

No abstract available

## 7868-20, Session 9

### The role of visualization and interaction in maritime anomaly detection

M. Riveiro, G. Falkman, Univ. of Skövde (Sweden)

The surveillance of large sea, air or land areas normally involves the analysis of large volumes of heterogeneous data from multiple sources. Timely detection and identification of anomalous behavior or any threat activity is an important objective for enabling homeland security. While it is worth acknowledging that many existing mining applications support identification of anomalous behavior, autonomous anomaly detection systems for area surveillance are rarely used in the real world. We argue that such capabilities and applications present two critical challenges: (1) they need to provide adequate user support and (2) they need to involve the user in the underlying detection process.

In order to encourage the use of anomaly detection capabilities in surveillance systems, this paper analyzes the challenges that existing anomaly detection and behavioral analysis approaches present regarding their use and maintenance by users. We analyze input parameters, detection process, model representation and outcomes. We discuss the role of visualization and interaction in the anomaly detection process. Practical examples from our current research within the maritime domain illustrate key aspects presented.

## 7868-21, Session 9

### Multiscale visual quality assessment for cluster analysis with self-organizing maps

J. Bernard, T. von Landesberger, S. Bremm, T. Schreck, Technische Univ. Darmstadt (Germany)

Cluster analysis is an important data mining technique for analyzing large amounts of data, by reduction to a limited number of clusters. Cluster visualization techniques aim at supporting the user in better understanding the characteristics and relationships among the found clusters. While promising approaches to visual cluster analysis already exist, these usually fall short of incorporating the quality of the obtained clustering results. However, due to the nature of the clustering process, quality plays an important aspect, as for most practical data sets, typically many different clusterings are possible. Being aware of clustering quality is important to judge the expressiveness of a given cluster visualization, or to adjust the clustering process with refined parameters, among others.

In this work, we present an encompassing suite of visual tools for quality assessment of an important visual cluster algorithm, namely, the Self-Organizing Map (SOM) technique. We define, measure, and visualize the notion of SOM cluster quality along a hierarchy of cluster abstractions. The quality abstractions range from simple scalar-valued quality scores up to the structural comparison of a given SOM clustering with output of additional supportive clustering methods. The suite of methods allows the user to assess the SOM quality on the appropriate abstraction level, and arrive at improved clustering results. We implement our tools in an integrated system, apply it on experimental data sets, and show its applicability.

**7868-22, Session 10****Privacy-preserving data visualization using parallel coordinates**

A. Dasgupta, R. Kosara, The Univ. of North Carolina at Charlotte (United States)

The proliferation of data in the past decade has created demand for innovative tools in different areas of exploratory data analysis, like data mining and information visualization. However, the problem with real-world datasets is that many of their attributes can identify individuals, or the data are proprietary and valuable. The data mining field has developed a variety of ways for dealing with such data, and has established an entire subfield for privacy-preserving data mining. Visualization, on the other hand, has seen little, if any, work on handling sensitive data. With the growing applicability of data visualization in real-world scenarios, the handling of sensitive data has become a non-trivial issue we need to address in developing visualization tools.

With this goal in mind, in this paper, we analyze the issue of privacy from a visualization perspective and propose a privacy-preserving data visualization technique based on clustering in parallel coordinates. We also outline the key differences in approach from the privacy-preserving data mining field and compare the advantages and drawbacks of our approach.

**7868-23, Session 10****A tri-linear visualization for network anomaly detection**

R. F. Erbacher, Northwest Security Institute (United States); R. B. Whitaker, Utah State Univ. (United States)

This research discusses a novel application of ternary plots to the visualization of network traffic data. These plots prove to be enormously effective at identifying anomalous network activity and can be valuable in monitoring network activity much more efficiently than can be done with existing techniques. The visualization was implemented in our existing visualization infrastructure to reduce development time. Testing was performed on actual network traffic data collected from a local network. Multiple anomalies were easily identifiable within the data set without any prior knowledge as to the contents of the test file. This paper discusses the ternary plot and its application to network traffic data, the formulas needed to calculate and display ternary coordinates, and the basic architecture for the visualization implementation.

**7868-24, Session 11****EmailTime: visual analytics and statistics for temporal email**

M. Erfani Joorabchi, J. Yim, C. D. Shaw, Simon Fraser Univ. (Canada)

Although the discovery and analysis of communication patterns in large and complex email datasets are difficult tasks, they can be a valuable source of information. We present EmailTime, a visual analysis tool of email correspondence patterns over the course of time that interactively portrays personal and interpersonal networks using the correspondence in the email dataset. Our approach is to put time as a primary variable of interest, and plot emails along a time line. EmailTime helps email dataset explorers interpret archived messages by providing zooming, panning, filtering and highlighting etc. To support analysis, it also measures and visualizes histograms, graph centrality and frequency on the communication graph that can be induced from the email collection. This paper describes EmailTime's capabilities, along with a large case study with Enron email dataset to explore the behaviors of email users within different organizational positions between January 2000 and December 2001. We defined email behavior as the email activity level of people regarding a series of measured metrics e.g. sent and received emails, numbers of email

addresses, etc. These metrics were calculated through EmailTime. Results showed specific patterns in the use email within different organizational positions. We suggest that integrating both statistics and visualizations in order to display information about the email datasets may simplify its evaluation.

**7868-25, Session 11****A web-enabled visualization toolkit for geovisual analytics**

Q. V. Ho, P. Lundblad, T. Åström, M. Jern, Linköping Univ. (Sweden)

We introduce a framework and class library (GAV Flash) implemented in Adobe's ActionScript, designed with the intention to significantly shorten the time and effort needed to develop customized web-enabled applications for visual analytics or geovisual analytics tasks. Through an atomic layered component architecture, GAV Flash provides a collection of common geo- and information visualization representations extended with motion behavior including scatter matrix, extended parallel coordinates, table lens, choropleth map and treemap, integrated in a multiple, time-linked layout. Versatile interaction methods are drawn from many data visualization research areas and optimized for dynamic web visualization of spatio-temporal and multivariate data. Based on an atomic layered component thinking and the use of programming interface mechanism the GAV Flash architecture is open and facilitates the creation of new or improved versions of existing components so that ideas can be tried out or optimized rapidly in a fully functional environment. Following the Visual Analytics mantra, a mechanism "snapshot" for saving the explorative results of a reasoning process is developed that aids collaboration and publication of gained insight and knowledge embedded as dynamic visualizations in blogs or web pages with associative metadata or "storytelling".

**7868-27, Poster Session****A 3D particle visualization system for temperature management**

B. Lange, N. Rodriguez, W. Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); H. Rey, X. Vasques, IBM (France)

This paper deals with a 3D visualization technique proposed to analyze and manage energy efficiency from a data center. Data are extracted from sensors located in the IBM green data center in Montpellier. These sensors measure different information such as hygrometry, pressure and temperature. We want to visualize in real time the large amount of data produced by these sensors. A visualization engine has been designed, based on particles system and a client server paradigm. In order to solve performance problems a Level Of Detail solution has been developed. These methods are based on the earlier work introduced by James Clark in 1976. In this paper we introduce the particle method used for this work and subsequently we explain different simplification methods that we have applied to improve our solution.

**7868-28, Poster Session****A digital topology-based method for the topological filtering of a reconstructed surface**

M. Allili, Bishop's Univ. (Canada); D. Li, Univ. de Sherbrooke (Canada); M. Allili, Univ. du Québec en Outaouais (Canada)

In this paper, we use concepts from digital topology for the topological filtering of reconstructed surfaces. Given a finite set  $S$  of sample points in 3D space, we use the voronoi-based algorithm of Amenta and Bern

to reconstruct a piecewise-linear approximation surface in the form of a triangular mesh with vertex set equal to  $S$ . A typical surface obtained by means of this algorithm often contains small holes that can be considered as noise. We propose a method to remove the unwanted holes that works as follows. We first embed the triangulated surface in a volumetric representation. Then, we use a 3D-hole closing algorithm of to filter the holes by their size and close the small holes that are in general irrelevant to the surface while the larger holes often represent topological features of the surface. We present some experimental results that show that this method allows automatically and effectively searching and suppressing unwanted holes in a 3D surface.

## 7868-30, Poster Session

### A meta-notation for data visualization

S. Y. Lee, U. Neumann, The Univ. of Southern California (United States)

We propose a notation devised to express major structural characteristics in widely-used data visualizations.

The notation consists of unary and binary operators and they can be combined together to describe a visualization.

By capturing significant structural features of a visualization, it can be applied in matching or comparing two visualizations in a conceptual level.

In this paper, we discuss the design of operators in our notation by presenting their concepts and usages.

Then we show how expressive our notation is by exploiting some of commonly-used data visualizations and study rules and relationships derived from them.

The advantage of the proposed approach is that the behaviors of operators provide a basic set of required capabilities with which an implementation can be organized.

Thus, it can be used to design or simulate a system, which interconnects and communicates with various types of data visualization tools by sending and receiving visualization requests between them.

## 7868-32, Poster Session

### Enhancing online timeline visualizations with events and images

A. Pandya, A. Mulye, S. T. Teoh, San José State Univ. (United States)

The use of timeline to visualize time-series data is one of the most intuitive and commonly used methods, and is used for widely-used applications such as stock market data visualization, and tracking of poll data of election candidates over time. While useful, these timeline visualizations are lacking in contextual information of events which are related or cause changes in the data. We have developed a system that enhances timeline visualization with display of relevant news events and their corresponding images, so that users can not only see the changes in the data, but also understand the reasons behind the changes. We have also conducted a user study to test the effectiveness of our ideas.

## 7868-33, Poster Session

### Multivariate data visualization via outdoor scenes

B. A. Hillery, R. P. Burton, Brigham Young Univ. (United States)

Abstract Visualization of multivariate data presents a challenge due to the sheer dimensionality and density of information. When presenting the data symbolically, this high information dimensionality and density makes it difficult to develop a symbology capable of displaying it in a single presentation. One approach to multivariate visualization involves creating symbols with higher dimensionality. Higher dimensional symbols can be problematic, since they typically require significant human attentive processing to interpret, offsetting their greater informational capacity. Although attempts have been made to develop higher-dimensional symbols that are processed in a preattentive fashion, success has proven elusive. Recent cognitive research indicates that outdoor scenes are processed in a preattentive manner. We evaluate outdoor scenes as a candidate for developing an effective higher-dimensional symbology by implementing them and comparing them to other methods, both standard and preattentive.

# Conference 7869: Computer Vision and Image Analysis of Art II

Wednesday 26 January 2011

Part of Proceedings of SPIE Vol. 7869 Computer Vision and Image Analysis of Art II

## 7869-15, Poster Session

### **Time and order estimation of paintings based on expert priors: applications in art history and curatorial treatment**

R. S. Cabral, J. P. Costeira, Univ. Técnica de Lisboa (Portugal); F. de La Torre, Carnegie Mellon Univ. (United States); A. Bernardino, G. Carneiro, Univ. Técnica de Lisboa (Portugal)

In this paper, we present a framework for estimating the ordering and date information of paintings and drawings. We formulate this problem as the embedding into a one dimension manifold, which aims to place paintings far or close to each other according to a measure of similarity. Our formulation can be seen as a manifold learning algorithm, albeit properly adapted to deal with existing questions in the art community.

To solve this problem, we propose a dynamic programming approach and a convex optimization formulation. Both methods are able to incorporate art expertise as priors to the estimation, in the form of constraints. Types of information include exact or approximate dating and partial orderings. We explore the use of soft penalty terms to allow for constraint violation to account for the fact that prior knowledge may contain small errors.

Since the proposed approach lies on the existence of statistics correlating image data with time variation, we provide a preliminary study of the features available in the image processing applied to the arts literature.

We describe possible applications where time information (and hence, this method) could be of use in art history, fake detection or curatorial treatment.

## 7869-17, Poster Session

### **Machine learning of multi-feature visual texture classifiers for the authentication of Jackson Pollock's drip paintings**

M. Al-Ayyoub, Stony Brook Univ. (United States); D. G. Stork, Ricoh Innovations, Inc. (United States); M. T. Irfan, Stony Brook Univ. (United States)

Jackson Pollock's action paintings are some of the most important works in American Abstract Expressionism. There are many works of doubtful attribution and outright fakes. Material studies of paint, support, priming, signatures and provenance are not always definitive in attribution studies, and so Taylor and his colleagues introduced a box-counting algorithm to estimate fractal and scale-space signatures of Pollock's works. They reported that such signatures generally differed from fake Pollocks that their method could be used as part of authentication protocol. Our current project builds upon prior work by training on more image data, and of higher resolution, of both genuine Pollocks and fakes, and aims at employing feature extraction, feature selection and classifier selection techniques commonly used in pattern recognition research. Here we present the results of several supervised classification frameworks, such as Support Vector Machines (SVM), decision trees (DT), and AdaBoost. We extract features from the fractality, multifractality, pink noise patterns, topological genus, and curvature properties of the images of candidate paintings, and address learning issues that have arisen due to the small number of examples. In our experiments, we have found that the unmodified classifiers like Support Vector Machines or Decision Tree alone do not give good results. In particular, Decision Tree alone gives an accuracy of roughly 60%. We used a Decision Tree as a weak learner in AdaBoost, we obtained accuracies of roughly 80%. Thus, although our set of observations is very small, boosting methods can significantly improve classification accuracy for Pollock authentication.

## 7869-18, Poster Session

### **Improved curvature-based inpainting applied to fine art: recovering van Gogh's partially hidden brush strokes**

D. G. Stork, Ricoh Innovations, Inc. (United States); Y. Kuang, F. Kahl, Lund Univ. (Sweden)

Underpaintings and pentimenti (revealed through x-ray imaging and infrared reflectography) comprise important evidence revealing the intermediate states of a work and thus the working methods of many artists. Based on digital image processing and statistical analysis, Shahram, Stork and Donoho introduced the De-pict algorithm, which recovers layers of brush strokes in paintings with open brush work where several layers are partially visible, such as van Gogh's "Self portrait with a grey felt hat." While that preliminary work served as a proof of concept that computer image analytic methods could recover some occluded images, the work needed further refinement before it could be a tool for art scholars. Our current work rectifies this omission. We extended that earlier method by refining the inpainting step through the inclusion of curvature-based constraints, in which a mathematical curvature penalty term for extracted chromatic level lines biases the reconstruction toward matching the artist's hand motion. We refine our methods using "ground truth" image data: passages of four layers of brush strokes in which the intermediate layers were recorded photographically. At each successive top layer (currently identified by the user), we used  $k$ -means clustering combined with graph cuts to obtain chromatically and spatially coherent segmentation of brush strokes. We then reconstructed strokes at the deeper layer with our curvature-based inpainting algorithm based on chromatic level lines. Our methods are clearly superior to previous versions of the De-pict algorithm on van Gogh's works, and could be applied to the classic drip paintings of Jackson Pollock, where the drip work is more open and the physics of splashing paint ensures that the curvature more uniform than in the hand created brush strokes of van Gogh.

## 7869-19, Poster Session

### **Did Caravaggio employ optical projections? An image analysis of the parity in the artist's paintings**

D. G. Stork, Ricoh Innovations, Inc. (United States)

We examine one class of evidence put forth in support of the recent claim that the Italian Baroque master Caravaggio secretly employed optical projectors as a direct drawing aid. Specifically, we test the claims that there is an "abnormal number" of left-handed figures in his works and, more specifically, that "During the Del Monte period he had too many left-handed models." We also test whether there was a reversal in the handedness of a given individual model in different paintings. Such evidence would be consistent with the claim that Caravaggio switched between using a convex lens projector to using a concave mirror projector and would support, but not prove, the claim that Caravaggio used optical projections. We estimate the parity (+ or -) of each of Caravaggio's 76 appropriate oil paintings based on the handedness of figures, the orientation of asymmetric objects, placement of scabbards, depicted text, and so on, and search for statistically significant changes in handedness in figures. We also track the direction of the illumination over time in the artist's oeuvre. We discuss some historical evidence as it relates to the question of his possible use of optics. We find the proportion of left-handed figures lower than that in the general population (not higher), and no significant change in estimated handedness even of individual models. Optical proponents have argued that "Bacchus" (1597) portrays a left-handed figure, but we give visual and cultural evidence and conclude

that this figure is instead right-handed, thereby rebutting this claim that the painting was executed using optical projections. Moreover, scholars recently re-discovered the image of the artist with easel and canvas reflected in the carafe of wine at the front left in the tableau in "Bacchus," showing that this painting was almost surely executed using traditional (non-optical) easel methods. We conclude that there is 1) no statistically significant abnormally high number of left-handed figures in Caravaggio's oeuvre, including during any limited working period, 2) no statistically significant change in handedness among all figures or even individual figures that might be consistent with a change in optical projector, and 3) the visual and cultural evidence in "Bacchus" shows the figure was right-handed and that the artist executed this work by traditional (non-optical) easel methods. We conclude that the general parity and handedness evidence does not support the claim that Caravaggio employed optical projections.

## 7869-20, Poster Session

### **A computer graphics reconstruction and optical analysis of scale anomalies in Caravaggio's "Supper at Emmaus"**

D. G. Stork, Ricoh Innovations, Inc. (United States); Y. Furuichi, Consultant (Japan)

David Hockney has argued that the right hand of the disciple, thrust to the rear in "Supper at Emmaus" (1606), is anomalously large as a result of Caravaggio refocusing a concave mirror projector. We show rigorously that to achieve such an anomalously large image, Caravaggio would have needed to make extremely large, conspicuous and implausible alterations to his studio setup, moving his purported mirror nearly one meter forward between projecting the disciple's left hand and then his right hand. Moreover, the 192-cm-wide canvas would have been off to the side of any aperture, so that the light from the subject could strike the mirror and canvas. But such a placement would mean that the light from figures at the extremes of the tableau would have struck the mirror at a large angle with respect to the optical axis, leading to blurry, useless images. To avoid such severe degradation in the projected images, Caravaggio would have likely had to switch his canvas from one side of the aperture to the other in order to capture the figures at the extremes of the tableau. All these major disruptions to his studio would have impeded---not aided---Caravaggio in his work. We use computer graphics reconstruction of Caravaggio's studio to explore and demonstrate these problems. We argue that Caravaggio most likely set the sizes of these hands "by eye" for artistic reasons. In this way we argue against the optical projection claim for this painting.

## 7869-21, Poster Session

### **Image analysis of the underdrawings in Lorenzo Lotto's "Husband and wife"**

D. G. Stork, Ricoh Innovations, Inc. (United States); A. J. Kossolapov, State Hermitage Museum (Russian Federation)

Underdrawings and pentimenti reveal intermediate states of a painting and thus the working methods of some artists. It has been claimed that Lorenzo Lotto used optical projections during the execution of "Husband and wife" (1543) and that underdrawings might reveal evidence of tracing of optical projections. We analyze x-ray and infra-red images of the underdrawings in this painting---captured under careful, museum-studio conditions and enhanced through digital image processing---with special attention to the possibility of evidence of tracings of optical projections in the keyhole portion of the depicted carpet. We also study the work in situ and in high-resolution macro optical images of the central portion of the carpet pattern. These photographs reveal that the top portion of the keyhole pattern is not "blurry, like an out-of-focus image," but instead was merely executed in a somewhat broader brush than neighboring passages. These photographs also show that the white portions were executed atop a broad layer of dark red, and thus that no record of an optical projection

would have been present when Lotto executed the visible portion---the portion that led to the optical claim. There is no evidence of tracing marks---in pencil or in any medium---in the top, visible portion of this passage either. As such, this visual, infra-red and x-ray evidence does not support the claim that this painting was executed under optical projections. We also review contemporary textual evidence in early 16th-century Venice that has been used to support the optical projection claim for Lotto and conclude that it also fails to support the projection claim for this painting.

## 7869-22, Poster Session

### **Automated classification of quilt photographs into crazy and non-crazy**

A. Gokhale, Indian Institute Of Technology, Kharagpur (India); P. Bajcsy, Univ. of Illinois at Urbana-Champaign (United States)

This work addresses the problem of automatic classification and labeling of 19th- and 20th-century quilts from photographs, which are classified according to the quilt patterns into crazy and non - crazy categories. The motivation of our work is in automated annotation of a large collection of quilt images for research purposes of humanists. The research value of annotations for humanists is in understanding the distinct characteristics of an individual quilt-maker or relevant quilt-making groups in terms of their choices of pattern selection, color choices, layout, and original deviations from traditional patterns. The current annotation method is manual and the assignment is achieved by visual inspection. According to our knowledge, there does not exist currently a clear definition of the level of crazy-ness, nor an automated method for classifying patterns as crazy and non-crazy.

We approach the problem by modeling the level of crazy-ness by the distribution of clusters of color-homogeneous connected image segments of similar shapes. The model is turned into a set of image features that are extracted and represent our model of crazy-ness. The features are input into a supervised classification method, such as the Support Vector Machine (SVM) with the radial basis function kernel and optimized using 10-fold cross validation, used in our work.

The classification methodology consists of four steps. In the first step, a color-homogeneous Region/Textel is detected by color based K-means clustering followed by connectivity analysis. This step leads to a cluster of color - homogeneous regions that represent quilt patches with similar colors. The second step uses a divide-and-conquer approach to identify sub-clusters that share similar shape properties such as area and perimeter. For each sub-cluster, statistics of nearest neighbor distances are computed for example, the mean and variance of distances. In the third step, a quilt signature per image is formed from parameters including the number of clusters containing a single color-homogeneous Region/Textel, number of clusters containing multiple color-homogeneous Region/Textel, the maximum and average number of color-homogeneous regions per cluster of regions and the minimum variance present in any cluster. Our selection of these parameters is based on the observation that crazy patterns have a small number of color-homogeneous and shape-similar regions in a cluster and a large number of clusters containing only a single region. They also have no symmetry and hence large variance in inter - Region nearest neighbor distance. In contrary, non-crazy patterns will have a small number of clusters and a large number of color-homogeneous and shape-similar regions in a cluster. Finally, a Support Vector Machine (SVM) model is trained using labeled quilt images and 10-fold cross validation is used.

We implemented the classification methodology using a combination of Java and Matlab code. The algorithm was applied to 40 quilt images from the MATRIX database at the Michigan State University. We report almost 90 percent classification accuracy over 40 images using SVM and its radial basis function. In the future, we plan on extending the categorical model of crazy-ness to a continuous function reflecting the level of craziness.

## 7869-23, Poster Session

### **Polarized light scanning for cultural heritage investigation**

J. A. O. Toque, Y. Murayama, A. Ide-Ektessabi, Kyoto Univ. (Japan)

Numerous cultural heritage art works have shiny surfaces resulting from gold, silver, and other metallic pigments. In addition varnish overlayer on oil paintings makes it challenging to retrieve true color information. This is due to the great effect of lighting condition when images are acquired and viewed. The reflection of light from such surfaces is a combination of the surface's specular and diffused light reflections. In this paper the specific problems encountered when digitizing cultural heritage were discussed. Experimental results using the images acquired with a high-resolution large flat bed scanner, together with a mathematical method for image capture were presented and discussed in detail. Focus was given in separating the diffused and specular components of the reflected light for the purpose of analytical imaging. The mathematical algorithm developed in this study enables imaging of cultural heritage efficiently within a practical time limit.

## 7869-24, Poster Session

### **After digital cleaning: visualization of dirt layer**

M. N. Soriano, C. M. Palomero, Univ. of the Philippines (Philippines)

We have previously shown that a neural network can be trained to perform digital cleaning on an oil painting by learning the transformation from dirty to clean pixels. Such a network was used to virtually clean an image of Filipino National Artist, Fernando Amorsolo's 1948 oil painting, Malacanang by the River. When the painting was removed from its original frame, it was observed that the parts that were previously covered by the frame were generally cleaner than the exposed parts. Using the cleaner unexposed parts as our basis for what the painting might have looked like had it not undergone dirtying, we trained a neural network to learn the transformation from dirty to clean segments of a painting. A total of 1,350 pairs of input and output pixels were manually selected from all around the edges of the painting image. Special care was taken to preserve texture information by taking the input and output pair from the same texture component (brushstroke, shadow of brush stroke, bumps and dips of canvas weave).

A comparison of the painting image before and after digital cleaning shows more vivid colors and a cleaner look for the digitally-cleaned painting. The mask-like boundary of dirt between the parts that were exposed and unexposed due to the frame is also less visible.

Exploiting the trained neural network's ability to solve the desired output for completely new inputs allowed us to perform whole painting-cleaning in a totally non-invasive manner.

In this work, we demonstrate two methods to visualize the dirt that the digital-cleaning procedure removed. First is the vector difference method, wherein we compute the color change between each original and cleaned pixel as a vector difference in different color spaces. The color difference vector is superimposed onto a neutral color and rendered for the whole image. In the second method we model the dirt as a transparency film that is superimposed onto the clean painting. A transparency function dependent on the color difference between original and cleaned pixel determines the opacity of each pixel in the virtual "dirt" film. Spectral measurements on known white portions of the painting provide an estimate of the transmittance as a function of wavelength. Both methods provide good visualization of the dirt removed and could offer insights on a painting's dirtying or discoloration process.

## 7869-01, Session 1

### **Documenting Van Eyck's Ghent Altarpiece: field work experiences from the crypt**

R. Spronk, Queen's Univ. (Canada) and Radboud Univ. Nijmegen (Netherlands)

Jan and Hubert van Eyck's famous Ghent Altarpiece (1432) in St. Bavo Cathedral in Ghent, Belgium, is the single most important work of Early Netherlandish painting in existence, and is generally accepted to be among the most important surviving art works in the world. In 2010, the polyptych was subjected to an urgent conservation treatment and to technical documentations, to establish whether a full restoration is necessary in the near future. The individual panels were documented with infrared reflectography, under ultraviolet light, and with high-resolution digital macro-photography in visible light and in the infrared. Some 20 details were documented with X-radiography and compared with such documents from the 1980s, to establish whether recent deteriorations can be observed at the craquelure-level. The supports of the four central panels were analysed with dendrochronology, to complement the already available findings for the panels in the left and right zones. The central panels were documented with multispectral infrared scanning and with non-destructive instrumental analyses such as XRF and XRD, among others. In this presentation, Ron Spronk will describe the project and some of its initial results.

## 7869-02, Session 1

### **Computer analysis of lighting style in fine art: an inter-artist study**

D. G. Stork, Ricoh Innovations, Inc. (United States)

Stylometry---the mathematical description of artists' styles---has been based on a number of properties of visual art, such as color, brush stroke shape, visual texture, and curvature measures of contours. We introduce lighting coherence (the agreement among lighting directions estimated throughout a painting) as a property of style. Surrealists such as Giorgio de Chirico worked from imagination rather than models and deliberately introduced incommensurate and contradictory lighting clues; artists of the high Renaissance who worked from nature, such as Leonardo, and photorealists who worked from photographs, such as Richard Estes, strive for great coherence and agreement among lighting clues. Perceptual studies show that observers are poor judges of lighting consistency in photographs and thus, by extension, paintings, while computer methods such as rigorous cast-shadow analysis, occluding contour analysis and spherical harmonic based estimation of light fields can be quite accurate. For this reasons, computer lighting analysis methods may provide a new tool for art historical studies. We define a scalar measure of lighting coherence based on the distribution of lighting directions estimated in a painting. We use this measure to describe the lighting in paintings previously analyzed (e.g., Vermeer's "Girl with a pearl earring," de la Tour's "Christ in the carpenter's studio," Caravaggio's "Magdalen with the smoking flame" and "Calling of St. Matthew") and extend our corpus to works where lighting coherence is of interest to art historians, such as Caravaggio's "Adoration of the Shepherds" (1609) for the Capuchin church of Santa Maria degli Angeli. Our measure of lighting coherence may help reveal the working methods of some artists, and in diachronic studies reveal changes in the working methods of a given artist. We speculate on artists and art historical questions that may ultimately profit from this new computational tool.

## 7869-03, Session 2

### **The automatic annotation and retrieval of digital images of prints and tile panels using network link analysis algorithms**

G. Carneiro, J. P. Costeira, Univ. Técnica de Lisboa (Portugal)

The study of the visual art of printmaking is fundamental for art history.



Printmaking methods have been used for centuries to replicate visual art works, and these works have influenced artists for centuries.

Particularly in this work, we are interested in the influence of these prints on artistic tile panel painters, who have produced an impressive body of work in Portugal. The study of such panels have gained interest from art historians, who essentially try to find links between prints and tile panels in order to comprehend the evolution of visual arts. Several databases of digitized art images have been used for such end, but the use of these databases relies on manual image annotations by art historians and an effective internal organization. We propose an automation of these tasks using statistical pattern recognition techniques. Specifically, we introduce a new method for the automatic analysis of databases containing digital images of prints and tile panels. The main contribution of our paper is a novel statistical pattern recognition method based on link analysis. The successful implementation of this system shall enable a more efficient study and research of art databases by making the analysis process faster and less dependent on expert users.

## 7869-04, Session 2

### Explaining scene composition using kinematic chains of humans: application to Portuguese tiles history

N. P. da Silva, M. Marques, G. Carneiro, J. P. Costeira, Univ. Técnica de Lisboa (Portugal)

Painted tile panels (Azulejo) are one of the most specific Portuguese forms of art. Most of these panels are inspired on, and sometimes are literal copies of, paintings and prints. In order to study the influences in Azulejo (tile), Art Historians need to trace these roots. To do that they manually search databases of prints searching for similarities between the tile panel and the paintings and prints database. This is an overwhelming task that should be automated as much as possible. Among several cues, the pose of humans and the general composition of people in a scene is quite discriminative.

This paper describes a human pose matching framework for composition analysis in 16th-18th century prints and engravings. The pose annotations in the prints are represented by a wire frame with the kinematic chain of the articulated body of a human. Modeling the annotated poses as subspaces (such as used in [1,2]) we can create a dictionary of poses. Given an image from a panel of tiles, we perform subspace comparison to find the prints that inspired the panel. In order to handle the deformations due to the different scaling of body parts and to the artistic interpretation of the original drawing, we explain the observed pose (subspace) as a sparse convex combination of the subspace dictionary [3,4]. This convex combination is interpreted as a nonparametric (empirical) probability distribution. The system retrieves the prints corresponding to the positive coefficients, together with the associated probability of having inspired the tiles' work.

[1] C. Tomasi, T. Kanade, "Shape and motion from image streams under orthography: a factorization method", *IJCV* 9(2):137-154, 1992.

[2] C. Bregler, A. Hertzmann, H. Biermann, "Recovering Non-Rigid 3D Shape from Image Streams", in *Proc. IEEE CVPR'00*, vol. 2, pp.2690, 2000.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso", *JRSS B* 58(1):267-288, 1996.

[4] C. Chennubhotla, A. Jepson, "Sparse PCA Extracting Multi-Scale Structure from Data", in *Proc. IEEE ICCV'01*, vol. 1, pp.641, 2001.

## 7869-05, Session 2

### Top-down analysis of low-level object relatedness leading to semantic understanding of medieval image collections

P. Yarlagadda, J. A. Monroy, B. Carque, B. Ommer, UniversitätsKlinikum Heidelberg (Germany)

Image understanding, an active research area in computer vision, deals with the analysis of low-level relationships among various object instances found in a collection of images. Such object relationships prove to be useful input for tasks such as understanding the different principles of artistic design and for identifying the characteristics of different teams of artists that have drawn a collection of images. In this contribution, we identify a suitable feature representation for the object instances and utilize a top-down approach to obtain low-level relationships among them based on the features. Finally, we analyze object relationships to identify different artistic styles such as the concise and accurate style of Hagenan workshop of Diebold Lauber or the delicate and sketchy style of the swabian workshop of Ludwig Henflin. Our work is based on a database consisting of 27 late medieval paper manuscripts from Upper German origin archived by Heidelberg University library. These codices are illustrated with more than 2,000 half or full-page tinted drawings.

## 7869-06, Session 2

### A framework for analysis of large database of old art paintings

J. Da Rugna, G. Chareyron, Pôle Univ. Léonard de Vinci (France)

For many years, a lot of museums and countries organize the high definition digitalization of their own collections. In consequence, they generate massive data for each object. In this paper, we only focus on art painting collections. Nevertheless, we faced a very large database with heterogeneous data. Indeed, image collection includes very old and recent scans of negative photos, digital photos, multi and hyper spectral acquisitions, X-ray acquisition, and also front, back and lateral photos. Moreover, we have noted that art paintings suffer from much degradation: craquelure, softening, artifact, human damages and, overtime corruption. Considering that, it appears necessary to develop specific approaches and methods dedicated to digital art painting analysis. Consequently, this paper presents a complete framework to evaluate, compare and benchmark devoted image processing algorithms.

In the first part we present images, diversity of acquisition methods and all underlying difficulties linked to our problematic. In the second part we discuss the overall schema of our framework. Finally, we introduce a new approach to craquelure detection designed to old painting.

To conclude, our framework is designed to analyze and benchmark image processing algorithm in the specific context of old art painting. It is highly upgradeable and customizable.

## 7869-07, Session 3

### Image fusion for art analysis

B. Zitova, M. Beneš, J. Blažek, Institute of Information Theory and Automation (Czech Republic)

Our paper addresses problem of multimodal data acquisition and following data visualization for an art analysis and interpretation. Various types of modalities for acquisition of digital images are used for art analysis. The data we can obtain using various modalities differ in two ways. The group of differences we are interested in are details or characteristics of an artwork, which are apparent just in the certain modality. Next to this the acquired images differ by their mutual geometry and by their radiometric quality. These difference classes represent two categories of image processing methods. The first one deals with effective ways how to combine the acquired information into one image - image fusion. The second category of methods covers data enhancement and restoration algorithms. In the proposed paper we will present the methodology for identification of objects different in individual modality channels, the comparison of applicability of image fusion algorithms for art analysis, and the contrast preserving image fusion method. From the second category we will present the image quality enhancement for blurred and noisy data with special attention to the multimodal case.

### 7869-08, Session 3

#### Recovery of handwritten text from the diaries and papers of David Livingstone

K. T. Knox, Air Force Research Lab. (United States); R. L. Easton, Jr., Rochester Institute of Technology (United States); W. A. Christens-Barry, Equipoise Imaging, LLC (United States); K. Boydston, Megavision Inc. (United States)

During his explorations of Africa, David Livingstone kept a diary and wrote letters about his experiences. Near the end of his travel he ran out of paper and ink and began recording his thoughts on leftover newspaper with ink made from seeds. These writings suffer from fading, from interference with the printed text and from bleedthrough of the handwriting on the other side of the paper, making them hard to read. New image processing techniques have been developed to deal with these papers to make Livingstone's handwriting available to the scholars to read.

A scan of the David Livingstone's papers was made using a twelve-wavelength, multispectral imaging system. The wavelengths ranged from the ultraviolet to the near infrared. In these wavelengths, the three different type of writing behave differently, making them distinguishable from each other. Because one of the writings has bled through from the other side of the paper, the scans of the reverse side can be used to enhance the effect. The result is a pseudocolor image that shows the desired text in a high-contrast color, while the two texts to be suppressed appear in low-contrast colors making the desired text legible to the scholars.

### 7869-09, Session 3

#### Automation of digital historical map analyses

T. Shaw, P. Bajcsy, Univ. of Illinois at Urbana-Champaign (United States)

This paper addresses the problem of automating analyses of historical maps. The problem is motivated by the lack of accuracy and consistency in the current comparison process of geographical objects found in historical maps by visual inspections. The objective of our work is to compare shape characteristics of the Great Lakes region created in the 17th through the 19th centuries in a dataset of 40 French and British maps. Our approach decomposes the visual inspection into steps such as object segmentation, spatial scale calibration, extraction of calibrated object descriptors and comparison of descriptors over time and multiple cartographer houses. The automation of object segmentation is achieved by template shape-based segmentation using Hu moments as shape descriptors and ball-based region growing. The automation of spatial calibration is accomplished by classification of lines along map borders and by mapping striped boundaries intersected by latitude and longitude lines into degrees of arc length. Thus, shape characteristics of segmentation results in pixels can be converted to geographical units, for example, an area of a lake in square miles. We report our experimental evaluations of automation accuracy based on the 40 French and British maps, as well as the knowledge obtained from the area comparisons.

### 7869-16, Session 3

#### Automatic multispectral ultraviolet, visible and near-infrared capturing system for the study of artwork

J. A. Herrera-Ramirez, M. Villaseca, J. Pujol, Univ. Politècnica de Catalunya (Spain)

The spectral imaging technology has proved its usefulness in a variety of sensing applications ranging from remote sensing, such as satellite or radar imaging, to artwork conservation where it has recently gained importance. In the case of the near-infrared region

(NIR) the spectral imaging helps in the analysis of paintings identifying pigments through the analysis of the spectra and their distribution over the artwork, besides it provides a good tool in the exploration and study of the underdrawings of the art pieces. In the visible region of the spectra several color related studies, like color imaging and archiving, can also be carried out. Besides, in the case of the UV range, its proper use serves for the detection of organic materials that have fluorescence properties. In this work a multispectral capturing prototype system intended for the analysis of artwork in the ultraviolet, visible and near-infrared range of the spectrum has been developed. Several aspects related to its construction as well as simulated and first experimental results are presented in this study. Specifically, the system captures the information in the range of 350nm to 1650nm. It is based in two imaging sensors, one CCD and an InGaAs camera, and a set of LEDs acting as a multiplexed light source that allow extracting the information of several spectral bands inside this mentioned range. The data can be analyzed as a cube of data, a stack of monochromatic images, or a set of reflectance spectra samples, one per each pixel over the field of view. To get this last information the automatic system is complemented by a set of computational routines for the acquisition of images and the study of them through several mathematical algorithms (Moore-Penrose pseudoinverse and Matrix R method), which allow the reflectance spectra of the samples imaged to be reconstructed from the multispectral images acquired if the system is previously trained with a set of known and calibrated samples. A distortion algorithm, based on the preliminary acquisition of a pattern consisting of a grid of circular spots, to correct the possible artifacts introduced in the capture process is utilized. In addition to this distortion correction, a flat field procedure which uses a gain and an offset matrices is carried out over the images to overcome the possible inhomogeneities in the illumination of the sample and jointly help in the performance of a subsequent stitching algorithm or mosaicing, based on the existing correlation between the digital levels of the multispectral images, for high spatial resolution images in large format paintings. This whole system underscores the potential of further developments in this field.

### 7869-11, Session 4

#### Automatic registration of multi-band reflectance and luminescence images

D. Conover, The George Washington Univ. (United States); J. Delaney, National Gallery of Art (United States); M. Loew, The George Washington Univ. (United States)

As high-resolution images of paintings, acquired using various imaging modalities, become more available, it is increasingly important to achieve accurate registration between the images in order to obtain a better understanding of how the painting was constructed. The goals of this project are to first accurately register reflectance and luminescence images, as well as true color, scanned x-rays, and infrared reflectograms. Then, using the registered set of images, automatically identify and emphasize information not visible at the surface of the painting. The registration algorithm will identify large sets of candidate fiducial points in a first image, pair each fiducial point with a point in a second image, select the best set of fiducial point pairs, and then transform the second image using the best set of pairs to bring the second image into alignment with the first.

### 7869-12, Session 4

#### Art documentation quality in function of 3D scanning resolution and precision

E. Bunsch, Museum Palace at Wilanow (Poland); R. Sitnik, Warsaw Univ. of Technology (Poland)

Currently, a lot of different 3D scanning devices are used for 3D acquisition of art artifact surface shape and color. Each of them has different technical parameters starting from measurement principle (structured light, laser triangulation, interferometry, holography) and ending on parameters like measurement volume size, spatial

resolution and precision of output data and color information. Some of the 3D scanners can grab additional information like surface normal vectors, BRDF distribution, multispectral color. In this paper, the problem of establishing of threshold for technical parameters of 3D scanning process in function of required information about the object is discussed. Only two main technical parameters are under consideration, due to cover as many different 3D scanning devices as possible, measurement sampling density (MSD - represented by number of points per square millimeter) and measurement uncertainty (MU -directly influencing final data accuracy). Also different materials and finishing techniques requires different thresholds of MSD and MU parameters to collect similar documentation (for example documentation of object state for art conservation department) of different objects. In this paper we consider exemplary painting and stone samples to visualize what object features can be observed within different values of MSD and MU parameters.

#### 7869-13, Session 4

### **Investigation of the degradation mechanism and discoloration of traditional Japanese pigments by multispectral imaging**

J. A. O. Toque, A. Ide-Ektessabi, Kyoto Univ. (Japan)

Pigment degradation has been a subject of interest among researchers in the field of cultural heritage studies. Knowing how pigments behave when subjected to different elements such as high temperature, humidity, electromagnetic radiation and many more others is of prime importance. In this study, the effects of subjecting Japanese pigments to high temperature were investigated. Focus was given on the effects in terms of pigment discoloration and the micromechanism of degradation. Multispectral images were used to track the changes in color and spectral reflectance by reconstructing colorimetric and spectral information from the images. The multispectral images were taken using a high-resolution flat-bed scanner equipped with a line-CMOS camera. In addition, the pigments were characterized using commercially available spectrometers, X-ray diffraction and X-ray fluorescence spectroscopy to ascertain the influence of high temperature exposure of the pigments. The high resolution multispectral scans gave the most valuable insights into the discoloration and micromechanism of pigment degradation since they provide both analytical and visual information.

#### 7869-14, Session 4

### **Improved methods for dewarping images in convex mirrors in fine art: applications to van Eyck and Parmigianino**

Y. Usami, Waseda Univ. (Japan); D. G. Stork, Ricoh Innovations, Inc. (United States); J. Fujiki, National Institute of Advanced Industrial Science and Technology (Japan); H. Hino, Waseda Univ. (Japan); S. Akaho, National Institute of Advanced Industrial Science and Technology (Japan); N. Murata, Waseda Univ. (Japan)

We derive and demonstrate new methods for dewarping images depicted in convex mirrors in artwork and for estimating the three-dimensional shapes of the mirrors themselves. Previous methods were based on the assumption that mirrors were spherical or paraboloidal, an assumption unlikely to hold for hand-blown glass spheres used in early Renaissance art. We assume merely that the mirror is radially symmetric and require merely that there be straight source lines in the actual scene. We express the mirror's shape locally as a mathematical series and pose the image dewarping task as that of estimating the coefficients in the series expansion. Central to our method is the "plumbline principle": that the optimal coefficients are those that dewarp the mirror image so as to straighten lines corresponding to straight lines in the source scene. We solve for these coefficients algebraically through principal component analysis, PCA. We find that it is important to select an appropriate set of basis functions, particularly when there are but few straight lines, so as to avoid overfitting. Our method relies on a global figure of merit to balance warping errors throughout the image and thereby reduces a reliance on the somewhat subjective criterion used in earlier methods. Once we have found the optimal image dewarping, we compute the mirror shape by solving a differential equation based on the estimated dewarping function. We demonstrate our methods on the Arnolfini mirror and reveal a dewarped image superior to those found in prior work---an image noticeably more rectilinear throughout and having a more coherent geometrical perspective and vanishing points. Moreover, we find the mirror deviated from spherical and paraboloidal shape; this implies that it would have been useless as a concave projection mirror, as has been claimed.

# Conference 7870: Image Processing: Algorithms and Systems IX

Monday-Tuesday 24-25 January 2011

Part of Proceedings of SPIE Vol. 7870 Image Processing: Algorithms and Systems IX

## 7870-01, Session 1

### An adaptive optimization of the polynomial wavelet threshold

D. Akopian, S. G. Sathyanarayana, S. S. Agaian, The Univ. of Texas at San Antonio (United States)

This paper presents a new class of polynomial threshold operators for denoising signals using wavelet transforms. The operators are parameterized to include classical soft- and hard- thresholding operators and have many degrees of freedom to optimally suppress undesired noise and preserve signal details. To avoid the complicated process of signal model identification for specific type of signals, an adaptive least mean squares (LMS) optimization method is proposed for the polynomial coefficients. It is shown that the optimal choice of the polynomial coefficients can be formulated as a least squares (LS) problem if the training sequences are available. Then an adaptive LMS approach for the optimization of wavelet coefficients is proposed and studied as an approach to reduce computational costs. This approach allows optimizing coefficients without matrix inversions and if needed optimally adapt the threshold polynomials for different wavelet transform bands without computational overheads.

## 7870-02, Session 1

### Video denoising using separable 4D nonlocal spatiotemporal transforms

M. T. Maggioni, Tampere Univ. of Technology (Finland); G. Boracchi, Politecnico di Milano (Italy); A. Foi, K. O. Egiazarian, Tampere Univ. of Technology (Finland)

We propose a powerful video filtering algorithm that exploits temporal and spatial redundancy characterizing natural video sequences. The algorithm implements the paradigm of nonlocal grouping and collaborative filtering, where a higher-dimensional transform-domain representation is leveraged to enforce sparsity and thus regularize the data. The proposed algorithm exploits the mutual similarity between 3-D spatiotemporal volumes constructed by tracking blocks along trajectories defined by the motion vectors. Mutually similar volumes are grouped together by stacking them along an additional fourth dimension, thus producing a 4-D structure, termed group, where different types of data correlation exist along the different dimensions: local correlation along the two dimensions of the blocks, temporal correlation along the motion trajectories, and nonlocal spatial correlation (i.e. self-similarity) along the fourth dimension. Collaborative filtering is realized by transforming each group through a decorrelating 4-D separable transform and then by shrinkage and inverse transformation.

In this way, collaborative filtering provides estimates for each volume stacked in the group, which are then returned and adaptively aggregated to their original position in the video.

The proposed filtering procedure is particularly effective at noise removal, for which it outperforms the state of the art.

## 7870-03, Session 1

### Intelligent edge enhancement using multilayer neural network based on multi-valued neurons

I. N. Aizenberg, S. Alexander, J. T. Jackson, T. Neal, J. Wilson, K. Kendrick, Texas A&M Univ.-Texarkana (United States)

In this paper, we suggest to solve the edge enhancement problem using an intelligent approach. We use here a multilayer neural network based on multi-valued neurons (MLMVN) as an intelligent edge enhancer. A problem of neural edge enhancement using a classical multilayer perceptron (MLP) was already considered by some authors. Since MLMVN significantly outperforms MLP in terms of learning speed, complexity and classification/prediction rate, when solving benchmark and different real-world problems, it is very attractive to apply it for solving the edge enhancement problem.

The main result, which is presented in the paper, is a proven ability of MLMVN to enhance edges corresponding to various edge detection operators. Moreover, it is possible to enhance edges on the noisy images ignoring a noisy texture. It is shown that to learn any edge detection operator using MLMVN, it is enough to use a single image for the learning purposes.

The most important conclusion is that a neural network can learn different edge detection operators from a single example and then it can process those images that did not participate in the learning process detecting edges definitely according to the learned operator with a high accuracy.

## 7870-04, Session 1

### New loss functions for ordered hypothesis machines

R. B. Porter, Los Alamos National Lab. (United States)

Just as linear models generalize the sample mean and weighted average, weighted order statistic models generalize the sample median and weighted median. This analogy can be continued informally to generalized additive models in the case of the mean, and Stack Filters in the case of the median. Both of these model classes have been extensively studied for signal and image processing, but it is surprising to find that for pattern classification, their treatment has been significantly one sided. Generalized additive models are now a major tool in pattern classification and many different learning algorithms have been developed to fit model parameters with finite training data. However Stack Filters remain largely confined to signal and image processing and application to classification is rarely seen.

In previous work we found that applying Stack Filter model classes to classification problems is interesting from both a theoretical and a practical perspective. Specifically, we found Stack Filter design by classification loss functions exhibited many desirable properties in terms of controlling bias and variance, as well as efficient optimization procedures based on Linear Programming. We call this new model class, and its associated optimization method, Ordered Hypothesis Machines. In this paper we will summarize this recent work and also present new results on quadratic and exponential loss functions using both synthetic and real-world data.

## 7870-05, Session 1

### Signal filtering of daily cloud types trends as derived from satellites images

J. R. Dim, H. Murakami, Japan Aerospace Exploration Agency (Japan)

The relationship between the intensity function of neighboring pixels of daily cloud satellite thermal images is used to extract the horizontal gradient. The images used are derived from the National Oceanic and Atmospheric Administration/Advanced Very-High-Resolution Radiometer (NOAA-AVHRR) satellite. The horizontal gradient is locally obtained for each 3\*3-pixel area. Connections between cloud faces as expressed by the magnitude of this gradient allow for the

distinction of various textural features whose interpretation leads to the discrimination of cloud types. The long-term variation of these cloud types amounts is then evaluated for consistency and reliability. The cloud type amount signal is derived from multiple satellite series which are known to have experienced orbital drift. The effect of this drift on the cloud amount signal is removed through a filtering process using the empirical mode decomposition (EMD). The EMD component, associated with the drift is filtered. The results obtained show an improvement of the signal by a substantial percentage according to the geographical area.

## 7870-06, Session 2

### **Analysing wear in carpets by detecting varying local binary patterns**

S. A. Orjuela Vargas, E. Vansteenkiste, F. Rooms, S. De Meulemeester, R. De Keyser, W. R. Philips, Univ. Gent (Belgium)

In this approach we propose a novel method for grouping LBP patterns based in detecting those that change along the transitional degrees of wear. The method consists in evaluating for each pattern the linear rank correlation related to the wear labels. For this, the same bins corresponding to transitional wear labels are sequentially placed and the Spearman rank correlation for each bin is computed.

Afterwards, a minimal amount of LBP varying patterns is established and a corresponding Spearman rank correlation threshold is automatically computed. LBP patterns that are less than the threshold referred as non varying patterns are grouped together. Thus, when computing the symmetric Kullback-Leibler it is assured to retain the maximum information corresponding to changes in appearance for a given threshold. In this approach we propose a novel method for grouping LBP patterns based in detecting those that change along the transitional degrees of wear. The method consists in evaluating for each pattern the linear rank correlation related to the wear labels. For this, the same bins corresponding to the 8 wear labels are sequentially placed and the Spearman rank correlation for each bin is computed.

## 7870-07, Session 2

### **Line and streak detection on polished and textured surfaces using line integrals**

M. S. Erkilinc, M. Jaber, E. Saber, Rochester Institute of Technology (United States)

In this paper, a framework for detecting lines in a given polished or textured substrate is proposed. Modules for image capture, rectification, enhancement, and line detection are included. If the surface being examined is specular (mirror-like), the image capture will be restricted, that is, the camera has to be fixed off-axis in the zenith direction. A module for image rectification and projection is included to overcome this limitation in order to yield an orthographic image. Moreover, a module for image enhancement that includes high-boost and bilateral filters has been employed to improve the edge sharpness and decrease the spatial noise in the image. Finally, a line-integral technique has been applied to find the confidence vectors that represent the spatial positions of the lines of interest. The Full-Width Half-Max approximation is applied to determine the corresponding lines in a target image. Experimental results show that our technique has an effective performance on synthetic and real images. Assessment of print quality is the main application of the proposed algorithm; however, it can be used to detect lines/ streak in prints, on substrate or any type of media where lines are visible.

## 7870-08, Session 2

### **Detecting photographic and computer-generated composites**

V. Conotter, M. Broilo, L. Cordin, Univ. degli Studi di Trento (Italy)

Nowadays, sophisticated computer graphics editors lead to a significant increase in the photorealism of images. Thus, computer generated (CG) images result to be convincing and hard to be distinguished from real ones at a first glance. Here, we propose an image forensics technique able to automatically detect local forgeries, i.e., objects generated via computer graphics software inserted in natural images, and vice versa. We develop a novel hybrid classifier based on wavelet based features and sophisticated pattern noise statistics. Experimental results show the effectiveness of the proposed approach.

## 7870-09, Session 2

### **Spatio-temporal analysis and forward modelling of solar polar plumes in white light**

A. Llebaria, O. Morillot, Observatoire Astronomique de Marseille-Provence (France)

The analysis of the data provided by LASCO-C2 coronagraph onboard the SOHO spatial observatory revealed the fractal characteristics of many outstanding structures of the solar corona, which is the tiny but extended envelope of plasma wrapping the Sun. A multiscale analysis of recent image sequences has brought a clearer view of the evolution and the local structure of these features which results from a two steps projection process of the 2D electronic distribution over the Sun polar caps. To get an insight in the volume density distribution over these caps and their evolution within time, we used the forward modelling approach based on the present knowledge about the plasma distribution, the physical process of diffusion and the projection geometry on the field of view. The analysis provides us with the multifractal characterization of the observed phenomena. In the forward modelling process the goal is to reconstruct the time sequence of 2D electronic distributions slowly evolving over the Sun polar caps. We used different methodologies: the inverse Fourier transform of 2D+1D (surface and time) frequency modelling, the evolving multiscale synthesis with Gaussian wavelets and the concealed Markov approach. Lately a procedure derivative of the Voss generation schema of fBm fractals has been successfully developed. These different methods are compared and their relative advantages and drawbacks discussed as well as the tools used to compare synthetic images to observed ones.

## 7870-10, Session 2

### **Imaging using synchrotron radiation for forensic science**

F. M. Cervelli, S. Carrato, Univ. degli Studi di Trieste (Italy); A. Mattei, Rep. Investigazioni Scientifiche (Italy); M. Jerian, Amped SRL - P.I. (Italy); L. Benevoli, L. Mancini, F. Zanini, L. Vaccari, A. Perucchi, G. Aquilanti, Sincrotrone Trieste S.C.p.A. (Italy)

Forensic Science has experienced an increasing interest in all research activities: all paths that allow the investigators to obtain more information about the crime scene dynamics and about the culprit are tracked to solve serious crimes like homicides and major crimes related to national security like terrorists' attack.

The aim of this research is to adopt a multi-technique approach, based on conventional and Synchrotron Radiation (SR) techniques, to study latent fingerprints (i.e. fingerprints not visible to the human eye) from the morphological and chemical point of view, offering to forensic science a comprehensive tool to be exploited for particularly complex criminal cases.

Here we address fingerprint analysis, performing a study on latent fingerprint visualization with a SR source.

Several human fingerprints were deposited on two substrates of different nature, ie. undoped silicon wafers and poly-ethylene-terephthalate (PET). We adopted two different deposition modalities, a set of clean fingerprints and a set of fingerprints contaminated by a different mixture of gunshot residue, in order to test the analysis and

visualization system in two common scenarios.

We characterized the morphology as well as the chemical composition of the fingerprints and of their contaminants by Fourier transform infrared (FTIR) spectroscopy, X-ray fluorescence and tomography, X-ray diffraction and scattering, and X-ray absorption fine structure analysis. These techniques were tested as alternative imaging techniques to be used in order to both preserve the collected items and to allow fingerprint analysis in those cases where all the other classical techniques fail.

## 7870-11, Session 3

### **PSO-based methods for medical image registration and change assessment of pigmented skin**

S. T. Kacenjari, Lockheed Martin Corp. (United States); M. Zook, Fox Chase Cancer Ctr. (United States); M. Balint, Lockheed Martin Corp. (United States)

In the field of dermatology, the classification of skin lesions is essential to the detection of melanoma, a malignant tumor of melanocytes. The observation of changes in the color, shape, and size of moles over time is especially important and may lead to the early discovery of malignant forms of skin cancer. However, this endeavor is complicated by the fact that the average patient has dozens of moles, all of which require a thorough examination over time. In this paper we propose a process for tracking changes in skin lesions by conducting periodic, regional, body scans of a patient and leveraging an algorithm to eliminate noise induced by differing camera angles, lighting, and general changes in skin tone. This algorithm consists of (1) a coarse alignment of time-sequenced imagery, (2) refined alignment of local skin topographies through the utilization of a Particle Swarm Optimizer (PSO), and (3) the assessment of local changes in lesion pigmentation between time-sequenced imagery. Once optimized, the differences in imagery will equip dermatologists with a better tool for detecting potential cases of melanoma.

## 7870-12, Session 3

### **Image-based segmentation for characterization and quantitative analysis of the spinal cord injuries by using diffusion patterns**

M. Hannula, A. Olubamiji, Tampere Univ. of Technology (Finland); I. Kunttu, Nokia Research Ctr. (Finland); P. Dastidar, Tampere Univ. Hospital (Finland); J. Hyttinen, Tampere Univ. of Technology (Finland)

Central nervous system injuries such as brain traumas and spinal cord injuries as well as neurodegenerative diseases are among the most common causes of death or serious disability in industrialized countries. Image based diagnosis and analysis methods for this field are escalating and developing. Modern clinical imaging gives data to develop methods for 3-D modelling and designing the cell transplantation therapy for humans. Various new magnetic resonance (MR) imaging sequences, such as diffusion imaging, provide us information of the damaged brain structure and the neuronal connections. These can be analyzed to form 3D models of the geometry and further including functional information of the neurons of the specific brain area to develop functional models. Modeling offers a tool which can be used for the modeling of brain trauma from MR-images of the patients and thus information to tailor the properties of the transplanted cells.

It is known that the water molecules of the white matter of the brain and spinal cord behaves in anisotropic nature forming an arranged pattern which can be analyzed by using Magnetic resonance diffusion tensor imaging (DTI). The DTI is a novel technology that can be used to study the diffusion pattern of the white matter fibers of the spinal cord. In this paper, we present a framework for the image based analysis of

the diffusion of water molecules in soft tissues. We are concentrating on the visualization of the water molecules in a perceptive meaningful manner by fiber tracking, characterization and quantitative analysis of the diffusion pattern. We analyze how the rate of diffusion changes in the cases when there is a spinal cord injury compared to the cases in which the spinal cord is normal. After this, we use a semi-automatic segmentation method to detect the actual location of the injury. The segmentation method is based on the changes in diffusion patterns.

## 7870-13, Session 3

### **Descreening using segmentation-based adaptive filtering**

M. N. Ahmed, A. H. Eid, Lexmark International, Inc. (United States)

In this paper, we present a new segmentation-based descreening technique. Scanned images are segmented into text, images and halftone classes using a multiresolution classification of edge features. The segmentation results guide a nonlinear, adaptive filter to favor sharpening or blurring of image pixels belonging to different classes.

Our experimental results show the ability of the non-linear, segmentation driven filter of successfully descreening halftone areas while sharpening small size text contents.

## 7870-14, Session 3

### **Novel parametric priors for the distribution of multivariate linear prediction errors**

I. Qazi, O. Alata, Univ. de Poitiers (France); J. Burie, Univ. de La Rochelle (France); A. Moussa, Abdelmalek Essaadi Univ. (Morocco); C. Fernandez-Maloigne, Univ. de Poitiers (France)

In this paper we present novel a priori parametric models to approximate the distribution of the two dimensional multichannel linear prediction errors. These parametric approximations and subsequently the discussed models, are used to improve the performance of the color texture segmentation algorithms. Two dimensional causal real (in RGB color space) and complex (in IHLS and  $L^*a^*b^*$  color spaces) multichannel linear prediction models are used to characterize the spatial structures in color images. Classically, the distribution of the multivariate linear prediction errors of these texture models are approximated with a multivariate Gaussian probability distribution. We use Wishart distribution and multivariate Gaussian mixture models to approximate the distribution of these errors. A novel color texture segmentation framework based on these models and Potts model for the spatial regularization of initial class label fields is presented. The framework also takes into account the region size characteristics of the labeled regions during the spatial regularization process. This paper also discusses the performance of this framework in the used color spaces i.e. RGB, IHLS and  $L^*a^*b^*$ . Experimental results show a better performance by the proposed method in terms of percentage segmentation error for the used color textures, as compared to the classical approach, in all three color spaces. The  $L^*a^*b^*$  color space show more stable results than the RGB and IHLS color spaces.

## 7870-16, Session 4

### **Secure annotation for medical images based on reversible watermarking in the Integer Fibonacci Haar transform**

F. Battisti, M. Carli, A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this contribution, a possible solution to secure medical image annotation is presented. The proposed framework is based on the joint use of a key-dependent wavelet transform, of a secure cryptographic scheme, and of a reversible watermarking scheme. The system allows: i) the insertion of the patient data into the encrypted image without

requiring the knowledge of the original image, ii) the encryption of annotated images without causing loss in the embedded information, and iii) due to the complete reversibility of the process, it allows recovering the exact original image once the mark is removed.

## 7870-17, Session 4

### Multi-seam carving via seamlets

D. D. Conger, Michigan State Univ. (United States); M. Kumar, Eastman Kodak Co. (United States); H. Radha, Michigan State Univ. (United States)

Seam carving [Avidan and Shamir 2007] is a powerful retargeting algorithm for mapping images to arbitrary sizes with arbitrary aspect ratios. Meanwhile, the seamlet transform [Conger et al. 2010] has been recently introduced as an efficient representation for seam-carving-based retargeting over heterogeneous multimedia devices with a broad range of display sizes. The original seamlet transform was developed using Haar filters, and hence, it enabled traditional seam carving by removing a single seam at a time in a recursive manner until the desired image size is reached. In this paper, we develop a more efficient approach for seam carving by enabling multi-seam carving, where at each step of the retargeting algorithm multiple seams are carved simultaneously. We achieve multi-seam carving by (a) extending the seamlet transform using more general wavelets than the Haar wavelets, and (b) employing local circular convolution in the vicinity of the selected seams. We show that popular filterbanks, such as the ones that are based on Daubechies wavelets, can achieve efficient multi-seam carving with equivalent visual quality when compared to single-seam carving using the Haar transform. Furthermore, with multi-seam carving, the number of iterations needed to achieve a given target size can be reduced significantly.

## 7870-18, Session 4

### A new DCT-based algorithm for numerical reconstruction of electronically recorded holograms

L. Bilevich, L. Yaroslavsky, Tel Aviv Univ. (Israel)

A new universal low computational complexity algorithm for numerical reconstruction of holograms recorded in near diffraction zone is presented. The algorithm implements digital convolution in DCT domain, which makes it virtually insensitive to boundary effects. It can be used for reconstruction of holograms for arbitrary ratios of hologram size to the object-to-hologram distance and wavelength to camera pitch and allows image reconstruction in arbitrary scale.

## 7870-20, Session 5

### User discrimination in automotive systems

A. Makrushin, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany); C. Vielhauer, Fachhochschule Brandenburg (Germany) and Otto-von-Guericke-Univ. Magdeburg (Germany); M. Leich, Otto-von-Guericke-Univ. Magdeburg (Germany)

The recently developed dual-view touch screens, which are announced to be installed in cars in a near future, give rise to completely new challenges in human-machine interaction. The automotive system should be able to identify if the driver or the passenger is currently interacting with the touch screen to provide a correct response to the touch. The optical devices, due to availability, acceptance by the users and multifunctional usage, approved to be the most appropriate sensing technology for driver/passenger discrimination. In this work the prototypic optical user discrimination system is implemented in the car simulator and evaluated in the laboratory environment with entirely controlled illumination. Three tests were done for this research. One of them examined if the near-infrared illumination should be switched on around the clock, the second one if there is a difference in

discrimination performance between day, twilight and night conditions, and the third one examined how the intensive directional lighting influences the performance of the implemented user discrimination algorithm. Despite the high error rates, the evaluation results show that very simple computer vision algorithms are able to solve complicated user discrimination task. The average error rate of 10.42% (daytime with near-infrared illumination) is a very promising result for optical systems.

## 7870-21, Session 5

### Study of radar system imaging with distributed architecture

L. Lei, J. Jiang, Ctr. for Space Science and Applied Research (China)

The concept of distributed radar has been proposed for several years for its significant advantages while it also has some problems since its geometry structure. Till now, researches have developed some algorithms, but there is still much work to do.

The paper is mainly on the theory and methods how to resolve the problems. With some new views and thoughts, it proposes new imaging theory and methods for distributed radars and can produce high resolution images with three-dimension, which are being developed with the idea that Spatio-temporal information can be combined in processing. The bases of the idea are electromagnetic wave propagation theory and radar resolution theory. The paper provides uniform echo model and ambiguity function and then gives uniform imaging process, which can apply to general distributed systems and targets. One important point in the process is the representation of different view angle in phase and then in frequency difference. The second important point is there are few approximations in the calculation, and the third is the use of interferometric theory in distinguishing the ambiguity targets.

At last, simulation work is presented for distributed radar imaging. And results are given to test the theory and method of the paper.

## 7870-22, Session 6

### Wiener crosses borders: interpolation based on second order models

A. Guevara, R. Mester, Johann Wolfgang Goethe-Univ. Frankfurt am Main (Germany)

Interpolation of signals (arbitrary dimension, here: 2D images) with missing data points is addressed from a statistical point of view.

We present a general framework for which a Wiener-style MMSE estimator can be seamlessly adapted to deal with problems such as image interpolation (inpainting), reconstruction from sparse samples, and image extrapolation.

The proposed method gives a precise answer on a) how arbitrary can linear filters can be applied to initially incomplete signals and b) shows the definite way to extend images beyond their borders such that no size reduction occurs if a linear filter/operator is to be applied to the image.

## 7870-23, Session 6

### Image interpolation based on a multi-resolution directional map

E. Van Reeth, STMicroelectronics (France); P. Bertolino, Gipsalab (France); M. M. Nicolas, STMicroelectronics (France)

This paper describes an interpolation method that takes into account the edge orientation in order to avoid typical interpolation artifacts (jagging, staircase effects...). It is first based on an edge orientation estimation, performed in the wavelet domain. The estimation uses the

multi-resolution features of wavelets to give an accurate and non-biased description of the frequency location of the edges, as well as their orientation. The interpolation is then performed, using the local information given by this directional map, to improve a reference interpolation (cubic-spline for instance). The improvement is carried out by filtering the edges with a gaussian kernel along their direction in order to smooth the contour in the direction parallel to the edge, which avoids disturbing variations across them (jagging and staircase effects). This technique also keeps the sharpness of the transition in the direction perpendicular to the contour to avoid blur.

Results will be presented on both synthetic and real images, showing the visual impact of the presented method on the quality of interpolated images. Comparisons will be made with the usual cubic-spline interpolation, and with other edge directed interpolation techniques to discuss the compromises that have been made in our method compared to others. Full references will also be given on state of the art orientation estimation and directed interpolation methods.

## 7870-24, Session 6

### Images reconstruction using modified exemplar based method

V. V. Voronin, South-Russian State Univ. of Economics and Service (Russian Federation) and Tampere Univ. of Technology (Finland); V. I. Marchuk, South-Russian State Univ. of Economics and Service (Russian Federation); K. O. Egiazarian, Tampere Univ. of Technology (Finland)

This paper describes a new image reconstruction method. The proposed approach uses modified exemplar based technique. Proposed modification allows to choose sub-optimally image-adaptive form and size of the block in order to find similar patches, number of which is further increased by rotation of these blocks. We show that the efficiency of image reconstruction depends on the choice of block size for the exemplar based method. Proposed adaptivity allows to obtain a smaller reconstruction error than that of the traditional method as well as other state-of-the-art image inpainting methods. We demonstrate the performance of a new approach via several examples, showing the effectiveness of our algorithm in removal of small and large objects on the test images.

## 7870-25, Session 7

### A graph non-tree representation of the topology of a gray scale image

P. Saveliev, Marshall Univ. (United States)

The paper provides a method of graph representation of gray scale images. For binary images, it is generally recognized that not only connected components must be captured, but also the holes. For gray scale images, there are two kinds of "connected components" - dark regions surrounded by lighter areas and light regions surrounded by darker areas. These regions are the lower and upper level sets of the gray level function, respectively. The proposed method represents the hierarchy of these sets, and the topology of the image, by means of a graph. This graph contains the well-known inclusion trees, but it is not a tree in general. Two standard topological tools are used. The first tool is cell decomposition: the image is represented as a combination of pixels as well as edges and vertices. The second tool is cycles: both the connected components and the holes are captured by circular sequences of edges.

## 7870-26, Session 7

### Colour processing in Runge space

A. Restrepo, Univ. de Los Andes (Colombia)

We do colour image processing in a space that is as intuitive as the most common spaces of the type hue-saturation-luminance yet it

avoids the caveats resulting from a normalization of the saturation by the luminance. We present applications in image correction in cases of high dynamic range images and faded photographs; also, we present applications that enhance the appeal of natural images. The space is called Runge space; it is spherical and the colour attributes there are hue, colourfulness and lightness; they are readily derived from RGB components.

## 7870-27, Session 7

### Robust image registration for multiple exposure high dynamic range image synthesis

S. Yao, Institute for Infocomm Research (Singapore)

Image registration is an important preprocessing technique in high dynamic range (HDR) image synthesis. This paper proposed a robust image registration method for aligning a group of low dynamic range images (LDR) that are captured with different exposure times. Illumination change and photometric distortion between two images would result in inaccurate registration. We propose to transform intensity image data into phase congruency to eliminate the effect of the changes in image brightness and use phase cross correlation in the Fourier transform domain to perform image registration. Considering the presence of non-overlapped regions due to photometric distortion, evolutionary programming is applied to search for the accurate translation parameters so that the accuracy of registration is able to be achieved at a hundredth of a pixel level. The proposed algorithm works well for under and over-exposed image registration. It has been applied to align LDR images for synthesizing high quality HDR images.

## 7870-28, Poster Session

### Efficiency analysis of DCT-based filters for color image database

V. V. Lukin, D. V. Fevralev, S. K. Abramov, N. N. Ponomarenko, National Aerospace Univ. (Ukraine); J. T. Astola, K. O. Egiazarian, Tampere Univ. of Technology (Finland)

Efficiency of DCT based image filtering is analyzed using a color image database TID2008 that contains images corrupted by i.i.d. and spatially correlated noise with four values of noise variance. It is shown that improvement of PSNR due to filtering is very close for R, G, and B components of color images and this improvement depends upon image content. Improvement of PSNR reaches 7 dB for quite simple images (that contain large homogeneous regions) and it is only 1.5 dB for highly textural images if initial PSNR=30 dB. The visual quality metric PSNR-HVS-M is studied as well. For spatially correlated noise, results of analysis clearly show that thresholding should be frequency dependent. This allows increasing PSNR and PSNR-HVS-M by about 2...3 dB compared to the case of fixed (frequency independent) threshold.

## 7870-29, Poster Session

### Color image lossy compression based on blind evaluation and prediction of noise characteristics

V. V. Lukin, N. N. Ponomarenko, National Aerospace Univ. (Ukraine); K. O. Egiazarian, Tampere Univ. of Technology (Finland); L. Lepisto, Nokia Research Ctr. (Finland)

Most of images formed by digital cameras are compressed before storage and/or transferring where JPEG lossy compression is a standard tool applied. Usually there are three or standards modes of lossy compression. We propose two approaches to automatic selection of lossy compression parameters that are able to take into account blind estimates or prediction of noise characteristics and



blur parameters as well as content of a given image and to carry out adaptive JPEG compression without introducing visually noticeable distortions. The designed approaches allow increasing compression ratio by, on the average, from 2 to 2.6 times.

## 7870-30, Poster Session

### **Unsupervised automated panorama creation for realistic surveillance scenes through weighted mutual information registration**

T. P. Keane, E. Saber, H. E. Rhody, A. E. Savakis, Rochester Institute of Technology (United States); J. Raj, Lenel Systems International Inc. (United States)

Automated panorama creation usually requires camera calibration or extensive knowledge of camera locations and relations to each other. Registration problems are often solved by these same camera parameters or the result of complex point matching schemes. This paper presents a novel automated panorama creation algorithm by using an affine transformation search based on maximized mutual information (MMI). MMI techniques are often limited to airborne and satellite imagery or medical images, but we can show that a simple MMI algorithm very well approximates realistic scenes of varying depth distortion. This study was performed on stationary color surveillance video cameras and proves extremely worthwhile in any system with limited or no a priori camera-to-camera parameters. This algorithm is quite robust on a very large range of strict- to nearly-affine related scenes, and provides a great approximation for the overlap regions in scenes related by a projective homography. Surprisingly significant practical considerations ultimately outweighed theoretical derivations in the development of this robust and versatile algorithm.

## 7870-31, Poster Session

### **Ellipse detection using an improved randomized Hough transformation**

Z. Teng, J. Kim, D. Kang, Pusan National Univ. (Korea, Republic of)

This paper proposes an ellipse detection algorithm based on the analytical solution to the parameters of ellipse in images. In the first instance, edge detection is processed, from which line segments are extracted. Then the method of finding the center coordinates of the ellipse is described based on the property of ellipse by using three points voting at a sense of Randomized Hough Transformation (RHT). Finally, an analytical solution of the other three parameters of the ellipse (semi-major axis length, semi-minor axis length and the angle between the X-axis and the major axis of the ellipse) are given via coordinate transformation. Based on this solution, we propose the separated parameter voting scheme for ellipse center and the other three parameters instead of 5 parameters voting scheme of RHT. The experiments show that the proposed algorithm performs well in various images.

## 7870-32, Poster Session

### **Detection of motion blur direction based on maxima locations for blind deconvolution**

R. M. Chong, T. Tanaka, Tokyo Univ. of Agriculture and Technology (Japan)

The blurs in images closely resemble an ideal point spread function (PSF) model. This similarity can be exploited in the deconvolution process by learning a model that best fits the estimated PSF. In order to achieve this, a model is selected from the provided training set and then integrated into the reconstruction cost function. In this paper, we propose to eliminate the need for a training set and instead use a reference PSF (RPSF) in its place. This eliminates the need for

specifying a training set as well as the dependence on estimated quantities. Furthermore, it is only dependent on the given degraded image assuming that it is uniformly blurred. We tested our method with motion blurs in different directions since it is one of the most commonly encountered problems when using consumer cameras. Using the blur support as a priori knowledge, the results show that the method is capable of determining the motion direction even in the presence of noise. The reconstruction of the image is achieved by using a modified cost function that also accounts for the contour of the estimated PSF. Results show that higher image quality and lower PSF estimation error can be obtained.

## 7870-33, Poster Session

### **EM algorithm-based hyperparameters estimator for Bayesian image denoising using BKF prior**

L. Boubchir, B. Durning, E. Petit, Univ. Paris 12 - Val de Marne (France)

This paper is devoted to a novel hyperparameter estimator for Bayesian denoising of images using the Bessel K Form prior which we recently developed [1,2]. More precisely, this approach is based on the EM algorithm. The simulation results show that this estimator offers good performances and is slightly better compared to the cumulant-based estimator suggested in [1,2]. A comparative study is carried to show the effectiveness of our Bayesian denoiser based on EM algorithm compared to other denoisers developed in both classical and Bayesian contexts. Our study has been effected on natural and medical images for gaussian and poisson noise removal.

#### REFERENCES

- [1] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities", IEEE Transactions on Image Processing, vol.14, no.2, pp.231-240, 2005.
- [2] L. Boubchir and J. M. Fadili, "Bayesian denoising based on the MAP estimation in wavelet-domain using Bessel K form prior", IEEE International Conference on Image Processing, vol.1, pp.113-116, 2005.

## 7870-34, Poster Session

### **Semantic analysis of facial gestures from video using a Bayesian framework**

G. Vashi, R. L. Canosa, Rochester Institute of Technology (United States)

The continuous growth of video technology has resulted in increased research into the semantic analysis of video. The multimodality property of the video has made this task very complex. The objective of this research is to research, implement and examine the underlying methods and concepts of semantic analysis of videos and also to show how to improve upon state of the art in gesture recognition by using semantic knowledge. The main domain of analysis is facial gesture recognition from video, including both visual and vocal aspects of facial gestures. A Bayesian network classification algorithm has been used to identify and understand facial expressions in video. The Bayesian network is an attractive choice because it provides a probabilistic environment and gives information about uncertainty from knowledge about the domain. The goal of this research is to determine if an expression on a person's face is happy, sad, angry, fearful or disgusted. This information will enhance the semantic understanding and interpretation of video data. Currently, it has not been established that two modalities are necessary for accurate interpretation of facial expressions in video. Therefore, this research is a contribution to the current knowledge by testing the hypothesis that combining the two modalities of vision and speech yields better classification results than either used alone.

## 7870-35, Poster Session

### **Color image enhancement algorithm based on logarithmic transform coefficient histogram shifting**

J. Xia, K. A. Panetta, Tufts Univ. (United States); S. S. Agaian, The Univ. of Texas at San Antonio (United States)

The goal of image enhancement techniques is to improve the characteristic or visual quality of an image for specific criteria. They can be classified as spatial domain enhancement and transform domain enhancement. Spatial domain techniques deal with the raw image data, altering the intensity values based on a specific algorithm for a set of criteria. Transform domain enhancement techniques involve transforming the image intensity data into a specific domain by using such methods as the Discrete Cosine, Fourier, and Wavelet transforms. These transforms are used to alter the frequency content of an image to improve desired traits, such as high frequency content. Combining spatial and transform techniques can produce powerful results which can compensate for weaknesses in individual algorithms. A color image enhancement algorithm based on logarithmic transform histogram shifting is proposed in this paper. Experimental results show that the proposed algorithm provides better contrast image with nearly no halo artifacts and good color consistency. Also compared with the retinex, which is a conventional color image enhancement algorithm, it is simple but more effective. This algorithm separates the color data into chromaticity and brightness and applies Log enhancement on brightness image only. Then all the Red, Green and Blue components are separately enhanced by logarithmic transform histogram shifting method, which is based on altering the transform coefficient histograms through shifting and mapping. Optimal parameter selection based on image quality measurement is also discussed in this algorithm.

## 7870-36, Poster Session

### **Neighbourhood-consensus message passing and its potentials in image processing applications**

T. Ružic, A. Pižurica, W. R. Philips, Univ. Gent (Belgium)

In this paper, a novel algorithm for inference in Markov Random Fields (MRFs) is presented. Its goal is to find approximate maximum a posteriori estimates in a simple manner by combining neighbourhood influence of iterated conditional modes (ICM) and message passing of loopy belief propagation (LBP). We call the proposed method neighbourhood-consensus message passing because a single joint message is sent from the specified neighbourhood to the central node. The message, as a function of beliefs, represents the agreement of all nodes within the neighbourhood regarding the labels of the central node. This way we are able to overcome the disadvantages of reference algorithms, ICM and LBP. On one hand, more information is propagated in comparison with ICM, while on the other hand, the huge amount of pairwise interactions is avoided in comparison with LBP by working with neighbourhoods. The idea is related to the previously developed iterated conditional expectations algorithm. Here we revisit it and redefine it in a message passing framework in a more general form. The results on three different benchmarks demonstrate that the proposed technique can perform well both for binary and multi-label MRFs without any limitations on the model definition. Furthermore, it manifests improved performance over related techniques either in terms of quality and/or speed.

## 7870-37, Poster Session

### **Alternative method for Hamilton-Jacobi PDEs in image processing**

C. Vachier-Mammar, A. Lagoutte, H. Salat, Ecole Normale Supérieure de Cachan (France)

Multiscale signal analysis has been used since the early 1990s as a powerful tool for image processing, notably in the linear case. However, nonlinear PDEs and associated nonlinear operators have advantages over linear operators, notably preserving important features such as edges in images. In this paper, we focus on nonlinear Hamilton-Jacobi PDEs defined with adaptive speeds or, alternatively, on adaptive morphological filters also called semi-flat morphological operators. Semi-flat morphology were introduced by H. Heijmans and studied only in the case where the speed (or equivalently the filtering parameter) is a decreasing function of the luminance. It is proposed to extend the definition suggested by H. Heijmans in the case of non decreasing speeds. We also prove that a central property for defining morphological filters, that is the adjunction property, is preserved while dealing with our extended definitions. Finally experimental applications are presented on actual images, including connection of thin lines by semi-flat dilations and image filtering by semi-flat openings.

## 7870-38, Poster Session

### **A novel dimming algorithm using local boosting algorithm for LED backlight system in LCD TVs**

J. Lee, LED-IT Fusion Technology Research Ctr. (Korea, Republic of)

In this paper, a novel dimming algorithm using the local boosting algorithm for LED backlight system in LCD TVs is proposed. The proposed dimming algorithm consists of two new algorithms: image classification and the local boost method. The proposed algorithm has higher contrast ratio and lower power consumption than the conventional methods

## 7870-39, Poster Session

### **Spatially adaptive alpha-rooting in BM3D sharpening**

M. Mäkitalo, A. Foi, Tampere Univ. of Technology (Finland)

The block-matching and 3-D filtering (BM3D) algorithm is currently one of the most powerful and effective image denoising procedures. It exploits a specific nonlocal image modeling through grouping and collaborative filtering. Grouping finds mutually similar 2-D image blocks and stacks them together in 3-D arrays. Collaborative filtering produces individual estimates of all grouped blocks by filtering them jointly, through transform-domain shrinkage of the 3-D arrays (groups).

BM3D can be combined with transform-domain alpha-rooting in order to simultaneously sharpen and denoise the image. Specifically, the thresholded 3-D transform-domain coefficients are modified by taking the alpha-root of their magnitude for some  $\alpha > 1$ , thus amplifying the differences both within and between the grouped blocks. While one can use a constant (global) alpha throughout the entire image, further performance may be potentially achieved by allowing different degrees of sharpening in different parts of the image, based on some content-dependent information.

We propose to adjust the value of alpha used for sharpening a group through weighted estimates of the edge and texture strengths of the average block in the group. This is shown to be a viable approach for image sharpening, and in particular it provides an improvement over its global non-adaptive alpha-rooting counterpart.

## 7870-40, Poster Session

### Extracting global salient open curves from cluttered backgrounds via Markov random fields

N. Durak, O. Nasraoui, Univ. of Louisville (United States)

Long portions of an ellipse can be more easily distinguished by the human eye than by a computer in cluttered images. We propose a method that automates the process of extracting principal loop shapes from cluttered images. First, curve segments are algorithmically traced and properties of curve segments are computed. Labels are assigned to each curve segment. Starting from the most salient curve segments, their neighborhood was searched for possible pairwise grouping. Candidate groups are checked with respect to ellipse fit error, smoothness, corner points, and angle similarity. We update the labels of the curve segments after each iteration till we obtain optimum results. We tested our system on synthetic and real images to show the effectiveness.

## 7870-41, Poster Session

### Joint distributed source-channel coding for 3D videos

V. Palma, M. Cancellaro, A. Neri, Univ. degli Studi di Roma Tre (Italy)

This paper presents a distributed joint source-channel 3D video coding system.

Our aim is the design of an efficient coding scheme for stereoscopic video communication over noisy channels that preserves the perceived visual quality while guaranteeing a low computational complexity.

The drawback in using stereo sequences is the increased amount of data to be transmitted. Several methods are being used in the literature for encoding stereoscopic video. A significantly different approach respect to traditional video coding has been represented by Distributed Video Coding (DVC), which introduces a flexible architecture with the design of low complex video encoders.

DVC states that it is theoretically possible to separately encode and joint decode two or more statistically dependent sources at the same rate obtained when the same sources are joint encoded and decoded. This approach considerably reduces the overall amount of transmission necessary from the cameras to the central decoder and simplifies the complexity of the video encoder by shifting all the complex video processing tasks to the decoder. Power/processing limited systems such as wireless camera sensors that have to compress and send video to a fixed base station in a power-efficient way can take advantage of this property.

Theoretically source and channel encoding is based on tandem of two separate encoding systems. In this contribution, we present the design of DVC-based Joint Source-Channel 3D video Coding scheme for noisy channel. We adopt a single source-channel encoder for both compression and protection resulting in a distributed 3D video coding scheme.

In this contribution, the mathematical framework will be fully detailed and tradeoff among redundancy and perceived quality and quality of experience will be analyzed with the aid of numerical experiments

## 7870-42, Poster Session

### Simulating images captured by superposition lens cameras

A. S. Thangarajan, R. Kakarala, Nanyang Technological Univ. (Singapore)

As the demand for reduction in thickness of the lenses in camera rises, the need to look for better solutions becomes a necessity. One such radical approach toward developing a thin lens was obtained

from nature's superposition principle used in the eyes of many insects. But generally the images obtained from these lenses are fuzzy, and require reconstruction algorithms to complete the imaging process. The existing literature does not provide realistic test images for such algorithms, and commercial ray-tracing software required to produce such images is costly. A solution for this problem is presented in this paper. Here a Gabor Super Lens (GSL) which is based on superposition principle is selected and the complete lens structure is simulated and is tested with a test image using the public-domain ray-tracing software POV-ray. The image obtained is as viewed through an actual GSL, and can be used to test algorithms to reconstruct those blurry images. The large computational time in rendering such images requires further optimization, and methods to do so are discussed.

## 7870-43, Poster Session

### Features extraction based on Fisher's information

L. Costantini, P. Sità, M. Carli, A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this paper we present a novel technique for detecting the flat background on images based on Fisher's information. Our goal is to improve the performances of a content based image retrieval (CBIR) system. Many CBIR systems are based on the low level feature extracted, such as texture, colour, and edges, on the whole images. The performances of the CBIR systems can be improved if the low level features are extracted on image areas containing relevant information. For selecting those regions, the local Fisher's information is computed and the regions characterized by a low information level are removed from the image. In the proposed CBIR scheme we first evaluate the local Fisher's information and then we characterize the image by using the low level features. The evaluation of the Fisher's information is based on the Zernike polynomials. To select only the connected peripheral areas of the image a region growing procedure is performed.

Experimental results show that the proposed algorithm improves both the retrieval rate and the performance of the image clustering algorithm.

## 7870-45, Poster Session

### An improved RANSAC algorithm using within-class scatter matrix for fast image stitching

L. Zhang, Z. Liu, J. Jiao, Graduate Univ. of the Chinese Academy of Sciences (China)

In this paper, we proposed an improved RANSAC algorithm using within-class scatter matrix for fast image stitching. The algorithm localizes and describes the features with SIFT (short for scale-invariant feature transform) firstly, ensuring the accuracy. Then the features are matched using Min-cost K-flow algorithm. Next we apply the improved RANSAC (short for random sample consensus) algorithm with the within-class scatter matrix to register the neighboring images. The within-class scatter matrix is used to describe the scattering of the sample to measure the random sample generated by RANSAC algorithm. The improved algorithm we proposed can accelerate RANSAC algorithm effectively while proving the accuracy and robustness. Finally, image blending can be done by multi-band blending.

We compared image stitching algorithm using improved RANSAC algorithm and using original RANSAC algorithm on 20 small image pairs (320\*240) and 8 bigger image pairs (800\*600) on accuracy and speed. The images are selected randomly from ICCV2005 Computer Vision Contest and our image tests. After matching the features with Min-cost K-flow algorithm, both the two algorithm are repeated 40 times for each pairs of images to get statistically representative results. Experiment results demonstrate that our algorithm is more effective. It can get the result of the same quality as original RANSAC algorithm while accelerating speed by about 20 %.

## 7870-46, Poster Session

### Edge-directed image zooming based on radial basis function interpolation

Y. J. Lee, KAIST (Korea, Republic of); J. Yoon, Ewha Womans Univ. (Korea, Republic of)

Image interpolation is a prime technique in many application.

In this study, we propose an edge-directed non-linear interpolation algorithm for image zooming based on moving least squares method.

The basic idea is first to use an initial estimate of pixel information at the resampling position.

Specifically, this initial estimate involves measuring the orientation of the local gradients in the image.

Next, the covariance estimates is used to correct the orientation of the edge direction. Finally, this orientation information is then used to adaptively steer the local kernel function, not accessing edge, which results in improving quality of the interpolated images over conventional linear interpolation.

## 7870-47, Poster Session

### Enhanced bleed through removal for scanned document images

A. Sharma, Hewlett-Packard Labs. India (India); S. Mahaldar, Shell Inc. (India); S. Banerjee, Hewlett-Packard Labs. India (India)

Back-to-front interference is a common problem in documents, printed on translucent pages with insufficient opacity and is referred to as bleed through. The present state-of-art algorithms address bleed through based on entropy, entropic correlation and discriminator analysis. However, a common drawback of such algorithms is their inefficient processing of documents that are either sparse in terms of content or have a very dark background. Our proposed algorithm, based on Otsu's binarization method and pixel level classification addresses these problems. Experiments indicate that our algorithm performs comparable to state-of-the-art for most of the images and better than state-of-the-art for the low contrast images.

## 7870-48, Poster Session

### Classification of texture features in pathological prostate images

A. Almuntashri, S. S. Agaian, The Univ. of Texas at San Antonio (United States)

In this paper, we propose a classification method for prostate pathological images based on texture features. The classification is based on Gleason method for histological grading of malignancy of cancerous tissues. The proposed algorithm has a superior performance in visualizing and classifying fine tissues details for an automatic detection and classification of cancer in prostate biopsy images.

## 7870-49, Poster Session

### Image segmentation refinement by modeling in turning function space

C. F. S. Volotao, Instituto Nacional de Pesquisas Espaciais (Brazil) and Instituto Militar de Engenharia (Brazil) and Diretoria de Serviço Geográfico do Exército (Brazil); R. D. C. Santos, G. J. Erthal, L. V. Dutra, Instituto Nacional de Pesquisas Espaciais (Brazil)

This work proposes a new approach for the use of turning function space to change shapes in accordance with shape descriptions and consistent with spectral information. The main steps are: (1)

segmentation; (2) contour extraction; (3) turning function space transform; (4) classification; (5) shape analysis; and (6) blob enhancement on image space. In the analysis of shape the boundary is modified based on both image and model and constraints are imposed to portions of the turning function. Shape modeling can be done by defining criteria such as linearity, angles and sizes. Results on synthetic examples are presented.

## 7870-50, Poster Session

### Integrating empirical mode decomposition and nonlinear diffusion method for noise reduction in underwater sonar images

S. Bakhtiari, S. S. Agaian, M. Jamshidi, The Univ. of Texas at San Antonio (United States)

Sonar images are susceptible to be affected by nonlinear noise which makes the detection or recognition process more complicated. The traditional noise removal techniques which involve linear stationary noise are not suitable for such images. Empirical Mode Decomposition (EMD) has proved to be a powerful technique for analyzing non-linear and non-stationary signals. This method is fully data driven that decomposes the signal into some oscillatory components called Intrinsic Mode Functions (IMFs) by sifting process. In this paper, a new EMD based approach is proposed to reduce the noise of underwater Side Scan Sonar (SSS) images. Combination of EMD and Nonlinear Diffusion technique has shown to be considerably effective for eliminating this type of noise. The images are de-noised by filtering each IMF components by Nonlinear Diffusion method and recombining the processed IMFs.

## 7870-51, Poster Session

### Extending JPEG-LS for low-complexity scalable video coding

A. Ukhanova, Technical Univ. of Denmark (Denmark); A. Sergeev, St. Petersburg State Univ. of Aerospace Instrumentation (Russian Federation); S. Forchhammer, Technical Univ. of Denmark (Denmark)

JPEG-LS, the well-known international standard for lossless and near-lossless image compression, was originally designed for non-wireless applications. In this paper we propose scalable modification of JPEG-LS and compare it with the leading video coding standards JPEG2000 and H.264/SVC for application to high-rate and low complexity wireless video coding and transmission.

# Conference 7871: Real-Time Image and Video Processing 2011

Monday-Tuesday 24-25 January 2011

Part of Proceedings of SPIE Vol. 7871 Real-Time Image and Video Processing 2011

## 7871-01, Session 1

### Towards real-time image quality assessment

B. Geary, C. Grecos, Univ. of the West of Scotland (United Kingdom)

We introduce a real-time implementation and evaluation of a new, fast, accurate, structurally based image quality metric. Structural approaches to image quality measurement are predicated on the notion that humans perceive image quality as a function of the integrity of local structure in an image after it has been subjected to degradation.

In this paper we outline the salient features of the derivation of the Rotated Gaussian Discrimination Metric (RGDM) and show how analyses of local statistics of distortion type necessitate variation in discrimination function width. Results obtained on the LIVE image database show tight banding of RGDM metric value when plotted against mean opinion score indicating the usefulness of this metric.

We explore a number of strategies for algorithmic speed-up of RGDM including the application of Integral Images for patch based computation optimisation, cost reduction for the evaluation of the discrimination function and general loop unrolling. We also employ fast SIMD intrinsics and explore multi-core data parallel decomposition on a multi-core Intel Processor. Our results show in excess of an order of magnitude speed-up over the un-optimised RGDM algorithm (depending on the number of cores employed, the data set evaluated and the extent of pre-processing) measured in terms of number of processor clock cycles obtained using the Intel V-Tune profiling tool. It is anticipated that this fast Image Quality Assessment (IQA) technique will be employed in bit error rate optimisation experiments in the near future.

## 7871-02, Session 1

### 2000 fps real-time target tracking vision system based on color histogram

I. Ishii, T. Tatebe, Q. Gu, T. Takaki, Hiroshima Univ. (Japan)

In this study, we develop a high-speed color-histogram-based target tracking system that can be applied to 512x512 pixel images at 2000 fps using the hardware implementation of an improved CAM-SHIFT method on a high-speed vision platform. In the improved CAM-SHIFT method, the size, position, and orientation of an object to be tracked can be extracted using only the hardware implementation of hue conversion and the moment feature calculation of 16 binary images quantized by color bins according to the hue histogram based on the additivity in moment feature calculation. By installing our target tracking system on a two-axis active vision platform, we present several mechanical target tracking results for high-speed moving objects: (1) a color pattern rotating at 15 rps and (2) a human hand moving rapidly at 4 Hz in a room. In the experiments, the pan and tilt motors on the active vision platform are controlled through feedback at 2000 fps to correspond to the calculated image centroid by using the extracted moment features for the center of the camera view. These results indicate that our color-histogram-based target tracking system can robustly track high-speed moving objects even under actual complex scenes for a vision system having a frame rate of up to 2000 fps.

## 7871-03, Session 1

### Real-time iris tracking with a smart camera

M. Mehrübeoglu, H. T. Bui, Texas A&M Univ. Corpus Christi (United States); L. McLauchlan, Texas A&M Univ.-Kingsville (United States)

This paper presents a real-time iris detection procedure for gray intensity images. Typical applications for iris detection utilize template and feature based methods. These methods are generally time and memory intensive and not applicable for all practical real-time embedded realizations with limited system resources that require high speed inspection rates. In this article, we propose a method that utilizes a simple algorithm that is time-efficient with high detection and low error rates.

The real-time image acquisition system for this research involves a 17xx series Smart Camera (NI) with LabVIEW Real-Time Module used for automated applications. First, the images are analyzed to determine the region of interest (face) before detecting the eye. Utilizing a convolution-based algorithm on the edge image and using Hough Transform, the iris of the eye is then determined. This edge based method is efficient, since the algorithm is less complex and less computationally expensive than if the full image was to be analyzed. In this approach, the first image is used to compute the location information of the iris. The initial computation in the first frame is the most time consuming aspect of the procedure. The acquired iris location information is then stored in the camera's image buffer, and used to model one specific eye pattern. The location of the iris thus determined is then used as a reference to reduce the search region used to detect the iris in the subsequent image frames with high accuracy and fast, appropriate for real-time implementations.

The iris detection algorithm has been applied at different frame rates. The results demonstrate the speed of this algorithm allows the tracking of the iris when the eyes or the subject is moving in front of the camera at reasonable speeds and with limited occlusions. The results of this project has applications in gaze tracking in automotive (driver warning), medical (patient monitoring; instrument control), computer (consumer behavior testing), or gaming (game control) industries and for surveillance.

## 7871-05, Session 1

### Optimization of image processing algorithms on mobile platforms

M. V. Shirvaikar, P. Poudel, The Univ. of Texas at Tyler (United States)

This paper presents a technique to optimize popular image processing algorithms on mobile platforms such as cell phones and PDAs. The target platform chosen for the development was the OMAP3530 processor which is widely used in embedded media systems. The basic image correlation algorithm is chosen as it finds widespread applications for various template matching tasks such as face-recognition and context-aware computing. As the correlation algorithm is computationally complex, it is necessary to optimize performance to meet real-time deadlines, especially under mobile scenarios. The basic algorithm prototypes conform to OpenCV, a popular computer vision library developed by Intel Corporation. A methodology to take advantage of the asymmetric dual-core processor, which includes an ARM and a DSP core supported by shared memory, is presented with implementation details. DSPLib, a highly optimized library provided by TI is used to perform basic digital signal processing tasks. The Codec-Engine framework provided by Texas Instruments (TI) is used for Inter Processor Communication (IPC) and Remote Procedure Call (RPC) functionality. The performance results presented measure the algorithm speedup obtained due to dual-core implementation. The procedures established can be applied to other algorithms that are part of any general purpose imaging library.

## 7871-06, Session 2

### Scalable software architecture for on-line multicamera video processing

M. Camplani, L. Salgado, Univ. Politécnic de Madrid (Spain)

Multi-camera systems development is a very active research area due to the increasing demand of efficient systems for several application domains. Systems based on smart cameras devices are well suited for real time image processing thanks to dedicated hardware. Moreover, they are widely used in large camera systems. However, they present some drawbacks: dedicated hardware presents lack of flexibility and the design of cooperative tasks is not straightforward.

The proposed architecture guarantees a good trade-off between computational power and flexibility. The system is composed by a network of Processing Units (PUs). Each PU manages several cameras. Data Acquisition and processing tasks are completely decoupled. The incoming images are copied in a shared memory. Each processing task is implemented in a pipeline fashion where each stage is executed by a different thread in order to take advantage of the multi-core architecture of the PU.

We present a system composed by one PU connected with three cameras. The cameras are synchronized with an external triggering system. In the context of real-time tracking we have implemented a background estimation algorithm. System performance has been evaluated under different load conditions such as number of cameras, image size and frame rate.

## 7871-07, Session 2

### Real-time implementation of logo detection on open source BeagleBoard

M. K. George, N. Kehtarnavaz, The Univ. of Texas at Dallas (United States); L. W. Estevez, Texas Instruments Inc. (United States)

This paper presents a follow-up to our previous work on logo detection and tracking algorithm which targets mobile phone users to obtain information or discounts associated with logos. The algorithm involves a hybrid approach by using a combination of SIFT, online color calibration and moment invariants in a video stream operation mode. After SIFT logo detection, online color calibration using k-means clustering in the Cr-Cb color space is performed to extract the prominent logo color. This information is used to track the logo in subsequent frames of the video stream. Moment invariants of all regions are calculated and the region which has the closest match is selected. This paper addresses the real-time porting or implementation of the above steps on BeagleBoard. The BeagleBoard is an open source, low cost, fan-less OMAP device available from Texas Instruments. The OMAP processor has an ARM Cortex GPP, a C64x+ TI DSP and a SGX GPU. The objective here is to leverage these engines toward achieving a real-time throughput of the algorithm. The main focus is to offload certain operations and native OpenCV functions onto the DSP in order to reach a real-time solution.

## 7871-08, Session 2

### Image orientation detection for real-time implementation on embedded devices

V. V. Appia, Georgia Institute of Technology (United States); R. Narasimha, Texas Instruments Inc. (United States)

In this paper we describe a low complexity image orientation detection algorithm which can be implemented in real-time on embedded devices such as low-cost digital cameras, mobile phone cameras and video surveillance cameras. Providing orientation information to tamper detection algorithm in surveillance cameras, color enhancement algorithm and various scene classifiers can help improve their performance. Various image orientation detection algorithms have been

developed in the last few years for image management systems, as a post processing tool. But, these techniques use certain high-level features and object classification to detect the orientation, thus they are not suitable for implementation on a capturing device in real-time. Our algorithm uses low-level features such as texture, lines and source of illumination to detect orientation. We implemented the algorithm on a mobile phone camera device with a 180 Mhz, ARM926 processor. The orientation detection takes ~10 ms for each frame which makes it suitable to use in image capture as well as video mode. It can be used efficiently in parallel with the other processes in the imaging pipeline of the device. On hardware the algorithm achieved an accuracy of ~88% with a false detection rate of 4% on outdoor images.

## 7871-09, Session 2

### Real-time topological image smoothing on shared memory parallel machines

R. Mahmoudi, M. Akil, Ecole Supérieure d'Ingénieurs en Electronique et Electrotechnique (France)

Smoothing filter is the method of choice for image preprocessing and pattern recognition. We present a new concurrent method for smoothing 2D object in binary case. Proposed method provides a parallel computation while preserving the topology by using homotopic transformations. We introduce an adapted parallelization strategy called split, distribute and merge (SDM) strategy which allows efficient parallelization of a large class of topological operators including, mainly, smoothing, skeletonization, and watershed algorithms. To achieve a good speedup, we cared about task scheduling. Distributed work during smoothing process is done by a variable number of threads. Tests on 2D binary image (512\*512), using shared memory parallel machine (SMPM) with 8 CPU cores (2x Xeon E5405 running at frequency of 2 GHz), showed an enhancement of 5.2.

## 7871-10, Session 2

### Multithreaded real-time 3D image processing software architecture and implementation

V. Ramachandra, K. Atanassov, M. Aleksic, S. R. Goma, Qualcomm Inc. (United States)

A real time 3D player was implemented on the GPU using CUDA and OpenGL. The player provides user interactive 3D video playback. Stereo images are first read by the player from a fast drive and then rectified. Further processing of the images determines the optimal convergence point in the 3D scene to reduce eye strain. The rationale for this convergence point selection takes into account scene depth and display geometry. The first step in this processing chain is identifying keypoints by detecting vertical edges within the left image. Regions surrounding reliable keypoints are then located on the right image through the use of block matching. The difference in position between corresponding regions are then used to calculate disparity. The extrema of a disparity histogram gives the scene disparity range. The left and right images are shifted based upon the calculated range.

All the above computations were performed on one CPU thread which calls CUDA functions. Image upsampling and shifting is performed in response to user zoom and pan.

The player also consists of a CPU display thread, which uses OpenGL rendering (quad buffers). This also gathers user input for digital zoom and pan and sends them to the processing thread.

### 7871-11, Session 3

#### **Real-time video streaming using H.264 scalable video coding (SVC) in multihomed mobile networks: a testbed approach**

J. M. Nightingale, Q. Wang, C. Grecos, Univ. of the West of Scotland (United Kingdom)

Users of the next generation wireless paradigm known as multihomed mobile networks expect satisfactory quality of service (QoS) when accessing streamed multimedia content. The recent H.264 Scalable Video Coding (SVC) extension to the Advanced Video Coding standard (AVC), offers the facility to adapt real-time video streams in response to the dynamic conditions of multiple network paths encountered in multihomed wireless mobile networks. Nevertheless, pre-existing streaming algorithms were mainly proposed for AVC delivery over multipath wired networks and were evaluated by software simulation.

This paper introduces a practical hardware-based testbed whereby we implement and evaluate real-time H.264 SVC streaming algorithms in multihomed wireless mobile networks. We propose an optimised streaming algorithm with multi-fold technical contributions. Firstly, we extended the AVC packet prioritisation schemes to reflect the greater granularity of SVC. Secondly, we designed a mechanism for the evaluation of the effects of different streamer 'read ahead window' sizes on real-time performance. Thirdly, we took account of the previously unconsidered path switching and mobile networks tunnelling overheads encountered in real-world deployments. Finally, we implemented a path condition monitoring and reporting scheme to facilitate the intelligent path switching. The proposed system has been experimentally shown to offer a significant improvement in PSNR of the received stream compared with representative existing algorithms.

### 7871-12, Session 3

#### **A new bitstream structure for parallel CAVLC decoding**

Y. Lee, K. Cho, Samsung Electronics Co., Ltd. (Korea, Republic of)

A CAVLC decoder cannot know the exact starting position of the k-th syntax element in a bit-stream until it finishes decoding of the (k-1)-th syntax element. It makes a parallel decoding difficult in hardware implementation. It significantly increase hardware cost to predict the starting position of a syntax element prior to decoding of its previous one. In this paper, we propose a new bitstream structure to concurrently access multiple syntax elements for parallel CAVLC decoding. The method divides a bit-stream into N kinds of segments whose size is M bits and puts syntax elements into the segments, based on a proposed rule. Then, a CAVLC decoder can simultaneously access N segments to read N syntax elements from a single bitstream and decode them in parallel. This technique increases the speed of CAVLC decoding by up to N times. Since the method just rearranges generated bit-stream, it does not affect coding efficiency. Simulation results show that speed-up is 80% with N = 2.

### 7871-13, Session 3

#### **3D video sequence reconstruction algorithms implemented on DSP**

V. I. Ponomaryov, E. Ramos-Diaz, Instituto Politécnico Nacional (Mexico)

Depth map usually serves as important information in several fields: video filtering, robot navigation, video editing, etc. in images and video sequences. Depth map computation has been studied widely; however, to obtain dense depth map information from video sequences is a difficult problem. A lot of algorithms have been proposed to address some of the aforementioned issues in stereo vision, however it is still relatively an open problem. Promising method to visualize 3D

information uses the anaglyph.

In this work, we proposed an algorithm to compute the depth map information employing the real video sequences. Obtained depth map information is then applied in the construction of 3D video sequence by means of anaglyph employment. In order to improve the anaglyphs, the depth map manipulation via P-th law compression is realized, then, the video construction should be done. We present a comparison between depth map results using different Wavelets and other techniques (Differential, Stereo Matching, Warping, etc.). The quantity of Bad Disparities as quantitative criterion in order to select the better depth map is applied. Visual reconstruction results are also compared with classical Photoshop algorithm in order to prove the efficiency of the proposed framework. Additionally, Digital Signal Processor TMS320DM642TM, Matlab 2009aTM in an Intel Core2 Quad ProcessorTM are used to implement the proposed method.

### 7871-14, Session 3

#### **Real-time patch sweeping for high-quality depth estimation in 3D videoconferencing applications**

W. Waizenegger, I. Feldmann, O. Schreer, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

In future 3D videoconferencing systems, depth estimation is required to support autostereoscopic displays and even more important, to provide eye contact. Real-time 3D video processing is currently possible, but within some limits. Since sub-pixel disparity estimation is computationally expensive, the depth resolution of fast stereo approaches is directly linked to pixel quantization and the selected stereo baseline. The computational load requires 4x4 sub-sampled disparity estimation and therefore a loss of fine details.

Plane sweeping offers the capability to increase the depth resolution, but a fronto-parallel surface assumption is made. Patch based approaches use oriented spatial patches for piecewise linear approximation of the real object surface, but current algorithms are computationally expensive and hard to parallelize. Hence, a novel real-time capable algorithm is presented, the so-called patch sweeping, which combines plane sweeping with patch based 3D reconstruction by exploiting the processing capabilities of standard GPUs.

Fortunately, the 3D videoconferencing scenario allows for significant simplifications. Physiognomic constraints and coarse depth estimation results induce a valid search range for depth refinement. The current implementation on a single GPU performs three pair wise depth estimations of a trifocal camera on a 256 x 256 block in real-time in high quality depth resolution.

### 7871-15, Session 4

#### **Real-time scene change detection assisted with camera 3A: auto exposure, auto white balance, and auto focus**

L. Liang, B. Hung, Y. Noyes, R. Velarde, QUALCOMM MEMS Technologies, Inc. (United States)

Many scene change detection techniques have been developed for scene cuts, fade in and fade out by analyzing video encoder input signals. For real time scene change detection, sensor input signals provide the first-hand information which can be used for scene change detection. In this paper, by analyzing camcorder front end sensor input signals with our proposed algorithms based on camera 3A (auto exposure, auto white balance and auto focus), a novel scene change detection technique has been developed. With the feature of the fast responses to the scene, camera 3A based scene change detection algorithm can detect scene changes in a timely manner and therefore fits well for real time scene change detection application. Experimental results show that this algorithm can detect scene changes with a satisfying accuracy. As utilizing the embedded 3A features, the proposed algorithm is computationally efficient and easy to be implemented.

#### 7871-16, Session 4

### Fast approximate 4D:3D discrete radon transform, from light field to focal stack with $O(N^4)$ sums

J. G. Marichal-Hernandez, J. P. Lüke, F. L. Rosa, J. M. Rodriguez-Ramos, Univ. de La Laguna (Spain)

In this work we develop a new algorithm, that extends the bidimensional Fast Digital Radon transform from Götz and Druckmüller (1996), to digitally simulate the refocusing of a 4D light field into a 3D volume of photographic planes, as previously done by Ren Ng et al. (2005), but with the minimum number of operations.

This new algorithm does not require multiplications, just sums, and its computational complexity is  $O(N^4)$  to achieve a volume consisting of 2N photographic planes focused at different depths, from a  $N^4$  plenoptic image.

This reduced complexity allows for the acquisition and processing of a plenoptic sequence with the purpose of estimating 3D shape at video rate. Examples are given of implementations on GPU and CPU platforms.

Finally, a modified version of the algorithm to deal with domains of sizes different than power of two, is proposed.

#### 7871-17, Session 4

### A cross-based filter for fast edge-preserving smoothing

K. Zhang, IMEC (Belgium) and Katholieke Univ. Leuven (Belgium); J. Lu, Advanced Digital Sciences Ctr. (Singapore); G. Lafruit, R. Lauwereins, IMEC (Belgium); L. J. Van Gool, Katholieke Univ. Leuven (Belgium)

No abstract available.

#### 7871-18, Session 4

### Human action recognition in a wide and complex environment

S. Kumar, Indian Institute of Technology Roorkee (India); S. Kumar Malik, Univ. degli Studi di Udine (Italy); B. Raman, N. Sukavanam, Indian Institute of Technology Roorkee (India)

In this paper, a direct fractional linear discriminant analysis (DF-LDA) based classifier employed in a tree structure is presented to recognize the human actions in a wide and complex environment. In particular, the proposed classifier is based on a supervised learning process and achieves the required classification in a multi-step process. This multi-step process is performed simply by adopting a tree structured which is built during the training phase. Hence, there is no need of any priori information like in other classifiers such as the number of hidden neurons or hidden layers in a multilayer neural network based classifier or an exhaustive search as used in training algorithms for decision trees. A skeleton based strategy is adopted to extract the features from a given video sequence representing any human action. A pan-tilt-zoom (PTZ) camera is used to monitor the wide and complex test environment. A background mosaic image is built offline and used to compute the background frame in real time for any given pan and tilt setting. A background subtraction strategy has been adopted for detecting the object in various frames and to extract their corresponding silhouette. A skeleton based process is used to extract attributes of a feature vector corresponding to a human action. Finally, the proposed framework is tested on various indoor and outdoor scenarios and encouraging results are achieved in terms of classification accuracy and computational time.

#### 7871-20, Poster Session

### Swimming behavior detection for Nitocra Spinipes in water quality evaluation

Z. Jia, W. Wang, Henan Polytechnic Univ. (China)

The Environmental Protection Agency has over the years suggested a number of biological tests for characterization of industrial waste water, among which a test with the brackish water crustacean Nitocra spinipes can be found. It is of substantial interest to design a low-cost early warning test apparatus for brackish water. The principle of a test is to monitor the swimming behaviour of Nitocra spinipes by the use of digitized video films in daylight or indoor illumination. In our study, grown animals are of a size between 0.6 to 0.8 mm, and it is empirically known that their swimming behaviour is affected by the amount of toxic substances in the water. The first processing step is to manually mark the position of each animal on a starting image, then find out the image difference between starting image and the image in the sequence, finally find out locations of animals on the new image. The processing result is satisfactory. For all the working sequence, a Windows program was developed, the software system can process the images easily.

#### 7871-21, Poster Session

### Human heart movement tracing on ultrasonic images

X. Yang, Fuzhou Univ. (China)

Currently, the accurate human heart diagnosing is more and more important since this kind of sickness increases year and year. The traditional method can only obtain the heart movement rate and acceleration, but without displacement or orientation information, in order to supplement the heart movement information, this paper presents an algorithm for collecting human heart information on two-dimensional ultrasonic images. The algorithm firstly enhances images for easily detecting key heart points that have obvious information for heart movements, then, the algorithm finds out the key points by detecting obvious displacements, subsequently it traces the heart movements and detects the movement directions, and finally it evaluates the main orientation and displacement of the heart movements, and calculates the average speed and acceleration of the movements based on the sequence ultrasonic images. The algorithm is tested by using a number of human ultrasonic images, and the experiments show the detection results are correct. When compared to other techniques - traditional method, this noninvasive approach can clearly yield more accurate results. In this way, the information of the heart movements can be obtained accurately, and the information can be used for human health evaluation by doctors. It provides quantitative foundation for heart diagnosing.

#### 7871-22, Poster Session

### Efficient object tracking in WAAS data streams

T. R. Clarke, Ball Aerospace & Technologies Corp. (United States) and Rochester Institute of Technology (United States)

Wide area airborne surveillance (WAAS) systems are a new class of remote sensing imagers which have many military and civilian applications. These systems are characterized by long loiter times (extended imaging time over fixed target areas) and large footprint target areas. These characteristics complicate moving object detection and tracking due to the large image size and high number of moving objects. This presentation evaluates existing object detection and tracking algorithms with WAAS data and provides enhancements to the processing chain which decrease processing time and maintain or increase tracking accuracy. Decreases in processing time are needed to perform real-time or near real-time tracking either on the WAAS sensor platform or in ground station processing centers. Increased tracking accuracy benefits real-time users and forensic (off-line) users.



## 7871-23, Poster Session

### How fast can one numerically reconstruct digitally recorded holograms?

L. Bilevich, L. Yaroslavsky, Tel Aviv Univ. (Israel)

Results of comparative study of the computational complexity of different algorithms for numerical reconstruction of electronically recorded holograms are presented and discussed. The following algorithms were compared: conventional Fourier and convolutional algorithms with and without scaling and a new universal DCT-based algorithm, in terms of the number of operations and required computer time. Based on the comparison results, the feasibility of real-time implementation of numerical reconstruction of holograms is evaluated.

## 7871-24, Poster Session

### Tracking flow of leukocytes in blood for drug analysis

A. Basharat, W. D. Turner, Kitware, Inc. (United States); G. Stephens, B. Badillo, R. Lumpkin, P. Andre, Portola Pharmaceuticals Inc. (United States); A. Perera, Kitware, Inc. (United States)

Modern microscopy techniques allow the imaging of blood components, including leukocytes, under flow conditions. The resulting video sequences provide unique insights into the behavior of blood under normal and restricted flow such as would be found within vasculature and they also allow for testing various drug therapies; however, manual analysis of these video sequences is intractable, requiring hours per 6 minute video clip. In this paper, we present an automated technique to analyze leukocyte flow through the microscope stage. Our techniques detect and track those leukocytes which slow or which adhere to the microscope flow chamber. We automatically count the detections, measure the velocity, and identify leukocytes which strongly adhere. From this, we calculate and graph statistics of leukocyte detections, velocity distributions, and adherence.

## 7871-25, Poster Session

### Phase correlation based adaptive mode decision for the H.264/AVC

A. Abdelazim, S. Mein, M. R. Varley, Univ. of Central Lancashire (United Kingdom); C. Grecos, Univ. of the West of Scotland (United Kingdom); D. Ait-Boudaoud, Univ. of Portsmouth (United Kingdom)

The H.264 video coding standard achieves high performance compression and image quality at the expense of increased encoding complexity, due to the very refined Motion Estimation (ME) and mode decision processes. This paper focuses on decreasing the complexity of the mode selection process by effectively applying a novel fast mode decision algorithm.

Firstly the phase correlation is analysed between a macroblock and its prediction obtained from the previously encoded adjacent block. Relationships are established between the correlation value and object size and also best fit motion vector.

From this a novel fast mode decision and motion estimation technique has been developed utilising pre-processing frequency domain ME in order to accurately predict the best mode and the search range. We measure the correlation between a macroblock and the corresponding prediction. Based on the result we select the best mode, or limit the mode selection process to a subset of modes. Moreover the correlation result is also used to select an appropriate search range for the ME stage.

Experimental results show that the proposed algorithm significantly reduces the motion estimation time whilst maintaining similar Rate Distortion performance, when compared to both the H.264/AVC Joint Model (JM) reference software and recently reported work.

## 7871-26, Poster Session

### Fast multilayered prediction algorithm for group of pictures in H.264/SVC

A. Abdelazim, S. Mein, M. R. Varley, Univ. of Central Lancashire (United Kingdom); C. Grecos, Univ. of the West of Scotland (United Kingdom); D. Ait-Boudaoud, Univ. of Portsmouth (United Kingdom)

The objective of scalable video coding is to enable the generation of a unique bitstream that can adapt to various bit-rates, transmission channels and display capabilities. The scalability is categorised in terms of temporal, spatial, and quality. To improve encoding efficiency, the SVC scheme incorporates inter-layer prediction mechanisms which increases complexity of overall encoding.

In this paper several conditional probabilities are established relating motion estimation characteristics and the mode distribution at different layers of the H264/SVC. An evaluation of these probabilities is used to structure a low-complexity prediction algorithm for Group of Pictures (GOP) in H.264/SVC, reducing computational complexity whilst maintaining similar performance.

When compared to the JSVM software, this algorithm achieves a significant reduction of encoding time, with a negligible average PSNR loss and bit-rate increase in temporal, spatial and SNR scalability. Experiments are conducted to provide a comparison between our method and a recently developed fast mode selection algorithm. These demonstrate our method achieves appreciable time savings for scalable spatial and scalable quality video coding, while maintaining better PSNR and lower bit rate.

## 7871-27, Poster Session

### X-Eye: a novel wearable vision system

Y. Wang, C. Fan, S. Chen, H. Chen, Fu-Jen Catholic Univ. (Taiwan)

This paper proposes a smart portable device, named the X-Eye, which provides a gesture interface with a small computing device and a large display for the application of photo capture and management. The small portable device can achieve the capture of photos at any time and any where and display captured photos on large screen up to 42 inches. The wearable vision system is implemented with a dual-core embedded system and can achieve real-time performance. The display device is a pico DLP projector which has a small volume size but can project large screen size. Five software modules are integrated into the embedded hardware. Color identification and gesture recognition are the core of the software technologies in the paper.

The dimensions of the X-Eye are optimized to be 8.7(W) x 8.5(H) x 3.2(D) cm<sup>3</sup>, and its weight is 170g. Total power consumption of the X-Eye is no more than 9.5W. The screen resolution is 640x480 pixels. The processing speed of the whole system including the gesture recognition is with the frame rate of 20FPS. Experimental results give 85% recognition rate. It demonstrates that this system has effective gesture interface with real-time performance, small size, but large screen.

## 7871-28, Poster Session

### Real-time vehicle matching for multi-camera tunnel surveillance

V. Jelaca, J. O. Nino-Castaneda, A. Frias-Velazquez, A. Pizurica, W. R. Philips, Univ. Gent (Belgium)

Tracking multiple vehicles with multiple cameras in tunnels is a challenging problem of great importance for tunnel safety. One of the main challenges is accurate vehicle matching across the cameras with non-overlapping fields of view. Since the cameras used in video surveillance are usually of substantially low to medium resolution, the motion blur and noise are significant and it is difficult to extract

informative features from the acquired vehicle images. Additionally, computational efficiency is essential because the systems dedicated to tunnel surveillance can contain hundreds of cameras which observe dozens of vehicles each. In this paper, we propose a low complexity, yet highly accurate method for vehicle matching using vehicle signatures composed of Radon transform like projections of the vehicle image. The proposed signatures can be calculated by a simple scan-line algorithm, by the camera software itself and transmitted to the central server or to the other cameras in a smart camera environment. The amount of data is drastically reduced compared to the whole image, which relaxes the data link capacity requirements. Experiments on real vehicle images, extracted from video sequences recorded in a tunnel by two distant security cameras, validate the proposed method.

## 7871-29, Poster Session

### **Differential coding of intra modes for high efficiency video coding**

E. Maani, Sony Electronics Inc. (United States); W. Liu, Hangzhou Dianzi Univ. (China)

Spatial domain directional intra prediction has been shown to be very effective to remove the correlation between the pixels in the current block and reconstructed neighbors. In AVC, 8 directional prediction modes (plus the DC prediction mode) are defined. The prediction mode number is signaled to the decoder using a simple predictive coding method. The current intra prediction has two major disadvantages: 1) the small number of directions does not provide sufficient precision to cover arbitrary directional patterns; and 2) the mode number prediction from neighbors is not accurate enough to exploit the geometric dependency between blocks. Increasing the number of directions typically results in a lower residual energy, however, the cost for signaling the prediction mode may also increase significantly such that little gain is observed. This is especially the case for small block sizes such as 4x4 or 8x8. To address this problem, in this submission, we propose a new method to accurately predict the intra directions from reconstructed neighboring pixels and differentially encode the intra directions. This allows a more precise directional prediction without the significant increase in the cost for transmitting the side information. Simulation results shows that the new intra prediction method can provide as much as 13% bitrate reduction compared to AVC intra prediction.

# Conference 7872: Parallel Processing for Imaging Applications

Monday-Tuesday 24-25 January 2011

Part of Proceedings of SPIE Vol. 7872 Parallel Processing for Imaging Applications

## 7872-01, Session 1

### Using a commercial graphical processing unit (GPU) and the CUDA programming language to accelerate image processing applications

R. P. Broussard, R. Ives, U.S. Naval Academy (United States)

The processing power available in current video graphics cards is approaching super computer levels. In the past two years the processing power in these cards has quadrupled. State-of-the-art graphical processing units (GPU) boast of computational performance in the range of 1.4 trillion floating point operations per second (1.4 Teraflops). This processing power is readily accessible to the scientific community at a relatively small cost. High level programming languages are now available that give access to the internal architecture of the graphics card allowing greater algorithm optimization. This research takes computationally expensive portions of an image-based iris identification algorithm and hosts it on a GPU using the C++ compatible CUDA language. The selected segmentation algorithm uses basic image processing techniques such as image inversion, value squaring, thresholding, dilation, erosion and the computationally intensive local kurtosis calculation (fourth standardized moment) and circular Hough transform. Strengths and limitations of the GPU Single Instruction Multiple Data architecture are discussed. The primary source of the graphical processing power, the multiple processing elements and layered memory system, are discussed in detail. Actual memory access and instruction execution times are provided. Impressive acceleration results were obtained. The iris segmentation algorithm was accelerated by a factor of 150 over the highly optimized C++ version hosted on the computer's central processing unit. Some parts of the algorithm ran at speeds that were over 400 times faster than their C++ counterpart. CUDA programming details and code samples are presented as part of the acceleration discussion.

## 7872-02, Session 1

### Automatic distribution of vision-tasks on computing clusters

T. Müller, A. Knoll, Technische Univ. München (Germany)

Distribution of computer vision tasks in parallel environments is essential considering the increasing demand for computational resources to accomplish advanced visual processing tasks.

Thus, a consistent and efficient but yet convenient system for parallel computer vision, and in fact also realtime actuator control is proposed. The system implements the multi-agent paradigm and a blackboard information storage. This, in combination with a generic interface for hardware abstraction and integration of external software components, is setup on basis of the message passing interface, which is the de facto standard for HPC environments and hence provides support for a large variety of platforms and has a huge user pool.

The system furthermore allows for data- and task-parallel processing, and supports both synchronous communication, as data exchange can be triggered by events, and asynchronous communication, as data can be polled, strategies. Also, by duplication of processing units (agents) redundant processing is possible to achieve greater robustness.

As the system automatically distributes the agents to available resources, and a monitoring concept allows for combination of tasks and their composition to complex processes, it is very easy to develop vision / robotics applications quickly.

Thus, for evaluation multiple vision based applications have already been implemented, e.g. an evolutionary approach for learning visual

saliency features, or a parallel active perception system for robotic recognition and handling of limp objects.

## 7872-03, Session 1

### Highly scalable digital front end architectures for digital publishing

D. Staas, Hewlett-Packard Co. (United States)

HP's digital printing presses consume a tremendous amount of data. The architectures of the Digital Front Ends (DFEs) that feed these large, very fast presses have evolved from basic, single-RIP (Raster Image Processor) systems to multi-rack, distributed systems that can take a PDF file and deliver data in excess of 1.1 Gigapixels per second to keep the presses printing at 2500+ pages per minute. This paper highlights some of the more interesting parallelism features of our DFE architecture.

The high-performance architecture developed over the last 5+ years can scale up to HP's largest digital press, out to multiple mid-range presses, and down into a very low-cost single box deployment for low-end devices as appropriate. Principles of parallelism pervade every aspect of the architecture, from the lowest-level elements of jobs to parallel imaging pipelines that feed multiple presses.

From cores to threads to arrays to network teams to distributed machines, we use a systematic approach to move bottlenecks. The ultimate goals of these efforts are: to take the best advantage of the prevailing hardware options at our disposal; to reduce power consumption and cooling requirements; and to ultimately reduce the cost of the solution to our customers.

## 7872-04, Session 1

### Parallel training and testing methods for complex image processing algorithms on distributed, heterogeneous, unreliable, and non-dedicated resources

R. Usamentiaga, D. F. García, J. Molleda, I. Sainz, F. G. Bulnes, Univ. de Oviedo (Spain)

Advances in the image processing field have brought new methods which are able to perform complex tasks robustly. However, in order to meet constraints on functionality and reliability, imaging application developers commonly design complex algorithms with many parameters which need to be finely tuned for each particular environment. The best approach to tune these algorithms is to use an automatic training method, but a major issue arises when designing this kind of training method: the computational cost. The execution of the training method can be completely prohibitive, even in powerful machines. The same problem shows up when designing testing procedures. This work presents methods to train and test complex image processing algorithms within parallel execution environments. The approach proposed in this work is to use existing resources, in offices or laboratories, rather than expensive clusters. These resources are typically non-dedicated, heterogeneous and unreliable. The proposed methods have been designed to deal with all these issues. Two different methods are proposed: intelligent training based on genetic algorithms and PVM, and a full factorial design based on grid computing which can be used for training or testing. These methods are capable of harnessing the available computational power resources, giving more work to more powerful machines, and also taking its unreliable nature into account. Both methods have been tested using real applications.

## 7872-05, Session 1

### Integrated parallel printing systems with hypermodular architecture

D. K. Biegelsen, L. Crawford, C. Eldershaw, M. Fromherz, Palo Alto Research Center, Inc. (United States); G. Kott, B. Mandel, S. Moore, Xerox Corporation (United States); B. Preas, L. Swartz, Palo Alto Research Center, Inc. (United States)

Printing systems composed of multiple, interconnected marking engines (MEs) provide many potential advantages compared with ad hoc single engine designs. Tightly integrated multi-ME systems enable, for example, near optimal system utilization and document throughput through multi-threaded job production. Composable printing systems allow reconfigurability to match user needs and redundancy to support high reliability. The work presented here describes a system of four MEs linked by a paper path that has a deep level of modularity. The paper path consists of a regular grid populated by a small number of module types - nip modules to provide bidirectional sheet motion and two types of directors for static and dynamic definition of path topology. Each module is capable of acting, sensing, computing and communicating. Modules including MEs, are hot-swappable, and the system is capable of auto-configuration. Real-time planning and control software, like the hardware, is designed to be modular, distributed, reconfigurable and scalable. The system can handle exceptions, such as sheet jams, while maintaining (reduced) throughput.

## 7872-06, Session 2

### Parallel processing considerations for image recognition tasks

S. J. Simske, Hewlett-Packard Labs. (United States)

Many image recognition tasks are well-suited to parallel processing. The most obvious example is that many imaging tasks require the analysis of multiple images. From this standpoint, then, parallel processing need be no more complicated than assigning individual images to individual processors. However, there are three less trivial categories of parallel processing that will be considered in this paper: parallel processing (1) by task; (2) by image region; and (3) by meta-algorithm.

Parallel processing by task allows the assignment of multiple workflows-as diverse as optical character recognition [OCR], document classification and barcode reading-to parallel pipelines. This can substantially decrease time to completion for the document tasks. For this approach, each parallel pipeline is generally performing a different task.

Parallel processing by image region allows a larger imaging task to be sub-divided into a set of parallel pipelines, each performing the same task but on a different data set. This type of image analysis is readily addressed by a map-reduce approach. Examples include document skew detection and multiple face detection and tracking.

Finally, parallel processing by meta-algorithm allows different algorithms to be deployed on the same image simultaneously. This approach may result in improved accuracy.

## 7872-07, Session 2

### GPGPU real-time texture analysis framework

M. A. Akhlooufi, Ctr. of Robotics and Vision (Canada) and Laval Univ. (Canada)

In recent years we assist to an increase of interest in using texture features in industrial applications. For this type of applications, real time processing of captured images is an important issue, particularly with the current increase in image resolutions used in real world applications. Different approaches are available to solve this problem: DSP processing, FPGA, specialized hardware, parallel systems, computer clusters. All these solutions come at a higher cost.

More recently a research community became interested in graphic processing units available in commercial graphic cards. This domain called GPGPU (General-Purpose computation on Graphics Processing Units) aim to using the processing power of the GPU in order to accelerate general processing like mathematics, 3D visualization, image processing, etc.

In this work we present the use of GPGPU technology for building a framework for parallel real time texture analysis in computer vision. The following techniques were developed: Local Binary Pattern (LBP), Local Ternary Pattern, Laws texture kernels, Gabor filter jets and Gray Level Co-Occurrence Matrix (GLCM).

For this work we chose to use CUDA technology for developing the proposed algorithms. CUDA (Compute Unified Device Architecture) is a parallel computing architecture developed by NVIDIA. It enables dramatic increases in computing performance by harnessing the power of the GPU (graphics processing unit) and parallel architecture programming. GPU optimizations are compared to CPU optimizations using MMX-SSE technologies and Multicore parallel programming. The experimental results show an important increase in the performance of the proposed algorithms when GPGPU is used particularly for large image sizes.

## 7872-08, Session 2

### A parallel implementation of 3D Zernike moment analysis

D. Berjón, S. Arnaldo, F. Morán, Univ. Politécnica de Madrid (Spain)

Zernike polynomials are a well known set of functions that find many applications in image or pattern characterization because they allow to construct shape descriptors that are invariant against translations, rotations or scale changes. The concepts behind them can be extended to higher dimension spaces, making them also fit to describe volumetric data. They have been less used than their properties might suggest due to their high computational cost.

We present a parallel implementation of 3D Zernike moments analysis, written in C with CUDA extensions, which makes it practical to employ Zernike descriptors in interactive applications, yielding a performance of several frames per second in voxel datasets about  $200^3$  in size.

In our contribution, we describe the challenges of implementing 3D Zernike analysis in a GPGPU. These include how to deal with numerical inaccuracies, due to the high precision demands of the algorithm, or how to deal with the high volume of input data so that it does not become a bottleneck for the system.

Our GPU-based implementation runs about ten times faster than our previous CPU-based one.

## 7872-09, Session 2

### A novel parallel algorithm for airport runway segmentation in satellite images using priority-directional region growing strategy based on ensemble learning

F. Duan, Y. Zhang, Tsinghua Univ. (China)

This paper addresses the problem of airport detection and runway segmentation in satellite images with complex background clutter. To this ends, we propose a novel ensemble learning based parallel runway segmentation algorithm. The contributions of our work can be summarized as follows: (a) we propose the concept of priority direction region growing. (b) We introduce the Bresenham's line generation algorithm into our segmentation task. (c) we adopt a two-stage strategy to better segment the regions corresponding to the airport runway by applying traditional region growing method and our priority directions (two orthogonal directions in our problem) growing method. (d) In our runway segmentation algorithm, ensemble-learning strategy is used to combine the growing results of each line segment. In addition, those thin branches, which have significantly different width, are eliminated.

To evaluate the effectiveness of our algorithm, extensive simulations are carried out on the testing images obtained from Google Map. Our experimental results show that our algorithm can effectively and efficiently segmented the airport region, and generate very neat boundaries of the runways, and have great superiority over the state-of-the-art methods.

## 7872-10, Session 2

### Visualization assisted by parallel processing

B. Lange, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); X. Vasques, H. Rey, IBM (France); N. Rodriguez, W. Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France)

This paper discusses the experimental results of our visualization model for data extracted from sensors. We have to find the faster and more efficient method to produce a real time rendering visualization for a large amount of data. We develop visualization methods to monitor temperature variance of a data center. Sensors are placed on three layers and do not cover all the room. We use particle paradigm to interpolate data sensors, as presented in cite[Latta2004]. Particles model the "space" in the room, as stated by the Industry Foundation Classes (IFC) model, a standard building semantic model. In this work we make a partition of the particle set using two mathematical methods: Delaunay triangulation and Vorono<sup>i</sup> cells. Avis and Bhattacharya present these algorithms in cite[Avis1983]. Particles carry information of the room temperature, and display the evolution of temperature in time. To locate and update particles data we have defined a computational cost function. To solve this function in an efficient way, we use a client server paradigm cite[Kapferer2008]. Server computes data, client display this data in a virtual screen composed from four video projectors in stereoscopic view. This paper is organized as follows. The first part presents related solutions used to visualize large flow of data. The second part presents our different platforms and methods used, which have been evaluated in order to determine the better solution for the task proposed. The benchmark use the computational function of our algorithm, it was based on located particles compared to sensors and on update of particles value. Figure ref[fig-10] and ref[fig-11] illustrate the inclusion method using ray tracing, each particle is tested on the nearest tetrahedron. The benchmark was done on a personal computer using CPU, multi core programming. However, real time rendering is hard to have. This work is performed in collaboration with IBM Montpellier, and within this collaboration, our work can be improved on High Performance Computer and on a hybrid CPU, GPU server. The first method is commonly used in data visualization (astronomy, physic, etc.) and the second one is growing in computer science. With this other different platform we want to have a real time rendering of our large data flow.

## 7872-11, Session 3

### A parallel impulse-noise detection algorithm based on ensemble learning for switching median filters

F. Duan, Y. Zhang, Tsinghua Univ. (China)

In this paper, a highly effective and efficient ensemble learning-based parallel algorithm for impulse noise detection is proposed. The contribution of this paper is three-fold. First, a novel intensity homogeneity metric, which has very powerful discriminative ability that has been proven, is proposed. Second, this proposed algorithm has high parallelism in feature extraction stage, classifier training and testing stage. Finally, instead of manually tuning the thresholds for each feature as most of the works in this research area do, Random Forests (RF) is used to make decision since it has been demonstrated to own better generalization ability and performance comparable to SVMs in classification problem. Another important reason why RF is adopted is that it has natural parallelism structure and very significant performance advantage (e.g. the overhead of training and testing the model) over other popular classifiers e.g. SVMs. To the best of our knowledge, this

is the first time that the ensemble learning strategies have been used in the area of switching median filtering. Extensive simulations are carried out on eight most common standard testing images. The experimental results show that our algorithm achieves zero-miss detection result while keeping the false alarm rate at a rather low level, and has great superiority over other state-of-the-art methods.

## 7872-12, Session 4

### GPU color space conversion

G. L. Vondran, Jr., P. Chase, Hewlett-Packard Co. (United States)

Tetrahedral interpolation is commonly used to implement continuous color space conversions from sparse 3D and 4D lookup tables. We investigate the implementation and optimization of tetrahedral interpolation algorithms for GPUs, and compare to the best known CPU implementations. We show that a \$350 NVIDIA GTX-470 GPU is 4x faster than a \$1000 Intel Core i7 980X CPU for 3D interpolation, and 16x faster for 4D interpolation.

Performance-relevant GPU attributes are explored including thread scheduling, local memory characteristics, global memory hierarchy, and cache behaviors. We consider existing tetrahedral interpolation algorithms and tune based on the structure and branching capabilities of current GPUs. Global memory performance is improved by reordering and expanding the lookup table to ensure optimal access behaviors. Per multiprocessor local memory is exploited to implement optimally coalesced global memory accesses, and local memory addressing is optimized to minimize bank conflicts. We explore the impacts of lookup table density upon computation and memory access costs.

We present a CPU-based 3D interpolator, using SSE vector operations, that is faster than any previously published solution.

## 7872-13, Session 4

### Acceleration of the Retinex algorithm for image restoration by GPGPU/CUDA

Y. Wang, W. Huang, Fu-Jen Catholic Univ. (Taiwan)

In this paper, a data parallel algorithm called GPURetinex is proposed to parallelize the Retinex algorithm on GPGPU. The computing of the Gaussian blur in the GPURetinex adopts separable convolution kernels to reduce the computation and the internal data transfer. The data distribution of parallel Gaussian blur convolution adopts a horizontal stripe method. Each thread reads pixels of a horizontal stripe of the image to implement the Gaussian blur convolution. The Gaussian blur utilizes texture memory and constant memory to improve efficiency. The data distribution in the log-domain subtraction and normalization steps uses a square subimage method. Each thread computes the two operations for all pixels within its square subimage. A parallel reduction method is devised to find the maximum and minimum values of the log-domain subtraction image. Threads within a grid communicate by global memory and shared memory.

In our experiments, the GT200 GPU and CUDA 3.0 are deployed. The experimental results show that the GPURetinex can gain 23x speedup compared with CPU-based implementation on the images with 2048 x 2048 resolution. Our experimental results indicate that using CUDA can move the Retinex algorithm to the GPU hardware and achieve acceleration to gain the real-time performance.

## 7872-14, Session 4

### Performance evaluation of Canny edge detection on a tiled multicore architecture

A. Z. Brethorst, N. Desai, D. Enright, R. Scrofano, The Aerospace Corp. (United States)

Because transistor size has continued to fall while processor clock

frequency has remained fairly static, the trend in computer architecture has been to develop multicore processors. To take full advantage of multicore processors, applications must be amenable to parallelization. Many imaging applications can benefit from multicore technology because they consist of localized operations. Hence, there are a multitude of multicore platforms and parallelization strategies that can be applied to imaging problems.

A variety of multicore architectures have been introduced, including more traditional shared-memory parallel architectures and more exotic tiled multicore architectures. Tiled multicore architectures have many processor cores connected by an on-chip interconnection network. With these features, tiled multicore architectures can support a variety of parallel programming patterns.

In this paper, we investigate the performance and scalability of a particular imaging application--Canny edge detection--on the Tiler Tile64 tiled multicore processor. We apply various parallel programming patterns, including divide and conquer and geometric decomposition, and draw conclusions about their suitability to the application and the target platform. As part of our study, we will develop implementations in which parallelism is explicitly managed in the program and implementations in which parallelism is managed by compiler directives and run-time systems.

### 7872-15, Session 5

#### Video transcoding using GPU accelerated decoder

W. Hsu, Advanced Micro Devices, Inc. (United States)

Combining consumer electronics, digital entertainment, and the personal computer platform has become one of the key driving forces of modern PC development. In this paper, we presented the implementation of a UVD-accelerated decoder MFT through Microsoft DirectX Video Acceleration (DXVA) interface for enhancing the performance of drag-and-drop transcoding of high-definition videos running on Windows 7 platforms. In which the uncompressed video is downsampled by a GPU-based resize to conserve memory bandwidth between the GPU and CPU. Experimental results show this UVD-accelerated decoder MFT is able to perform real-time playback of high-definition video on a PC with extremely low CPU usage. Combining the GPU-based resolution scaler and the driver feedback, the proposed technology can double the speed of video transcoding of 1080p (1920x1080) video content on a processor with an integrated graphics unit by using less than half CPU capability compared to software VC1 and H.264 decoders. Although modern high-speed multi-core CPUs with advanced single-instruction multiple-data (SIMD) architecture may outperform the UVD unit in video decoding, UVD acceleration will always be helpful in off-loading the CPU's task and increasing the computational capacity of PC platforms.

### 7872-16, Session 5

#### Real-time image deconvolution on the GPUs

J. T. Klosowski, S. Krishnan, AT&T Labs. Research (United States)

2D image deconvolution is an important and well-studied problem with applications to image deblurring and restoration. Most of the best deconvolution algorithms use natural image statistics that act as priors to regularize the problem. Recently, Krishnan and Fergus provide a fast deconvolution algorithm that yield results comparable to the current state of the art. They use a hyper-Laplacian image prior to regularize the problem. The resulting optimization problem is solved using alternating minimization in conjunction with a half-quadratic penalty function. In this paper, we provide an efficient CUDA implementation of their algorithm on the GPU.

Our implementation leverages many well-known CUDA optimization techniques, as well as several others that have a significant impact on this particular algorithm. We discuss each of these, as well as make a few observations regarding the CUFFT library. Our experiments were run on an nVidia GeForce GTX 260 GPU. For a single channel image of

size 710 x 470, we obtain 40.6 fps, while on a larger image of size 1900 x 1266, we get 5.8 fps (without counting disk I/O). In addition to linear performance, we believe ours is the first implementation to perform deconvolutions at video rates.

### 7872-17, Session 5

#### Stitching giga pixel images using parallel computing

R. Kooper, P. Bajcsy, Univ. of Illinois at Urbana-Champaign (United States)

This paper addresses the problem of stitching Giga Pixel images from airborne images acquired over multiple flight paths of Costa Rica in 2005. The set of input images contains about 10,158 images, each of size around 4072x4072 pixels, with very coarse georeferencing information (latitude and longitude of each image). Given the spatial coverage and resolution of the input images, the final stitched image is 294,847 by 269,195 pixels (79.3 Giga Pixels). Our approach is to utilize the coarse georeferencing information for initial image grouping followed by an intensity-based stitching of groups of images. This group-based stitching is highly parallelizable. The stitching process results in image patches that can be cropped to fit a tile of an image pyramid frequently used as a data structure for fast image access and retrieval. We report our preliminary experimental results obtained when stitching and pyramid tiling a 4 Giga Pixel image from the input images at one fourth of their original spatial resolution using a single core on our eight core server. As the processing requires parallel computing approaches in order to generate the full resolution image, we are collecting actual benchmarks on parallel hardware platforms.

### 7872-18, Session 6

#### GPU-completeness: concept and implications

I. Lin, Hewlett-Packard Labs. (United States)

This paper formalizes one of a major insight into a class of algorithms that relate between parallelism and performance. The purpose of this paper is to define a class of algorithms that trades off parallelism for quality of result (e.g. visual quality, compression rate), and we propose a similar method for algorithmic classification based on NP-Completeness techniques, applied toward parallel acceleration. We will define this class of algorithm as "GPU-Complete" and will postulate the necessary properties of the algorithms for admission into this class. We will also formally relate his algorithmic space and imaging algorithms space.

While GPUs are merely one type of architecture for parallelization, we show that their introduction into the design space of printing systems demonstrate the trade-offs against competing multi-core, FPGA, and ASIC architectures. While each architecture has its own optimal application, we believe that the selection of architecture can be defined in terms of properties of GPU-Completeness.

For a well-defined subset of algorithms, GPU-Completeness is intended to connect the parallelism, algorithms and efficient architectures into a unified framework to show that multiple layers of parallel implementation are guided by the same underlying trade-off.

### 7872-20, Session 6

#### A parallel error diffusion implementation on a GPU

Y. Zhang, Univ. of California, Davis (United States); J. Recker, R. A. Ulichney, G. B. Beretta, I. Tastl, I. Lin, Hewlett-Packard Labs. (United States); J. D. Owens, Univ. of California, Davis (United States)

With the ever-increasing printing resolution and speed, digital presses

present high demands for the processing power of the Raster Imaging Process (RIP). Today's massively parallel GPUs can potentially provide a high-performance and cost-effective solution for RIPs. However, error diffusion, as a major stage in the printer imaging pipeline, is inherently serial in its original form. In this paper, we investigate the suitability of the GPU for a parallel implementation of the error diffusion algorithm. We demonstrate a high-performance GPU implementation achieved by efficiently managing the memory usage to meet the hardware constraints. Our GPU implementation achieves a 10-20x speedup over a sequential CPU error diffusion with comparable image quality. We conduct various experiments to study the performance and quality tradeoffs for differences in parallelism, randomization, block and image sizes.

## 7872-21, Session 6

### Optimization of imaging algorithms on multiple core CPUs

R. J. Moore, 3M Co. (United States)

With the release of an eight core Xeon processor by Intel and a twelve core Opteron processor by AMD in the spring of 2010, the increase of multiple cores per chip package continues. Multiple core processors are common place in most workstations sold today and are an attractive option for increasing imaging performance. Most imaging algorithms, especially large difference of Gaussian filters, segmentation, and region finding are very compute intensive. In this paper we present our work in optimizing the performance of different imaging algorithms to run on standard multi-core Windows workstations. Our work leverages the OpenMP libraries in C++ to create parallel for loops. We will present our experience in getting the best performance for imaging algorithms from these libraries.

## 7872-22, Session 7

### Evaluation of CPU and GPU architectures for spectral image analysis algorithms

V. Fresse, Univ. Jean Monnet Saint-Etienne (France); D. Houzet, Gipsa-lab (France); C. Gravier, Telecom Saint Etienne (France)

Graphical Processing Units (GPU) architectures are massively used for resource-intensive computation. Initially dedicated to imaging, vision and graphics, these architectures serve nowadays a wide range of multi-purpose applications. The GPU structure, however, does not suit all applications. This can lead to performance shortage. Among several applications, the aim of this work is to analyze GPU structures for image analysis applications in multispectral to ultraspectral imaging. Algorithms used for the experiments are multispectral and hyperspectral imaging dedicated to art authentication. Such algorithms use a high number of spatial and spectral data, along with both a high number of memory accesses and a need for high storage capacity. Timing performances are compared with CPU architecture and a global analysis is made according to the algorithms and GPU architecture. This paper shows that GPU architectures are suitable to complex image analysis algorithm in multispectral.

## 7872-23, Session 7

### Computational scalability of large size image dissemination

R. Kooper, P. Bajcsy, Univ. of Illinois at Urbana-Champaign (United States)

We have investigated the computational scalability of image pyramid building needed for dissemination of very large image data. The sources of large images include high resolution microscopes and telescopes, remote sensing and airborne imaging, and high resolution scanners. The term 'large' is understood from a user perspective which means either larger than a display size or larger than a memory/

disk to hold the image data. The application drivers for our work are digitization projects such as the Lincoln Papers project (each image scan is about 100-150MB or about 5000x8000 pixels with the total number to be around 200,000) and the UIUC library scanning project for historical maps from 17th and 18th century (smaller number but larger images). The goal of our work is understand computational scalability of the web-based dissemination using image pyramids for these large image scans, as well as the preservation aspects of the data. We report our computational scalability benchmarks using the Microsoft Seadragon library for building image pyramids, hyper-threading for computation execution and various hard drive configurations such as RAID drives for input/output operations. The benchmarks are obtained with a map (334.61 MB, JPEG format, 17591x15014 pixels). The discussion combines the speed and preservation objectives.

## 7872-25, Session 8

### Real-time 3D flash lidar imaging through GPU data processing

C. M. Wong, C. Bracikowski, B. Baldauf, S. Havstad, Northrop Grumman Aerospace Systems (United States)

We present real-time 3D image processing of flash lidar data using our recently developed GPU parallel processor kernels. Our laboratory and airborne experiences with flash lidar focal planes have shown that per laser flash, typically only a small fraction of the pixels on the FPA actually produce a meaningful range signal. Therefore, to optimize overall data processing speed, this large quantity of uninformative data should be filtered out and removed from the data stream prior to the mathematically intensive data processing. This front-end pre-processing, which largely consists of control flow instructions, is specific to the exact type of flash lidar focal plane array used. The valid signals along with their corresponding inertial navigational metadata are then transferred to a GPU device to perform range-correction, geo-location, and ortho-rectification on each 3D data point so that data from multiple frames can be properly tiled together either to create a wide-area map or to reconstruct an object from multiple look angles. GPU parallel processor kernels were developed using the OpenCL application programming interface. Post-processing to perform fine registration between data frames via complex iterative steps also benefits greatly from this type of high-performance computing. The performance improvements obtained using GPU processing to create corrected 3D images and for frame-to-frame fine-registration are presented.

## 7872-26, Session 8

### Advanced MRI reconstruction toolbox with accelerating on GPU

X. Wu, Y. Zhuo, J. Gai, F. Lam, M. Fu, J. P. Haldar, W. Hwu, Z. Liang, B. P. Sutton, Univ. of Illinois at Urbana-Champaign (United States)

In this paper, we present a fast iterative MR image reconstruction algorithm taking advantage of the prevailing GPGPU programming paradigm. In clinical environment, MR image reconstruction is usually performed via fast Fourier transform (FFT). However, imaging artifacts (signal loss and signal distortions) resulting from susceptibility-induced magnetic field inhomogeneities degrade the quality of reconstructed images. These artifacts must be addressed using accurate modeling of the physics of the system coupled with iterative reconstruction. We have developed a reconstruction algorithm with improved image quality at the expense of computation time. Hence, an implementation on GPUs is proposed, achieving significant speedup. The proposed algorithm implements a conjugate gradient reconstruction using explicit Fourier transform (FT) in order to model the field inhomogeneity and its gradients. In addition, a smoothing constraint is included in the form of sparse matrix regularization in order to reduce noise in reconstructed images. We apply the compilation optimizations from levels of algorithm, program code structures, and specific architecture

performance tuning, featuring both our MRI reconstruction algorithm and GPU hardware specifics. The current GPU implementation produces accurate image estimates while accelerating the reconstruction by two orders of magnitudes. Future directions include further optimization of current and higher-dimension approach.

## 7872-27, Session 8

### Accelerating image recognition on mobile devices using GPGPU

M. Bordallo Lopez, H. Nykänen, J. Hannuksela, O. J. Silvén, Univ. of Oulu (Finland); M. Vehviläinen, Nokia Research Ctr. (Finland)

The future multimodal user interfaces of battery-powered mobile devices are expected to require energy-efficient computationally costly image analysis techniques.

GPU computing is well suited for parallel processing. The addition of programmable stages and high precision arithmetic provide for opportunities to implement energy-efficient complete algorithms on GPU. At the moment the first mobile graphics accelerators with programmable pipelines are available, enabling the GPGPU implementation of several image processing algorithms.

In this context, we consider a face tracking approach that uses efficient gray-scale invariant texture features and boosting. The solution is based on the Local Binary Pattern (LBP) features and makes use of the GPU on the pre-processing and feature extraction phase.

We have implemented a series of image processing techniques in the shader language of OpenGL ES 2.0, compiled them for a mobile graphics processing unit and performed tests on a mobile application processor platform (OMAP3530).

In our contribution, we describe the challenges of the design on a mobile platform, present the performance achieved and provide measurement results for the actual power consumption in comparison to using the CPU (ARMv7) on the same platform.

Our experiments show how a considerable speedup of image processing applications can be achieved by the simultaneous use of the GPU and the CPU on mobile devices.

## 7872-28, Session 8

### Multi-view stereo reconstruction via voxels clustering and parallel volumetric graph-cut optimization

Y. Zhu, Y. Zhang, Tsinghua Univ. (China)

Traditional methods of multi-view stereo (MVS) via volumetric graph-cut formulate the multi-view scene reconstruction problem into a computationally tractable global optimization using graph-cut. The optimal surface can be obtained as the max-flow/min-cut solution of the weighted graph. With the resolution of cubical voxels become larger, the reconstruction accuracy is increased gradually, however, the photo consistency estimation for voxels and the graph-cut computation increasing rapidly, if more edges between neighbor voxel are added into graph, the desktop computer nearly can't handle the problem. Therefore, the contradiction between computation and accuracy needs to be reduced urgently. In our paper, we are focusing on improving the performance of graph-cut based MVS algorithm with the help of multi-core CPUs. We have developed a system to demonstrate the clustering and it can utilize the parallel optimization algorithm for this problem. A key technical contribution is the voxels clustering algorithm, it divides the voxels into several overlapping clusters, after that graph-cut algorithm can be used to reconstruct each area of the surface in parallel. Finally, the optimization result in each subgraph is collected together and the labels on overlapped voxels are forced to be consistent interactively. This technique demonstrated faster graph based multi-view volumetric reconstruction computations when multiple CPU cores are available. The reconstruction results are comparable with the state-of-art MVS methods.

## 7872-29, Session 8

### A GPU accelerated PDF transparency engine

J. Recker, I. Lin, I. Tastl, Hewlett-Packard Labs. (United States)

As commercial printing presses become faster, cheaper and more efficient, so too must the Raster Image Processors (RIP) that process and feed them data to print. Digital press RIPs, however, have been challenged to both meet the ever increasing print performance of the latest digital presses, and the more common use of advanced pixel processing such as ICC color profile specified color spaces and transparent image layers. As a result, the cost of the RIPs deployed at some of the more demanding Print Service Providers (PSP) can exceed \$250,000.

This paper explores the challenges encountered when implementing a GPU accelerated driver for the open source Ghostscript Adobe PostScript and PDF language interpreter targeted at accelerating PDF transparency for high speed, large page size commercial presses. It further describes our solution, including an image memory manager for tiling input and output images, a cache of dynamically compiled GPU programs for an accelerated ICC v4 compatible color transformation engine, and an Adobe PDF compatible multiple image layer blending engine. The result, we believe, is the foundation for a scalable, efficient, distributed RIP system that can meet current and future RIP requirements for a wide range of commercial digital presses.

## 7872-30, Poster Session

### Infrared small target tracking based on SOPC

T. Hu, Electronic Engineering Institute of Hefei (China); X. Fan, Univ. of Science and Technology of China (China); Y. Zhang, Tsinghua Univ. (China); Z. Chen, B. Zhu, Electronic Engineering Institute of Hefei (China)

The tracking of infrared small targets has been a key technology in the field of satellite early warning, precise guidance, surveillance and detection. The paper presents a low cost FPGA based solution for a real-time infrared small target tracking system. A specialized architecture is presented based on a soft RISC processor capable of running kernel based mean shift tracking algorithm. Mean shift tracking algorithm is realized in Nios II soft-core With SOPC technology. Though mean shift algorithm is widely used for target tracking, the original mean shift algorithm can not be directed used for infrared small target tracking. As infrared small target only has intensity information. So an improved mean shift algorithm is presented in this paper. How to describe target will determine whether target can be tracked by mean shift algorithm. Because color target can be tracked well by mean shift algorithm, imitating color image expression, spatial component and temporal component are advanced to describe target, which forms pseudo-color image. The experimental results show that infrared small target is tracked stably in complicated background. In order to improve the processing speed parallel technology and pipeline technology is taken. Two RAM are taken to store images separately by ping-pong technology. A FLASH is used to store mass temp data.

## 7872-31, Poster Session

### A novel method for multi-view synthesis using relative affine structure

Z. Huo, Nanjing Univ. of Posts and Telecommunications (China)

Nowadays, there are increasing research interests in Free viewpoint Television (FTV) that offers arbitrary views of 3D scene. View synthesis is important for prediction-based compression and novel view display in FTV and other multiview imaging application such as 3DTV.

View synthesis consists in rendering images of a scene as if they were taken from a virtual viewpoint different from all the viewpoints of the real views. This paper proposes a novel approach for multiview synthesis in which the process of view synthesis is based on the



## Conference 7872: Parallel Processing for Imaging Applications

relative affine structure. The advantage is that photographs of real scenes can be used as a basis to create very realistic images, and rendering time is decoupled from the complexity of the scene. The virtual camera position is specified in an uncalibrated setting-up via the interpolation of the motion information among the reference views. This method yields a parametric family of camera poses that describes a smooth trajectory in the Euclidean space as the parameters vary continuously. The new synthesized images can be rendered from virtual cameras moving on the given parameterized trajectories.

Synthetic and real experimental results illustrate that the method has advantages of low computational complexity without the error produced from depth estimation.

## 7873-02, Session 2

### **Myopic reconstruction and its application to MRFM data**

S. U. Park, Univ. of Michigan (United States); N. Dobigeon, Univ. de Toulouse (France); A. O. Hero III, Univ. of Michigan (United States)

We propose a solution to the image deconvolution problem where the convolution operator or point spread function (PSF) is assumed to be only partially known. Small perturbations generated from the model are exploited to produce a few principal components explaining the uncertainty in a high dimensional space. Specifically, we assume the image is sparse since we focus on recovering magnetic resonance force microscopy (MRFM) data.

Unlike previous trials, our approach is stochastic, within the Bayesian Metropolis-within-Gibbs sampling procedure for estimation of the convolution kernel and the images. Our algorithm is compared to show the performance with previous Bayesian approach and alternating minimization (AM) algorithm as blind deconvolution method on a sparse synthetic image, and it is applied to MRFM data.

## 7873-04, Session 2

### **Seismic imaging of transmission overhead line structure foundations**

D. Vautrin, Institut de Recherche en Communications et en Cybernétique de Nantes (France); M. Voorons, Ecole Polytechnique de Montréal (Canada); J. Idier, Institut de Recherche en Communications et en Cybernétique de Nantes (France); Y. Goussard, Ecole Polytechnique de Montréal (Canada); S. Kerzale, Apside Technologies (France); N. Paul, EDF Recherche & Développement (France)

The objective of the presented work is to retrieve the shape of transmission line structure foundations using a seismic imaging approach. This non-destructive testing problem is formulated as a seismic inverse scattering problem where two-dimensional maps of the pressure- and shear-wave velocities are estimated from the measured data set.

The inversion amounts to a large-scale, nonlinear programming problem. It is rendered all the more difficult given the large dimensions of the scattering object and the large velocity contrasts. In this context, our goal is to propose an inversion scheme that produces precise images with an acceptable computational effort.

This goal is met by combining the following elements: we minimize a penalized least-square criterion under the positivity constraint with the L-BFGS-B algorithm, we work in the frequency domain to introduce the measured data in a progressive way and we introduce a logarithmic variable substitution.

Our main contribution is the introduction of a new working variable. It counterbalances the lack of sensibility of the criterion and results in a significant acceleration of the inversion process. This is confirmed on synthetic examples: the convergence is reached two to eight times faster than with the initial working variables.

## 7873-05, Session 2

### **Inverse problems for cryo electron microscopy of viruses: randomly oriented projection images of random 3D structures in noise**

Q. Wang, P. C. Doerschuk, Cornell Univ. (United States)

Instances of biological macromolecular complexes that have identical chemical constituents may not have the same geometry due to, for example, flexibility. Cryo electron microscopy provides one noisy projection image of each of many instances of a complex where the projection directions for the different instances are random. The noise is sufficient severe ( $\text{SNR} \ll 1$ ) that the projection direction for a particular image cannot be easily estimated from the individual image. The goal is to determine the 3-D geometry of the complex (the 3-D distribution of electron scattering intensity) which requires fusing information from these many images of many complexes. In order to describe the geometric heterogeneity of the complexes, the complex is described as a weighted sum of basis functions where the weights are random. In order to get tractable algorithms, the weights are modeled as Gaussian random variables with unknown statistics and the noise is modeled as additive Gaussian random variables with unknown covariance. The statistics of the weights and the statistics of the noise are jointly estimated by maximum likelihood by a generalized expectation maximization algorithm. An example using these ideas on images of Flock House Virus is described.

## 7873-06, Session 2

### **Inverse problems arising in different synthetic aperture radar imaging and a general Bayesian approach for them**

S. Zhu, A. Mohammad-Djafari, Lab. des Signaux et Systèmes (France)

Synthetic Aperture Radar (SAR) imaging systems are nowadays very common technics of imaging in remote sensing and environment survey. There are different acquisition modes: Spotlight, Stripbands, Interferometric, Polarimetric and different geometries: mono-, bi- and multi-static.

In a first approximation, the relation between the measured data and the scene can be modelled by a linear relation. In this paper, first, using this linear forward model, a common inverse problem framework for all of them is given and then a general probabilistic Bayesian estimation method is presented for image reconstruction problems. In particular, we consider two priors which permit parsimonious modeling of the scene, one for point sources directly in the image domain and the second by representing the scene on a dictionary based elementary functions. For Bayesian computations we will consider and compare the MAP and the posterior mean estimates. We will show the performances of the proposed methods on some simulated and real data.

## 7873-07, Session 2

### **Medical image enhancement using resolution synthesis**

T. Wong, C. A. Bouman, Purdue Univ. (United States); J. Thibault, GE Healthcare (United States); K. D. Sauer, Univ. of Notre Dame (United States)

No abstract available

## 7873-08, Session 3

### An open level set framework for image segmentation and restoration using the Mumford and Shah model

R. Mohieddine, L. A. Vese, Univ. of California, Los Angeles (United States)

In two dimensions, the Mumford and Shah functional allows for minimizers  $u$  (the image) and  $K$  (the edge set) such that the set  $K$  could include both closed and open curves. The current level set based segmentation algorithms can only detect objects with closed edges, which are boundaries that are formed by closed loops. We propose an efficient level set based algorithm for segmenting images with edges which are made up of open curves or crack-tips. By adapting P. Smereka's open level set formulation to variational problems, we are able to extend the current level-set based image segmentation methods, such as Chan-Vese. The algorithm retains many of the advantages of using level sets, such as a well-defined boundaries and ability to change topologies. We solve the resulting evolution equations using Sobolev gradients, avoiding the need for regularization or re-initialization of the level sets functions while also accelerating convergence to the reconstructed image. Another model derived by Mumford and Shah solely for the detection of the edge set is presented and is reformulated. This second model is useful when one only desires information on object boundaries rather than a reconstructed image, such as pathology detection in medical imaging, object detection and location in computer vision, object tracking in multiple images, etc. Finally, we present the numerical implementation with various examples comparing this method to both closed level set and general edge detection methods.

## 7873-09, Session 3

### Video indexing and retrieval using Fisher information nonlinear embedding

X. Chen, A. O. Hero III, Univ. of Michigan (United States)

In this paper, we present a novel information embedding based approach for video indexing and retrieval. The high dimensionality and complexity of video sequences poses a major challenge to video indexing and retrieval. Different from the traditional dimensionality reduction techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), we embed the video data into a low dimensional statistical manifold obtained by applying manifold learning techniques to the information geometry of video feature probability distributions (PDF). We estimate the feature PDF of the video features using histogram estimation and Gaussian mixture models (GMM), respectively. By calculating the similarities between the embedded trajectories, we demonstrate that the proposed approach outperforms traditional approaches to video indexing and retrieval with real world data in terms of precision and recall.

## 7873-10, Session 3

### Segmentation assisted food classification for dietary assessment

F. Zhu, M. Bosch, T. R. Schap, N. Khanna, D. S. Ebert, C. J. Boushey, E. J. Delp III, Purdue Univ. (United States)

Accurate methods and tools to assess food and nutrient intake are essential for research on the association between diet and health. Preliminary studies have indicated that the use of a mobile device with a built-in camera to obtain images of the food consumed may provide a less burdensome and more accurate method for dietary assessment. We have developed methods to identify food items using a single image acquired from the mobile device. Our goal is to automatically determine the regions in an image where a particular food is located (segmentation) and correctly identify the food type

based on its features (classification or food labeling). Images of foods are segmented using Normalized Cuts based on intensity and color. Local features are extracted from image patches around key-points detected by Scale-invariant feature transform (SIFT), we also extract global features of each segmented food region. Classification decision for each classifier (or feature space) is made by choosing the food item which is at the minimum distance in that feature space. The final decision is made by fusing the candidate outputs of separate classifiers by a majority vote rule. Segmentation of each food region is refined based on feedback from the output of classifiers to provide more accurate estimation of the quantity of food consumed.

## 7873-46, Session 3

### Joint pose estimation and image segmentation for monocular articulated tracking

L. M. Huffman, I. Pollak, Purdue Univ. (United States)

The need for automated tracking of a human's pose from video data has arisen in applications from surveillance to human-machine interaction. Pose estimation is particularly difficult in monocular applications which are rife with troublesome self-occlusions. We propose a novel estimation method for articulated human tracking based on jointly modeling the observed video sequence, the articulated body and the field of background and foreground labels, which aid in the resolution of self-occlusions. We develop a statistical appearance model of the human in a scene and perform graphical inference using nonparametric belief propagation to reliably track a single human in video sequences captured using a single stationary camera.

## 7873-34, Session 4

### Sparse Fisher linear discriminant analysis

H. A. Siddiqui, H. Hwang, Qualcomm Inc. (United States)

No abstract available.

## 7873-13, Session 5

### Characterization of moving dust particles

B. J. Bos, S. R. Antonille, N. Memarsadeghi, NASA Goddard Space Flight Ctr. (United States)

A large depth-of-field Particle Image Velocimeter (PIV) has been developed at NASA GSFC to characterize dynamic dust environments on planetary surfaces. This instrument detects and senses lofted dust particles. To characterize a dynamic planetary dust environment, the instrument would have to operate for at least several minutes during an observation period, easily producing more than a terabyte of data per observation. Given current technology, this amount of data would be very difficult to store onboard a spacecraft and downlink to Earth. We have been developing an autonomous image analysis algorithm architecture for the PIV instrument to greatly reduce the amount of data that it has to store and downlink. The algorithm analyzes PIV images and reduces the image information down to only the particle measurement data we are interested in receiving on the ground - typically reducing the amount of data to be handled by more than two orders of magnitude.

We give a general description of the PIV algorithms and describe only the algorithm for estimating the direction and velocity of the traveling particles in more detail, which was done by taking advantage of the optical properties of moving dust particles and image processing techniques. Our experiments on simulated particles imply an average absolute error of less than 4 degrees for direction estimation of traveling particles when the blurring filter length was greater than 3 pixels long. Our algorithms performed well for simulated data in presence of small amounts of noise.

## 7873-14, Session 5

**A super-resolution algorithm for enhancement of flash lidar data**

A. Bulyshev, Analytical Mechanics Associates, Inc. (United States); M. D. Vanek, F. Amzajerdian, NASA Langley Research Ctr. (United States); D. F. Pierrotet, Coherent Applications, Inc. (United States); G. D. Hines, R. A. Reisse, NASA Langley Research Ctr. (United States)

A novel method of the enhancement of the spatial resolution of Flash Lidar images in application to the surface map generation is proposed. The ability of the Flash LIDAR to generate 3-dimensional maps of the landing site area during the final stages of the descent phase for detection of hazardous terrain features is under study in the frame of ALHAT project. Major goals of this algorithm are to create a Digital Elevation Map (DEM) covering a sufficiently large area and with acceptable accuracy and precision and to retrieve the relative trajectory. The algorithm is utilizing an iterative scheme which updates DEM related to the high resolution grid and finds the next trajectory point using one Lidar frame at a time. A back projection algorithm is used for the DEM creation and 6-d generalization of Lucas-Kanade method is utilized to retrieve the trajectory of the space craft and Lidar attitude. Performance of the super-resolution algorithm has been analyzed through a series of simulation runs. The results show that achieved level of accuracy and precision in generating an elevation map of the landing site is adequate for detecting hazardous terrain features and identifying safe areas.

## 7873-17, Session 5

**Image registration for stability testing of MEMS**

N. Memarsadeghi, J. Le Moigne, P. N. Blake, NASA Goddard Space Flight Ctr. (United States); P. A. Morey, Ball Aerospace & Technologies Corp. (United States); W. B. Landsman, Adnet Systems Inc. (United States); V. J. Chambers, S. H. Moseley, NASA Goddard Space Flight Ctr. (United States)

Image registration, or alignment of two or more images covering the same scenes or objects, is of great interest in many disciplines such as remote sensing, medical imaging, astronomy, and computer vision. In this paper, we introduce a new application of image registration algorithms. We demonstrate how through a wavelet based image registration algorithm, engineers can evaluate stability of Micro-Electro-Mechanical Systems (MEMS). In particular, we applied image registration algorithms to assess alignment stability of the MicroShutters Subsystem (MSS) of the Near Infrared Spectrograph (NIRSpec) instrument of the James Webb Space Telescope (JWST). This work introduces a new methodology for evaluating stability of MEMS devices to engineers as well as a new application of image registration algorithms to computer scientists.

## 7873-35, Session 5

**Shape-based segmentation of alloy micrographs using matching pursuits**

L. M. Huffman, I. Pollak, Purdue Univ. (United States); J. P. Simmons, Air Force Research Lab. (United States); M. De Graef, Carnegie Mellon Univ. (United States)

Computerized analysis of alloy micrographs is expected to revolutionize materials development by replacing laborious physical testing with computer simulations. These simulations utilize segmentations of alloy micrographs which indicate the arrangement of material precipitates.

Abundant prior information is typically available regarding the shape of the precipitates, and any viable segmentation method must properly account for such information. We propose a novel application of

matching pursuits to construct a micrograph segmentation consisting of rectangles which match the variably sized and oriented rectangular material precipitates.

## 7873-18, Session 6

**Capacitive touch sensing: signal and image processing algorithms**

Z. I. Baharav, Corning Inc. (United States); R. Kakarala, Nanyang Technological Univ. (Singapore)

In this work we will analyze the most common method of projected capacitive sensing, that of absolute capacitive sensing, together with the most common sensing pattern, that of diamond-shaped sensors.

We formulate the problem as a reconstruction from projections, and consider several aspects. The first aspect is that of usage scenarios, which give rise to issues like working with small and big fingers, proximity detection, stability of interpolated location, linearity of line drawing, and alike. The second aspect relates to manufacturing variations, and include things like variation in the thickness of cover glass (or PET), uniformity of ITO layer, minimum feature sizes on ITO, background capacitance, etc. The last aspect we consider relates to noise sources, like LCD noise, electronic sampling noise, RF noise coupled in, finger-coupled noise, and many more. These factors (usage mode, physical variations, and noise) guide the evaluation criteria for various algorithms used, in addition to the usual constraints of computation power and memory. Algorithms considered include simple linear interpolation (global and local), curve fitting (bilinear interpolation), filtering, general look-up-table, and combinations thereof. We discuss the merits of using separable algorithms (as are most commonly used today), and compare them to optimal ones.

## 7873-19, Session 6

**Denosing, deblurring, and super-resolution in mobile phones**

F. Sroubek, J. Kamenicky, J. Flusser, Institute of Information Theory and Automation (Czech Republic)

Current mobile phones are equipped with low-budget digital cameras and very poor optics. Consequently, images acquired by these cameras have effective resolution lower than the number of pixels and contain a considerable amount of noise and/or blur if light conditions are poor. We propose a novel algorithm which takes a set of acquired images from such cameras and performs simultaneously three tasks: denosing, deblurring and resolution enhancement. The amount of each depends on the characteristics of the input set. The algorithm implements a single framework, which we formulate as an energy minimization problem, where the energy function comprises three terms. The first one is a data term that models the acquisition process. The second and third are regularization terms that act as image and blur priors, respectively. Since we work with more than one image, a critical preprocessing step is accurate image registration. For this purpose, we propose to use a method based on optical flow, which provides sub-pixel accuracy. We demonstrate the performance of the proposed method on a system with a camera in a mobile phone (or web camera) and a PC.

## 7873-20, Session 6

**Arabic word recognizer for mobile applications**

N. Khanna, G. Abdollahian, B. Brame, M. Boutin, E. J. Delp III, Purdue Univ. (United States)

When traveling in a region where the local language is not written using the Roman alphabet, translating written text (e.g., documents, road signs, or placards) is a particularly difficult problem since the text

cannot be easily entered into a translation device or searched using a dictionary. To address this problem, we are developing the "Rosetta Phone," a handheld device (e.g., PDA or mobile telephone) capable of acquiring a picture of the text, locating the region (word) of interest within the image, and producing both an audible and a visual English interpretation of the text. The present system is targeted for interpreting words written in Arabic character set. We have tested our autonomous, segmentation-free phrase recognizer on a dictionary of five thousand Pashto words. Present system offers close to perfect uniqueness and more than 75% recognition accuracy on different noisy versions of the words in this dictionary. Given the limited image quality and resolution for iPhone images, recognition accuracy is also a function of font sizes as demonstrated by experiments on documents scanned at different resolutions (150DPI to 300DPI).

### 7873-21, Session 6

#### **Volume estimation using food specific shape templates in mobile image-based dietary assessment**

J. Chae, I. Woo, S. Kim, R. Maciejewski, F. Zhu, E. J. Delp III, C. J. Boushey, D. S. Ebert, Purdue Univ. (United States)

Given the mounting concerns of childhood and adult obesity, methods are being developed for prevention and intervention. Key components in these methods include the recording, cataloging and analysis of daily dietary records to monitor energy and nutrient intakes. Given the ubiquity of mobile devices with built-in cameras, one possible means of improving dietary assessment is through photographing food items and inputting these images into a system that can determine the nutrient content of the food in the images. A key problem with such image-based dietary assessment tools is the accurate and consistent estimation of food portion size. We propose a method to automatically estimate food volumes through the use of food specific shape templates. We are able to reconstruct properties of the 3D scene from a single image. Each classified food item by image segmentation methods corresponds to a particular food template shape. For volume computation, we use either a feature point extraction algorithm or the active contour methodology to size our shape templates accordingly. By applying this template-based approach, we are able to automatically estimate food portion size. This leads to reduced burden on users having to estimate portions consumed and provides a consistent method for estimating food volume.

### 7873-37, Session 8

#### **Spectral x-ray CT imaging using energy sensitive photon counting detectors**

K. Taguchi, The Johns Hopkins Outpatient Ctr. (United States)

No abstract available.

### 7873-38, Session 8

#### **Toward material characterization using dual energy x-ray CT**

J. A. O'Sullivan, B. R. Whiting, D. G. Politte, Washington Univ. in St. Louis (United States); J. F. Williamson, Virginia Commonwealth Univ. (United States)

No abstract available

### 7873-39, Session 8

#### **A hybrid approach to imaging and anomaly characterization from dual energy CT data**

E. L. Miller, O. Semerici, Tufts Univ. (United States)

No abstract available

### 7873-40, Session 8

#### **Robust multifrequency inversion in terahertz diffraction tomography**

D. A. Castañón, K. A. Chen, Boston Univ. (United States)

Multi-frequency terahertz imaging has received much attention in recent years due to its ability to observe unique spectral characteristics of chemicals, which can be used in numerous applications such as explosives detection. Short-pulse THz sources can provide broadband excitation, but current approaches for image formation based on diffraction tomography construct images independently for each frequency. This results in a lack of resolution at lower frequencies, and lower signal-to-noise reconstructions. In this paper, we explore different techniques for joint image formation using multiple frequencies for enhanced detection. Among these are techniques that use prior information on spectral characteristics of materials of interest to coherently combine information from multiple frequencies, as well as robust techniques that assume incomplete or inaccurate prior knowledge of spectral signatures. We explore the relative performance of these techniques on image reconstruction and object recognition tasks using numerical simulations.

### 7873-41, Session 8

#### **Classification-aware dimensionality reduction methods for explosives detection using multi-energy x-ray computed tomography**

W. C. Karl, P. Ishwar, L. Eger, Boston Univ. (United States)

Multi-Energy X-ray Computed Tomography (MECT) is a non-destructive scanning technology in which multiple energy-selective measurements of the X-ray attenuation can be obtained. This provides more information about the chemical composition of the scanned materials than single-energy technologies and potential for more reliable detection of explosives. We study the problem of discriminating between explosives and non-explosives using low-dimensional features extracted from the high-dimensional attenuation versus energy curves of materials. We study various classification-aware dimensionality reduction methods and demonstrate that the detection performance can be significantly improved by using more than two features and when using features different than the standard photoelectric and Compton coefficients. This suggests the potential for improved detection performance relative to conventional dual-energy X-ray systems.

### 7873-42, Session 8

#### **Robustness of spectral CT for explosives detection**

S. Basu, Morpho Detection Inc. (United States)

No abstract available

## 7873-03, Session 9

### Bayesian estimation with Gauss-Markov-Potts priors in optical diffraction tomography

H. Ayasso, B. Duchêne, A. Mohammad-Djafari, Lab. des Signaux et Systèmes (France)

In this paper, we consider the Optical Diffraction Tomography (ODT) as an inverse scattering problem and propose to use the Bayesian estimation framework for the image reconstructing problem.

A Gauss-Markov-Potts prior translates appropriately the a priori knowledge that the object under test is composed of compact homogeneous regions made of a finite number of homogeneous materials. We propose two such models and use them for proposing two image reconstruction algorithms based on the MCMC sampling schemes. Some preliminary results, obtained by applying the inversion algorithm to experimental laboratory controlled data, will illustrate the performances of the proposed method.

## 7873-43, Session 9

### Constrain static target kinetic iterative image reconstruction for 4D cardiac CT imaging

A. M. Alessio, Univ. of Washington Medical Ctr. (United States); P. J. La Rivière, The Univ. of Chicago Medical Ctr. (United States)

Iterative image reconstruction offers improved signal to noise properties for CT imaging. The primary challenge with iterative methods is the substantial computation time. This computation time is even more prohibitive in 4D imaging applications, such as cardiac gated or dynamic acquisition sequences. In this work, we propose only updating the time-varying elements of a 4D image sequence while constraining the static elements to be fixed or slowly varying in time. We test the method with simulations of 4D acquisitions based on measured cardiac patient data from a) a retrospective cardiac-gated CT acquisition and b) a dynamic perfusion CT acquisition. We target the kinetic elements with one of two methods: 1) position a circular ROI on the heart, assuming area outside ROI is essentially static throughout imaging time; and 2) select varying elements from coefficient of variation image formed from fast analytic reconstruction of all time frames. Targeted kinetic elements are updated with each iteration, while static elements remain fixed at initial image values formed from reconstruction of data from all time frames. Results confirm that the computation time is proportional to the number of targeted elements; our simulations suggest that <25% of elements need to be updated in each frame leading to >4x reductions in reconstruction time. The images reconstructed with the proposed method have matched mean square error with full 4D reconstruction. The proposed method is amenable to most optimization algorithms and offers the potential for significant computation improvements, which could be traded off for more sophisticated system models or penalty terms.

## 7873-44, Session 9

### Model based motion artifact reduction for computed tomography

Z. Yu, J. Thibault, GE Healthcare (United States); J. Wang, K. D. Sauer, Univ. of Notre Dame (United States); C. A. Bouman, Purdue Univ. (United States)

Model based iterative reconstruction (MBIR) algorithms have recently been applied to computed tomography and demonstrated superior image quality. Typical reconstruction algorithms assume the voxels are constant over time. This assumption is not true when patient motion is present, which, in turn, results in motion artifacts. In this paper, we present a method that models the voxel values as a function of time in the MBIR framework. Our results on phantom study and clinical data

show that the proposed method can significantly reduce the motion artifacts in the reconstruction.

## 7873-24, Session 10

### An expectation maximization solution for fusing 2D and 3D ladar data

P. F. Dolce, S. C. Cain, Air Force Institute of Technology (United States)

FLASH 3-D LADAR (LAsER Detection and Ranging) systems represent an important advancement in imaging technology in that they allow an entire scene to be captured simultaneously as opposed to scanning systems. 3-D FLASH systems suffer from spatial resolution problems due to pixel pitch fabrication limitations. A 2-D system can produce high spatial resolution images without range data. One method for obtaining better spatial and range resolution from 3-D LADAR systems is to interpolate the images through various techniques. Interpolation may introduce errors due to aliasing effects. This paper proposes an expectation maximization (EM) solution that corrects these problems through fusing 2-D and 3-D LADAR data. The EM solution is shown to produce 3-D images with improved resolution over those produced with standard interpolation techniques. The combination of 2-D high spatial resolution images and 3-D FLASH LADAR images produces a new LADAR system with improved resolution over current realizable FLASH 3-D sensors. This work proves improvement using a simple simulated target and assumes a known point spread function (PSF) for the optics and atmosphere. Using the proposed algorithm on real 3-D LADAR data would further the evidence that the algorithm works with all data.

## 7873-25, Session 10

### Superresolution with the focused plenoptic camera

A. Lumsdaine, G. N. Chunev, Indiana Univ. (United States); T. G. Georgiev, Adobe Systems Inc. (United States)

This work is based on the focused plenoptic camera, which differs from the traditional plenoptic camera by capturing an array of micro images focused on the object. It has been shown that the focused plenoptic camera data can be used to generate final rendered images of much higher resolution than the plenoptic camera. In this paper we show that this very fact makes it possible to use the camera data with super-resolution techniques, which enables even higher spatial resolution to be obtained from the focused plenoptic camera. We derive the conditions under which the focused plenoptic camera can capture radiance data suitable for super-resolution and develop an algorithm for super resolving those images. Experimental results are presented that show a 9x increase in spatial resolution compared to the basic focused plenoptic rendering approach.

## 7873-26, Session 10

### Content-preserving zoom-in view generation for surveillance videos

K. Watanabe, N. Nitta, N. Babaguchi, Osaka Univ. (Japan)

In order to generate a zoom-in view of certain regions of interest (ROIs), most existing methods such as full-zoom and fisheye view discard or distort the remaining regions of the input frame without considering their content. In this paper, we propose a method for generating a content-preserving zoom-in view which provides magnified ROIs and at the same time preserves the content of the remaining regions. Targeting on stationary surveillance videos, our method firstly extracts moving objects from every input frame as ROIs. Then, the importance score is calculated for each pixel in the input frame based on its content in order to determine where the deformation, which may cause the destruction of the content, should be avoided. Finally, a mapping

problem from the input frame to the zoom-in view with respect to the importance scores is formulated to deform less important regions more than the important ones. A content-preserving zoom-in view that allows viewers to see both the magnified ROIs and an overview of the whole area under monitoring in a single view is generated by solving a mapping problem by optimization. Experiments are conducted to study the effectiveness of considering the content importance by comparing the results when changing the importance scores of the remaining regions.

### 7873-45, Session 10

#### Accelerating sparse reconstruction for fast and precomputable system matrix inverses

S. J. Reeves, Auburn Univ. (United States)

Signal reconstruction using an  $l_1$ -norm penalty has proven to be valuable in edge-preserving regularization as well as in sparse reconstruction problems. The developing field of compressed sensing typically exploits this approach to yield sparse solutions in the face of incoherent measurements. Unfortunately, sparse reconstruction generally requires significantly more computation because of the nonlinear nature of the problem and because the most common solutions damage any structure that may otherwise exist in the system matrix. In this work we adopt a majorizing function for the absolute value term that can be used with structured system matrices so that the regularization term in the matrix to be inverted does not destroy the structure of the original matrix. As a result, a system inverse can be precomputed and applied efficiently at each iteration to speed the estimation process. We demonstrate that this method can yield significant computational advantages when the original system matrix can be represented or decomposed into an efficiently applied singular value decomposition.

### 7873-27, Poster Session

#### Color image compression by gray-to-color mapping

M. S. Drew, Simon Fraser Univ. (Canada); G. D. Finlayson, Univ. of East Anglia Norwich (United Kingdom); A. Jindal, Indian Institute of Technology Kanpur (India)

Instead of de-correlating image luminance from chrominance, some use has been made of the correlation between the luminance component of an image and its chromatic components, or the correlation of one color component.

In one approach, Green was taken as a base, and the other color channels or their DCT subbands were approximated as polynomial functions of the base inside image windows.

This paper points out that we can do better if we introduce an addressing scheme into the image description such that similar colors are grouped together spatially. With a Luminance component base, we test several color spaces and rearrangement schemes, including segmentation, and settle on a log-geometric-mean color space. Along with PSNR versus bits-per-pixel, we found that spatially-keyed s-CIELAB color error better identifies problem regions. Instead of segmentation, we found that rearranging on sorted chromatic components has almost equal performance.

Here, we sort on each of the chromatic components and separately encode windows of each.

The result consists of the original grayscale plane plus the polynomial coefficients of windows of rearranged chromatic values, which are then quantized. The simplicity of the method produces a fast and simple scheme for color image and video compression, with excellent results.

### 7873-28, Poster Session

#### Study of recognizing human motion observed from an arbitrary viewpoint based on decomposition of a tensor containing multiple view motions

T. Hori, J. Ohya, Waseda Univ. (Japan); J. Kurumisawa, Chiba Univ. of Commerce (Japan)

This paper proposes a method of human motion recognition based on Tensor Decomposition using a single-camera video sequence and a multiple viewpoint image database. The video sequence is of a person performing a motion. The viewpoint image database consists of three viewpoints of a human model image: front, side, and 45 degree. The aim of this paper is to accurately classify the person's motion from the database of multiple viewpoint images, using a computer vision based approach. The study of motion in image sequences is a common research topic in computer vision. Motion is a powerful feature of image sequences, revealing situational information not available with still pictures by relating spatial image features to temporal changes. The task of motion analysis, in particular human motion recognition, remains a challenging and fundamental problem of computer vision. The video sequence, viewpoints database, and possible action classes that the motion can be classified as, are mapped into tensor space. The multi-linear analysis of the tensor space (tensor decomposition) described in this paper subsumes as special cases the simple, linear (1-factor) analysis associated with conventional SVD (singular value decomposition), as well as the incrementally more general bilinear (2-factor) analysis that has recently been investigated in the context of computer vision. In this paper, we explore the effectiveness of the above-mentioned recognition method. Our method was compared with the Nearest Neighbor method, which is a very common recognition algorithm, and achieved higher accuracy.

### 7873-29, Poster Session

#### Visual real-time detection, recognition, and tracking of ground and airborne targets

L. Kovács, C. Benedek, Computer and Automation Research Institute (Hungary)

This paper presents methods and algorithms for real-time target detection, recognition and tracking, both in the case of ground-based objects (surveyed from a moving airborne imaging sensor) and flying targets (observed from a ground-based or vehicle mounted sensor). The methods are highly parallelized and partially implemented on GPU, with the goal of real-time speeds even in the case of multiple target observations. Target segmentations include robust foreground and object extraction steps involving multi-layer Gaussian Mixture Modeling and shape/texture-based object segmentation, view registration based on invariant feature point detection and Hidden Markov Models for ground object recognition and segmentation. Tracking is implemented on GPU for high resolution real-time processing. Recognition of the extracted targets is based on fusing shape, texture and motion information, providing constant estimations for the classification of the observed targets. Recognition evaluation is presented, by using a database of extracted and categorized object shapes, collected from real video sources. Real-time applicability is in focus. The methods use single camera observations, providing a passive and expendable alternative for expensive and/or active sensors. Use cases involve perimeter defense and surveillance situations, where passive detection and observation is a priority (e.g. aerial surveillance of a compound, detection of reconnaissance drones, etc.).

## 7873-30, Poster Session

### Illuminant color estimation by hue categorization based on gray world assumption

H. Kawamura, Nippon Telegraph and Telephone Corp. (Japan); S. Yonemura, J. Ohya, Waseda Univ. (Japan); N. Matsuura, Nippon Telegraph and Telephone Corp. (Japan)

This paper proposes a gray world assumption based method for estimating an illuminant color from an image by hue categorization. The gray world assumption hypothesizes that the average color of all the objects in a scene is gray. However, it is difficult to estimate an illuminant color correctly if the colors of the objects in a scene are dominated by certain colors. To solve this problem, our method roughly categorizes the colors in the image by hue and selects them one by one to decide whether the average of the selected colors can be regarded as an illuminant color or not. We use a surface reflectance set obtained from the object color spectra database and three illuminants, i.e., CIE standard illuminants A and D65, and blackbody radiation with 15000 K of color temperature, as an experiment for estimating an illuminant color. Experiment results show that estimated illuminants are closer to the correct one than that of the conventional one, and the estimation error by our method is within the just noticeable difference in human color perception.

## 7873-31, Poster Session

### Super-resolved refocusing with a plenoptic camera

Z. Zhou, Univ. of Science and Technology of China (China); Y. Yuan, BeiHang Univ. (China); X. Bin, Univ. of Science and Technology of China (China) and Academy of Opto-electronics, Chinese Academy of Sciences (China); L. Qian, Univ. of Science and Technology of China (China)

This paper presents an approach to enhance the resolution of refocused images by super resolution methods. In plenoptic imaging, we demonstrate that the raw sensor image can be divided to a number of low-resolution angular images with sub-pixel shifts between each other. The sub-pixel shift, which defines the super resolving ability, is mathematically derived by considering the plenoptic camera as equivalent camera arrays. We implement simulation to demonstrate the imaging process of a plenoptic camera, as well as that of the equivalent camera arrays. A high-resolution image is then reconstructed using maximum a posteriori (MAP) super resolution algorithms. Without other degradation effects in simulation, the super resolved image achieves a resolution as high as predicted by the proposed model. We also build an experimental setup to acquire light fields. With traditional refocusing methods, the image is rendered at a rather low resolution. In contrast, we implement the super-resolved refocusing methods and recover an image with much more spatial details. To evaluate the performance of the proposed method, we finally compare the reconstructed images using image quality metrics like peak signal to noise ration (PSNR).

## 7873-32, Poster Session

### Plenoptic rendering with interactive performance using GPUs

G. N. Chunev, A. Lumsdaine, Indiana Univ. (United States); T. G. Georgiev, Adobe Systems Inc. (United States)

Processing and rendering of plenoptic camera data requires significant computational power and memory bandwidth. At the same time, interactive rendering performance is highly desirable so that users can explore the infinite variety of images that can be rendered from a single plenoptic image. In this paper we describe a GPU-based approach for lightfield processing and rendering, with which we are able to achieve interactive performance for focused plenoptic rendering tasks such

as refocusing and novel-view generation. We present a progression of rendering approaches for focused plenoptic camera data and analyze their performance on popular GPU-based systems. Our analyses are validated with experimental results on commercially available GPU hardware. Even for complicated rendering algorithms, we are able to render 39Mpixel plenoptic data to 2Mpixel images with frame rates in excess of 500 frames per second.

## 7873-33, Poster Session

### Compressive through-focus wavefield imaging

E. A. Marengo, O. Mangoubi, Northeastern Univ. (United States)

Optical sensing and imaging applications often suffer from a combination of low resolution object reconstructions and a large number of sensors which, depending on frequency, can be quite expensive or bulky. It is desirable to minimize the number of sensors (which reduces cost) for a given target resolution level (image quality) and permissible total sensor array size (compactness). Equivalently, for a given imaging hardware one seeks to maximize image quality, which in turn means fully exploiting the available sensors as well as all priors about the properties of the sought-after objects such as sparsity properties, and other, which can be incorporated into reconstruction schemes. This paper proposes a compressive-sensing-based method to process through-focus optical field data captured at a sensor array. The proposed approach treats in-focus and out-of-focus data as projective measurements for compressive sensing, and assumes that the objects are sparse under known transformations applied to them. The proposed compressive through-focus imaging is illustrated for both coherent and incoherent light. The results illustrate the combined use of through-focus imaging and compressive sensing techniques, and provide insight on the information in in-focus and out-of-focus data.



# Conference 7874: Document Recognition and Retrieval XVIII

Wednesday-Thursday 26-27 January 2011

Part of Proceedings of SPIE Vol. 7874 Document Recognition and Retrieval XVIII

## 7874-12, Poster Session

### Improved document image segmentation algorithm using multiresolution morphology

S. S. Bukhari, Technische Univ. Kaiserslautern (Germany); F. Shafait, DFKI GmbH (Germany); T. M. Breuel, Technische Univ. Kaiserslautern (Germany)

Page segmentation into text and non-text components is an essential preprocessing step before OCR operation. If this is not done properly, an OCR classification engine produces garbage text due to the presence of non-text components. This paper describes improvements to the text/image segmentation algorithm described by Bloomberg[1], which is also available in his open source Leptonica library[2]. The modifications result in significant improvements over Bloomberg's algorithm on UW-III, UNLV, ICDAR 2009 page segmentation competition test images and circuit diagram datasets.

## 7874-28, Poster Session

### A simple and effective figure caption detection system for old-style documents

Z. Liu, H. Zhou, Amazon.com, Inc. (United States)

Identifying figure captions has wide applications in producing high quality e-books such as kindle books or ipad books. In this paper, we present a rule-based system to detect horizontal figure captions in old-style documents. Our algorithm consists of three steps: (i) segment images into regions of different types such as texts and figures, (ii) search the best caption region candidate based on heuristic rules such as region alignments and distances, and (iii) expand caption regions identified in step (ii) with its neighboring text-regions in order to correct over-segmentation errors.

We test our algorithm using 81 images collected from old-style books, with each image containing at least one figure area. We show that the approach is able to correctly detect figure captions from images with different layouts, and we also measure its performances in terms of both precision rate and recall rate.

## 7874-29, Poster Session

### Reflowing-driven paragraph recognition for electronic books in PDF

J. Fang, Z. Tang, L. Gao, Peking Univ. (China)

When reading electronic books on handheld devices, content sometimes should be reflowed and recomposed to adapt for small-screen mobile devices. According to people's reading practice, it is reasonable to reflow the text content based on paragraphs. Hence, this paper addresses the requirement and proposes a set of novel methods on paragraph recognition for electronic books in PDF. The proposed methods consist of three steps, namely, physical structure analysis, paragraph segmentation, and reading order detection. We make use of locally ordered property of PDF documents and layout style of books to improve traditional page recognition results. In addition, we employ the optimal matching of Bipartite Graph technology to detect paragraphs' reading order. Experiments show that our methods achieve high accuracy. It is noteworthy that, the research has been applied in a commercial software package for Chinese E-book production.

## 7874-30, Poster Session

### Ruling line detection and removal

E. Kavallieratou, Univ. of the Aegean (Greece); D. P. Lopresti, J. Chen, Lehigh Univ. (United States)

In this paper we present a procedure for removing ruling lines from a handwritten document image that does not require any preprocessing or postprocessing tasks and it does not break existing characters. We take advantage of common ruling line properties such as uniform width, predictable spacing, position vs. text, etc. The proposed procedure can also detect the existence of ruling lines in page, so it has no effect on document images without ruling lines. The system is evaluated on synthetic page images in five different languages and is compared to a previous methodology.

## 7874-31, Poster Session

### Natural scene logo recognition by joint boosting feature selection in salient regions

W. Fan, J. Sun, S. Naoi, Fujitsu Research and Development Center Co., Ltd. (China); A. Minagawa, Y. Hotta, Fujitsu Labs., Ltd. (Japan)

Logos are considered valuable intellectual properties and a key component of the goodwill of a business. In this paper, we propose a natural scene logo recognition method which is segmentation-free and capable of processing images extremely rapidly and achieving high recognition rates. The classifiers for each logo are trained jointly, rather than independently. In this way, common features can be shared across multiple classes for better generalization. To deal with large range of aspect ratio of different logos, a set of salient regions of interest (ROI) are extracted to describe each class. We ensure the selected ROIs to be both individually informative and two-by-two weakly dependant by a Class Conditional Entropy Maximization criteria. Experimental results on a large logo database demonstrate the effectiveness and efficiency of our proposed method.

## 7874-32, Poster Session

### A framework to improve digital corpus uses: image-mode navigation

L. Eynard, V. Malleron, H. Emptoz, Univ. Claude Bernard Lyon 1 (France)

In this paper, we propose a new system to enhance navigation inside digital corporuses.

This system is based on an automatic indexation in image mode and provides the user intuitive navigation in interactive time.

Keywords and containers are extracted directly from the document images to create an Image Mode Index, which shows the keywords as cut-out images of their actual appearances.

Our approach recreates a summary of the structured documents, following indications given by the creators of the document themselves.

Our system is detailed in the general case and sample applications on a 19th century handwritten corpus and a 18th century machine printed text corpus are provided.

This approach, developed for documents inaccessible otherwise, can be applied on any corpus where keywords and containers can be identified.

### 7874-33, Poster Session

#### **Parameter calibration for synthesizing realistic-looking variability in offline handwriting**

W. Cheng, D. P. Lopresti, Lehigh Univ. (United States)

Being motivated by the widely accepted principle that the more training data, the better performance the recognition system has, we conducted experiments asking human subjects to do test on a mixture of real English handwritten textlines and textlines altered from existing handwriting with various distortion degrees. The idea of generating synthetic handwriting is based on a perturbation method by T. Varga and H. Bunke that distorts an entire textline. There are two purposes of our experiments. First, we want to calibrate optimal distortion parameter settings for Varga and Bunke's perturbation model. Second, we intend to compare the effects of parameter settings on different writing styles, block, cursive and mixed. From the preliminary experimental results, we determined appropriate ranges for parameter amplitude, and found that parameter settings should change for different handwriting styles. Once the proper parameter settings are found, we will generate large amount of training and testing sets for building better off-line handwriting recognition systems.

### 7874-34, Poster Session

#### **Automatic segmentation of subfigure image panels for multimodal biomedical document retrieval**

B. Cheng, Missouri Univ. of Science and Technology (United States); S. K. Antani, National Library of Medicine (United States); R. J. Stanley, Missouri Univ. of Science and Technology (United States); D. Demner-Fushman, G. R. Thoma, National Library of Medicine (United States)

No abstract available.

### 7874-35, Poster Session

#### **A new method for perspective correction of document images**

J. Rodríguez-Pinheiro, P. Comesaña-Alfaro, F. Pérez-González, Univ. de Vigo (Spain); A. Malvido-García, Bit Oceans Research, S.L. (Spain)

In this paper we propose a method for perspective distortion correction of rectangular documents. This scheme exploits the orthogonality of the document edges, allowing to recover the aspect ratio of the original document. The results obtained after correcting the perspective of several document images captured with a mobile phone are compared with those achieved by digitizing the same documents with several scanner models.

### 7874-36, Poster Session

#### **Robust keyword retrieval method for OCR'd text**

Y. Fujii, H. Takebe, H. Tanaka, Y. Hotta, Fujitsu Labs., Ltd. (Japan)

Document management systems have become important because of the growing popularity of electronic filing of documents and scanning of books, magazines, manuals, etc., through a scanner or a digital camera, for storage or reading on a PC or an electronic book. Text information acquired by optical character recognition (OCR) is usually added to the electronic documents for document retrieval. Since texts generated by OCR generally include character recognition errors,

robust retrieval methods have been introduced to overcome this problem. In this paper, we propose a retrieval method that is robust against both character segmentation and recognition errors. In the proposed method, the insertion of noise characters and dropping of characters in the keyword retrieval enables robustness against character segmentation errors, and character substitution in the keyword of the recognition candidate for each character in OCR or any other character enables robustness against character recognition errors. The recall rate of the proposed method was 15% higher than that of the conventional method. However, the precision rate was 64% lower.

### 7874-37, Poster Session

#### **Online medical symbol recognition using a tablet PC**

A. Kundu, Q. Hu, S. Boykin, R. Fish, C. Clark, S. Jones, S. Moore, MITRE Corp. (United States)

In this paper we describe a scheme to enhance the usability of a Tablet PC's handwriting recognition system by including symbols that are not a part of the Tablet PCs symbol library. The goal of this work is to make handwriting recognition more useful for medical professionals accustomed to using medical symbols in medical records. The fact that medical abbreviations look similar to symbols makes this a difficult task. The paper also describes our effort to create a corpus of medical symbols and non-symbols which could be potentially identified as symbols. Using the data from this corpus, we demonstrate that this new symbol recognition module is robust and expandable as we report good results on both a medical symbol set and an expanded symbol test set which includes selected mathematical symbols. Finally, we have shown that using a multi-classifier architecture provides robust performance.

### 7874-38, Poster Session

#### **Characterizing challenged 2008 Minnesota ballots**

G. Nagy, Rensselaer Polytechnic Institute (United States); D. P. Lopresti, Lehigh Univ. (United States); E. H. Barney Smith, Boise State Univ. (United States); Z. Wu, Rensselaer Polytechnic Institute (United States)

No abstract available.

### 7874-40, Poster Session

#### **Document image retrieval with morphology-based segmentation and features combination**

T. Bockholt, Sr., G. Darmiton, Sr., C. Mello, Sr., Univ. Federal de Pernambuco (Brazil)

Digital libraries need more than just a retrieval based on keywords, which can be inefficient for some applications. Thus, a document retrieval based on content of the digitized image version of the document can be a more appropriated approach. This paper discusses the retrieval of document images by means of identifying a variety of elements present in the document's image body. We propose a new strategy to identify and combine features extracted from a document image. We also consider the task of constructing an optimized feature set to improve the retrieval performance and to validate our experiments on an assorted database. Experimental results show that the proposed segmentation together with a wisely feature combination increase the overall retrieval performance. Moreover the retrieved images demonstrate the generality and effectiveness of our approach for an efficient segmentation and classification of document images.

## 7874-41, Poster Session

### Boosting based text and non-text region classification

B. Xie, G. Agam, Illinois Institute of Technology (United States)

Layout analysis is a crucial process for document image understanding and information retrieval. Document layout analysis depends on page segmentation and block classification. This paper describes an algorithms for to extracting blocks from document images and a boosting based method to classify those blocks as text or not. The feature vector which is feed into the boosting classifier consists of a four direction run-length histogram, and connected components features, both background and foreground. Using a combination of features through a boosting classifier, we obtain accuracy of 99.5% on our test collection.

## 7874-42, Poster Session

### OMR of early plainchant manuscripts in square notation: a two-stage system

C. Ramirez, J. Ohya, Waseda Univ. (Japan)

While Optical Music Recognition (OMR) of modern printed and handwritten documents is considered a solved problem, with many commercial systems available today, the OMR of ancient musical manuscripts still remains an open problem. In this paper we present a system for the OMR of degraded western plainchant manuscripts in square notation from the XIV to XVI centuries. The system has two main blocks, the first one deals with symbol extraction and recognition, while the second one acts as an error detection stage for the first block outputs. For symbol extraction we use widely known image-processing techniques, such as Sobel filtering and Hough Transform, and SVM for classification. The error detection stage is implemented with a hidden Markov model (HMM), which takes advantage of a priori knowledge for this specific kind of music.

## 7874-01, Session 1

### Scientific challenges underlying production document processing

E. Saund, Palo Alto Research Center, Inc. (United States)

The field of Document Recognition is bipolar. On one end lies the excellent work of academic institutions engaging in original research on scientifically interesting topics. On the other end lies the document recognition industry which services needs for high-volume data capture for transaction and back-office applications. These realms seem to seldom meet, yet the need is great to address technical hurdles for practical problems using modern approaches from the Document Recognition, Computer Vision, and Machine Learning communities.

This talk will reflect on three categories of problems we have encountered which are both scientifically challenging and of high practical value. These are Doctype Classification, Functional Role Labeling, and Document Sets. Doctype Classification asks, "What is this page I'm looking at?" Functional Role Labeling asks, "What is the status of text and graphical elements in a model of document structure?" Document Sets asks, "How are pages and their contents related to one another?" Each of these has ad hoc engineering approaches that provide 40-80% solutions, and each of them begs for a deeply grounded formulation both to provide understanding and to support capture of the remaining 20-60% of practical value. The practical need is not purely technical but also revolves around User Experience and therefore, the art of Design.

## 7874-02, Session 2

### Automated identification of biomedical article type using support vector machines

I. Kim, National Institutes of Health (United States); D. X. Le, G. R. Thoma, National Library of Medicine (United States)

Authors of short papers such as letters or editorials often express complementary opinions, and sometimes contradictory ones, on related work in previously published articles. The MEDLINE® citations for such short papers are required to list bibliographic data on these "commented on" articles in a "CON" field. The challenge is to automatically identify the CON articles referred to by the author of the short paper (called "Comment-in" or CIN paper). Our approach is to use support vector machines (SVM) to first classify a paper as either a CIN or a regular full-length article (which is exempt from this requirement), and then to extract from the CIN paper the bibliographic data of the CON articles. A solution to the first part of the problem, identifying CIN articles, is addressed here. We implement and compare the performance of two types of SVM, one with a linear kernel function and the other with a radial basis kernel function (RBF). Input feature vectors for the SVMs are created by combining four types of features based on statistics of words in the article title, words that suggest the article type (letter, correspondence, editorial), size of body text, and cue phrases. Experiments conducted on a set of online biomedical articles show that the SVM with a linear kernel function yields a significantly lower false negative error rate than the one with an RBF. Our experiments also show that the SVM with a linear kernel function achieves a significantly higher level of accuracy, and lower false positive and false negative error rates by using input feature vectors created by combining all four types of features rather than any single type.

## 7874-03, Session 2

### Introduction of statistical information in a syntactic analyzer for document image recognition

A. Oliveira Maroneze, B. Couasnon, Institut National des Sciences Appliquées de Rennes (France); A. Lemaitre, Institut de Recherche en Informatique et Systèmes Aléatoires (France)

This paper presents an improvement to a document layout analysis system, offering a possible solution to Sayre's paradox ("a letter must be recognized before it can be segmented; and it must be segmented before it can be recognized"). This improvement, based on stochastic parsing, allows integration of statistical information, obtained from recognizers, during syntactic layout analysis. We present how this fusion of numeric and symbolic information in a feedback loop can be applied to syntactic methods to simplify document description. To limit combinatorial explosion during exploration of solutions, we devised an operator that allows optional activation of the stochastic parsing mechanism. Our evaluation on 1250 handwritten business letters shows this method allows the improvement of global recognition scores.

## 7874-04, Session 2

### High recall document content extraction

C. An, Lehigh Univ. (United States)

We report methodologies for computing high-recall masks for document image content extraction, that is, the location and segmentation of regions containing handwriting, machine-printed text, photographs, blank space, etc. The resulting segmentation is pixel-accurate, which accommodates arbitrary zone shapes (not merely rectangles). We describe experiments showing that iterated classifiers can increase recall of all content types, with little loss of precision. We also introduce two methodological enhancements: (1) a multi-stage voting rule; and (2) a scoring policy that views blank pixels as a "don't

care" classes with other content classes. These enhancements improve both recall and precision, achieving at least 89% recall and at least 87% precision among three content types: machine-print, handwriting, and photo.

### 7874-05, Session 2

#### **Shape codebook based handwritten and machine printed text zone extraction**

J. Kumar, Univ. of Maryland, College Park (United States); R. Prasad, H. Cao, BBN Technologies (United States); W. Abd-Almageed, D. S. Doermann, Univ. of Maryland, College Park (United States); P. S. Natarajan, BBN Technologies (United States)

We present a novel method for extracting handwritten and printed text zones from noisy document images with mixed content. We use Triple-Adjacent-Segment (TAS) based features which encode local shape characteristics of text in a consistent manner. We first construct two different codebooks of the shape features extracted from a set of handwritten and printed text documents. In the next step, we compute the normalized histogram of codewords for each segmented zone and use it to train Support Vector Machine (SVM) classifier. Due to a codebook based approach, our method is robust to the background noise present in the image. The TAS features used are invariant to translation, scale and rotation of text. In our experimental results, we show that a pixel-weighted zone classification accuracy of 98% can be achieved for noisy Arabic documents.

Further, we demonstrate the effectiveness of our method in document page classification and show that a high precision can be achieved for machine printed documents.

The proposed method is robust to the size of zones, which may contain text content at word, line or paragraph level.

### 7874-06, Session 3

#### **A MRF model with parameters optimization by CRF for on-line recognition of handwritten Japanese characters**

B. Zhu, M. Nakagawa, Tokyo Univ. of Agriculture and Technology (Japan)

This paper describes a Markov random field (MRF) model with weighting parameters optimized by conditional random field (CRF) for on-line recognition of handwritten Japanese characters. It extracts feature points along the pen-tip trace from pen-down to pen-up, and then sets each feature point from an input pattern as a site and each state from a character class as a label. It employs the coordinates of feature points as unary features and the differences of the coordinates between the neighboring feature points as binary features. The weighting parameters are estimated by CRF or the minimum classification error (MCE) method. In experiments using the TUAT Kuchibue database, the proposed method achieves the character recognition rate of 92.77%, which is higher than the previous model, and the method estimating the weighting parameters by CRF brings higher recognition accuracy than MCE.

### 7874-07, Session 3

#### **Improving a HMM-based off-line handwriting recognition system using MME-PSO optimization**

M. Hamdani, Ecole Nationale d'Ingénieurs de Sfax (Tunisia); H. El Abed, Technische Univ. Braunschweig (Germany); T. M. Hamdani, Ecole Nationale d'Ingénieurs de Sfax (Tunisia); V. Märgner, Technische Univ. Braunschweig (Germany); A. M. Alimi,

Ecole Nationale d'Ingénieurs de Sfax (Tunisia)

One of the trivial steps in the development of a classifier is the design of its architecture. This paper presents a new algorithm, Multi Models Evolvement (MME) using Particle Swarm Optimization (PSO). This algorithm is a modified version of the basic PSO, which is used to the unsupervised design of Hidden Markov Model (HMM) based architectures. For instance, the proposed algorithm is applied to an Arabic handwriting recognizer based on discrete probability HMMs. After the optimization of their architectures, HMMs are trained with the Baum-Welch algorithm. The validation of the system is based on the IfN/ENIT database. The performance of the developed approach is compared to the participating systems at the 2005 competition organized on Arabic handwriting recognition on the International Conference on Document Analysis and Recognition (ICDAR). An absolute improvement of 6% of word recognition rate with about 81% is presented. The proposed system outperforms also most of the known state-of-the-art systems.

### 7874-08, Session 3

#### **SemiBoost-based Arabic character recognition method**

B. Su, L. Peng, X. Ding, Tsinghua Univ. (China)

Traditionally, supervised learning method is adopted to build character recognition system. However, the distribution of training labeled samples and the distribution of practical samples may not match, due to the variation of character image quality. To solve this problem, it would be feasible to incorporate unlabeled practical samples in the training stage. In this paper, we present a Semiboost-based Arabic character recognition method. SVM is adopted as the base classifier. At each iteration, unlabeled examples are selected and assigned labels. The selected samples are used along with the original labeled samples to train a new SVM classifier. An empirical study on several similar character pairs with different similarities shows that the proposed method improves the performance when unlabeled samples reveal the underlying structure of data.

### 7874-09, Session 3

#### **First experiments on a new online handwritten flowchart database**

A. M. Awal, Univ. de Nantes (France); G. Feng, Nanjing Univ. (China); H. Mouchère, C. Viard-Gaudin, Univ. de Nantes (France)

We propose in this paper a new online handwritten flowchart database and perform some first experiments to have a baseline benchmark on this dataset. The collected database consists of 78 flowcharts labeled at the stroke and symbol levels. In addition, an isolated database of graphical and text symbols was extracted from these collected flowcharts. Then, we tackle the problem of online handwritten flowchart recognition from two different points of view. Firstly, we consider that flowcharts are correctly segmented, and we propose different classifiers to perform two tasks, text/non-text separation and graphical symbol recognition. Tested with the extracted isolated test database, we achieve up to 99% and 96% in text/non-text separation and up to 81.3% in graphical symbols recognition. Secondly, we propose a global approach to perform flowchart segmentation and recognition. For this latter, we adopt a global learning schema and a recognition architecture that considers a simultaneous segmentation and recognition. Global architecture is trained and tested directly with flowcharts. Results show the interest of such global approach, but regarding the complexity of flowchart segmentation problem, there is still lot of space to improve the global learning and recognition methods.

#### 7874-10, Session 4

### Segmenting texts from outdoor images taken by mobile phones using color features

Z. Liu, H. Zhou, Amazon.com, Inc. (United States)

Recognizing texts from images taken by mobile phones with low resolution has wide applications. It has shown that a good image binarization can substantially help the accuracy of downstream OCRs.

In this paper, we present a framework to segment texts from outdoor images taken by mobile phones using color features. The framework consists of three steps: (i) the initial process including image enhancement, binarization and noise filtering, where we binarize the input images in each RGB channel, and apply component level noise filtering; (ii) grouping components into blocks using color features, where we compute the component similarities by dynamically adjusting the weights of RGB channels, and merge groups hierarchically, and (iii) blocks selection, where we use the run-length features and choose the Support Vector Machine-(SVM) as the classifier.

We tested the algorithm using 13 outdoor images taken by an old-style LG-64693 mobile phone with 640x480 resolution. We compared the segmentation results with Tsar's algorithm, a state-of-the-art camera text detection algorithm, and show that our algorithm is more robust, particularly in terms of noise removal. In addition, we also evaluated the impacts of our algorithm on the Abbyy's FineReader, one of the most popular commercial OCR engines in the market.

#### 7874-11, Session 4

### A perceptive method for handwritten text segmentation

A. Lemaître, Institut de Recherche en Informatique et Systèmes Aléatoires (France); B. Coüasnon, Institut National des Sciences Appliquées de Rennes (France)

This paper presents a new method to address the problem of handwritten text segmentation into text lines and words. Thus, we propose a method based on the cooperation between points of view that enables to localize the text lines in a low resolution image, and then to associate the pixels at a higher level of resolution. Thanks to the combination of levels of vision, we can detect overlapping characters and re-segment the connected components during the analysis. Then, we propose a segmentation of lines into words based on the cooperation between digital data and symbolic knowledge. The digital data are obtained from distances inside a Delaunay graph, which gives a precise distance between connected components, at the pixel level. Then, we introduce structural rules in order to take into account some generic knowledge about the organization of a text page. This cooperation between information gives a bigger power of expression. We validate this work using the metrics and the database proposed for the segmentation contest of ICDAR 2009. Thus, we show that our method obtains very interesting results, compared to the other methods of the literature. More precisely, we are able to deal with slope and curvature, overlapping text lines and varied kinds of writings, which are the main difficulties met by the other methods.

#### 7874-39, Session 4

### A masked-based enhancement method for historical documents

E. H. Barney Smith, Boise State Univ. (United States); J. Darbon, Ecole Normale Supérieure de Cachan (France); L. Likforman-Sulem, Telecom ParisTech (France)

This paper proposes a novel method for document enhancement. The method is based on the combination of two state-of-the-art filters through the construction of a mask. The mask is applied to a TV (Total Variation) - regularized image where background noise has been reduced. The masked image is then filtered by NLmeans (Non Local

Means) which reduces the noise in the text areas located by the mask. The document images to be enhanced are real historical documents from several periods which include several defects in their background. These defects result from scanning, paper aging and bleed-through. We observe the improvement of this enhancement method through OCR accuracy.

#### 7874-13, Session 5

### Example-centric document design and development

S. R. Klemmer, Stanford Univ. (United States)

Designers often leverage examples when creating new work. Using successful elements from prior ideas can be more efficient than reinventing them from scratch. Moreover, examples can play an inspirational role, helping designers see the space of existing solutions, and illustrating how desirable design effects may be implemented. The more than one trillion pages on the web today provide a corpus of design examples unparalleled in human history. Today, working with existing Web designs require manually manipulating the HTML source. To enable novices and experts alike to more creatively use examples, Bricolage introduces an automatic method for transferring layout and style between Web pages. This transfer is guided by mappings learned using structured prediction methods and other machine learning techniques. In particular, our algorithm learns to identify visually and semantically similar regions between pages by training on a set of human-generated mappings. The end result is an automatic system that enables designers to view their content in the layout and style of any HTML page on the Web.

#### 7874-14, Session 6

### Feature relevance analysis for writer identification

I. Siddiqi, René Descartes Univ. (France) and National Univ. of Sciences and Technology (Pakistan); K. Khurshid, René Descartes Univ. (France) and Institute of Space Technology, Islamabad (Pakistan); N. Vincent, René Descartes Univ. (France)

This work presents an analytical study on the relevance of features in an existing framework for writer identification from offline handwritten document images. The identification system comprises a set of 15 features combining the orientation and curvature information in a writing with the well-known codebook based approach. This study aims to find the optimal feature subset for the task of identifying the author of a questioned document while maintaining acceptable identification rates. Employing a genetic algorithm with a wrapper method we carry out a feature selection mechanism and identify the most relevant features in characterizing the writer of a handwritten text.

#### 7874-15, Session 6

### Using perturbed handwriting to support writer identification in the presence of severe data constraints

J. Chen, W. Cheng, D. P. Lopresti, Lehigh Univ. (United States)

Since real data is time-consuming and expensive to collect, label, and use, researchers have proposed approaches using synthetic variations for the tasks of signature verification, speaker authentication, handwriting recognition, keyword spotting, etc. However, the limitation of real data is particularly critical in the field of writer identification in that in forensics, enemies or criminals usually leave little amount of real data. Therefore, it is unrealistic to always assume sufficient real data for writer identification. In addition, this field differs from many others in that we strive to preserve as much inter-writer variations, but model perturbed handwriting might break such discriminability among writers.

In this work, we started by conducting user studies where human subjects were involved in calibrating realistic-looking transformations. Next, we measured the effects of incorporating perturbed handwriting into the real training dataset. Experimental results justified our hypothesis that with limited real data, model perturbed handwriting improved the performance of writer identification. In addition, we justified by experiments that it was beneficial to search for better performance in the parameter subspaces.

## 7874-16, Session 6

### Statistical characterization of handwriting characteristics using automated tools

G. R. Ball, S. N. Srihari, Univ. at Buffalo (United States)

We provide a statistical basis for reporting the results of handwriting examination by questioned document (QD) examiners. As a facet of Questioned Document (QD) examination, the analysis and reporting of handwriting examination suffers from the lack of statistical data concerning the frequency of occurrence of combinations of particular handwriting characteristics. QD examiners tend to assign probative values to specific handwriting characteristics and their combinations based entirely on the examiner's experience and power of recall. The research uses data bases of handwriting samples that are representative of the US population. Feature lists of characteristics provided by QD examiners, are used to determine as to what frequencies need to be evaluated. Algorithms are used to automatically extract those characteristics, e.g., a software tool for extracting most of the characteristics from the most common letter pair th, is functional. For each letter combination the marginal and conditional frequencies of their characteristics are evaluated. Based on statistical dependencies of the characteristics the probability of any given letter formation is computed. The resulting algorithms are incorporated a system for writer verification known as CEDAR-FOX.

## 7874-17, Session 7

### Keyword and image-based retrieval of mathematical expressions

R. Zanibbi, B. Yuan, Rochester Institute of Technology (United States)

Two new methods for retrieving mathematical expressions using conventional keyword search and expression images are presented. An expression-level TF-IDF (term frequency-inverse document frequency) approach is used for keyword search, where queries and indexed expressions are represented by keywords taken from LaTeX strings. TF-IDF is computed at the level of individual expressions, rather than documents to increase the precision of matching. The second retrieval technique is a form of Content-Based Image Retrieval (CBIR), using a bag of visual words. Expressions are segmented into subregions using the XY-cutting algorithm, after which visual words are generated for each node of the resulting XY-tree. Matching of visual words is based on expression shape (from contour features), aspect ratio, and the histogram of node depths for the sub-tree associated with a node in an XY-tree. Preliminary results using LaTeX documents from the online arXiv repository suggest that the two methods are individually effective, and may be profitably combined.

## 7874-18, Session 7

### Word spotting for handwritten documents using Chamfer distance and dynamic time warping

R. M. Saabni, J. A. El-Sana, Ben-Gurion Univ. of the Negev (Israel)

A large amount of handwritten historical documents are located in

libraries around the world. The desire to access, search, and explore these documents paves the way for a new age of knowledge sharing and promotes collaboration and understanding between human societies. Currently, the indexes for these documents are generated manually, which is very tedious and time consuming. Results produced by state of the art techniques, for converting complete images of handwritten documents into textual representations, are not yet sufficient. Therefore, word-spotting methods have been developed to archive and index images of handwritten documents in order to enable efficient searching within documents. In this paper, we present a new matching algorithm to be used in word-spotting tasks for historical Arabic documents. We present a novel algorithm based on the Chamfer Distance to compute the similarity between shapes of word-parts. Matching results are used to cluster images of Arabic word-parts into different classes using the Nearest Neighbor rule. To compute the distance between two word-part images, the algorithm subdivides each image into equal-sized slices (windows). A modified version of the Chamfer Distance, incorporating geometric gradient features and distance transform data, is used as a similarity distance between the different slices. Finally, the Dynamic Time Warping (DTW) algorithm is used to measure the distance between two images of word-parts. By using the DTW we enabled our system to cluster similar word-parts, even though they are transformed non-linearly due to the nature of handwriting. We tested our implementation of the presented methods using various documents in different writing styles, taken from Juma'a Al Majid Center - Dubai, and obtained encouraging results.

## 7874-19, Session 7

### Automatic identification of ROI in figure images toward improving hybrid (text and image) biomedical document retrieval

D. You, Univ. at Buffalo (United States); S. K. Antani, D. Demner-Fushman, M. M. Rahman, National Library of Medicine (United States); V. Govindaraju, Univ. at Buffalo (United States); G. R. Thoma, National Library of Medicine (United States)

Biomedical images are often referenced for clinical decision support (CDS), educational purposes, and research. They appear in specialized databases or in biomedical publications and are not meaningfully retrievable using primarily text-based retrieval systems. The task of automatically finding the images in an article that are most useful for the purpose of determining relevance to a clinical situation is quite challenging. This task can be done by automatically annotating images extracted from scientific publications with respect to their usefulness for CDS. As an important step toward achieving the goal, we proposed figure image analysis for localizing pointers (arrows, symbols) to extract regions of interest (ROI) that can then be used to obtain meaningful local image content. Content-based image retrieval (CBIR) techniques can then associate local image ROIs with identified biomedical concepts in figure captions for improved hybrid (text and image) retrieval of biomedical articles.

In this work we present methods that make robust our previous Markov random field (MRF)-based approach for pointer recognition and ROI extraction. These include use of Active Shape Models (ASM) to overcome problems in recognizing distorted pointer shapes and a region segmentation method for ROI extraction.

We measure the performance of our methods on two criteria: (i) effectiveness in recognizing pointers in images, and (ii) improved document retrieval through use of extracted ROIs. Preliminary tests on three test sets have shown 87% accuracy in the first criterion. Further, the quality of document retrieval using local visual features and text is shown to be better than using visual features alone. More intensive tests are in progress to evaluate impact of pointer localization and use of ROI in image annotation and retrieval.

7874-20, Session 7

## Automatic extraction of numeric strings in unconstrained handwritten document images

M. M. Haji, T. D. Bui, C. Y. Suen, Concordia Univ. (Canada)

Numeric strings such as identification numbers or dates carry vital pieces of information in documents. Applications concerning the processing of numeric strings can be categorized into two types based on whether the location of the numeral string to be recognized is known or not. In many applications, such as reading back cheques, the location of the numeric strings to be recognized is fixed and known, and the main challenge is how to recognize them. Consequently, plenty of methods have been proposed for recognition of numeric strings. However, in some applications, we have to spot the numeric strings in the first place, as they may appear at arbitrary locations inside the document. This is the case in a document retrieval application where we have a large collection of documents and we wish the user to be able to fetch the one that contains a specific numeric string (which is for example a reference or identification number). There are very few studies that have addressed this aspect of the problem. In this paper, we present a novel algorithm for automatic extraction of numeric strings in unconstrained handwritten document images. The algorithm has two main phases: pruning and verification. In the pruning phase, the algorithm first performs a new segment-merge procedure on each text line, and then using a new regularity measure, it prunes all sequences of characters that are unlikely to be numeric strings. The segment-merge procedure is composed of two modules: a new explicit character segmentation algorithm which is based on analysis of skeletal graphs and a merging algorithm which is based on graph partitioning. All the candidate sequences that pass the pruning phase are sent to a recognition-based verification phase for the final decision. The recognition is based on a coarse-to-fine approach using probabilistic RBF networks. We developed our algorithm for the processing of real-world documents where letter and digits may be connected or broken in a document. We have addressed all steps of a complete system from margin removal and line extraction to evaluation. In order to evaluate the performance of the algorithm, we have created a comprehensive database of handwritten and machine-printed documents including numeric strings of different types, lengths and writing styles. To the best of our knowledge, this is the first work to report a quantitative evaluation of a numeral extraction algorithm on a real-world dataset. The effectiveness of the proposed approach is shown by extensive experiments done on this database which contain over six hundred documents with different types of layouts and levels of noise.

7874-21, Session 8

## Unsupervised method to generate page templates

H. Déjean, Xerox Research Ctr. Europe Grenoble (France)

In this paper, we propose a method for automatically inferring the different page templates used to layout the document content. After the identification of labeled elements through a logical analysis, geometric relations are computed between these labeled elements, and page templates candidates are generated using frequent related elements. A fuzzy matching operation allows for selecting the most frequent and relevant page templates for a given document. Such page templates can be used to correct errors produced during the different previous steps of the document analysis: zoning, OCR, and logical analysis. Evaluation has been performed using the INEX book track collection.

7874-22, Session 8

## Font group identification using reconstructed fonts

M. P. Cutter, J. van Beusekom, Technische Univ. Kaiserslautern (Germany); F. Shafait, DFKI GmbH (Germany); T. M. Breuel, Technische Univ. Kaiserslautern (Germany)

Ideally, digital versions of scanned originals should be represented in a format that is searchable, compressed, highly readable, and faithful to the original. These goals can theoretically be achieved through OCR and font recognition, re-typesetting the document text with original fonts. However, OCR and font recognition remain hard problems, and many historical documents use fonts that are not available in digital forms. It is desirable to be able to reconstruct fonts with vector glyphs that approximate the shapes of the letters that form a font. In this work, we address the grouping tokens in a token-compressed document into candidate fonts. This permits us to incorporate font information into token-compressed images even when the original fonts are unknown or unavailable in digital format. This paper extends previous work in font reconstruction by proposing and evaluating an algorithm to assign a font to every character within a document. This is a necessary step to represent a scanned document image with a reconstructed font. Through our evaluation method, we have measured a 98.4% accuracy for the assignment of letters to candidate fonts in multi-font documents.

7874-23, Session 8

## How carefully designed open resource sharing can help and expand document analysis research

B. Lamiroy, Univ. Nancy 2 (France) and Lehigh Univ. (United States); D. P. Lopresti, J. Heflin, H. F. Korth, Lehigh Univ. (United States)

Making datasets available for peer reviewing of published document analysis methods or distributing large commonly used document corpora for benchmarking are extremely useful and sound practices and initiatives. This paper shows that they cover only a very tiny segment of the uses shared and commonly available research data may have. We develop a completely new paradigm for sharing and accessing common data sets, benchmarks and other tools that is based on a very open and free community based contribution model. The model is operational and has been implemented so that it can be tested on a broad scale. The new interactions that will arise from its use may spark innovative ways of conducting document analysis research on the one hand, but create very challenging interactions with other research domains as well.

7874-24, Session 8

## Multiple-agent adaptation in whole-book recognition

P. Xiu, H. S. Baird, Lehigh Univ. (United States)

In order to accurately recognize textual images of a book, we often employ various models including iconic model (for character classification), dictionary (for word recognition), character segmentation model, etc., which are derived from prior knowledge. Imperfections in these models affect recognition performance inevitably. In this paper, we propose an unsupervised learning technique that adapts multiple models on-the-fly on a homogeneous input data set to achieve a better overall recognition accuracy fully automatically. The major challenge for this unsupervised learning process is, how to make models improve rather than damage one another? In our framework, models measure disagreements between their input data and output data. We propose a policy based on disagreements to adapt multiple models simultaneously (or alternately) safely. We will construct a book

recognition system based on this framework, and demonstrate its feasibility.

#### 7874-25, Session 9

### **Ancient documents bleed-through evaluation and its application for predicting OCR error rates**

V. Rabeux, J. Nicholas, J. Domenger, Univ. Bordeaux 1 (France)

This article presents a way to evaluate the bleed-through defect on very old document images. We design measures to quantify and evaluate the verso ink bleeding through the paper onto the recto side. Measuring the bleed-through defect allows us to perform statistical analysis that are able to predict the feasibility of different post-scan tasks. In this article we choose to illustrate our measures by creating two OCR error rate predicting models based bleed-through evaluation. Two models are proposed, one for Abbyy FineReader which is a very power-full commercial OCR and OCRopus which is sponsored by Google. Both prediction models appears to be very accurate when calculating various statistic indicators.

#### 7874-26, Session 9

### **Binarization of camera-captured document using A MAP approach**

X. Peng, S. Setlur, V. Govindaraju, Univ. at Buffalo (United States); R. Sitaram, Hewlett-Packard Labs. India (India)

Document binarization is one of the initial and critical steps for many document analysis systems. Nowadays, with the success and popularity of hand-held devices, large efforts are motivated to convert documents into digital format by using hand-held cameras. In this paper, we propose a Bayesian based maximum a posteriori (MAP) estimation algorithm to binarize the camera-captured document images. A novel adaptive segmentation surface estimation and normalization method is proposed as the preprocessing step in our work and followed by a Markov Random Field based refine procedure to remove noises and smooth binarized result. Experimental results show that our method has better performance than other algorithms on bad or uneven illumination document images.

#### 7874-27, Session 9

### **Statistical multiresolution schemes for historical document binarization**

T. Obafemi-Ajayi, G. Agam, Illinois Institute of Technology (United States)

In previous work , we proposed the application of the Expectation-Maximization (EM) algorithm in the binarization of historical documents by defining a multi-resolution framework.

In this work, we extend the multi-resolution framework to the Otsu algorithm for effective binarization of historical documents.

We compare the effectiveness of the EM based binarization technique to the Otsu thresholding algorithm on historical documents. We demonstrate how the EM can be extended to perform an effective segmentation of historical documents by taking into account multiple features beyond the intensity of the document image.

Experimental results, analysis and comparisons to known techniques are presented using the document image collection from the DIBCO 2009 contest in addition to document images from the Frieder collection.



# Conference 7875: Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII

Tuesday-Thursday 25-27 January 2011

Part of Proceedings of SPIE Vol. 7875 Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII

## 7875-26, Poster Session

### Approach to quantitative detection of CD146 with the biosensor based on imaging ellipsometry

Y. Niu, L. Liu, Institute of Mechanics (China); X. Yan, Institute of Biophysics (China); G. Jin, Institute of Mechanics (China)

CD146 glycoprotein, a member of cell adhesion molecule (CAMs), is considered to be a novel target on endothelial cell involved in tumor angiogenesis. The biosensor based on imaging ellipsometry (BIE) was used for CD146 detection as a trial by the following steps. Firstly, CD146 antibody as ligand was tempted to immobilize orientally on modified silicon substrate by Protein G. Then, CD146 detection was performed and its calibration curve was established for the need of quantitative detection. Finally, 18 serum samples were tested quantitatively, and their results were compared to ELISA's. The sensitivity for CD146 detection reaches the order of ng/mL and the relationship between BIE signal  $y$  (grayscale value) and CD146 concentration  $x$  (ng/mL) is  $y=3.3\ln(x) + 91.3$ . It agrees mostly with ELISA's results, and the correlation coefficient is 0.714, which indicates that results of two approaches have significant statistic relevance. In addition, biosensor with the total internal reflection imaging ellipsometry (TIRIE) was applied for real-time monitoring of the detection processes. To conclude, BIE provides a simple and effective approach for CD146 detection, which shows a potential for further clinical applications.

## 7875-27, Poster Session

### Dynamic range extension of a CMOS active pixel sensor by in-pixel charge mixing

S. Jo, M. Bae, J. Kong, J. Shin, Kyungpook National Univ. (Korea, Republic of)

Various approaches have been utilized to extend the dynamic range of the CMOS image sensor, which are based on a linear-logarithmic CIS, overflow integration capacitor, and multiple sampling or individual pixel resetting. These approaches, however, suffer from noise, nonlinearity, lower sensitivity, reduced operating speed, and lower resolution. In order to overcome these problems, we have previously proposed a dynamic range extension method by combining output signals from two photodiodes with different sensitivities, such as a high-sensitivity photodiode and a low-sensitivity photodiode.

The proposed active pixel sensor has been fabricated by using 2-poly 4-metal standard CMOS process and its characteristics have been measured. It is found that charges in the high- and low-sensitivity photodiodes could be mixed each other and the lost image information of the high-sensitivity photodiode could be regenerated using the charges in the low-sensitivity photodiode, as shown by simulation results. Also, dynamic range extension of the proposed active pixel sensor has been experimentally verified. Detailed experimental results will be presented in the paper.

## 7875-28, Poster Session

### A novel 3D architecture for high dynamic range image sensor and on-chip data compression

G. M. Fadoua, Lab. d'Electronique de Technologie de l'Information (France); A. Dupret, Ecole Supérieure d'Ingénieurs en Electronique et Electrotechnique (France); A. Peizerat, Lab.

d'Electronique de Technologie de l'Information (France); Y. Blanchard, Ecole Supérieure d'Ingénieurs en Electronique et Electrotechnique (France)

The intensity of light of natural scenes has a dynamic range that can be over 120 dB. Classical 3T or 4T pixel architectures cover only 60-70 dB. Current works on CMOS image High Dynamic Range (HDR) sensor have led to dynamic range over 120dB at the expense of more complex architectures. In some cases, this leads to lower Fill Factor or larger pixel pitch. The emergence of 3D circuits may help to overcome those limitations. Moreover large scale image sensor must face the increase in required bandwidth and this problem becomes more acute with HDR images. In this paper, we propose an original architecture for extending the image sensor dynamic range together with a local compression of data for a 3D circuit image sensor. The targeted circuit is composed of 2 vertically stacked wafers with a pixel size below  $5\mu\text{m}\times 5\mu\text{m}$ . The proposed technique for HDR is based-on a floating point coding. A first data reduction is obtained by applying a common 4-bit exponent to each block of pixels, referred to as macro-pixel. For each macro-pixel, the optimal exposure is set by a dynamic adaptation of the integration time according to the received photon quantity. It theoretically allows reaching a dynamic range equivalent to about 20 bits. Simulation results show images with very few artefacts. In order to further reduce the amount of data, an on-chip data compression is performed at the macro-pixel level. Indeed, a compact compression architecture implements a compression algorithm on each block of macro-pixels. Only the mantissa array is compressed and the reduced exponent array with an exponent per macro-pixel is stored. This new concept features a good image quality (PSNR of about 40 dB) and a high dynamic range (120 dB) and shows a compression ratio over 75%, while maintaining a complexity compatible with 3D circuits. Finally, further work such as A/D conversion is discussed.

## 7875-29, Poster Session

### Improvement for sensitivity of biosensor with total internal reflection imaging ellipsometry (TIRIE)

L. Liu, Institute of Mechanics (China); Y. Chen, Suzhou Institute of Nano-tech and Nano-bionics (China); Y. Meng, S. Chen, G. Jin, Institute of Mechanics (China)

The biosensor based on the total internal reflection imaging ellipsometry (TIRIE) is realized as an automatic analysis method for protein interaction processes in real-time, with high throughput and label-free. An evanescent wave is used as the optical probe to monitor bio-molecular interactions on a chip surface with a high sensitivity due to its phase sensitive property. In this paper, the technique is optimized with a polarization setting, a spectroscopic light source and a low noise CCD detector to improve the performance of the biosensor in sensitivity and detection limit, as evidenced by a quantitative detection of hepatitis B virus surface antigen (HbsAg) with concentrations of 8, 16, 32, 64, 125 and 250ng/mL. The sensitivity is increased by one order of magnitude and the detection limit has been extended more than 50 times for HbsAg detection.

## 7875-01, Session 1

### Single-chip color imaging for UHDTV camera with a 33M-pixel CMOS image sensor

R. Funatsu, T. Yamashita, K. Mitani, Y. Nojiri, NHK Science & Technical Research Labs. (Japan)

We have been researching an ultrahigh-definition television (UHDTV) camera with a resolution 16 times higher than that of HDTV resolution. To develop a UHDTV camera that is compact and has high mobility, we investigated the use of a 33M-pixel CMOS image sensor to provide single-chip color imaging. The sensor has a Bayer color filter array (CFA) and its output signal format is compatible with the conventional UHDTV camera that uses four 8M-pixel CMOS image sensors. We first calculated the theoretical MTF characteristics of the single-chip camera and of the conventional four-8M-pixel CMOS camera. We then studied the Bayer CFA demosaicing used for the single-chip UHDTV camera. Finally, we developed an experimental pick-up system for single-chip imaging with a 33M-pixel color CMOS image sensor. The experimental results showed that the resolution is equivalent to or surpasses that of the conventional four-8M-pixel CMOS camera. We confirmed the possibility of a practical compact UHDTV camera that makes use of single-chip color imaging.

## 7875-02, Session 1

### On the design of multispectral color filter arrays

J. Y. Hardeberg, R. Khan, R. Shrestha, Gjøvik Univ. College (Norway)

In the past few years there has been a significant volume of research work carried out in the field of multispectral image acquisition. The focus of most of this research has been to facilitate a type of multispectral image acquisition systems that usually requires multiple subsequent shots (e.g. systems based on filter wheels, liquid crystal tunable filters, or active lighting).

Recently, an alternative approach for one-shot multispectral image acquisition has been proposed, based on an extension of the much used color filter array (CFA) to using more than the conventional three RGB channels, we can thus introduce the concept of multispectral color filter array (MCFAs). But this field has not been much explored, particularly little focus has been given to developing systems which focuses on the reconstruction of scene spectral reflectances.

Baone and Qi [1] made a proposal of such an MCFA based imaging system, whose main purpose was classification, where they used 4 bands in the middle and low wave infrared region in addition to the conventional three in the visible spectrum. They also proposed a new demosaicing algorithm that tried to better restore the image by maximizing a-posteriori probability. A very recent work by Lu et al. [2] also focused on constructing MCFAs to capture a NIR band along with the visible bands. The purpose of this work was the simultaneous capture of high quality visible and NIR image pair. Another work by Brauers et al. [3] has been carried out, which proposes an MCFA with narrow band filters in the visible range. Here also, a demosaicing algorithm has been proposed, which attempts to make use of the inter-band correlation by low pass filtering of the channel differences.

A MCFA based multispectral camera introduces several design issues that need to be handled, even without worrying about possible and probable issues related to the eventual real production of imaging sensors and systems. Notable ones include the choice of the number of filters and their selection for the acquisition system, the spatial arrangement of the filters, and the demosaicing algorithm. In the present work we focus on the issue which has probably received the least attention so far, namely the spatial arrangement of the filter array.

For conventional CFAs the well-known Bayer matrix [4] has enjoyed a tremendous success, although alternative approaches do exist. For MCFAs, recently Miao et al. [5] proposed a method for the spatial arrangement of the filters based on the probability of appearance of the corresponding bands.

In this paper we have used the algorithm proposed in [5] to construct MCFAs of different sizes. We have simulated acquisitions of several spectral scenes using 6, 8, and 10-channel systems, and compared the results with those obtained by the conventional regular MCFA arrangement proposed in [3], evaluating the precision of the reconstructed scene spectral reflectances in terms of spectral RMS error, goodness-of-fit coefficient (GFC) and colorimetric CIEDE2000 color differences. Using the proposed approach we significantly improve the precision, in particular for an eight-channel MCFA we

reduce the average CIEDE2000 color difference by up to 50%.

In conclusion, we believe that MCFA-based systems can be a viable alternative for affordable acquisition of multispectral color images, in particular for applications where spatial resolution can be traded off for spectral resolution. We have shown that the spatial arrangement of the array is an important design issue.

[1] Gaurav A. Baone and Hairong Qi. "Demosaicing methods for multispectral cameras using mosaic focal plane array technology," in Spectral Imaging: Eighth International Symposium on Multispectral Color Science, SPIE Proceedings, volume 6062, pages 75-87, January 2006.

[2] Yue M. Lu, Clément Fredembach, Martin Vetterli, and Sabine Süsstrunk, "Designing color filter arrays for the joint capture of visible and near-infrared images," in Proc. IEEE International Conference on Image Processing (ICIP), November 2009.

[3] Johannes Brauers and Til Aach. A color filter array based multispectral camera. In Proc. 12. Workshop Farbbildverarbeitung, German Color Group, Ilmenau, October 2006.

[4] Bryce E. Bayer, "Color Imaging Array," US Patent 3,971,065, 1976

[5] L. Miao, H. Qi, and W. Snyder, "A generic method for generating multi-spectral filter arrays," IEEE International Conference on Image Processing (ICIP), October 2003.

## 7875-03, Session 1

### Spectral-based calorimetric calibration of a 3CCD color camera for fast and accurate characterization and calibration of LCD displays

R. Safaee-Rad, Qualcomm Inc. (Canada); M. Aleksic, Qualcomm Inc. (United States)

LCD displays exhibit significant amount of variability in their tone-responses, color responses and backlight-modulation responses. As a result, a fast and efficient system for a full LCD display (not just the panel center) characterization and calibration is required. Herein, a system based on a 3CCD calorimetrically-calibrated camera is presented which can be used for full characterization and calibration of LCD displays. The camera can provide tri-stimulus measurements over thousands of locations on a LCD display (as high as camera total pixel count) in real time--camera frame rate (33 ms). To achieve high-degree of accuracy, calorimetric calibration of camera is carried out based on spectral method.

## 7875-04, Session 2

### Optimizing quantum efficiency in a stacked CMOS sensor

R. S. Hannebauer, Lumiense Photonics, Inc. (Canada); S. Yoo, HanVision Co. Ltd. (Korea, Republic of); D. L. Gilblom, A. D. Gilblom, Alternative Vision Corp. (United States)

Optimizing quantum efficiency of image sensors, whether CCD or CMOS, has usually required backside thinning to bring the photon receiving surface close to the charge generation elements. A new CMOS sensor architecture has been developed that permits high-fill-factor photodiodes to be placed at the silicon surface without the need for backside thinning. The photodiode access provided by this architecture permits ultra-shallow front side implants, the application of highly-effective anti-reflection coatings on the input surface and construction of a mirror inside the silicon below the photodiodes to effectively double the thickness of the silicon charge generation volume. Secondary benefits of this architecture include prevention of light from reaching the CMOS circuitry under the photodiodes and improvement of overall quantum efficiency.

A sensor was constructed with 4096 x 4096 pixels 4.8  $\mu\text{m}$  square with 95% fill factor and 100,000 electron full-well capacity backed with a mirror tuned to the 400-700 nm visible band. A multi-layer anti-

reflectance coating was applied to the input surface with a reflectivity in the visible of less than 2%. The result was measured quantum efficiency exceeding 85% through the visible. The blocking action of the mirror resulted in an extinction ratio for the global shutter exceeding 1,000,000:1.

### 7875-06, Session 2

#### **Detailed characterisation of a new large area CCD manufactured on high resistivity silicon**

M. S. Robbins, P. Mistry, P. Jorden, e2v technologies plc (United Kingdom)

e2v technologies has developed "Hi-Rho" devices manufactured on very high resistivity silicon. Special design features have been included that enable extremely high gate to substrate potentials to be applied without significant current leakage between back and front substrate connections. The approach taken allows the usual design rules for low noise output amplifier circuitry to be followed thus low noise devices very sensitive to red and near infrared wavelengths can be manufactured. This paper reports on the detailed characterisation of the large format "Hi-Rho" sensor designed for astronomical applications and extends the data previously reported to include detailed assessment of the CTE, spatial resolution, dark signal and cosmetic quality. The influence of the base material has also been investigated with devices having been manufactured on silicon from two different manufacturers. New, detailed measurements of the quantum efficiency of devices utilising a newly developed antireflection coating process are presented.

### 7875-07, Session 2

#### **Simulating enhanced photo carrier collection in the multifinger photogate active pixel sensors**

P. V. R. Kalyanam, G. H. Chapman, A. M. Parameswaran, Simon Fraser Univ. (Canada)

In our current work we use an extensive set of optical tools provided by the Sentaurus device simulator suite to simulate the multifinger photogate designs with optical illumination. First the optical generation profile is extracted for all the layers of the device. This profile is implemented when solving for the electrical characteristics of the device in device simulations. Carrier collection and accumulation is studied by integrating the generated photocarriers over time. By showing the optical generation and charge collection for different wavelengths of light, we observe the behaviour of the 7 finger design in the 0.5  $\mu\text{m}$  case, 9-finger for 0.25  $\mu\text{m}$  and 11-finger for 0.18  $\mu\text{m}$ . These designs which were estimated of having higher sensitivity ratios in our previous works are observed under optical illumination to find the exact optimum multifinger design with the maximum efficiency.

### 7875-08, Session 3

#### **An introduction to the atmospheric imaging assembly (AIA) on the Solar Dynamics Observatory (SDO)**

A. M. Title, Lockheed Martin Space Systems Co. (United States)

No abstract available

### 7875-10, Session 3

#### **Correcting distortion and braiding of micro-images from multi-aperture imaging systems**

A. Oberdörster, A. Brückner, F. C. Wippermann, A. Bräuer, Fraunhofer-Institut für Angewandte Optik und Feinmechanik (Germany)

Multi-aperture imaging systems inspired by insect compound eyes promise advances in both miniaturization and cost reduction of digital camera systems. Instead of a single lens stack with size and sag in the order of a few millimeters, the optical system consists of an array of microlenses. At a given field of view of the complete system, the focal length of the microlenses is a fraction of the focal length of a single-aperture system, reducing track length and increasing depth of field significantly. As each microimage only spans a small field of view, the optical systems can be simple. Because the microlenses have a diameter of hundreds of microns and a sag of tens of microns, they can be manufactured cost-effectively on wafer scale and with high precision. However, reaching a sufficient resolution for applications such as camera phones has been a challenge so far.

We demonstrate a multi-aperture color camera system with approximately VGA resolution (700x550 pixels) and a track length of 1.4 mm. The MTF of the complete system (optics and image processing) is comparable to current commercial miniaturized VGA camera modules. The algorithm for correcting optical distortion of the microlenses and combining the microimages into a single image is the focus of this presentation.

### 7875-11, Session 3

#### **An analog logarithmic number system subtractor for edge detection in logarithmic CMOS image sensors**

D. R. Desai, The Univ. of Akron (United States); F. Hassan, Ohio Northern Univ. (United States); R. Veillette, J. Carletta, The Univ. of Akron (United States)

This paper describes the design of analog circuitry to implement logarithmic number system (LNS) subtraction. Such circuitry, if incorporated in the readout circuitry of a logarithmic CMOS image sensor, would allow for the on-chip calculation of spatial derivatives, while operating directly on logarithmically-scaled pixels. The circuit was implemented for a 1.2  $\mu\text{m}$  CMOS process. The maximum relative error at the output of the LNS subtractor for pixel currents that correspond to an illumination range of more than four decades is 6.25%.

### 7875-12, Session 4

#### **A CMOS image sensor with draining only modulation pixels for fluorescence lifetime imaging**

Z. Li, K. Yasutomi, T. Takasawa, S. Itoh, S. Kawahito, Shizuoka Univ. (Japan)

Fluorescence lifetime imaging is becoming a powerful tool in biology. A charge-domain CMOS FLIM chip using a pinned photo diode (PPD) and the pinned storage diode (PSD) with different depth of potential wells has been developed by the authors. However, a transfer gate between PPD and PSD causes charge transfer noise due to traps at the channel surface. This paper presents a time-resolved CMOS image sensor with draining only modulation pixel for fluorescence lifetime imaging, which removes the transfer gate between PPD and PSD. The time windowing is done by draining with a draining gate only, which is attached along the carrier path from PPD to PSD. This allows us to realize a trapping less charge transfer between PPD and PSD, leading to a very low-noise time-resolved signal detection. A video-rate

CMOS FLIM chip has been fabricated using 0.18 $\mu\text{m}$  standard CMOS pinned diode image sensor process. The pixel array has 200(Row)  $\times$  256(Column) pixels and the pixel pitch is 7.5 $\mu\text{m}$ . The signal intensity of the PSD as a function of the TD gate voltage is measured. The ratio of the signal for the TD off to the signal for the TD on is 212:1.

#### 7875-14, Session 4

### Development of biosensor based on imaging ellipsometry and its applications

G. Jin, Institute of Mechanics (China)

So far, combined with a microfluidic reactor array system, a serviceable engineering system of biosensor based on imaging ellipsometry is installed for biomedical applications, such as antibody screen, hepatitis B markers detection, tumor markers spectrum and virus recognition, etc. Furthermore, the biosensor in total internal reflection (TIR) mode has been improved by a spectroscopic light, optimization settings of polarization and low noise CCD which brings an obvious improvement of 10 times increase in the sensitivity and SNR, and 50 times lower concentration in the detection limit with a throughput of 48 independent channels and the time resolution of 0.08 S.

#### 7875-15, Session 4

### Study on colony image acquisition and analysis system

Z. Jia, W. Wang, Henan Polytechnic Univ. (China)

For counting of both colonies and plaques, there is a large number of applications including food, dairy, beverages, hygiene, environmental monitoring, water, toxicology, sterility testing, AMES testing, pharmaceuticals, paints, sterile fluids and fungal contamination. Recently, many researchers and developers have made efforts for this kind of systems. By investigation, some existing systems have some problems since they belong to a new technology product. The main problems are image acquisition and image segmentation. In order to acquire colony images with good quality, an illumination box was constructed as: the box includes front lighting and back lighting, which can be selected by users based on properties of colony dishes. With the illumination box, lighting can be uniform; colony dish can be put in the same place every time, which makes image processing easy. A digital camera in the top of the box connected to a PC computer with a USB cable, all the camera functions are controlled by the computer. In this paper, the developed colony image segmentation algorithm consists of the sub-algorithms: (1) image classification; (2) image processing; and (3) colony delineation. The colony delineation algorithm mainly contains: the procedures based on grey level similarity, on boundary tracing, on shape information and colony excluding. In addition, a number of algorithms are developed for colony.

#### 7875-16, Session 5

### Aging effects on image sensors due to terrestrial cosmic radiation

G. Gangadharan Nampoothiri, A. J. P. Theuwissen, Technische Univ. Delft (Netherlands); M. Horemans, Consultant (Belgium)

We analyze the "ageing" effects on image sensors introduced by neutrons present in terrestrial cosmic environment. Defects develop during the lifetime of imagers and do not disappear, limiting the imaging performance. It is hypothesized that the ageing phenomenon is due to the influence of terrestrial cosmic rays, which are the result of very high energy particles created in space or by the sun. In a previous work we have compared post-flight measurements at aviation altitudes to that of sea level and presented activation energy analysis of the sensors. For the first time, hot pixel development at sea level (terrestrial cosmic radiation environment) is corroborated successfully with accelerated neutron beam tests for various image sensor operating conditions.

Group of image sensors were irradiated in the ANITA (Atmospheric-like Neutrons from Thick Target) beam at The Svedberg Laboratory (TSL), Sweden to further understand the underlying mechanisms. Influence of neutron flux (dose rate) and biasing on hot pixel development is also reported. These experiments provide further validation to the hypothesis that the prominent cause of hot pixels is displacement damage in the silicon bulk due to neutron radiation, introduced by secondary cosmic rays.

#### 7875-17, Session 5

### Nonlinear time dependence of dark current in charge-coupled devices

R. Widenhorn, J. Dunlap, E. Bodegom, Portland State Univ. (United States)

It is generally assumed that charge-coupled device (CCD) imagers produce a linear response of dark current versus exposure time except near saturation. We found a large number of pixels with nonlinear dark current response to exposure time to be present in two scientific CCD imagers. These pixels are found to exhibit distinguishable behavior with other analogous pixels and therefore can be characterized in groupings. Data from two Kodak CCD sensors are presented for exposure times from a few seconds up to two hours. Linear behavior is traditionally taken for granted when carrying out dark current correction and as a result, pixels with nonlinear behavior will be corrected inaccurately.

#### 7875-18, Session 5

### Tradeoffs in imager design parameters for sensor reliability

G. H. Chapman, J. Leung, Simon Fraser Univ. (Canada); Z. Koren, I. Koren, Univ. of Massachusetts Amherst (United States)

Image sensors are continuously subject to the development of in-field permanent defects in the form of hot pixels. Based on laboratory measurements of defect rates in 23 DSLRs, 2 midsize cameras and 11 cell phone cameras, we show in this paper that the rate of these defects depends on the technology (APS or CCD) and on design parameters like imager area, pixel size, and gain (ISO). Increasing the image sensitivity (ISO) (from 400 up to 25,600 ISO range) causes the defects to be more noticeable, with some going into saturation and at the same time increases the defect rate. Partially stuck hot pixels, which have an offset independent of exposure time, make up >40% of the defects and are particularly affected by ISO changes. Comparing different sensor sizes has shown that the defect rate does not scale entirely linearly. Measuring imagers with different pixel sizes (from 7.5 to 2.2 microns) has demonstrated that defect rates grow rapidly as pixel area shrinks. These defect rate trends result in interesting tradeoffs in imager design, allowing the designer to determine the specific imager parameters based on the imager's designated function and reliability requirements.

#### 7875-19, Session 5

### Dark noise in a CMOS imager pixel with negative bias on transfer gate

H. Yamashita, Toshiba Corp. (Japan); M. Maeda, S. Furuya, T. Yagami, Toshiba Materials Co., Ltd. (Japan)

Several reports on negative bias on transfer gate in 4-transistor CMOS imager pixel have already been published [1] [2]. The advantage of the negative bias on transfer gate is a drastic dark current reduction due to a transfer gate channel surface pinning by accumulated holes. But it was also reported that further lowering of negative gate bias beyond -0.9V in turn increases dark current [2]. The analysis of the dark current increase caused by the negative gate bias on the transfer gate was reported, by investigating hot pixel dark output dependency

both on bias and on temperature with test pixel array [3]. The proposed mechanism for the dark current increase in the report is Gate-Induced-Leak (GIL) Trap-Assisted-Tunneling (TAT). The dark current observed when large negative gate bias is applied follows  $I_{\text{dark}} \sim A \cdot \exp(V_m/V_0)^2$ , where A is a proportionality factor,  $V_0$  is threshold voltage, and  $V_m$  is difference between floating diffusion (FD) voltage and gate voltage [3]. The exponential dependence on  $V_m$  suggests that the TAT generated by high electric field in FD near read gate is the cause of the dark current. Electric field in a FD tends to be larger when pixel size is reduced, because a transfer gate scale down is necessary in a small size pixel. Thus both full understanding of the generation mechanism of TAT dark current in FD and a proposal to reduce the dark current is of value for future pixel size reduction. In this contribution, the detailed analysis both with experimental results and with device simulation shows the dark current is dominated by the electric field generated at the transfer gate edge overlapped by a FD n+. It also is reported that the reduction of the electric field in FD by changing the dopant profile of FD drastically reduces the dark current.

The key parameter dominating the TAT dark current is electric field in FD. The device simulation results show that maximum electric field after FD reset is induced in silicon surface at the FD edge below transfer gate. The generation of high electric field is attributed to the large voltage difference between the high concentration hole layer under transfer gate and high concentration n+ layer at the FD edge below transfer gate. The reduction of dark output level of hot pixel has been observed when the dopant concentration at the surface of FD is lowered. The device simulation shows that maximum electric field at the FD edge below transfer gate is reduced when the FD dopant concentration is lowered. It also is shown that the dark output level of hot pixel can be scaled with local maximum electric field. The results prove that TAT is the dominant cause of the dark output increase when negative gate bias is applied on transfer gate.

[1] H. Han, et al, "Evaluation of a Small Negative Transfer Gate Bias on the Performance of 4T CMOS Image Sensor Pixel", International Image Sensor Workshop, 2007 pp238

[2] B. Mheen, et al, "Negative Offset Operation for Four-Transistor CMOS Image Pixels for Increased Well Capacity and Suppressed Dark Current" Electron Device Letters Vol.29, No.4, April, 2008, pp347

[3] H. Yamashita et al., "Analysis of Dark Current in 4-Transistor CMOS Imager Pixel with Negative Transfer-Gate Bias Operation", International Image Sensor Workshop 2009, Session 04-1

## 7875-20, Session 5

### Image sensor noise: you love it or you hate it!

A. J. P. Theuwissen, Harvest Imaging (Belgium)

A software tool is developed that allows to simulate solid-state image sensors based on their specification. The output of the simulator is a set of IMAGES. By means of the tool the camera engineer gets a direct view of the images the sensor is able to create. A great advantage of this simulation tool is to check the influence of all different fixed-pattern noise sources, temporal noise sources and performance parameters, as well as their impact on the image quality.

Next to the simulation tool, a second tool is developed that uses IMAGES as the input and extracts the sensor parameters. Examples of extracted data : fixed-pattern noise components, temporal noise component, conversion gain, quantum efficiency, dark current, etc. The input images can be the simulated ones generated by the simulator tool, or can be real images from existing sensors or cameras.

Both software tools can handle CCD and CMOS devices, color as well as monochrome devices.

## 7875-21, Session 6

### The early history of CCDs

M. M. Blouke, Portland State Univ. (United States)

No abstract available

## 7875-22, Session 6

### 3D ranging with a single-photon imaging array

S. Bellisai, F. Guerrieri, Politecnico di Milano (Italy); S. Tisa, Micro Photon Devices S.r.l. (Italy); F. Zappa, Politecnico di Milano (Italy) and Micro Photon Devices S.r.l. (Italy)

Several applications require systems for 3D ranging acquisition, where both high frame-rate and high sensitivity (for either very dark environments or opaque objects) are a must. We exploited a monolithic chip with 32x32 Single-Photon Avalanche Diode smart-pixels for 3D ranging applications based on an Indirect Time-of-Flight (iTOF) technique. The scene is illuminated by a sinusoidal modulated LED and the reflected light is acquired by the imager in different time-slots, for measuring the phase delay of outgoing vs. incoming signal, hence computing the distance between chip and objects in the scene.

All 1024 array pixels are synchronously enabled by a global gate signal, which allows photon counting in well-defined time-slots within each frame. The frame duration is set in accordance to the desired SNR. We report on measurements performed on chips fabricated in a standard high-voltage 0.35  $\mu\text{m}$  CMOS technology, which feature 40% photon detection efficiency at 450 nm and 20% at 650nm. The single-photon sensitivity allowed the use of just one single LED at 650 nm and 20MHz for acquiring a scene with a maximum distance of 7.5 m, with better than 10 cm distance resolution and higher than 50 frames/s frame-rate.

## 7875-23, Session 6

### Linear arrays of single-photon detectors for photon counting and timing

F. Guerrieri, Politecnico di Milano (Italy); S. Tisa, Micro Photon Devices S.r.l. (Italy); A. Tosi, S. Bellisai, B. Markovic, Politecnico di Milano (Italy); F. Zappa, Politecnico di Milano (Italy) and Micro Photon Devices S.r.l. (Italy)

Scientific experiments often demand the detection of very weak light signals at high-speed or to precisely measure the time of arrival of single photons. Arrays of Single-Photon Avalanche Diodes (SPAD) are ideal candidates when high sensitivity is required together with high frame-rate or precise photon-timing resolution. We designed a linear 32x1 SPAD array using a high-voltage CMOS technology able to provide both good SPAD performance and fast electronics.

During frame acquisition all pixels work in parallel, each of them being equipped with anything necessary for photon counting. The array architecture is capable of fully-parallel operation of all pixels allowing free-running acquisition at high frame-rate. With a low-speed 10 MHz clock frequency, one pixel is read out in 100 ns while the whole array is readout in 320 ns, corresponding to a frame-rate of 312.5 kframe/s. The frame-rate can top to 4 Mframe/s with a clock of 128 MHz.

The photon timing modality employs the photon time-of-arrival information provided by each of the 32 outputs. All 32 "timing" outputs feed external Time-Correlated Photon Counting boards. The Full-Width at Half -Maximum using very short laser pulses is 55 ps with few kcps counting rate.

7875-24, Session 6

**A single photon sensitive fast ebCMOS camera system for multitarget tracking of single fluorophores: applications to nano-photonics**

T. Cajgfinger, R. Barbier, A. Dominjon, E. Chabanat, D. Quang Tuyen, C. Guérin, J. Houles, Institut de Physique Nucléaire de Lyon (France)

Our development of a camera system based on electron bombarded CMOS (ebCMOS) device is consistent with the demands of applications such as fast real-time multi target tracking of fluorescent dyes used in fluorescence microscopy and nano-photonics. The design and fabrication of a Back Side Illuminated (BSI) CMOS (160 kPixels -10 micron pitch) is optimized for single photoelectron detection. The ebCMOS shows a high resolution with single photon sensitivity at a 500 Hz frame rate and is used as a proof of concept of real time tracking at single photon sensitivity of fluorescent nano-particles. The performances of the ebCMOS itself in terms of spatial resolution, dark count rate, single photon sensitivity and true counting capability are presented. The full camera system is described. Then we present the measurement based on single photon detection capability for spatio-temporal identification of fluorescent targets (Quantum Dots). The accuracy of localization for different noise and signal conditions is measured on Quantum Dots and on a dedicated optical test bench. We conclude by giving the measurement of the speed limit of a target that can be followed by the camera system as a function of photon signal and photon noise.

7875-25, Session 6

**Monolithic single-photon detectors and time-to-digital converters for picoseconds time-of-flight ranging**

B. Markovic, Politecnico di Milano (Italy); S. Tisa, Micro Photon Devices S.r.l. (Italy); A. Tosi, F. Zappa, Politecnico di Milano (Italy)

We present a novel "smart-pixel" able to measure and record in-pixel the time delay (photon timing) between a START (e.g. given by the laser excitation, the cell stimulus, or the lidar flash) and a STOP (e.g. the arrival of the first returning photon from the fluorescence decay signal or back reflection from an object). Such smart-pixel relies of a SPAD detector and a Time-to-Digital Converter monolithically designed and manufactured in the same chip. Many pixels can be laid out in a rows by columns architecture, to give birth to expandable 2D imaging arrays for picoseconds-level single-photon timing applications. Distance measurements, by means of the direct TOF detection (the same used in lidar systems) provided by each smart-pixel, can open the way to the fabrication of single-chip 3D ranging arrays for scene reconstruction and intelligent object recognition.

We report on the design and characterization of prototype circuits, fabricated in a 0.35 $\mu$ m standard CMOS technology containing complete conversion channels, smart-pixels and ancillary electronics with 20 $\mu$ m active area diameter SPAD detectors and related quenching circuitry. With a 100MHz reference clock, the TDC provides a time-resolution of 10ps, a dynamic range of 160ns and very high conversion linearity.

7875-30, Session 6

**Human-technology interaction for IED detection**

A. Zhang, Y. Zou, L. Wu, EYZtek, Inc. (United States); J. E. Fulton, Naval Surface Warfare Ctr. Crane Div. (United States)

No abstract available

## 7876-01, Session 1

### High dynamic range image sensor architectures

B. Fowler, Fairchild Imaging (United States)

Digital photographers continuously demand more performance from their equipment. Digital camera performance is defined by a set of parameters including dynamic range, noise, frame rate, resolution, and color. Amongst these parameters dynamic range is becoming increasingly more important. This is true because the human eye typically has a wider dynamic range than a digital camera. In this paper we define dynamic range as the ratio of the maximum to the minimum signal that can be detected. At the heart of all digital cameras is either a CCD or a CMOS image sensor (CIS). The dynamic range of the sensor typically limits the dynamic range of the camera.

In this paper we review five CIS architectures that are designed to improved dynamic range. We start by reviewing standard CCD and CIS architectures and then present a simple sensor model. Using this model we show how signal to noise ratio (SNR) can be used to evaluate different wide dynamic range (WDR) sensor architectures. Then we sequentially review five different wide dynamic range techniques. The first WDR technique is multiple gains, and the second technique is non-linear pixel response. The third technique is variable exposure, and the fourth technique is well capacity recycling. The fifth and final technique is time to saturation. For each of these techniques we present the pixel level circuitry and its advantages and disadvantages. Furthermore, all of these techniques are compared based on SNR and implementation complexity. We discuss how implementation complexity affects signal processing in a digital camera, and other parameters in the sensor such as quantum efficiency and read noise. We conclude with a few summary comments.

## 7876-02, Session 2

### Bayer and panchromatic color filter array demosaicing by sparse recovery

M. Aghagolzadeh, A. Abdolhosseini Moghadam, H. Radha, Michigan State Univ. (United States); M. Kumar, Eastman Kodak Co. (United States)

The utility of Compressed Sensing (CS) for demosaicing of digital images have been explored by several recent efforts [1, 2 and 3]. Most recently, a Compressive Demosaicing (CD) [4] framework, based on employing a random panchromatic Color Filter Array (CFA) at the sensing stage, has provided compelling CS-based demosaicing results by visually outperforming other leading techniques. Meanwhile, it is well known that the Bayer pattern is arguably the most popular CFA used in low-cost consumer digital cameras. In this paper, we explore and compare the Bayer and random panchromatic CFA structures using a generic approach for demosaicing of images based on recent advances in the field of CS. In particular, a key objective of this work is to provide a comparative analysis between these two CFA patterns (Bayer and random panchromatic) under the general umbrella of sparse recovery, which represents the cornerstone of CS-based decoding. We demonstrate the viability of the Bayer pattern under certain CS conditions. Meanwhile, we show that a random panchromatic CFA, which meets certain incoherence constraints, can visually outperform a Bayer based sparse recovery. As illustrated in our simulation results, a panchromatic CFA is more consistent in terms of providing better visual quality when tested on a wide range of color images.

#### REFERENCES

- [1] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," IEEE Transactions on Image Processing, Vol. 17, 2008.
- [2] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman, "Non-Local Sparse Models for Image Restoration," International Conference

on Computer Vision, Tokyo, Japan, 2009.

[3] P. Nagesh and B. Li, "Compressive Imaging of Color Images". IEEE ICASSP, Taipei, Taiwan, 09.

[4] A.A. Moghadam, M. Aghagolzadeh, M. Kumar and H. Radha, "Compressive demosaicing," accepted in IEEE International Workshop on Multimedia Signal Processing, Saint-Malo, France, 2010.

## 7876-07, Session 3

### Implementation of a multispectral color imaging device without color filter array

G. Langfelder, A. F. Longoni, F. Zaraga, Politecnico di Milano (Italy)

Multispectral acquisition of digital images is interesting for several applications as a way to improve the accuracy in color reproduction. With respect to colorimetric-based imaging, where typically a set of three color filter arrays is used, multichannel acquisition allows estimating the scene spectral reflectance.

In this work we first review the working principle of the Transverse Field Detector (TFD), a proposed sensor for color acquisition without CFA. In its simplest geometrical and biasing configuration, a TFD pixel implements a set of three spectral responses. Thanks to its basic working principle, the spectral responses implemented on the device depend on the biasing configuration and can be tuned by changing the voltages applied to the collecting electrodes. Experimental results on a TFD, in good agreement with electron device simulations, are shown.

We then present detailed simulation results on an improved structure, where temporal or spatial pixel tunability can be used to generate different spectral responses. We then show a TFD pixel tuned in a non-symmetric configuration, a concept that allows increasing the number of different spectral responses of the device, avoiding the need for a spatially or temporally separated acquisition. A 5 to 6 micron wide TFD pixel can be used to implement 5 to 7 different spectral responses in a single acquisition at full resolution.

## 7876-08, Session 3

### One shot multispectral color imaging with a stereo camera

R. Shrestha, J. Y. Hardeberg, Gjøvik Univ. College (Norway); A. Mansouri, Univ. de Bourgogne (France)

Multispectral color imaging is a promising technology, which can solve many of the problems of traditional RGB color imaging. However, it still lacks widespread and general use because of its own limitations. State of the art multispectral imaging systems need multiple shots making them not only slow but also incapable of capturing scenes in motion. Moreover, the systems are mostly costly and complex to operate. This purpose of the work described in this paper is to conceive a fast and practical six-channel multispectral color image acquisition system using a stereo camera and a pair of optical filters. The best pair of filters is selected from among readily available filters such that they modify the sensitivities of the two cameras in such a way that they get spread reasonably well spaced throughout the visible spectrum. As the cameras are in a stereoscopic configuration, the system is capable of acquiring 3D images as well, and stereo matching algorithms provide a solution to the image alignment problem. Thus the system can be used as a "two-in-one" multispectral-stereo system. Both simulations and experiments have shown that the proposed system performs better than the RGB system in scene spectral reflectance reconstruction as well as in scene color reproduction.

## 7876-09, Session 3

**Multispectral image invariant to illumination colour, strength, and shading**

M. S. Drew, A. Yazdani, Simon Fraser Univ. (Canada)

We present here a method that makes use of multispectral image data and generates a novel “photometric-invariant multispectral image” for this type of data. For RGB, an “invariant image” has been constructed independent of the colour and intensity of the illuminant and to shading [ECCV04]. To generate this image either a set of calibration images is required, or entropy information taken from a single image can be used to develop the parameters necessary to produce the invariant [JCV09]. Nonetheless, generating an invariant image remains a complex and error-prone task for RGB image data. For multispectral images, we show that photometric-invariant image formation is in essence greatly simplified: one of the requirements for forming an invariant is the necessity of narrowband-sensor sensors. Here this is the case, and we show that with the simple knowledge of peak sensor wavelengths we can generate a high-D multispectral invariant: the PSNR is shown to be high between the respective invariant multispectral features for multispectral images taken under different illumination conditions, showing lighting invariance for a per-pixel measure; and the s-CIELAB error measure shows that the colour error between the 3-D colour images used to visualize the output invariant high-D data is also small.

## 7876-10, Session 3

**Methods for spectral characterization of multispectral cameras**

J. Klein, J. Brauers, T. Aach, RWTH Aachen (Germany)

High fidelity color image acquisition requires an accurate characterization of the camera’s spectral sensitivity curves to perform color calibration or spectral estimation. Several methods have been proposed to perform this task; these include characterizations via test charts, narrowband filters and methods utilizing a monochromator. In most publications, RGB cameras are characterized. In this paper, we describe the characterization of the spectral sensitivity curves of a multispectral camera featuring seven optical bandpass filters. We show two different methods for the calibration using a monochromator - either by measuring the grayscale sensor of the camera and the filters separately or by characterizing the multispectral camera as a complete system. A comparison of both methods validates the measurement results.

We furthermore develop different reconstruction methods (maximum value method, principal eigenvector method, linear or Wiener estimation). We perform also simulations of the characterization process to evaluate the methods and show the impact of the bandwidth of the monochromator stimuli on the reconstruction.

## 7876-11, Session 4

**Evaluation of a hyper-spectral image database for color filter array design and demosaicking algorithms**

M. Larabi, Univ. de Poitiers (France); S. E. Süssstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The design of color filter arrays (CFA) and associated demosaicking algorithms is still a hot topic today in digital photography, as the perfect spatial arrangement of the filters and their spectral characteristics have a large influence on image quality. In this work, we propose to study the applicability of the hyperspectral image database proposed by Foster et al. [1] for CFA and demosaicking design testing. The evaluation of the demosaicking algorithms is studied by using different well-known metrics such as CPSNR, s-CIELAB but also subjectively by running a psychovisual experiment where observers are asked to judge the demosaicked images and give their visual preference.

## 7876-12, Session 4

**Automatic annotation of outdoor photographs**

C. Cusano, R. Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

We propose here a strategy for the automatic annotation of outdoor photographs. Images are segmented in homogeneous regions which are then assigned to six different classes: sky, vegetation, snow, water, ground, and sand. These visual categories allows for content-aware processing strategies. For instance, the knowledge about the presence of uniformly colored regions (such as the sky) can be used to drive color balancing algorithms which should ignore those regions. Another example is provided by edge sharpening algorithms which should avoid boosting the edges in high frequency regions, such as those typically found in the vegetation class. Our strategy describe regions, obtained using a normalized cut segmentation strategy, using a joint histogram of color and texture information. The classification is performed by a multi-class Support Vector Machine. The strategy has been evaluated on images taken from the LabelMe dataset.

## 7876-14, Session 5

**How many pixels does it take to make a good 4”x6” print? Pixel count wars revisited**

M. A. Kriss, Consultant (United States)

The fallacy of implying that more, small pixels, produces better images than fewer, larger pixels for a given sensor size is explored in detail using photographic models developed in the 1970’s and modified for digital images and by experiments using a consistent set of digital cameras ranging from 6 million pixels to 14 million pixels in compact cameras and using 6 million and 12 million pixels in SLR format digital cameras. Both the model metrics and experimental results clearly demonstrate that smaller pixels, as small as 1.4 microns, produce lower quality images than their larger counterparts even if the pixel count is lower for the larger sensors. The smaller pixels also introduce greater noise, lower true ISO speed (no camera gain) and significant loss in exposure latitude. The high pixel count, small pixel cameras also show a much larger tendency for JPEG artifacts. The results indicate that digital photographers should pick cameras that meet their needs, be it studio work or sports photography, by purchasing the right combination of pixel count and pixel size.

## 7876-15, Session 5

**A prototype high-speed CMOS image sensor with 10,000,000 burst-frame rate and 10,000 continuous-frame rate**

Y. Tochigi, K. Hanzawa, Y. Kato, N. Akahane, R. Kuroda, S. Sugawa, Tohoku Univ. (Japan)

In this paper, a high-speed CMOS image sensor having a new architecture and a new operating principle has been developed. The image sensor achieves both the continuous capturing and the burst capturing by a single chip, and this image sensor has low power consumption, low heat generation, high sensitivity and high S/N ratio. This image sensor consist of mainly four blocks, two dimensional pixel array of 4-trangister CMOS active pixel, analog memory arrays connected independently to the pixel array by each pixel output line, scanning circuits and multiple number of output amplifiers. A prototype image sensor was fabricated using a 0.18um 2-Poly 3-Metal CMOS technology with the die size of 5550um(H) x 4575um(V), the pixel size of 48um(H) x 48um(V), the number of pixels of 72(H) x 32(V), the number of analog memories of 104 memories per pixel and the 6 parallel horizontal output circuits and output amplifiers. The aperture ratio is 35% and the conversion gain is 60uV/e-(input referred). It has been confirmed that this image sensor achieves 10,000,000 fps during burst capturing mode and 10,000 fps during the continuous



capturing mode through the image capture experiments of high speed phenomena such as rotating object and discharge phenomenon.

## 7876-16, Session 5

### Two-dimensional measurement of the lens optical transfer function from a digital image

D. P. Morgan-Mar, M. R. Arnison, C. A. Deller, P. A. Fletcher, K. G. Larkin, Canon Information Systems Research Australia Pty. Ltd. (Australia)

The lens optical transfer function (OTF) describes the resolution and sharpness of images formed through a lens. We present a novel method for accurately measuring the OTF of a camera lens by digitally imaging a tartan test pattern containing sinusoidal functions with multiple frequencies and orientations. The tartan pattern can be tuned to optimise the measurement accuracy for an adjustable set of sparse spatial frequencies. The measurement method is designed to be accurate, reliable, and fast in a wide range of measurement conditions, including uncontrolled lighting. We describe the design of the tartan pattern and the algorithm for estimating the OTF accurately from a captured digital image. We present simulation results which show that the tartan method has significantly better accuracy for measuring the modulus of the OTF (the modulation transfer function, or MTF) than the ISO 12233 standard slanted-edge method, especially at high spatial frequencies. With 1% simulated imaging noise, the root mean square (RMS) error of the tartan method is on average 5 times smaller than the RMS error of the slanted-edge method.

## 7876-17, Session 6

### Efficient defect pixel cluster detection and correction for Bayer CFA image sequences

T. Tajbakhsh, Technische Univ. Hamburg-Harburg (Germany)

No abstract available.

## 7876-20, Session 7

### Comparison of objective metrics for image sensor crosstalk characterization

A. Dokoutchaev, Aptina Imaging Corp. (United States); H. Eliasson, Sony Ericsson Mobile Communications AB (Sweden); F. Li, Aptina Imaging Corp. (United States)

Image sensor crosstalk can be divided into spectral crosstalk and pixel crosstalk. This paper focuses on the pixel crosstalk and its effect on signal to noise ratio (SNR). Pixel crosstalk occurs in the spatial domain and is due to the signal leakage between adjacent pixels either by imperfect optical isolation or diffusion of electrons. This will have a negative impact on image quality mainly in two ways: spatial blurring and decreased SNR due to more aggressive color correction required. A method for modeling the spectral broadening due to the pixel crosstalk is presented where a matrix is calculated from crosstalk kernels representing the spatial leakage between neighboring pixels. In order to quantify the amount of crosstalk we present a method in which ratios of integrals of the same color channel but within different wavelength intervals are calculated. This provides a metric that is more robust with respect to color channel scaling. To study the impact on SNR due to pixel crosstalk, a number of SNR metrics are compared to results from a limited psychophysical study. The studied SNR metrics are the metric used for calculating the SNR10 value in mobile imaging, the ISO 12232 noise metric and a metric where the signal is transformed into orthogonal color opponent channels, thereby enabling the analysis of the luminance noise separate from the chrominance noises. The results indicate that the ISO total noise and SNR10 metric yield very similar results and that the green channel has the largest individual impact on the crosstalk.

## 7876-21, Session 7

### An image quality evaluation tool simulating image sensors including quantum efficiency off-axis effect

C. Mornet, J. M. Vaillant, T. Decroux, N. Virollet, D. Herault, STMicroelectronics (France); I. Schanen, Institut de Microélectronique Électromagnétisme et Photonique (France)

The image quality evaluation of CMOS sensors is a big challenge for camera phone manufacturers. In this paper, we present an update of the Image Quality Evaluation Tool, a graphics user interface simulating image sensors to assess the performance of a pixel. The simulated images are computed from operating conditions and sensor's characteristic data like Quantum Efficiency including off-axis effect. Simulation of QE off-axis impact has been based on characterization data. The method does not require optics, making it suitable for early design phases as for optimizations and investigations. Both measurement and implementation in the tool will be explained. The QE degradation with angle effect, especially the noise in corners, will be highlighted on simulated images. A uniform gray scene or colored image simulation from QE off-axis measurement will help engineers to calculate post-processing digital correction like color shading correction or color correction matrix versus pixel position

## 7876-22, Session 7

### Image quality assessment based on edge

X. Mou, M. Zhang, W. Xue, Xi'an Jiaotong Univ. (China); L. Zhang, The Hong Kong Polytechnic Univ. (China)

The research on image quality assessment (IQA) has been become a hot topic in most area concerning image processing. Seeking for the efficient IQA model with the neurophysiology support is naturally the goal people put the efforts to pursue. In this paper, we argue that comparing the edges position of reference and distorted image can well measure the image structural distortion and become an efficient IQA metric, while the edge is detected from the primitive structures of image convolving with LOG filters. The so-called NSER metric is designed following a simple logic based on the cosine distance of the primitive structures and two accessible improvements. Validation is taken by comparison of the well known state-of-the-art IQA metrics: VIF, MS-SSIM, VSNR over the six IQA databases: LIVE, TID2008, MICT, IVC, A57, and CSIQ. Experiments show that NSER works stably across all the six databases and achieves the good performance.

## 7876-23, Session 8

### Method for evaluating tone mapping operators for natural high dynamic range images

M. Kuhna, M. Nuutinen, P. Oittinen, Aalto Univ. School of Science and Technology (Finland)

The dynamic range of digital cameras has been increasing in recent years. High Dynamic Range (HDR) and especially tone mapping has been an active field of research for years. With current image sensor technologies there is high feasibility for consumer level single shot HDR cameras. Tone mapping operators are needed for rendering HDR images on consumer displays. This study focuses on establishing a method for evaluating tone mapping operators in terms of image quality. The study is based on the observation that the test images used in similar studies in the past lack important features such as human skin tone for evaluating image quality using subjective methods. The method consists of image capture and processing as well as subjective evaluation in controlled conditions. The current level of objective quality metrics for HDR images was also benchmarked. Objective quality metrics have been developed and published widely to ease image quality evaluation, which is often performed with extremely time consuming subjective tests.

## 7876-24, Session 8

**High dynamic range imaging of non-static scenes**

I. Hossain, B. Gunturk, Louisiana State Univ. (United States)

No abstract available.

## 7876-25, Poster Session

**Toward a quantitative visual noise evaluation of sensors and image processing pipes to improve color reconstruction**

C. Mornet, STMicroelectronics (France); D. J. Baxter, STMicroelectronics (R&amp;D) Ltd. (United Kingdom); J. M. Vaillant, T. Decroux, D. Herault, STMicroelectronics (France); I. Schanen, Institut de Microélectronique Électromagnétisme et Photonique (France)

The evaluation of sensor's performance in terms of signal-to-noise ratio (SNR) is a big challenge for both camera phone manufacturers and customers. The first ones want to predict and assess the performance of their pixel while the seconds requires being able to benchmark raw sensors and processing pipes. The SNR10 metric is very sensitive to crosstalk whereas for low-light issue, the weight of sensitivity should be increased. To evaluate noise on final image, the analytical calculation of SNR on luminance channel has been performed by taking into account noise correlation due to the processing pipe. However, this luminance noise does not match the perception of human eye which is also sensitive to chromatic noise. Alternative metrics have been investigated to find a visual noise metric closer to the human visual system. They have been used to improve color reconstruction by optimizing the color correction matrix: trade-off between these metrics, color accuracy and saturation has been explored.

## 7876-26, Poster Session

**Fidelity tolerance analysis for computational imaging system**

C. Chang, Y. Chen, K. Chen, H. Tsao, H. Sung, C. Chang, P. Chen, H. Chang, Industrial Technology Research Institute (Taiwan)

In the paper, we present an analysis method using for computational imaging system which including affects of optical aberrations and finite sampling form image sensor, and fidelity tolerance analysis will be archived by suitable image metric (peak signal to noise ratio, PSNR). Tolerance analysis for computational imaging system which considering surface error of cubic phase mask is illustrated and behavior of PSNR and point spread function (PSF) similarity in such kind of system is discussed. Finally, by using PSNR, the capability of extension depth of focus and surface tolerance of phase mask can be determined.

## 7876-27, Poster Session

**Noise-robust image deblurring by blending regular- and short-exposure images**

Y. Tsuda, H. Hatanaka, S. Fukumoto, M. Ueda, SANYO Electric Co., Ltd. (Japan); K. Chihara, Nara Institute of Science and Technology (Japan)

No abstract available.

## 7876-28, Poster Session

**Improving the sensitometric and OECF standards: recognizing the photosensitive exposure range**

M. G. Prais, Consultant (United States)

This article demonstrates that and how sensitometric and opto-electronic characteristic function (OECF) standards should be changed: The sensitivity  $S$  of all photosensitive arrays is--and, in standards, should be--determined by the midtone photosensitive exposure of the array  $H_{mid}$ , the binary logarithm of which is  $[\log_2(H_{max}) + \log_2(H_{min})]/2$ . These quantities are dependent on the width of the photosensitive exposure range  $\Delta$ , which is determined by the measured minimum and maximum usable photosensitive exposures,  $H_{max}$  and  $H_{min}$ . The reference exposure  $H_0 = S/H_{sp}$  of a photosensitive array is (and, in standards, should be) determined by  $\Delta$  of the array. Nevertheless,  $H_0$ , the speed point exposure  $H_{sp}$  and the safety factor are no longer needed in the face of knowledge of  $H_{max}$  and  $H_{min}$  or  $H_{mid}$  and  $\Delta$  and should be eliminated. The sensitometric standard for solid-state arrays, ISO 12232-2006, and the OECF standard, ISO 14524-1999, should be changed because they use a photosensitive exposure range which is inappropriate for solid-state arrays. Finally, it shows that all current standards establish midtone reflectances  $R_{mid}$  for standard photosensitive arrays that are much less than oft-touted 18% making most references to 18% inappropriate.

## 7876-29, Poster Session

**Image enhancement technique using color and edge features for mobile imaging systems**

W. Cho, T. Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

The paper provides a method of selectively controlling the strength of image noise reduction (NR) and sharpening in regions associated with specific colors that are visually impact on the human visual system. The subjective color quality is often judged by how regions classified by specific colors look in the images through color characterization by the viewer. Our method controls the subjective quality of a specific color region by determining the strength of sharpening and NR operation. In addition, the proposed algorithm carries out the regional segmentation so that the algorithm selectively can control the strength of NR and sharpening for each region. However, since CIS SoC (CMOS image sensor System on Chip) products cannot use affordable memory due to the cost issue, the proposed algorithm suggests how the regional segmentation can be done without any line memory. In the proposed method, pixels are labeled and then, clustered by using the regional information and the color proximity. The edge information along with the color coded pixels is used for effective NR and sharpening of images. The main contribution of the proposed method includes (i) a memory color classifier in normalized color space for efficient hardware optimization, (ii) an edge-directed run-length filter (iii) for noise reduction and sharpening for visual appearance.

## 7876-30, Poster Session

**Rectangular pixels for efficient color image sampling**

T. Singh, M. Singh, Consultant (United States)

We present CFA designs that faithfully capture images with specified luminance and chrominance bandwidths. Previous academic research has mostly been concerned with maximizing PSNR of reconstructed images without regard to chrominance bandwidth and cross-talk. Commercial systems, on the other hand, pay close attention to both these parameters as well as to the visual quality of reconstructed images. They commonly sacrifice resolution by using a sufficiently

aggressive OLPF to achieve low cross-talk and artifact free images.

We introduce the Chrominance Bandwidth Ratio model that captures both the chrominance bandwidth and the cross-talk between the various signals. Next, we examine the effect of tuning photosite aspect ratio, a hitherto neglected design parameter. We derive panchromatic CFA patterns with provably minimum photo-site count for all values of the Chrominance Bandwidth Ratio.

An interesting outcome is a CFA design that captures full chrominance bandwidth, yet uses fewer photo-sites than the venerable color-stripe design. Another interesting outcome is a practical CFA design that captures chrominance at half the resolution of luminance using only 4 unique filter colors, that lends itself to efficient linear demosaicking, and yet vastly outperforms the Bayer CFA with the same photosite count, demosaicked with state of the art nonlinear algorithms.

## 7876-31, Poster Session

### A robust color signal processing with wide dynamic range WRGB CMOS image sensor

S. Kawada, R. Kuroda, S. Sugawa, Tohoku Univ. (Japan)

We have developed a highly accuracy and robust color reproduction by a simple calculation with a new color linear matrix using the formerly developed wide dynamic range WRGB LOFIC CMOS image sensor.

The image sensor was fabricated through a 0.18um 2-poly 3-Metal CMOS technology and has a 45 degrees oblique pixel array, the 4.2um effective pixel pitch and the W pixels. A W pixel was formed by replacing one of the two G pixels in the Bayer RGB color filter. The W pixel has a high sensitivity through the visible light waveband. An emerald-green and yellow (EGY) signals are generated from the difference between the W signal and the sum of RGB signals. This EGY signals mainly include emerald-green and yellow ingredients. These colors could not be reproduced accurately by the conventional linear matrix because their wave lengths are in the valley of the spectral sensitivity characteristics of RGB pixels. A new linear matrix based on the EGY-RGB signal was developed. Using this linear matrix, a highly accurate color processing with a large margin to the sensitivity fluctuation and noise has been achieved.

## 7876-32, Poster Session

### Adaptive contrast enhancement for underexposed images

S. Corchs, F. Gasparini, R. Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

In the present article we focus on enhancing the contrast of images with low illumination that present large underexposed regions. For these particular images, when applying the contrast enhancement techniques, we also introduce noise over-enhancement within the darker regions. Even if both the contrast enhancement and denoising problem have been widely addressed within the literature, these two processing steps are, in general, independently considered in the processing pipeline. Therefore, the goal of this work is to integrate contrast enhancement and denoise algorithms to proper enhance the above described type of images (for example night images or indoor images acquired with a short exposure time and/or high ISO setting). After applying the contrast enhanced method, and in order to selectively enhance the different regions of the underexposed images, we evaluate the saliency map of the image. At this point, the local increase of noise is estimated applying a proper noise measure. In a subsequent module, the denoise and final contrast correction are tuned with respect the strength of the contrast and consequent noise increase and the local saliency as well. An edge enhancement module is also included at the end of the pipeline to obtain the final enhanced image. The method has been applied to a proper database of underexposed images and compared with Retinex results.

## 7876-34, Poster Session

### Moving refractive optical low pass filter for digital cameras

M. Schöberl, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany); J. D. Ernst, W. Schnurrer, S. Föbel, Fraunhofer-Institut für Integrierte Schaltungen (Germany); A. Kaup, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany)

No abstract available.

## 7876-35, Poster Session

### A JPEG-like algorithm for compression of camera sensors images

O. Benahmed Daho, XLIM-SIC (France); M. Larabi, Univ. de Poitiers (France)

To reduce costs, digital cameras use a single sensor per pixel. A Bayer CFA filter (Color Filter Array) is generally used to recover only one color component per pixel. Subsequently, the images are first interpolated with a demosaicing process to reconstruct the full color picture prior to the compression stage for storage. This scheme is called the interpolation-first scheme.

In this work, we introduce the problem of CFA data compression. We propose to adapt the compression scheme to the demosaicing process, where the decoded data are directly used to reconstruct the full resolution color image.

## 7876-36, Poster Session

### Reduced reference image quality assessment based on statistics of edge

M. Zhang, W. Xue, X. Mou, Xi'an Jiaotong Univ. (China)

Objective Image Quality Assessment (IQA) model investigation is a hot topic in recent times. This paper proposed a novel and efficient universal Reduced Reference (RR) image quality assessment method based upon the statistics of edge discrimination. Firstly, binary edge maps created from the multi-scale wavelet transform modulus maxima were used as the low level feature to discriminate the difference between the reference and distorted image for IQA purpose. Then the gradient operator was applied on the binary map to produce the so called edge pattern map. The histogram of edge pattern map was used to verify the pattern of the edges of reference and distorted image, respectively. The RR features extracted from the histogram was used to discriminate the difference of edge pattern maps, and then form a new RR IQA model. Comparing to the typical RR model (Zhou Wang's method, 2005), only 12 features (96 bits) are needed instead of 18 features (162 bits) in Zhou Wang et al.'s method with better overall performance.

## 7876-37, Poster Session

### Evaluation of LED flash performance for camera phones

J. Pincenti, C. Sheldon, B. Richards, G. John, Motorola, Inc. (United States)

In this work, LED based flash solutions are evaluated for use in a camera phone application. The performance of a given flash solution is measured in terms of color accuracy and signal to noise ratio (SNR), both of which are standard test methods used in industry. Early in a camera phone design before completed camera modules are available, color accuracy and SNR are evaluated through a model which is based on knowledge of a given image sensor's color response as well as the power spectral distribution of the flash. Later in the design,

when working camera modules become available, the evaluation is performed through direct measurement. These direct measurements are also used to verify the results of the aforementioned model. Color accuracy, SNR, how they are related, and the compromise between the two are discussed as well as the efficiency at which electrical power is converted to light that is detectable by the image sensor. Though many issues remain to be investigated, measuring color accuracy and SNR provides an evaluation method that builds on developed techniques and provides a practical foundation for flash evaluation as it applies to the camera phone industry.

## 7876-39, Poster Session

### **Characterization of pixel crosstalk and impact of Bayer patterning by quantum efficiency measurement**

J. M. Vaillant, STMicroelectronics (France); C. Mornet, STMicroelectronics (France) and IMEP (France); T. Decroux, D. Herault, STMicroelectronics (France); I. Schanen, IMEP (France)

Development of small pixels for high resolution sensors implies a lot of challenges. A high level of performance should be guaranteed whereas the overall size must be reduced and so the degree of freedom in design and process. One key parameter of this constant improvement is the knowledge and the control of the crosstalk between pixels. In this paper, we present an advance in crosstalk characterization method based on the design of specific color patterns and the measurement of quantum efficiency. In a first part, we describe the color patterns designed to isolate one pixel or to simulate un-patterned colored pixels. These patterns have been implemented on test-chip and characterized. The second part deals with the characterization setup for quantum efficiency. Indeed, the use of spectral measurements allows us to discriminate pixels based on the color filter placed on top of them and to probe the crosstalk as a function of the depth in Silicon, thanks to the photon absorption length variation with the wavelength. In the last part, results are presented showing the impact of color filters patterning, i.e. pixels in a Bayer pattern versus un-patterned pixels. The crosstalk directions and amplitudes are also analyzed in relation to pixel layout.

# Conference 7877: Image Processing: Machine Vision Applications IV

Tuesday-Thursday 25-27 January 2011

Part of Proceedings of SPIE Vol. 7877 Image Processing: Machine Vision Applications IV

## 7877-08, Poster Session

### Vehicle detection using new AdaBoost features

H. Park, J. Kim, C. Lee, J. Jang, LED-IT Fusion Technology Research Ctr. (Korea, Republic of)

This paper presents an improvement to the object detection method of Viola and Jones, using the example of vehicle detection. Our method for training detectors is AdaBoost (adaptive boosting) using a new type of visual features which is based on maximally extremal point. Our features are faster and robust to partial visibility and clutter. The proposed method is expected to have various applications, such as vehicle detection, Road environment recognition, real-time image processing and so on.

## 7877-29, Poster Session

### Vehicle detection using DOM-FAST and support vector machine

J. Kim, LED-IT Fusion Technology Research Ctr. (Korea, Republic of); C. Lee, J. Jang, Yeungnam Univ. (Korea, Republic of); H. Park, LED-IT Fusion Technology Research Ctr. (Korea, Republic of)

In this paper, we present a novel vehicle detection algorithm using the difference of mean-features from accelerated segment test (DOM-FAST) and the support vector machine (SVM). In given test images, we detect firstly the interest feature points by the DOM-FAST algorithm. And then the local descriptor is computed by the contourlet transform (CT) at each feature point. The local descriptor based on the coefficients of CT represents the most significant information of the image patch around the feature points efficiently. Then the combinations of coefficients are applied as study samples to the SVM classifiers. Finally, the coefficients of testing images are used to test classifiers, and the vehicle detection results are obtained. The experiments are performed on various databases with our database.

## 7877-30, Poster Session

### Failures on atmospheric visibility measurements using digital image processing

A. Restrepo-Martínez, F. E. Lopez, Instituto Tecnológico Metropolitano (Colombia)

This paper studies some problems of visibility of the atmospheric visibility measurements when artificial vision is used to apply Beer's Laws and edges analysis.

Daytime digital images of Medellín downtown city were captured using a Canon EOS Rebel XSI camera. Regions of interest were extracted for each image, then using strategies of Beer's Laws the visibility were measured. Edges Analysis Sobel and Canny were done too.

Different values for the shutter speed, sensibility (ISO number) and f-number, were used to evaluate how the visibility values are affected. We found that for the same scene, change in these parameters produces variations in the visibility values; this fact is a serious problem. Then some strategies about how defined these parameters are necessary to implement for future studies.

The images with best contrast have many edges and highest values of visibility. In these cases lowest levels of matter particulate in the city could be hope.

In spite of the limitations of measurements obtained, the strategy monitoring the visibility with artificial vision has a high potential due to the low cost and the wide possibilities of the automation when machine vision is used.

## 7877-31, Poster Session

### Segmentation and visualization of anatomical structures from volumetric medical images

J. Park, S. Park, Mokpo National Univ. (Korea, Republic of); W. Cho, Chonnam National Univ. (Korea, Republic of)

This paper presents a method that can extract and visualize anatomical structures from volumetric medical images by using a 3D level set segmentation method and a hybrid volume rendering technique. First, the segmentation using the level set method was conducted through a surface evolution framework based on the geometric variation principle. This approach addresses the topological changes in the deformable surface by using the geometric integral measures and level set theory. These integral measures contain a robust alignment term, an active region term, and a mean curvature term. By using the level set method with a new hybrid speed function derived from the geometric integral measures, the accurate deformable surface can be extracted from a volumetric medical data set. Second, we employed a hybrid volume rendering approach to visualize the extracted deformable structures. Our method combines indirect and direct volume rendering techniques. Segmented objects within the data set are rendered locally by surface rendering on an object-by-object basis. Globally, all the results of subsequent object rendering are obtained by direct volume rendering (DVR). Then the two rendered results are finally combined in a merging step. This is especially useful when inner structures should be visualized together with semi-transparent outer parts. This merging step is similar to the focus-plus-context approach known from information visualization. Finally, we verified the accuracy and robustness of the proposed segmentation method for various medical volume images. The volume rendering results of segmented 3D objects show that our proposed method can accurately extract and visualize human organs from various multimodality medical volume images.

## 7877-32, Poster Session

### Extraction and fusion of spectral parameters for face recognition

Z. Abdessalem, B. Billiot, P. Gouton, J. Y. Hardeberg, Univ. de Bourgogne (France)

I. Introduction: Many methods have been developed in image processing for face recognition, especially in recent years with the increase of biometric technologies. However, all these techniques are mainly used on grayscale images acquired in the visible range of the electromagnetic spectrum.

The aims of our study are to improve existing tools and to develop new methods for face recognition. The techniques used takes advantages of the different spectral ranges, the visible, optical infrared and thermal infrared, by either combining them or analyzing them separately in order to extract the most appropriate information for face recognition.

II. Methods: Our study use three acquisition devices: a digital SLR camera, an infrared camera (800-2600 nm) and a thermal camera (12000 nm).

Several face recognition techniques are based on algorithms of facial features, in order to characterize them. Firstly, we apply the SIFT algorithm [1] (Lowe, 2004) (Scale Invariant Feature Transform) on a

database of visible, infrared and thermal images of faces acquired with our equipment. Such algorithm determines the characteristic points of the images and defines them by characteristic vectors.

To validate the results, the ASIFT algorithm [2] (Morel and Yu, 2009), a derivative of SIFT invariant to affine transformation, is applied to the same database.

III. Results: By analyzing results, we notice that the number of characteristic points in the infrared range is more important (an average of 152 points/kind of person (SIFT) and 242 points/ kind of person (ASIFT)), this number is half reduced in the visible spectrum (an average of 75 points/person (SIFT) and 151 points/person (ASIFT)). In the thermal range, the characteristic points are lower compare to the previous spectrum range (an average of 43 points/person (SIFT) and 57 points/person (ASIFT)).

The other results will be obtain by a comparison between 2 different faces in same range. We can note that in the infra-red spectrum and thermal the number of characteristic points between two different person is reduced (an average of 1 point/even (SIFT) and 1 point/ even (ASIFT)) on the other hand the visible spectrum presents more characteristic points (an average of 3 points/even (SIFT) and 11 points/ even (ASIFT)).

IV. Conclusion: These initial results show well that the infra-red spectrum is most adequate to ensure a higher rate of recognition with the extraction of characteristic point method. However the use of the thermal band is not well adapted for this kind of algorithm.

Currently, several research groups study the use of methods such as segmentation and face recognition (Gabor wavelet) are being implemented. The results aim to determine whether the contribution of information contained in images acquired in infrared range can improve the performance of methods used.

The fusion of information contained in each type of image is also planned to characterize the face in the most discriminating possible.

## 7877-33, Poster Session

### Monitoring plant growth using high resolution micro-CT images

V. C. Paquit, S. S. Gleason, U. C. Kalluri, Oak Ridge National Lab. (United States)

A multidisciplinary research conducted at the Oak Ridge National Laboratory aims at understanding the molecular controls of partitioning, transport and fate of carbon fixed by photosynthesis in plants and its correlation with other measured plant system properties. Ultimately, we intend to develop a modeling framework to assess, correlate and predict as to which spatiotemporal changes in system dynamics are key to predicting emergent properties of system.

Within this research, this paper relates to the quantitative morphological measurements of the main structures forming a plant (stem, roots, and leaves), their internal sub-structures, and changes occurring overtime.

## 7877-35, Poster Session

### Automating the estimation of coating thickness measurements in the ball crater technique

J. Huang, The City Univ. (United Kingdom) and Teer Coatings Ltd. Berry Hill Industrial Estate (United Kingdom); P. Liatsis, The City Univ. (United Kingdom); K. Cooke, D. Teer, Teer Coatings Ltd. (United Kingdom)

No abstract available

## 7877-01, Session 1

### Lipschitz exponents based signal restoration

B. Jalil, O. Beya, E. Fauvet, O. Laligant, Univ. de Bourgogne (France)

In this work, we attempt to propose a signal restoration technique from the noise corrupted signal. The scope of the work has also been expanded to the identification of singularities inside the signal as well. As these singularities induce strong coefficients, therefore the adjacent small singularity cannot be detected during the denoising process. The main aim of the propose method is to address such types of problem without introducing further spurious peaks or oscillations. In order to overcome this problem we are proposing a multi segmentation criterion to separate the noise elements from these significant singular points. At the same time, the problem of the influence of strong singularities on adjacent weaker singular points can be solved by using multi segmentation approach. In multi segmentation approach, we are proposing a recursive algorithm which identifies the weakest require singular point in the signal.

## 7877-02, Session 1

### Real-time wavelet-based inline banknote-in-bundle counting for cut-and-bundle machines

V. Lohweg, Ostwestfalen-Lippe Univ. of Applied Sciences (Germany); J. Schaede, T. Türke, KBA-GIORI S.A. (Switzerland); E. Gillich, Ostwestfalen-Lippe Univ. of Applied Sciences (Germany); D. Petker, OWITA GmbH (Germany); H. Willeke, KBA-Bielefeld (Germany)

Automatic banknote sheet cut-and-bundle machines are widely used within the scope of banknote production. Beside the cutting and bundling features which are today a mature technology, image-processing-based quality inspection for this type of machine becomes attractive. We present in this work a new real-time touchless counting and cutting blade quality insurance system, based on a Color-CMOS-Camera and a dual-core Computer, for cut-and-bundle applications in banknote production. The system which applies Wavelet-based multiscale filtering is able to count banknotes inside a 100-bundle within 200-300 ms depending on the window size.

## 7877-03, Session 1

### A robust segmentation and tracking method for characterizing GNSS signals reception environment

A. Cohen, C. Meurie, Y. Ruichek, Univ. de Technologie de Belfort-Montbéliard (France); J. Marais, Univ. Lille Nord de France (France)

This paper is focused on the characterization of GNSS signals reception state by new image processing techniques. The main aim of the application consists to detect satellites situated in sky region (with direct reception state) in fish-eye images. This proposed strategy is composed by four steps: 1/ a new adaptive and automatic segmentation method combining color and texture information. 2/ a classification step by the k-means algorithm for determining the sky and non-sky regions. 3/ a calibration and rectification stage; 4/ a region-tracking method based on a block-matching estimation that reduces the execution time of the application in order to approach the real-time constraints. The tracking results are compared to the results of the classification method on a large real database. The evaluation shows that the proposed method has a very low error, reaching a good classification rate of 90% (vs 94.2% obtained in previous works) and decreasing the execution time of the application by ten times.

## 7877-04, Session 2

### Accurate, fast, and robust centre localisation for images of semiconductor components

F. Timm, Univ. zu Lübeck (Germany) and Pattern Recognition Co. GmbH (Germany); E. Barth, Univ. zu Lübeck (Germany)

We propose two novel approaches for the precise centre localisation of circular objects, e.g. p-electrodes of light emitting diodes.

The first approach is based on image gradients for which we provide an objective function that is solely based on dot products and thus being maximised by gradient ascend.

The second approach is inspired by the concept of isophotes for which we derive an objective function that is based on the definition of radial symmetry.

We compare the accuracy and the runtime to the Hough transform for artificial images with several kinds of noise and images of semiconductor components with occlusions and strong image noise.

The radial symmetry approach prove to be the most robust one, especially for low contrast images with strong noise with a mean error rate of 0.86 pixel for artificial images and 0.98 pixel for real world images. The gradient approach yields more accurate results for almost all images (mean error 4 pixel) compared to the Hough transform (mean error 8 pixel).

Concerning the runtime, the gradient-based approach significantly outperforms the other approaches with a reduction of 79% compared to the Hough transform (100%), whereas the radial symmetry approach yields a reduction of still 12%.

## 7877-09, Session 3

### Vision based forest smoke detection using analyzing of temporal patterns of smoke and their probability models

S. Ham, B. Ko, J. Nam, Keimyung Univ. (Korea, Republic of)

In general, since smoke appears before flames, smoke detection is particularly important for early fire detection systems. To detect fire-smoke using video camera is a difficult work because main characteristics of a smoke are uncertain, vague, constant patterns of shape and color. Thus, this paper proposes a new fire-smoke detection method, especially forest smoke using analyzing of temporal patterns of smoke and Fuzzy Finite Automata (FFA). To consider the smoke characteristics over time, the temporal patterns of intensity entropy, wavelet energy and motion orientation have been used for generating, multivariate Gaussian probability density functions (PDFs) are applied Fuzzy Finite Automata (FFA) for smoke verification. The proposed FFA consist of a set of fuzzy states (VH, H, L, VL), and a transition mapping that describes what event can occur at which state and resulting new state. For smoke verification, FFA is most appropriate method in case variables are time-dependent and uncertain. The proposed algorithm is successfully applied to various fire-smoke videos and shows a better detection performance.

## 7877-10, Session 3

### Estimation of fire volume by stereovision

T. Molinier, L. Rossi, A. Pieri, Univ. di Corsica Pasquale Paoli (France); M. A. Akhloufi, Ctr. of Robotics and Vision (Canada); Y. Tison, Univ. di Corsica Pasquale Paoli (France)

This paper presents the estimation of fire front volume in the context of laboratory experiments of fire spreading on inclinable table. The method is based on the use of two pre calibrated synchronized stereovision systems positioned respectively in a back position and in a front position of the fire propagation direction. The two vision systems are used to obtain complementary 3D fire points that are expressed in

a single reference system and that will be used to obtain a global form of the fire. To determine the coordinate transformations of each system with respect to a single reference system a calibration procedure is carried out with a cube pattern. The same technique is used to perform a 3d mesh with the several sets of 3d points and to connect those sets between them. 3Dshape reconstruction and estimation of the volume of the fire are finally estimated.

## 7877-11, Session 3

### Pavement distress detection and severity analysis

E. Salari, G. Bao, The Univ. of Toledo (United States)

#### Summary:

Automatic recognition of road distresses is a challenging research area since it reduces economic loses before cracks and potholes become too severe. However, due to factors such as complex texture, uneven illumination, and non-uniform background, pavement distress detection turns out to be a very difficult problem rather than a simple edge detection process. In this paper, a novel automatic pavement crack detection approach based on advanced image processing techniques is proposed. The proposed method can provide real-time pavement distress detection and its evaluation based on images captured from a camera installed at the front of a testing vehicle. The entire detection process consists of two main phases. The first phase is pavement surface extraction and the second phase involves pavement distress detection and evaluation. In pavement surface extraction, a novel color segmentation method based on a neural network is applied to separate the road surface from background features, such as houses, bushes, grass and trees. After road segmentation is accomplished, a pavement distress detection algorithm based on probabilistic relaxation is executed to obtain the skeleton of the cracks. Then a new pavement distress classification algorithm based on neural networks is introduced to assign the cracks into different types and severity groups according to the geometrical and topological parameters obtained in the crack detection step. The proposed road inspection system can precisely detect various distress features and estimate its severity from a regular outdoor scene image. Simulation results show the method is effective and robust in the extraction of cracks in a variety of pavement images.

## 7877-28, Session 3

### Line segment based structure and motion from two views

S. Mosaddegh, A. Fazlollahi, D. Fofi, Univ. de Bourgogne (France); P. Vasseur, Univ. de Picardie Jules Verne (France)

We present an efficient measure of overlap between two co-linear segments which considerably decreases the overall computational time of a Segment-based motion estimation and reconstruction algorithm already exist in literature. We also discuss the special cases where sparse sampling of the motion space for initialization of the algorithm does not result in a good solution and suggest to use dense sampling instead to overcome the problem. Finally, we demonstrate our work on two real data sets.

## 7877-14, Session 4

### Multi-frame face recognition with a discriminant analysis and decision level fusion

S. Yeom, H. Lee, Daegu Univ. (Korea, Republic of)

Currently, CCTV and DVR systems are widely installed for security and surveillance. However, the video stream transferred from the widely used monitoring system has comparably low resolution and poor quality since it usually operates in harsh conditions. The system covers

certain region of interest from a distance. The image information is acquired without the consideration of desirable imaging conditions such as illumination and focusing. Moreover, dramatic change of real words often generates blurring effects on the image. Therefore, face recognition with the CCTV system is a difficult task. Although there have been research to increase the low resolution using multi-frames for face recognition, it usually requires high computational load and capacity.

In this paper, a face recognition method based on photon-counting linear discriminant analysis and decision level fusion is discussed. The photon-counting linear discriminant analysis realizes the Fisher's criterion with the Poisson probability model without suffering from the singularity problem of Fisher linear discriminant analysis. A video surveillance system provides multi-frame images of low quality. Therefore, we can utilize a sequence of images by performing decision level fusion to improve the face recognition performance. This face recognition technique is shown to be robust to low resolution and noise environments. In the experiments, simulated data and low resolution facial images are tested to verify the performance of the proposed method. The accuracy rate and the false alarm rate are obtained to compare the results with conventional techniques. The proposed system shows potentials of using photon-counting linear discriminant analysis and decision level fusion combined for widely available low-end surveillance systems.

#### 7877-15, Session 4

### Pose-robust face recognition using shape-adapted texture features

T. Gernoth, A. Goossen, R. Grigat, Technische Univ. Hamburg-Harburg (Germany)

Unconstrained environments with variable ambient illumination and changes of head pose are still challenging for many face recognition systems. To recognize a person independent of pose, we first fit an active appearance model to a given facial image. Shape information is used to transform the face into a pose-normalized representation. We decompose the transformed face into local regions and extract texture features from these not necessarily rectangular regions using a shape-adapted discrete cosine transform. We show that these features contain sufficient discriminative information to recognize persons across changes in pose. Furthermore, our experimental results show a significant improvement in face recognition performance on faces with pose variations when compared with a block-DCT based feature extraction technique in an access control scenario.

#### 7877-16, Session 5

### A novel framework for white blood cell segmentation based on stepwise rules and morphological features

J. Gim, J. Park, J. Lee, B. Ko, J. Nam, Keimyung Univ. (Korea, Republic of)

In automatic cell analysis using image processing, WBC segmentation is the most important procedure, where the ultimate goal is to extract all the WBCs from a complicated background and then only segment the WBCs into morphological components, such as the nucleus and cytoplasm.

This study proposes a new WBC segmentation method using region merging scheme and GVF (Gradient Vector Flow) snake. WBC segmentation consists of two schemes; nuclei segmentation and cytoplasm segmentation. For nuclei segmentation, we create a probability map using probability density function estimated from samples of WBC's nuclei and crop the sub-images to include nucleus by using the fact that nuclei have salient color against background and red blood cells. Then, mean-shift clustering is performed for region segmentation and stepwise rule is applied to merge particle clusters to nucleus. For cytoplasm segmentation, we create saturation map within the sub-images based on the fact that cytoplasm has higher saturation

than red blood cell. Then, the GVF snake is applied to gradient obtained from the saturation map. Finally, GVF forces to guide snakes to deform to cytoplasm boundary edges.

#### 7877-17, Session 5

### Contour extraction and amendments of left ventricle short axis from heart ultrasonic image sequences

X. Yang, Chongqing Univ. of Posts and Telecommunications (China)

In the diagnosis for heart disease, people are usually very concerned about the movement situation of ventricle, atria and vascular cavity. The diastolic and systolic movement of heart and vascular cavity represents each component's performance of heart body, and it offers an important reference for disease diagnosis. In the section of ultrasonic images, the conformation of heart structure's contour precisely describes their movement features. However, the extraction of these contour lines is very difficult.

The ventricular movement is also called contour moving of ventricle. This article presents an automatic detection method according to the characteristics of ultrasonic images. The method can also use artificial intervention algorithm to get image contour if necessary.

At first, for the first frame from the selected images, the characteristic points O and P on the outline are manually designated. And then, the point P is treated as a starting position, and the contour is searched. The radial gradient-based searching method is used to detect the borderline, after finding the discrete points on the outline, then, the smooth contour curves are obtained by the method of curve fitting.

To get the outline of the following frame, one needs to consider continuity of heart's movement. The outline's difference between adjacent frames will not be too great, so that the following frame's outline after the second one is extracted on the base of the former one.

In addition to the cardiac diastolic and systolic movement, it also includes rotational and translational motion caused by blood shock and muscle traction.

In recent years, clinicians pay more attention to detect ventricular diastolic function of patients suffered with coronary heart disease. In order to provide a more intuitive reference for doctors, one should amend heart's translation.

#### 7877-18, Session 6

### Non-parametric texture defect detection using Weibull features

F. Timm, Univ. zu Lübeck (Germany) and Pattern Recognition Co. GmbH (Germany); E. Barth, Univ. zu Lübeck (Germany)

We propose a novel, non-parametric, local approach for defect detection in texture images with only two features. We compute the two parameters of a Weibull fit of the gradient magnitudes within a local window. Then, we perform a simple novelty detection algorithm to detect arbitrary deviations of the normal texture.

Therefore, we compute the Euclidean distance of the parameters of the local windows to a reference point that is learnt during training.

Our approach is independent of the present type of texture and also independent of the defect type.

For performance evaluation we use the highly challenging dataset proposed at the DAGM 2007 with different classes of textures and different types of defects.

The Weibull parameters can detect local deviations within different types of textures with an error rate of less than 5% using defect-free images for training. Compared to existing approaches such as Gabor filters or grey level statistics, our novel approach is not only powerful but also very efficient.



## 7877-19, Session 6

### Machine vision applied to industrial quality control of artificial teeth: lighting methodology and image enhancement

J. W. Branch, Univ. Nacional de Colombia Sede Medellín (Colombia); A. Restrepo-Martínez, Instituto Tecnológico Metropolitano (Colombia); E. Mesa-Múnera, J. F. Ramírez-Salazar, P. Atencio, E. Franco, Univ. Nacional de Colombia Sede Medellín (Colombia); O. Franco, R. Carmona, H. Rodríguez, New Stetic S.A. (Colombia)

This paper considers the quality control of polymeric resin artificial teeth, which are manufactured from two different layers. It will be taken into account only dark particles detection in teeth surfaces, which are called particle defects. A special lighting system was designed with features that consider the size and shape of teeth. For that purpose, some LED arrays were assembled in various configurations: Direct diffused light and backlight, which allow us to define the Region of Interest (ROI), to avoid brightness on the surface and to generate a homogeneous illumination of the teeth. However, in order to avoid saturation of the sensor, it was necessary to implement an intensity control system according to the size of the tooth. In the stage of image acquisition, different amateur cameras were evaluated due to its low cost, but also the performance of the processing tools of a SMART camera from National Instruments was compared. Additionally to lighting, an adaptive contrast enhancement stage was performed in order to improve the condition of some images in which lighting methods used do not improve the scene. As future work it will be necessary to integrate some of the implemented techniques.

## 7877-25, Session 6

### Quantitative measurement by artificial vision of small bubbles in flowing mercury

V. C. Paquit, M. W. Wendel, D. K. Felde, Oak Ridge National Lab. (United States)

At the Spallation Neutron Source (SNS), an accelerator-based neutron source located at the Oak Ridge National Laboratory (Tennessee, USA), the production of neutrons is obtained by accelerating protons against a mercury target. This self cooling target, however, suffers rapid heat deposition by the beam pulse leading to large pressure changes and thus to cavitations that may be damaging to the container. In order to locally compensate for pressure increases, a small-bubble population is added to the mercury flow using gas bubblers. The geometry of the bubblers being unknown, we are testing several bubblers' configurations and are using machine vision techniques to characterize their efficiency by quantitative measurement of the created bubble population.

In this paper we thoroughly detail the experimental setup and the image processing techniques used to quantitatively assess the bubble population. To support this approach we are comparing our preliminary results for different bubblers and operating modes, compare the efficiency of our method to fluid dynamics theory, and discuss potential improvements.

## 7877-21, Session 7

### Coded source neutron imaging

P. R. Bingham, Oak Ridge National Lab. (United States)

While neutron radiography in general is not a new technology with first images taken in the 1930s, recent innovations in neutron imaging techniques have increased the applicability of neutron radiography to complex science and engineering problems such as stress measurements [1], magnetic field measurements [2], and fuel cell research and development [3] to name a few. These extensions have improved the applicability of neutron radiography, but system

resolution is still limited by the detection systems to around 50um with an integrating detector. Conventional neutron radiography is also limited today by the count rate at the detector restricting time resolved measurements. Since neutron beams are marginally-diffracting or refracting at microscopic and macroscopic scales, neutron optics that can magnify or de-magnify (i.e., focus) imaged objects are difficult to create and require expensive designs and materials [4]. Current improvements in detector technology are pushing resolution down to the 10-15um mark [5] with high-cost detectors, but there is a clear demand for resolutions of 1 um or less that would dramatically extend the application of neutron imaging to micro-scale structures such as microchannel heat exchangers, fuel cell components, biological microscopy for pharmacology, drug delivery research, fuel injector sprayers for efficient diesel engine technology, and biofuels research.

One possible route to achieve resolutions on this scale is the novel application of coded aperture imaging to neutron radiography. Coded aperture imaging is a technique for imaging non-diffracting and refracting sources that has been implemented by the Astronomy community since the 1960s. Pinhole cameras, such as those commonly employed in neutron radiography, have limited flux due to the small size of the pinhole. A coded aperture system exhibits the resolution of a pinhole-style camera but with collection efficiency proportional to the number of pinholes in the aperture. The effect of a coded aperture can be deconvolved from the measured image allowing high collection efficiency and high spatial resolution [6, 7] in a magnifying imaging configuration.

In this presentation, we will investigate the theoretical limitations for coded source neutron radiography, present a coded source design implemented at a prototype neutron imaging instrument at Oak Ridge National Laboratory (ORNL), discuss image reconstruction methods for this type of imager, and show initial resolution measurement results from the ORNL instrument.

1. Penumadu, D., "Material Science and Engineering with Neutron Imaging," Neutron Imaging and Applications, Springer, New York, 2009.
2. Manke, I., Kardjilov, N., Hilger, A., Stroble, M., Dawson, M., and Banhart, J., "Polarized neutron imaging at the CONRAD instrument at Helmholtz Centre Berlin," Nuclear Instruments and Methods in Physics Research A, 605 (2009) 26-29.
3. Arif, M., Hussey, D.S., and Jacobson, D.L., "Neutron Imaging for the Hydrogen Economy," Neutron Imaging and Applications, Springer, New York, 2009.
4. Beguiristain, H. R., Anderson, I.S., et al. "A simple neutron microscope using a compound refractive lens," Applied Physics Letters 81(22) (2002) 4290-4292.
5. Tremsin, A.S., Vallerger, J.V., McPhate, J.B., Siegmund, O.H.W., Feller, W.B., Crow, L., and Cooper, R.G., "On the possibility to image thermal and cold neutrons with sub-15 m spatial resolution," Nuclear Instruments and Methods in Physics Research A, 592 (2008) 374-384.
6. Damato, A.L., Binns, P., and Lanza, R.C., "Progress Report on a Neutron Coded Source Phase Contrast Imaging System at the MIT Reactor," IEEE Nuclear Science Symposium Conference Record, (2007) 1725-28.
7. Coakley, K.J. and Hussey, D.S., "Feasibility of single-view coded source neutron transmission tomography," Meas. Sci. Technol. 18 (2007) 3391-3398.

## 7877-22, Session 7

### Toward autonomic computing in machine vision applications: techniques and strategies for in-line 3D reconstruction in harsh industrial environments

J. Molleda, R. Usamentiaga, D. F. García, F. G. Bulnes, Univ. de Oviedo (Spain)

Nowadays machine vision applications require skilled users to configure, tune, and maintain. Because such users are rarely found, this usually means that robustness and reliability of applications is significantly affected. Autonomic computing offers a set of principles which can be used to partially overcome these problems, such as self-

monitoring, self-regulation, and self-repair.

Systems which include self-monitoring property observe its internal state, and extract features about it. Systems with self-regulation are capable of regulating its internal parameters to provide the best quality of service depending on the operational conditions and environment. Finally, self-repairing systems are able to detect anomalous working behavior and to provide strategies to deal with such conditions.

Machine vision applications are the perfect field to apply the autonomic computing techniques. This type of applications has strong constraints on reliability and robustness, especially when working in industrial environments, and must provide accurate results even under changing conditions such as variable luminance, or noise.

In order to exploit the autonomic approach of a machine vision application, we believe the architecture of the system must be designed using a set or orthogonal modules.

In this paper, we describe how autonomic computing techniques can be applied to machine vision systems, using as example a real application: 3D reconstruction in harsh industrial environments based on laser range finding.

The application is designed based on modules with different responsibilities at three layers: image processing (low level), monitoring (middle level) and supervision (high level).

High level modules supervise the execution of lower level modules and, based on the information gathered by middle level modules, regulate lower level modules in order to optimize the global quality of service, and tune the module parameters based on operational conditions and the environment.

Regulation actions involve modifying the exposure time of the image sensor due to changes in lighting conditions, or modifying the laser extraction method to meet continuous deadlines due to changes in speed manufacturing.

## 7877-23, Session 7

### Evaluating distances using a coded lens camera and blur metrics

L. Angot, C. Chang, Y. Chen, Industrial Technology Research Institute (Taiwan)

A method and a system are proposed to measure distances from a reference point to a target and at the same time obtaining an image of the target. The system is based on a wavefront coded lens and an image processing unit. The method consists in capturing a series of images of the target, computing a blur metric of the images in order to obtain calibration data, and obtaining the distance to the target for any position of the later. The method and system are easy to manufacture and provide an alternative to other distance measuring methods and devices, while also producing an image of the scene. The target is a printed image of pseudo random black and white elements which can be stucked to objects for distance evaluation. Further investigations are under way to evaluate distance to particular objects. The distance precision is less than 3cm over a 16cm to 120cm range. Other ranges can be selected by changing the focal distance of the lens.

## 7877-24, Session 7

### Automatic firearm class identification from cartridge cases

S. Kamalakannan, Texas Tech Univ. (United States); C. J. Mann, P. R. Bingham, T. P. Karnowski, S. S. Gleason, Oak Ridge National Lab. (United States); H. Sari-Sarraf, Texas Tech Univ. (United States)

We present a machine vision system for automatic identification of the class of firearms by extracting and analyzing two significant properties from spent cartridge cases, namely the Firing Pin Impression (FPI) and the Firing Pin Aperture Outline (FPAO). Within the framework of the proposed machine vision system, a white light interferometer is employed to image the head of the spent cartridge cases. As a first

step of the algorithmic procedure, the Primer Surface Area (PSA) is detected using a circular Hough transform. Once the PSA is detected, a customized statistical region-based parametric active contour model is initialized around the centre of the PSA and evolved to segment the FPI. Subsequently, the scaled version of the segmented FPI is used to initialize a customized Mumford-Shah based level set model in order to segment the FPAO. Once the shapes of FPI and FPAO are extracted, a shape-based level set method is used in order to compare these extracted shapes to an annotated dataset of FPIs and FPAOs from varied firearm types. A total of 74 cartridge case images non-uniformly distributed over five different firearms are processed using the aforementioned scheme and the promising nature of the results (95% classification accuracy) demonstrate the efficacy of the proposed approach.

## 7877-34, Session 7

### Generation of biologically motivated artificial retina tessellations $\log(z)$ and $\log(z+a)$ and Point based matching performance evaluation on backprojected response (V1) on retina domain

I. Ram, P. Siebert, Univ. of Glasgow (United Kingdom)

We present the results of an investigation that compares matching local SIFT-like image features; extracted from a software-based retina model, to matching standard SIFT image features. Our retinas, conformal and non-conformal is sampled by receptive fields(RF) which are organised at a high density in the central foveal region of the retina and at a sparse resolution in the surrounding periphery similar to that found in biological vision. We have also shown the point based and variable Gaussian kernel based overlapping sampling response quality. Multi-resolution, space-variant visual information is extracted on a scale-space continuum and SIFT-like interest point descriptors are extracted that represent the visual appearance of local regions. This paper also describes the design, implementation and initial evaluation of space variant artificial  $\log(z)$  and  $\log(z+\alpha)$ retina tessellations comprising circular overlapping RF model. We compare the matching performance of the backprojected response on standard SIFT by plotting receiver operating character (ROC) curve. While the primary objective of retina SIFT is to reduce feature data rates while focusing attention in the context of visual search, our preliminary matching results indicate that standard SIFT in fact outperformed on  $\log(z+\alpha)$ response by 5% and 4% at false alarm rates set to 10% and 20% respectively.

# Conference 7878: Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques

Monday-Tuesday 24-25 January 2011

Part of Proceedings of SPIE Vol. 7878 Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques

## 7878-01, Session 1

### Software framework for nano and micro scale measurement applications

J. Rönning, V. Tuhkanen, R. Sipola, T. J. Vallius, Univ. of Oulu (Finland)

Development of new instruments and measurement methods has advanced research in field of nanotechnology. Development of measurement systems used in research requires support from reconfigurable software.

Application frameworks can be used to develop domain-specific application skeletons. New applications are specialized from framework by filling its extension points.

This paper presents application framework for nano and micro scale applications. Framework consists of implementation of robotic control architecture and components that implement features available in measurement applications. To ease the development of user interfaces for measurement systems, framework also contains ready-to-use user interface components.

The goal of the framework was to ease the development of new applications for measurement systems. Features of implemented framework were examined through two test cases. Benefits gained by using the framework were analyzed by determining work needed to specialize new applications from the framework. Also degree of reusability of specialized applications was examined.

The work shows that developed framework can be used to implement software for measurement systems and that major part of the software can be implemented by using reusable components of the framework. When developing new software, developer only needs to develop components related to hardware used and performing measurement task. Using the framework developing new software takes less time. The framework also unifies structure of developed software.

## 7878-02, Session 1

### A traffic situation analysis system

O. Sidla, SLR Engineering OG (Austria); M. Ulm, Austrian Institute of Technology (Austria); M. Rosner, SLR Engineering OG (Austria); N. Braendle, Austrian Institute of Technology (Austria)

The observation and monitoring of traffic with smart visions systems for the purpose of improving traffic safety has a big potential. For example embedded vision systems built into vehicles can be used as early warning systems, or stationary camera systems can modify the switching frequency of signals at crossings or warn vehicles of pedestrian traffic at intersections.

The automatic analysis of traffic patterns is still in its infancy - the complexity of vehicle motion and pedestrian flow in a complex environment is too complex to be fully understood by a vision system.

We present steps towards such a traffic monitoring a system which is designed to detect potentially dangerous traffic situations, especially incidents in which the interaction of pedestrians and vehicles might lead to dangerous or even critical encounters.

The proposed system consists of a cluster of 3 smart cameras which are based on very compact PC hardware running a Linux operating system. Two cameras run vehicle detection software including license plate detection and recognition, one camera runs a complex pedestrian detection and tracking module:

Cameras 1,2: Real-time (25 Hz) license plate detection with fast segmentation and character recognition based on the HOG principle.

We describe the outline of our very fast license plate detector including statistical results based on a large dataset of characters and testing

images.

Camera 3: Pedestrian tracking with a tracking-by-detection approach based on a cascaded HOG detector. Training of well suited feature sets is crucial in order to achieve good detection rates and fast run-time performance, but also an optimized implementation contributes to an effective use of available hardware resources. We describe the processing and training pipeline of the pedestrian tracking system which is real-time capable on standard PC hardware.

No need of GPU or FPGA processing is required by our approach in order to achieve usable tracking frame rates on VGA Resolution even on low profile Intel Atom processors.

The remaining paper concentrates on the system architecture and describes results of our experiments during extensive trials and tests in an outdoor environment in the City of Vienna, Austria: the traffic monitoring system is installed at a double street intersection location which is especially dangerous for pedestrians. The operation of the system is assessed in extensive ground truth evaluation campaign which uses the same video data as the system itself.

## 7878-03, Session 1

### The 18TH Annual Intelligent Ground Vehicle Competition: trends and influences for intelligent ground vehicle control

B. L. Theisen, W. Smuda, P. A. Frederick, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States)

The Intelligent Ground Vehicle Competition (IGVC) is one of four, unmanned systems, student competitions that were founded by the Association for Unmanned Vehicle Systems International (AUUVSI). The IGVC is a multidisciplinary exercise in product realization that challenges college engineering student teams to integrate advanced control theory, machine vision, vehicular electronics and mobile platform fundamentals to design and build an unmanned system. Teams from around the world focus on developing a suite of dual-use technologies to equip ground vehicles of the future with intelligent driving capabilities. Over the past 18 years, the competition has challenged undergraduate, graduate and Ph.D. students with real world applications in intelligent transportation systems, the military and manufacturing automation. To date, teams from over 75 universities and colleges have participated. This paper describes some of the applications of the technologies required by this competition and discusses the educational benefits. The primary goal of the IGVC is to advance engineering education in intelligent vehicles and related technologies. The employment and professional networking opportunities created for students and industrial sponsors through a series of technical events over the four-day competition are highlighted. Finally, an assessment of the competition based on participation is presented.

## 7878-04, Session 2

### Stereo matching based on two cameras and one 3D image sensor

L. Yang, X. Shao, R. Shibasaki, The Univ. of Tokyo (Japan); R. Wang, Jilin Univ. (China)

Due to the problems of noise, textureless region and depth discontinuity in stereo matching, a new matching method based on two cameras and one 3D image sensor is proposed in this paper. The 3D image sensor can offer an intensity image and a depth map. The intensity image is used in calibration of image pairs and depth map.

After calibration, the depth map is transformed to an initial disparity map. With the constraint of this initial disparity map, the left and right images are matched by using the normalized covariance operator. The disparity searching range can be reduced from 80 pixels to 10 pixels. It can largely improve the matching accuracy and decrease the running time. Furthermore, the disparity maps with left and right image views are checked by left-and-right consistency. The experiment results indicate that the proposed algorithm performs well and the disparity map has more accuracy comparing with existing methods. The research achievement has a good prospect in application.

## 7878-06, Session 2

### Linear stereo vision based objects detection and tracking using spectral clustering

S. Moqqaddem, Y. Ruichek, Univ. of Technology of Belfort-Montbéliard (France); R. Touahni, A. Sbihi, Ibn Tofail Univ. of Kénitra (Morocco)

Objects detection and tracking is a key function for many applications like video surveillance, robotic, intelligent transportation systems, etc. This problem is widely treated in the literature in terms of sensors (video cameras, laser range finder, Radar) and methodologies. This paper proposes a new approach for detecting and tracking objects using stereo vision with linear cameras. After the matching process applied to edge points extracted from the images, the reconstructed points in the scene are clustered using spectral analysis. The obtained clusters are then tracked throughout their center of gravity using a Kalman filter and a NN based data association algorithm. The approach is tested and evaluated on real data to demonstrate its effectiveness for obstacle detection and tracking in front of a vehicle. This work is a part of a project that aims to develop advanced driving aid systems, supported by the CPER, STIC and Volubilis programs.

## 7878-07, Session 2

### Implementation of stereo vision on GPU for intelligent ground vehicle navigation in the presence of obstacles

C. Gamache, T. Padir, Worcester Polytechnic Institute (United States)

This paper discusses an image segmentation algorithm that uses a self organizing map (SOM) based color reduction with simulated annealing (SA) based color clustering implemented on a Graphics Processing Unit (GPU) for intelligent ground vehicle navigation in the presence of obstacles. The algorithm uses a neural network variant called a self-organizing map which is trained to perform a nonlinear color reduction. The simulated annealing is then used to group the SOM into clusters to produce the segmented image. The paper specifically discusses the computational method used to modify the SOM-SA algorithm to be run on the GPU. The original algorithm requires a new SOM to be generated for each image increasing the computational burden. To avoid this serial operation at runtime, one master SOM on different images under various lighting conditions can be trained and color clustered. The experimental results show that this modification improved the runtime of the algorithm on the GPU from approximately 3 minutes to 8 ms. The implementation of the computationally heavy segmentation algorithm on the GPU improves the autonomous navigation for the intelligent ground vehicle being developed at Worcester Polytechnic Institute in the presence of obstacles .

## 7878-08, Session 2

### Probabilistic recognition of person reoccurrence for visual surveillance of pedestrian flows

L. Paletta, G. Fritz, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria)

Surveillance tasks are ubiquitous in security services and in public transportation. However, the coverage of the sensor network is often far from being complete due to unobservable regions. Recognition of person reoccurrence is mandatory to enable wider areas with continuous coverage for tracking. The proposed work combines the already proved use of color information, makes use of segmentation of person information to weight the ambiguity in the contribution of different body parts, and applies Bayesian information fusion with weighted contributions for overall identity hypotheses. This innovative combination of successfully applied components clearly enables a higher degree of discrimination and robustness and therefore better recognition rates.

We performed two experiments to evaluate the performance in an extensive study at a bus station in Austria. Firstly, the reoccurrence of 29 persons was evaluated by a recognition rate of 82%, with images showing significant changes in person orientation, pose, and illumination conditions. The second experiment evaluated an automated person detector that has been post-processed by our method resulting in a recognition rate of 72%. We conclude that this method demonstrates robust performance under normal conditions and is therefore highly useful for connecting distributed camera networks into wide surveillance areas.

## 7878-09, Session 3

### A multimodal eye tracking system for studies of embodied attention

L. Paletta, A. Almer, G. Fritz, K. Amlacher, P. Luley, S. Ladstätter, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria)

Measurements of eye gaze on the environment are one of the most important indicators of visual attention. In mobile eye tracking test persons interact within a task specific environment. Until recently, results have mostly been generated by manual video annotation. Attention has to be considered in a framework of embodiment: bodies select location and orientation, eyes focus on objects of interest, and interactions decide about environment changes.

This work presents an innovative, multimodal system for the outdoor study of embodied attention. Visual attention is captured from the mobile eye tracking videos and semantically interpreted via object detection tools. The location of the test person's body is estimated from the analysis of multisensory data from GPS, accelerometers and a digital compass. Image based localisation supports multisensory processing for the estimation of a geo-referenced human gaze. The system is able to map the human gaze together with the extracted visual semantics using world coordinates into a 3D model of the environment.

In a reconstruction of human gaze we achieved an accuracy of 20-80 cm target localisation and 95% in logo recognition in world coordinates. Position trajectories, views, gaze and targets can be visualized in a 3D model of the urban shopping zone.

## 7878-10, Session 3

### Real-time car detection system

M. Rosner, SLR Engineering OG (Austria)

This paper presents a car detection system that is able to work in close to real-time on a smart camera. A cascade of histograms of oriented gradients was used as a detector. The algorithm and code were optimized for speed to meet the real-time constraints, without losing

too much on detection quality. The system is now able to process 10 frames per second on an Atom Z530 (1.6 GHz) processor used in the smart camera. On videos used for benchmarking only 1 false positive per 5 frames and detection rate of 80 % was observed.

Because there is no adequate car data set known to the author, a new car data set was introduced (SLR Car Dataset). It consists of car images scaled to a certain size, images with cars and other object, images without cars and videos with cars. The application on which the paper is based is ready to detect cars in real world scenarios. It is planned to extend it to also track and analyse the driver behaviour patterns.

### 7878-11, Session 3

#### **Real-time people and vehicle detection from UAV imagery**

A. Gaszczak, T. Breckon, J. Han, Cranfield Univ. (United Kingdom)

A generic and robust approach for the real-time detection of people and vehicles from an Unmanned Aerial Vehicle (UAV) is an important goal within the framework of fully autonomous UAV deployment for aerial reconnaissance and surveillance. Here we present an approach to the automatic detection of vehicles based on using multiple trained cascaded Haar classifiers (a disjunctive set of cascades) with secondary confirmation in thermal imagery as well as approach for people detection in thermal imagery using also multiple trained cascaded Haar classifiers with multi-variant Gaussian shape matching. The results presented show the successful detection of vehicle and people under varying conditions in both isolated rural and cluttered urban environments with minimal false positive detection. Performance of the detector is optimized to reduce the overall false positive rate by aiming at the detection of each object of interest (vehicle/person) at least once in the environment (i.e. per flight) rather than every object in each frame. Currently the detection rate for people is ~70% and cars ~80% although the overall episodic object detection rate for each flight pattern exceeds 90%.

### 7878-12, Session 3

#### **Real-time pose invariant logo and pattern detection**

O. Sidla, SLR Engineering OG (Austria)

The detection of pose invariant planar patterns has many practical applications in computer vision and surveillance systems. The recognition of company logos is used in market study analysis to examine the visibility and frequency of logos in advertisement or danger signs on vehicles could be detected to trigger warning systems in tunnels.

We present the results of a study on logo detection which is based on the detection of N invariant 2d features and subsequent matching and clustering.

Specifically we look at the following feature types:

- SURF
- Compact signatures + random ferns
- One way descriptor

which are combined with the following point detectors:

- Lowe DoG as from the SURF implementation
- Harris Corner Detector
- FAST Corner Detector

For application oriented tests we first generate a set of testing images which are used to examine the limits of the 2d feature descriptors under pose, perspective, and resolution variations.

A real-world test tries to detect vehicles with a distinctive logo in an outdoor environment under different lighting and weather conditions: a camera is mounted to observe a gate so that incoming trucks can be

monitored - sequences of incoming vehicles with a specific brand logo are detected, logged and stored for manual evaluation.

### 7878-13, Session 3

#### **FirstAidAssistanceSystem: improvement of first aid measures by using Car2Car-communication**

S. Tuchscheerer, T. Hoppe, C. Kraetzer, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

This work's goal is the enhancement of first aid measures directly after car accidents by determining suited first aiders via C2C communication and to provide detailed support instructions. The concept combines upcoming car2car (C2C) communication with established technology as GPS and GSM. After a crash, the proposed FirstAidAssistanceSystem (FAAS) sends a broadcast message via the 802.11p C2C standard. All nearby cars as potential first aiders are located and at least one nearest candidate (we suggest 3-5 as discussed in final paper) driving towards the accident is chosen and notified as first aider. A support guide on his multipurpose display (e.g. the navigation system) provides the helper with detailed instructions and illustrative tutorials. The paper presents the concept in detail with a practical evaluation using a first implementation.

### 7878-14, Session 4

#### **The report of estimating the egomotion of the moving stereo cameras in the environment including moving objects and reconstructing the observed space in 3D**

N. Tatematsu, J. Ohya, Waseda Univ. (Japan)

This paper proposes a temporal modified RANSAC based method that can discriminate each moving object from the still background, can compute the stereo cameras' egomotion, and can reconstruct 3D structure of each moving object. We computed 3D optical flows from the depth map and the tracking feature point. In general, flows from different objects have different orientations and lengths, while flows from a same object have uniform orientation and length. We define "flow region" as a set of connected pixels whose 3D optical flows have uniform orientation and length. Our temporal modified RANSAC segments the detected 3D optical flows into flow regions and computes the rotation and translation matrix for each flow region. The modified RANSAC estimates multiple models from the data and can find clusters. The temporal modified RANSAC performs the modified RANSAC to each of the flow region. Finally, the 3D points computed from the depth map in all the frames are registered using each flow region's matrix to the initial positions in the initial frame so that the 3D structures of the moving objects and still background are reconstructed. Experiments using multiple moving objects and real stereo sequences demonstrate the effectiveness of our proposed method.

### 7878-15, Session 4

#### **A multiple feature based particle filter using mutual information maximization**

K. Hong, K. Han, Purdue Univ. (United States)

In designing a tracking algorithm, utilizing several different features, e.g., color histogram, gradient histogram and other object descriptors, is preferable to increase robustness of tracking performance. In this paper, we propose a multiple feature fusion framework to improve the tracking by assigning appropriate weights to individual features. The feature weights are optimally obtained by a waterfilling procedure that maximizes mutual information between target object features and query features. Especially, in this paper, we focus on a particle

filter tracking implementation of the multiple feature fusion framework. Our experiments show that object tracking with multiple features outperforms single feature based tracking methods and illustrates that the proposed optimal feature weighting increases robustness of multiple-feature based tracking performance.

#### 7878-16, Session 4

### High precision object segmentation and tracking for use in super-resolution video reconstruction

T. N. Mundhenk, R. N. Sundareswara, Y. Chen, HRL Labs., LLC (United States)

Super resolution image reconstruction allows for the enhancement of images in a video sequence that is better than the original pixel resolution of the imager. Difficulty arises when there are foreground objects that move differently than the background. A common example of this is a car in motion in a video. Given the common occurrence of this, super resolution reconstruction becomes non-trivial. One method for dealing with this is to segment out foreground objects and quantify their pixel motion differently. First we estimate local pixel motion using a standard block motion algorithm common to MPEG encoding. This is then combined with the image itself into a six dimensional mean-shift kernel density estimation based image segmentation with mixed motion and color image feature information. This results in a tight segmentation of objects in terms of both motion and visible image features. The next step is to combine segments into a single master object. Statistically common motion and proximity are used to merge segments into master objects. To account for inconsistencies that can arise when tracking objects, we compute statistics over the object and fit it with a generalized linear model. Using the Kullback-Leibler divergence, we have a metric for the goodness of the track for an object between frames.

#### 7878-17, Session 4

### Robust pedestrian detection and tracking from a moving vehicle

N. X. Tuong, Nanyang Technological Univ. (Singapore); T. Müller, A. Knoll, Technische Univ. München (Germany)

In this paper, we address the problem of multi-person detection, tracking and distance estimation in a complex scenario using multi-cameras. Specifically, we are interested in a vision system for supporting the driver in avoiding any unwanted collision with the pedestrian.

We propose an approach using Histograms of Oriented Gradients (HOG) to detect pedestrians on static images and a particle filter as a robust tracking technique to follow targets from frame to frame.

Because the depth map requires expensive computation, we extract depth information of targets using Direct Linear Transformation (DLT) to reconstruct 3D-coordinates of correspondent points found by running Speeded Up Robust Features (SURF) on two input images. Using the particle filter the proposed tracker can efficiently handle target occlusions in a simple background environment. However, to achieve reliable performance in complex scenarios with frequent target occlusions and complex cluttered background, results from the detection module are integrated to create feedback and recover the tracker from tracking failures due to the complexity of the environment and target appearance model variability.

The proposed approach is evaluated on different data sets both in a simple background scenario and a cluttered background environment. The result shows that, by integrating detector and tracker, a reliable and stable performance is possible even if occlusion occurs frequently in highly complex environment.

A vision-based collision avoidance system for an intelligent car, as a result, can be achieved.

#### 7878-18, Session 5

### Design and evaluation of security multimedia warnings for interaction between human and industrial robots

J. Fruth, J. Dittmann, C. Krätzer, Otto-von-Guericke-Univ. Magdeburg (Germany)

In this article a security-warning design concept for production scenarios with direct human-machine interaction is introduced. The use of standard desktop IT network technologies for production networks could introduce the same vulnerabilities known from desktop IT, e.g. the infection with malicious code. A malware infection of a control computer of industrial robots could lead to potential indirect impacts to safety, e.g. humans could be harmed and/or work pieces could be damaged. So far only safety warnings in the production domain are designed and used. In our opinion security warnings are needed to warn the operator for special threats come from malware infections of robot control systems, because of their potential safety impacts. Therefore we designed security warnings based on standard design concepts of safety warnings, common antivirus-warnings and own ideas. The warning includes two classes of warning icons, the property of the malicious code and instructions to the operator and a confirmation button. The criticality of the event is communicated via signal light and acoustical warnings. In this article an evaluation approach, operationalisation and test results of the security warnings, using usability testing techniques, advancements of the design approach and future work are presented.

#### 7878-19, Session 5

### Fingertip guiding manipulator for blind persons to create mental images by switch passive/active hybrid line-drawing explorations

S. N. Syed Yusoh, Mie Univ. (Japan)

Blind people encounter many obstacles daily. One such obstacle is to visualize, mentally, graphics that are presented to them. Therefore, support systems have been developed to help improve communication and quality of life for the visually impaired people. A fingertip guiding manipulator has been developed as a haptic graphic display to help blind persons create mental images of line drawings. The manipulator has two characteristic functions: (1) an indicating function of slider-moving-direction by a stepping-motor attached at the arm end, and (2) a fingertip traction function by a manipulator. The former two functions are utilized to represent planer 2D figures. Some preliminary experiments were carried out, and the human perceptual thresholds with the fingertip rotation, and the human kinesthetic following-performance were clarified. These thresholds shall be used as the back data for designing mechanical specifications of the fingertip guiding manipulator. This paper describes our investigation on the dual mode fingertip-guiding function allowing either passive or active exploration. When using this manipulator, the person is assumed to touch a slider by his/her fingertip: the slider is attached at the end of the manipulator. In the active mode, the fingertip guiding manipulator pulls his/her fingertip along line drawings. In the passive mode, it provides a kind of selective compliance, and allows the persons to move his/her fingertip freely just only in the direction of the line drawings. For the sake of the dual mode function, it is expected that the efficiency would be improved a lot comparing to the former model being equipped with the passive mode alone. Thus, a haptic graphic display based on the novel concept, i.e., a dual mode fingertip guiding manipulator was presented, and some basic characteristics were confirmed. It helps blind persons create mental images of line drawings. The preliminary results are summarized as follows: (1) A passive mode fingertip guiding manipulator was developed. It was able to teach persons the line drawings of route maps. In the passive mode, the position of the slider traces the strokes of the target line drawing. Since a person pinches the slider, the person's fingertips were pulled along the target line drawings. During this process, the person is assumed

to perceive the line drawings through his/her kinesthetic positional sense. (2) A prototype of an active mode fingertip guiding manipulator was examined. It is also able to teach graphics. In the active mode, the position of the slider should be moved with free of force in an appropriate direction alone. Therefore, the person can move his/her fingertips along the target trajectory by utilizing his/her kinesthetic force sense with much higher velocity. During this process, the person is expected to perceive the line drawing through his/her kinesthetic positional sense. In the future, the authors are planning to implement the dual mode fingertip guiding manipulator in a more concrete form, and are going to carry out some experiments with blind people.

## 7878-20, Session 5

### Augmented reality user interface for mobile ground robots with manipulator arms

S. Vozar, D. Tilbury, Univ. of Michigan (United States)

Augmented Reality (AR) is a technology in which real-world visual data is combined with an overlay of computer graphics, enhancing the original feed. AR is an attractive tool for teleoperated UGV UIs as it can improve communication between robots and users via an intuitive spatial and visual dialogue, thereby increasing operator situational awareness. The successful operation of UGVs often relies upon both chassis navigation and manipulator arm control, and since existing literature usually focuses on one task or the other, there is a gap in mobile robot UIs that take advantage of AR for teleoperating the robot manipulator and base together.

This work describes the implementation of an AR UI system for a UGV with an attached manipulator arm, along with the results of preliminary user tests. The system supplements a video feed shown to an operator with information about geometric relationships within the robot task space to improve the operator's situational awareness.

Previous studies on AR systems and preliminary analyses indicate that such an implementation of AR for a mobile robot with a manipulator arm is anticipated to improve operator performance. A full user-study can determine if this hypothesis is supported by performing an analysis of variance on common test metrics associated with UGV teleoperation.

## 7878-21, Session 5

### An embedded omnidirectional vision navigator for automatic guided vehicles

W. Feng, Tianjin Univ. (China); B. Zhang, Z. Cao, X. Zong, Tianjin Univ. of Technology (China); J. Rönning, Univ. of Oulu (Finland)

Omnidirectional vision appears the definite significance since its advantage of acquiring full 360° horizontal field of vision information simultaneously. In this paper, an embedded original omnidirectional vision navigator (EOVN) based on fish-eye lens and embedded technology has been implemented. Fish-eye lens is one of the special ways to establish omnidirectional vision, however, it appears with an unavoidable inherent and enormous distortion. A unique integrated navigation method which is conducted on the basis of targets tracking has been researched. It is composed of targets recognition, multi-target tracking, distortion rectification, spatial location and navigation control. In order to adapt to the different indoor and outdoor navigation environments, we implant mean shift and dynamic threshold adjustment into the particle filter algorithm to improve the efficiency and robustness of tracking capability. RTRLN has been implanted in an independent research embedded platform which is composed of COMS+FPGA+DSP. It is like a smart crammer to guide various vehicles in different environments by tracking the diverse landmarks hanging in the air or on the roof. The experiments prove that the EOVN is particularly suitable for the guidance applications which have high requirements on precision, repeatability and long distance. The research achievement has a good prospect in application.

## 7878-23, Session 6

### Detecting stationary human targets in FLIR imagery

A. L. Chan, U.S. Army Research Lab. (United States)

In the military arena, intelligent unmanned ground vehicles (UGVs), weighing 10 tons or more, may be designed and used for transportation or combat purposes. To ensure safe operations among civilians and friendly combatants, it is crucial for these UGVs to detect and avoid humans who might be injured unintentionally. In this paper, a multi-stage detection algorithm for stationary humans in forward-looking infrared (FLIR) imagery is proposed. This algorithm first applies an efficient feature-based anomalies detection algorithm to search the entire input image, which is followed by an eigen-neural-based clutter rejecter that examines only the portions of the input image identified by the first algorithm, and culminates with a simple evidence integrator that combines the results from the two previous stages. The proposed algorithm was evaluated using a large set of challenging FLIR images and the results support the usefulness of this multi-stage architecture.

## 7878-24, Session 6

### Spectrally queued feature selection for robotic visual odometry

P. A. Frederick, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States); D. Pirozzo, Booz Allen Hamilton (United States); M. S. Del Rose, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States)

Over the last two decades, research in Unmanned Vehicles (UV) has rapidly progressed and become more influenced by the field of biological sciences. Researchers have been investigating mechanical aspects of varying species to improve UV air and ground intrinsic mobility, they have been exploring the computational aspects of the brain for the development of pattern recognition and decision algorithms, and they have been exploring perception capabilities of numerous animals and insects. This paper describes a 3 month exploratory applied research effort performed at the US ARMY Research, Development and Engineering Command's (RDECOM) Tank Automotive Research, Development and Engineering Center (TARDEC) in the area of biologically inspired spectrally augmented feature selection for robotic visual odometry. The motivation for this applied research was to develop a feasibility analysis on multi-spectrally queued feature selection, with improved temporal stability, for the purposes of visual odometry. The intended application is future semi-autonomous UGV control as the richness of data sets required to enable human like behavior in these systems has yet to be defined.

## 7878-25, Session 6

### Intuitive control of robotic manipulators

D. Rusbarsky, RE2, Inc. (United States); J. P. Gray, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States); D. J. Peters, RE2, Inc. (United States)

In the field of unmanned ground vehicles with dexterous manipulators, current control systems require a high cognitive load and training to properly position the manipulator and have it effectively interact with its environment. As robotic manipulators grow more capable through additional degrees of freedom and as Explosive Ordinance Disposal (EOD) robots are developed that take advantage of multiple manipulators on the same platform, the demand for more intuitive control and enhanced situational awareness will also increase. As part of the Modular Intelligent Manipulation system with Intuitive Control (MIMIC) program, industry is working with the U.S. Army to explore technologies that will allow a user to intuitively control multiple degree of freedom robotic arms and maintain better awareness of the operating environment through haptic feedback. In addition to reporting

resistance, haptic feedback can help make operators feel like they are actually there with the robot. Coupled with intuitive controls and advanced video feedback, the goal of this program is to provide users with the sensation that robots are an extension of their bodies. This paper presents on-going research in the area of intuitive control and haptic feedback along with the results of some of our early design and prototyping work.

## 7878-26, Session 6

### **Vision based low cost, precise, and robust localization method in GPS denied environments**

J. Walter, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States); D. C. Bentivegna, Seegrid Corp. (United States)

A low cost, precise and robust vision-based localization system exists in the commercial warehouse industry. The system uses a series of low-cost stereo cameras to collect image data along an autonomous pallet-jack's route. This image data is compiled into 3D maps, known as evidence grids, which correlate images to positions in the vehicle's route. During subsequent traverses of this route, visual information is checked against the evidence grid to pinpoint the vehicle's location. The indoor localization system has shown an accuracy of less than one centimeter. This paper shares promising research results when this indoor system was fielded with little modification to an outside, GPS-denied urban canyon like environment.

## 7878-27, Session 7

### **Curved solid and dotted line characters segmentation and classification**

K. Mohammad, S. S. Agaian, H. Saleh, The Univ. of Texas at San Antonio (United States)

Segmentation is a crucial step in a vision based recognition systems because it extracts meaningful regions for analysis. The segmentation process attempts to decompose the text image into classifiable images called characters. A poor segmentation process produces less accurate recognition. The potential application for vision based character recognition is huge, and there are many challenges to design a single system that is capable of performing automatic recognition for curved text. Issues like curvature, character connectivity, varying text formats, and rotation all have strong negative impact on the accuracy of the system. These effects reduce the accuracy of segmentation step.

This paper presents and implements two new algorithms for handling the rotation issues. The first is based on the Hough transform and is used to draw a line through the text, mimicking the rotation angle. This algorithm shows excellent results for dotted line characters in an image. The second algorithm is based on locating the corners of the text box and extracting them to determine rotation. This algorithm show excellent results for solid line characters. After the text is localized, the rotation is determined and orientation is corrected. The text is then segmented and the characters in the image are identified. A separate filling algorithm is used to deal with the dotted character.

Feature vectors based on character shape are proposed for classification. These vectors are divided into four groups: 1) uses the 4 levels of the Haar transform. 2) Uses the density of 'ones' in the skeleton image of the character. 3) Uses the density of 'ones' in the area of the binary image, and 4) grouping of the 'ones' within the same set. Different weights are applied based on their ability to distinguish the characters. The (SVM) classifier estimates similarities between the examined target characters and the training set of characters.

Images from Ozarka and Desani water bottles were tested using the new algorithms and showed excellent results and improved run times compared to conventional segmentation methods.

The rotation handling, segmentation and classification using the new set of feature vectors were combined and tested using a different

set of images for water bottles. Experimental results for these sets of more than 300 different character images with different text show an average of 93 % successful using the new set of feature vectors for classification and a new algorithm of rotation, filling and segmentation of the text. The proposed solution improves run time (less than 5second), accuracy (99%) and better for Hardware implementation comparing to the current state of art.

## 7878-28, Session 7

### **Accelerating robust 3D pose estimation utilizing a graphics processing unit**

A. R. Gerlach, B. K. Walker, Univ. of Cincinnati (United States)

The spin-image pose estimation algorithm is an accurate method for estimating pose of three-dimensional objects while being both robust to clutter and sensor noise. Unfortunately, the algorithm has a high computational complexity, thus preventing its use in applications that require a robotic system to interact with a dynamic environment. Upon inspection, the spin-image algorithm can be broken down into five portions where a single portion called spin-image matching commands 96% of the computation time in estimating pose. Because, the matching of individual spin-images can be performed independently regardless of order, this portion of the algorithm is ideal for the massively parallel architecture of the graphics processing unit (GPU).

This paper introduces a GPU implementation of the spin-image matching portion of the spin-image algorithm which makes no modifications to the spin-image algorithm, thus not compromising its robustness and accuracy. This implementation results in a speed-up in spin-image matching of 515x and a total algorithmic speed-up of 24.6x out of a theoretical maximum of 26.0x over a MATLAB implementation. This GPU implementation extends the use of the spin-image algorithm towards practical real-time robotic applications.

## 7878-29, Session 7

### **Calibration and rectification research for fish-eye lens application**

W. Feng, Tianjin Univ. (China); B. Zhang, Z. Cao, X. Zong, Tianjin Univ. of Technology (China); J. Rönning, Univ. of Oulu (Finland)

Accurate parameters calibration and effective distortion rectification of an imaging device is of utmost importance in computer vision. Fish-eye lens produces a hemispherical field of view of an environment, which appears definite significant since its advantage of panoramic sight with a single compact visual scene. But the images taken by fish-eye lens have an unavoidable inherent severe distortion. A novel calibration method is proposed to estimating the internal and intrinsic parameters of the vision system, which employs a transformation between points in the world coordinate system and their corresponding location on the image using the special calibration patterns. Then the calibration results are employed to rectify the image distortion. Support Vector Machine (SVM) and Spherical Equidistance Projection Algorithm (SEPA) are integrated to replace ordinary rectification model. SVM is a machine learning method based on the theory of statistics, which have good capabilities of imitating, regression and classification. The approach using SVM provides a mapping between the fisheye image and the standard image for human eyes, and utilizing SEPA reproduces the information about the edge of the fish-eye lens image. The validity and effectiveness of our calibration and rectification procedures are demonstrated by processing the real images.



## 7878-30, Session 7

### **A hardware-software co-design approach to a JPEG encoder design for a planetary micro-rover application**

S. Sarma, K. Parameswaran, S. Udupa, K. M. Bharadwaj, Indian Space Research Organisation (India)

There is a great interest amongst the scientific communities to explore various planets for scientific studies using robotic missions with the leverage to hostile environments at reduced risk, cost, and greater mobility than that is possible with manned explorations. Admittedly, the primary objective of many of such robotic missions are to collect science data, for instance, site exploration using visual images or videos, chemical analysis of surface terrain, and collection of samples to mention a few. However, such missions are characterized by severe power and bandwidth constraints due to long distance and lack of abundant power sources. Large set of visual images collected by various cameras on-board a rover is to be processed and transmitted to Earth for scientific studies which are mostly achieved by using various data compression techniques. In particular, the JPEG image compression standard that is developed by the Joint Photographic Experts Group committee for use in compressing digital images and full color photographic images is very popular. It is one of the primary formats used for exchanging pictures on the World Wide Web, and it is commonly used in digital cameras as the storage format. In this paper, a hardware-software based co-design approach presented with the aim to implement a JPEG encoder for a micro-rover for two processor systems namely, MIL-STD-1750 and PowerPC under serious power and bandwidth constraints. Two methods, one mostly using software implementation and the other- a FPGA based pipelined hardware architecture, are compared for their performance and resource utilization using planetary terrain images of various sizes and quality settings for both these processor architecture. In fact, the results presented are extensively substantiated by simulation and practical implementation in FPGA. Based on these studies, suitable guidelines are elucidated to arrive at an effectual architecture for a planetary micro-rover for future exploration by an Indian Moon mission.

## 7878-31, Session 8

### **Phobator: Princeton University's entry in the 2010 Intelligent Ground Vehicle Competition**

J. Newman, S. O. Abiola, R. M. Corey, S. A. Suresh, L. J. Szocs, B. A. Partridge, D. D. Yu, H. Zhu, Princeton Univ. (United States)

In this paper we present Phobator, our entry in the 2010 Intelligent Ground Vehicle Competition (IGVC). We describe revised vision and navigation software that improve the robustness and speed of our robot, and our newly-constructed platform designed to address previous years' concerns such as water tightness and serviceability.

Our vision software uses color stereo images. We use RANSAC to locally estimate the ground plane and classify points that lie above that plane as obstacles. This robust method detects a variety of obstacles on uneven terrain. To detect lanes, we process the image with an edge detection filter and reject edges if the corresponding image pixels are not lane-colored, greatly reducing false positives and noise. We use RANSAC to identify separate lanes and model each with a rotated parabola.

For path planning Phobator uses the Anytime Dynamic A\* algorithm. The algorithm allows incremental re-planning, for efficiency. It generates optimal paths given enough time, but first quickly generates a valid, sub-optimal path so the robot never needs to wait. We augment the cost map of the environment with a potential field which addresses the problem of "wall-hugging" and smoothes the path.

## 7878-32, Session 8

### **Application of parallelized software architecture to an autonomous ground vehicle**

R. Shakya, A. Wright, Y. H. Shin, O. Momin, S. Petkovsek, P. Wortman, P. Gautam, A. Norton, Trinity College (United States)

This paper presents improvements made to Q, an autonomous ground vehicle designed to participate in the Intelligent Ground Vehicle Competition ([www.igvc.org](http://www.igvc.org)). The IGVC has two main challenges, called the autonomous challenge and the navigation challenge. In the autonomous challenge the vehicle is required to follow a course while avoiding obstacles and staying within the course boundaries, which are marked by white lines. For the navigation challenge, the vehicle is required to reach a set of target destinations, known as way points, with given GPS coordinates and avoid obstacles that it encounters in the process. For the 2010 IGVC, Q was upgraded with a new parallelized software architecture and a new dual-core vision processor. The new software architecture modularizes all the necessary tasks such as motor control, navigation and sensor data collection and executes them in parallel, providing considerable flexibility and facilitating efficient use of processing power. The vision processor increased speed and reliability of the image processing algorithm on Q. With all these improvements Q was able to navigate through two S-curves and travel almost 300 feet, which is just past the halfway mark of the autonomous challenge course. As a result Q placed 2nd in the autonomous challenge and 3rd overall among 57 participating entries.

## 7878-33, Session 8

### **WOAH: an obstacle avoidance technique for high speed path following**

N. Tuck, M. McGuinness, F. W. Martin, Univ. of Massachusetts Lowell (United States)

This paper presents WOA, a Working Obstacle Avoidance Heuristic. WOA is a real-time reactive obstacle avoidance technique for mobile robots designed to leverage polar ranging data from a single laser ranging device. Unlike most current techniques, this method allows a robot to travel quickly past obstacles without slowing down, resulting in consistent fast progress towards a specified goal.

In testing obstacle avoidance techniques that work with a single LIDAR, we discovered that existing methods did not traverse paths as quickly as they could. Algorithms we tested notably included Smooth Nearness Diagram and Vector Field Histogram. Essentially, we found these techniques to be excessively cautious near obstacles even when the obstacles were not in the robot's path. WOA was designed specifically to avoid this problem.

Preliminary results have been promising. A robot using a prototype version of WOA achieved 3rd place in the 2010 IGVC Navigation Challenge, visiting seven of nine waypoints and doing so at the fastest pace of all competitors that visited at least two. In the paper we will present results comparing WOA to previous techniques both in simulation and in live testing on our outdoor test course.

## 7878-35, Poster Session

### **Continuous target tracking based on multiple views**

Y. Liu, Beijing Univ. of Posts and Telecommunications (China)

In order to solve the problem of target handoff in helping service robot to track targets based on ceiling cameras in smart space, the paper puts forward an algorithm of combining projective invariant with histogram, which can be used in non-overlapping or overlapping conditions. Meanwhile, it analyzes multi-target motion trajectories. The experiment indicates that the method may track objects effectively and it is robust to occlusion, which can meet actual needs.

### 7878-36, Poster Session

#### **A target detection method in multimodal images with complex backgrounds and different views**

Z. He, X. Ding, Tsinghua Univ. (China)

Target detection in multimodal (multisensor) images is a difficult problem especially with different views and complex backgrounds. In this paper, we propose a target detection method based on ground object region extraction and graph model matching to solve it. First, the extrinsic parameters of camera are used to transform the images to reduce the impact of viewpoints differences. Then the stable object regions are extracted to describe the object shapes in multimodal images by MSER which can effectively reduce the impact of noise and multimodal. Those regions of ground objects whose shapes are less affected by the view transformations are used to build a graph model to describe the target in the reference image with spatial constraints to reduce the impact of complex backgrounds. At last, a graph model registration algorithm is developed using round region matching and spatial constraints to find the target in the sensed images. The algorithm is based on the idea of RANSAC and obtained a satisfied experiment result in our data set of two visible reference images in top view and four groups of infrared sensed images in side view. The final detection rates of group 1, 2 and 4 are all above 95% while group 3 is 83.61%.

### 7878-37, Poster Session

#### **Gender classification robust to face pose variation and partial occlusion**

P. Zeng, Nanchang Univ. (China); Y. Zhang, Tsinghua Univ. (China)

A new approach for automatic gender classification is proposed to balance the algorithm complexity and accuracy rates. Three key contributions in the paper are presented:

Firstly, a concept to mend a partly occluded face with the symmetric pixels or a standard neural face is put forward according to the occluded position, which makes all images a complete uniform face go through the same classifier. In this way, the requirements for many classifiers and training datum are avoided for different occluded parts.

Second, a relative simple character triangle structured by eyes' and nose's positions is tapped to rotate the in-depth head rotation, in order to compensate pose variation.

Finally, an unevenly sampled Gabor Wavelet Mask (GWM), according to each part of face devoting to gender classification, is applied. This provides a considerably wider range of tolerance against head rotation.

The comparative experiments are conducting on several publicly available databases, such as FERET, AR face databases and UTech company face images. The results demonstrate that the proposed approach has a considerably tolerance against both in-depth head rotation and partial occlusion.

### 7878-38, Poster Session

#### **Lane marking detection by extracting white regions with predefined width from bird's-eye road images**

S. Abe, Utsunomiya Univ. (Japan)

Detecting lane markings on roads from in-vehicle camera images is very important because it is one of the fundamental tasks for autonomous running technology and safety driving support system. There are several lane markings detection methods using the width information, but most of these are considered to be insufficient for oblique markings. So, the primary intent of this paper is to propose a detecting lane markings method robust to orientation of markings. In this work, we focus on the width of lane markings standardized by road act in Japan, and propose a method for detecting white lane markings by extracting white regions with constant predefined width from bird's-eye road images after segmentation such as categorical color area one. The proposed method is based on the constrained Delaunay triangulation. The proposed method has a merit that can be measure an exact width for oblique markings on the bird's-eye images because it can be obtained perpendicular width for edge. The effectiveness of the proposed method was shown by experimental results for 187 actual road images taken from an in-vehicle camera.

### 7878-39, Poster Session

#### **Selective locality preserving projections for face recognition**

F. Dornaika, Univ. of the Basque Country (Spain); A. Assoum, Lebanese Univ. (Lebanon)

Recently, a graph-based method was proposed for Linear Dimensionality Reduction (LDR). It is based on Locality Preserving Projections (LPP). LPP is a typical linear graph-based dimensionality reduction (DR) method that has been successfully applied in many practical problems such as face recognition. LPP is essentially a linear extension of Laplacian Eigenmaps. When dealing with face recognition problems, LPP is preceded by a Principal Component Analysis (PCA) step in order to avoid possible singularities. Both PCA and LPP are computed by solving an eigen decomposition problem. In this paper, we propose a novel approach called "Selective Locality Preserving Projections" that goes beyond the combination of the principles of LPP with Feature Selection paradigm. It performs a simultaneous eigenvector selection and construction associated with PCA and LPP. We have tested our proposed approach on several public face datasets. Experiments on ORL, UMIST, and YALE Face Databases show significant performance improvements in recognition over the classical LPP. The proposed approach lends itself nicely to many biometric applications.

# Conference 7879: Imaging and Printing in a Web 2.0 World II

Wednesday-Thursday 26-27 January 2011

Part of Proceedings of SPIE Vol. 7879 Imaging and Printing in a Web 2.0 World II

## 7879-01, Session 1

### Web-based magazine design for self publishers

A. A. Hunter, D. N. Slatter, Hewlett-Packard Labs. (United Kingdom)

Short run printing technology and web services such as MagCloud provide new opportunities for long-tail magazine publishing. They enable self publishers to supply magazines to a wide range of communities, including groups that are too small to be viable as target communities for conventional publishers.

In a Web 2.0 world where users constantly discover new services and where they may be infrequent patrons of any single service, it is unreasonable to expect users to learn the complex service behaviors. Furthermore, we want to open up publishing opportunities to novices who are unlikely to have prior experience of publishing and who lack design expertise.

Magazine design automation is an ambitious goal, but recent progress with another web service, Autophotobook, proves that some level of automation of publication design is feasible. This paper describes our current research effort to extend the automation capabilities of Autophotobook to address the issues of magazine design so that we can provide a service to support professional-quality self publishing by novice users for a wide range of community types and sizes.

## 7879-02, Session 1

### Improve artwork design through data tracking system

W. H. Wang, R. Muzzolini, Shutterfly, Inc. (United States)

In personalized digital printing, such as greeting cards, calendars and photo books, people select artworks to match their photos at their preference. Art work design elements are often categorized by occasions, styles, and product. The amount of design grows significantly, as customers demand more choices and the trends of popular designs rise and fade season by season. It is crucial to manage and understand how design elements are used in order to create most desirable productions. In this paper, we analyze and compare different design tracking systems. Art work designs are labeled, ranked, and cross referenced. For each system, we demonstrate the scale of applications, data collection techniques and its advantages and disadvantages.

## 7879-03, Session 1

### DOM-based print-link detection for web article extraction

S. J. Liu, S. H. Lim, J. Liu, Hewlett-Packard Labs. (United States)

Web pages from some Web sites provide a hyperlink (or link) that leads to a print-friendly web page that contains mainly the article itself. Content extraction using these print-friendly pages is generally easier and more reliable. But there are many variations of the print-link representations that made the print link detection more difficult than it first appeared. First, the link can be text-based, image-based, or both. For example, there is a lexicon of phrases used to represent the print links, such as "print", "print article", "print-friendly version", etc. In addition, some pages use printer-resembling image icons with or without a print phrase present. To complicate the matter further, not all the links contain a valid URL, but instead the pages are dynamically generated either by the client Javascript or by the server, which are not

retrievable using the DOM-based extraction technique since no valid URL is available in the DOM. Our solution to the print-link extraction problem takes on two stages: (1) the detection of the print-link, (2) the retrieval of the print-friendly URL from the link attributes, including the test for its validity and the conversion of relative to absolute URL. Experimental results based on roughly 2000 web article pages suggest our solution is capable of achieving over 99% precision and over 97% recall.

## 7879-05, Session 2

### A web-based troubleshooting tool to help customers self-solve color issues with a digital printing workflow

H. J. Santos-Villalobos, Purdue Univ. (United States); V. Loewen, Hewlett-Packard Co. (United States); M. R. Letho, J. P. Allebach, Purdue Univ. (United States)

Current printing technologies enable customers to reproduce high quality, realistic, and colorful hard copies of their digital documents. Although the activity of printing is transparent to the customers, the progression of a customer's document through the digital printing workflow (DPW) is a complex process that may alter the colors in the print job. Given the complexity of the DPW, it is a difficult problem to diagnose the source of the color issue. Novel tools and methods that address this challenge are beneficial for both the manufacturer and its customers. We propose a web-based troubleshooting tool that helps customers to self-solve color issues with electrophotographic laser printers when printing solid colors in graphics and text. The tool helps the customer to reconfigure his/her DPW following printing best practices. If the issue is still unresolved, the tool guides the user to search the gamut of the printer for his/her color preference. The usability of the tool was carefully evaluated with human subject experiments. Also, the description and organization of the troubleshooting tasks were continuously reviewed and improved in regular meetings of the development team. In this paper, we describe the troubleshooting strategy, the color preference search algorithm, and the results of the usability experiments.

## 7879-06, Session 2

### Language-based color editing for mobile device

Y. Zhao, R. Bala, K. M. Braun, Z. Langford, R. J. Rolleston, Xerox Corp. (United States)

Natural language color was initially developed as a desktop application and then deployed in one Xerox print driver. NLC changes the image-editing paradigm from the use of curves, sliders, and knobs, to the use of verbal text-based commands such as "Make light greens much less yellowish". The technology appeals to a common user who has no expert knowledge in color science, and this naturally leads one to think about its use in mobile devices. A prototype GUI design for a language text-based color editing iPhone application will be presented that uses several of its haptic interfaces (e.g. "slot-machine", shaking, swiping, etc.). A textual interface is provided to select a color to be modified within the image and a direction of change for the modification. A swipe interface is provided to select a magnitude and polarity for the modification. Actions on the textual and swipe interface are converted to natural language commands that are in turn used to derive a color transformation that is applied to relevant portions of the image to yield a modified image. The modifications are displayed in real time for a user to observe as they are inputted.

## 7879-07, Session 2

### Personalized imaging: moving closer to reality

H. Ding, Purdue Univ. (United States); R. Bala, Z. Fan, Xerox Corp. (United States); C. A. Bouman, J. P. Allebach, Purdue Univ. (United States)

No abstract available

## 7879-16, Session 2

### How Web 2.0 technologies lead to more tangible printed output

R. Fageth, CeWe Color AG & Co. OHG (Germany)

No abstract available

## 7879-08, Session 3

### Document distance measures and document browsing

I. Ahmadullin, Purdue Univ. (United States); J. Fan, N. Damera-Venkata, S. H. Lim, Q. Lin, J. Liu, S. J. Liu, E. O'Brien-Strain, Hewlett-Packard Labs. (United States); J. P. Allebach, Purdue Univ. (United States)

Managing large document databases is an important task today. Being able to automatically compare the document layouts, and classify and search documents with respect to their layouts proves to be desirable in many applications. We measure document similarity with respect to the document layout components. Documents are initially segmented to identify regions of four classes: text, header, image and background. To simplify the document layout comparison task we represent the regions as bounding blocks that enclose document components of different classes. The document similarity measure is then calculated with respect to position, size, and color histogram of the regions. Using this document similarity measure we propose a browsing mechanism operating on a document dataset. For these purposes, we use a hierarchical browsing environment which we call the document similarity pyramid. It allows the user to browse large document dataset and search for documents in the dataset that are similar to the query document. Documents close to each other in the dataset are grouped together to represent document clusters that are sequentially merged in a form of a quad-tree. Each cluster is represented by one of the documents it contains, which allows us to create a similarity pyramid of the document dataset. The user is allowed to browse the dataset on different levels of the pyramid and reconstruct the pyramid with respect to the documents that are of interest. One application of the algorithm is browsing design templates. The software performance is tested on a document design templates dataset.

## 7879-09, Session 3

### Adaptive removal of background and white space from document images using seam categorization

C. S. Fillion, Z. Fan, Xerox Corp. (United States)

Document images are obtained regularly by rasterization of document content and as scans of printed documents. Resizing via background and white space removal is often desired for better consumption of these images whether on displays or in print. While white space and background are easy to identify in images, existing methods such as naïve removal and content aware resizing (seam carving) each have limitations that can lead to undesirable artifacts, such as uneven spacing between lines of text or poor arrangement of content. An

adaptive method based on image content is hence needed.

In this paper we propose an adaptive method to intelligently remove white space and background content from document images. Document images are different from pictorial images in structure. They typically contain objects (text letters, pictures and graphics) separated by uniform background, which include both white paper space and other uniform color background. Pixels in uniform background regions are excellent candidates for deletion if resizing is required, as they introduce less change in document content and style, compared with deletion of object pixels.

We propose a background deletion method that exploits both local and global context. The method aims to retain the document structural information and image quality. The algorithm will be illustrated with experimental examples.

## 7879-10, Session 3

### Aesthetic role of transparency and layering in the creation of a photo layout

M. V. Ortiz Segovia, Purdue Univ. (United States); N. Damera-Venkata, E. O'Brien-Strain, J. Fan, S. H. Lim, S. J. Liu, J. Liu, Q. Lin, Hewlett-Packard Labs. (United States); J. P. Allebach, Purdue Univ. (United States)

Even though technology has allowed us to measure many different aspects of images, it is still a challenge to objectively measure their aesthetic appeal. A more complex challenge is presented when an arrangement of images is to be analyzed, such as in a photo-book page. Several approaches have been proposed to measure the appeal of a document layout that, in general, make use of geometric features such the position and size of a single object relative to the overall layout. Even fewer efforts have been made to include in a metric the influence of the content and composition of images in the layout. Many of the aesthetic characteristics that graphic designers and artists use in their daily work have been either left out of the analysis or partially quantized in the effort to materialize the concepts.

Moreover, graphic design tools such as transparency and layering play an important role in the professional creation of layouts for documents such as posters and flyers. The main goals of our study are to apply similar techniques within an automated photo-book generation tool, and to further evaluate the aesthetic characteristics of the resulting layouts. Among other design techniques, the tool encourages the use of layering and transparency in the layout to produce a professional looking arrangement of the pictures. Two series of experiments with people from different levels of expertise with graphic design provided us with the tools to make the results of our system more appealing. For the first experiment, we entrusted 12 graphic designers, professionals and senior students, the task of creating a single-page layout with a given photo collection under some specific constraints. The constraints imposed to the designers were meant to guide their work. In the second experiment, we asked people with low or no level of expertise in graphic design to use our computer-based tool to generate a photo-layout. Both groups were required to report a detailed description about their creative process. In this paper, we discuss the results of our experiments and examine the underlying aesthetics of the resulting photo layouts in the context of distinct graphic design concepts.

## 7879-11, Session 3

### Automatic picture orientation detection based on classifier combination

C. Liu, Y. Sun, X. Ding, Tsinghua Univ. (China)

Automatic picture orientation recognition is of great significance in many applications such as consumer gallery management, webpage browsing, content-based searching or web printing. We try to solve this high-level classification problem by relatively low-level features including Spatial Color Moment (CM) and Edge Direction Histogram (EDH). An improved distance-based classification scheme is adopted as our classifier. We propose an input-vector-rotating strategy

instead of collecting and training samples for all four classes, which is computationally more efficient than several conventional schemes. Then we research on the classifier combination algorithm to make full use of the complementarity between different features and classifiers. Our classifiers combination methods include two levels: feature-level and measurement-level. And we present two classifier combination structures (parallel and cascaded) at measurement-level with a rejection option. As the precondition of measurement-level methods, the theory of Classifier's Confidence Analysis (CCA) is introduced with the definition of concepts such as classifier's confidence and generalized confidence. The classification system finally approached 90% recognition accuracy on a wide unconstrained consumer picture set.

#### 7879-12, Session 4

### Whiteboard sharing: capture, process, and print or email

M. J. Gormish, B. Erol, D. G. Van Olst, T. Li, A. Mariotti, Ricoh Innovations, Inc. (United States)

Whiteboards continue to be used to support meetings by facilitating the sharing of ideas, focusing attention, and summarizing. However, at the end of the meeting participants are left without a tangible, or even electronic summary. We consider the capture of the information on a whiteboard, improving the image quality, and sharing the results. This paper describes the initial algorithm for improving whiteboard image quality and changes made to reduce computation. The algorithm has been provided freely as a web widget and as an application on the iPhone and Android phones. User feedback and analytics on usage has led to further changes in the image processing and the user interface.

#### 7879-13, Session 4

### Building a print on demand web service

P. D. Reddy, Hewlett-Packard Labs. (United States); B. Rozario, V. Anil Dev, S. Dudekula, Hewlett-Packard Labs. India (India)

More than 91 million books have been printed in multiple languages using a variety of printing technologies and styles. There are currently about 32M books that are out of print in the United States. There is considerable effort underway to digitize all books that have ever been printed. There is need for a service that can take raw book scans and convert them into Print on Demand (POD) books. Such a service definitely augments the digitization effort and enables broader access to a wider audience. To make this service practical we have identified three key challenges that needed to be addressed. These are: a) produce high quality image images by eliminating artifacts that exist due to the age of the document or those that are introduced during the scanning process b) develop an efficient automated system to process book scans with minimum human intervention; and c) build an eco system which allows us the target audience to discover these books.

The thrust of this paper is to discuss our approach and the progress we have made in addressing each of the challenges outlined above.

#### 7879-14, Session 4

### An unsupervised fusion method for large scale cross-media meta-search engine with clickthrough data

Y. Cao, Y. Tian, T. Huang, W. Gao, Peking Univ. (China)

Supported by the China-US Million Book Project (MBP), we have developed and opened the first Universal Digital Library in Nov. 2007. The digital library, which offers four searching engines concerning different type of queries, provides great facility for users via the Internet. Nevertheless, its large amount and various form of data raise new challenge for current meta-search engine result fusion methods like Borda Count Model and CombModel, since these methods are

mainly designed for text-based retrieval systems only. Thus, we propose an unsupervised fusion method based on user feedback (i.e. clickthrough) to solve the problem. The proposed method takes into account cross-media conditions. It provides a global and local fusion weight based result reranking scheme to fuse results from different member search systems. A sliding window is also employed to solve the lack of uniformity within clickthrough and overlapped results in order to provide more reasonable results rank. Experiments carried on the WikipediaMM database demonstrated that the proposed method outperforms traditional fusion methods in terms of Mean Average Precision, B-Pre, R-Pre and other evaluation measurements.

#### 7879-15, Session 4

### iULib: where UDL and Wikipedia could meet

Y. Tian, T. Huang, W. Gao, Peking Univ. (China)

Empowering the group collaboration and knowledge-sharing capabilities for the Universal Digital Library (UDL) is definitely an important work after more than 1.5 million digitalized books were open to access online. One of the key motivations of the development of such a platform is the emergence of Web 2.0 in recent years, especially with the rapidly increased popularity of Wikipedia, an incentive application of Web 2.0 for users' strong involvement and knowledge sharing. This paper presents our vision, which we call iULib, about where and how UDL and Wikipedia could meet. In the first phase, we directly apply the Wiki architecture and software in UDL to upgrade the digital library as an interactive platform that facilitates community and collaboration. Preliminary implementation shows the feasibility and reliability of our design. Furthermore, as a free encyclopedia that assembles contributions from different users, Wikipedia may also be used as a knowledge base for UDL. As a result, UDL can be upgraded as an intelligent platform for information retrieval and knowledge sharing. Our practice at the WikipediaMM task in the ImgeCLEF 2008 shows that the knowledge network constructed from Wikipedia can be used to effectively expand the query semantics of image retrieval. It is expected that Wikipedia and digital library can integrate each other's valuable results and best practices to benefit each other.

#### 7879-04, Session 5

### Book Widget: embedding automated photo-document publication on the web and in mobile devices

E. O'Brien-Strain, Hewlett-Packard Labs. (United States); A. A. Hunter, Hewlett-Packard Labs. (United Kingdom); J. Liu, Q. Lin, D. Tretter, Hewlett-Packard Labs. (United States); J. Wang, Hewlett-Packard Labs. China (China); X. Zhang, Hewlett-Packard Labs. (United States)

We describe a cloud-based automated-publishing platform that allows third party developers to embed our software components into their applications, enabling their users to rapidly create documents for interactive viewing, or fulfillment via mail or retail printing. We also describe how applications built on this platform can integrate with a variety of different consumer digital ecosystems, and how we will address the quality and scaling challenges.

The platform will provide content transformation algorithms such as photo triage, clustering, and photo-book layout. The platform will provide temporary transactional storage of content (photos and text), editable documents (photo-books), and generated artifacts (PDFs). It will be easily embeddable in an application, accessed via secure RESTful URL resources. It will provide an elegantly powerful document-oriented data model with no concept of a "user" that would cause problems integrating. It will have an embeddable widget/iframe user interface for interactive document editing.

One differentiation of this platform is that it facilitates separating three properties of applications: the user interface context, the source content, and the generated artifacts. This allows for the ready building of a wide variety of applications.

## 7879-17, Session 5

### **Semantic photo books: leveraging blogs and social media for photo book creation**

M. Rabbath, P. Sandhaus, OFFIS e.V. (Germany); S. C. J. Boll, Carl von Ossietzky Univ. Oldenburg (Germany)

In this work we introduce an approach for creating photo books from Web 2.0 resources. We concentrate on two kinds of online shared media: (a) Blogs especially travel blogs (b) Social community websites like Facebook. We introduce an approach to select media elements including photos, geographical maps and texts, and then use these elements to create a printable photo book. Because the selected media elements can be too many, we choose the most proper. Additionally we add external media elements such as geographical maps, texts and externally hosted photos from linked resources. Important media are chosen according to several criteria including the social importance of the persons in the photos to the user and the level of user-media interactivity. Having selected the important media, our approach introduces a genetic algorithm to create an appealing layout using aesthetic rules, where some of the photos are chosen as background photos, and other media are positioned in the foreground. We implemented our approach as web services connected to a Facebook application, and a tool to choose entries from personal blogs. As a result, the output of our implemented application is a photo book in Cewe MCF format.

## 7879-18, Session 5

### **Automatic image selection scheme utilizing comments for insertion of images into weblogs**

T. Konno, E. Myodo, K. Takagi, R. Kawada, KDDI R&D Labs., Inc. (Japan)

This paper proposes a scheme which utilizes comments given to images on an image sharing site in order to find an appropriate image for insertion into poem-like weblogs (blogs) as a way to represent their atmosphere (impression). The result shows that utilizing comments is effective. To achieve this purpose, there are two issues: how impression words are extracted from blogs and how images representing the impression words are obtained. Assuming that it is important to obtain images representing the impression words, this paper focuses on the latter issue only. We hypothesize that comments and tags extracted from an image sharing site can be adequate for obtaining images corresponding to impression words at low cost. In particular, utilizing comments can be more adequate for the image search with impression words than utilizing tags because the impression words are often used in comments. Therefore, we propose a scheme which utilizes comments to find appropriate images. In order to investigate the effectiveness of utilizing comments, conformance between impression words and the images was evaluated. The rating for conformance is 3.5 on a scale of 1 to 5 when utilizing comments, which is 0.6 higher than when utilizing tags.

## 7879-19, Session 5

### **Title identification of web article pages using html and visual features**

J. Fan, Hewlett-Packard Labs. (United States); P. Luo, Hewlett-Packard Labs. China (China); P. Joshi, Hewlett-Packard Labs. (United States)

Extracting informative content from Web article pages has many applications such as printing and content reuse. Title is a significant and unique component of an article. However, identifying the true title is a non-trivial problem even for human readers. In this paper, we present a title identification method that taking into account of the title field of the html page and html tag of a DOM node as well as font size

and horizontal alignment. We test our method on a ground truth data set consisting of 2000 pages from 100 web sites and achieved 97.2% precision and 96.9% recall.

## 7879-20, Session 5

### **Creating 3D realistic head: from two orthogonal photos to multiview face contents**

Y. Lin, Tsinghua Univ. (China); Q. Lin, F. Tang, Hewlett-Packard Labs. (United States); L. Tang, Hewlett-Packard Labs. China (China); S. H. Lim, Hewlett-Packard Labs. (United States); S. Wang, Tsinghua Univ. (China)

3D Head models have many applications, such as virtual conference, 3D web game, Biometrics, and so on. There are several web-based face modeling solutions that can create a 3D face model from one or two user uploaded face images. With the created 3D model, realistic animations can be generated. The existing approaches are limited to generating the 3D model of the face region. The accuracy of such reconstruction is very limited for side views, as well as hair regions. The goal of our research is to develop a framework for reconstructing the realistic 3D human head based on two approximate orthogonal images.

Our framework takes a frontal head image and a side-view head image, and goes through segmentation, feature points detection, feature points matching, and texture mapping to create a 3D head model. The main contribution of the paper is that the processing steps are applied to both the face region as well as the hair region.

We will show examples of the reconstruction in the paper. We will also compare the reconstruction with the 3D model of head obtained using a commercial 3D scanner. Finally, we will discuss potential applications.

## 7879-21, Session 6

### **Mobile multimedia understanding applications: an overview**

X. Lin, Vobile, Inc. (United States)

In recent years, mobile devices are quickly reaching almost every corner of our daily life in a variety of forms: personal media players, smart phones, netbooks, and tablets. Besides the more powerful, smaller, and more versatile hardware, another driving force is the vast number of software applications ("apps") on those mobile devices. A number of mobile apps employ intelligent multimedia understanding (MU) technologies. This paper gives an overview of such apps. The focus is not on the underlying MU techniques, which are already covered in huge amount of literature. Instead, it attempts to shed some light on the junction of mobile apps and MU. For this purpose, it addresses a number of important aspects: unique requirements and characteristics of MU-related apps, values brought in by MU, typical MU technologies involved, comparison of alternative system architectures, and available development tools.

## 7879-22, Session 6

### **Learning object detectors from web image search**

F. Tang, D. Tretter, Hewlett-Packard Labs. (United States)

Being able to detect distinguishable objects is very useful in many high level computer vision applications. Traditional methods for building such a detector require a large amount of carefully collected and cleaned data. For example to build a face detector, a large number of face images need to be collected and faces in each image need to be cropped and aligned as the data for training. This process is tedious and error-prone. Recently more and more people are sharing their photos on the internet, if we could leverage these data for building a detector, it will save tremendous amount of effort to collect training

data. Popular internet search engines and community photo websites like Google image search, Picassa, Flickr make it possible to harvesting online images for image understanding tasks. In this paper, we develop a method leveraging images obtained from Google image search to build an object detector. The proposed method can automatically identify the most distinguishable features across the downloaded images. Using these learned features, the algorithm can detect the object in a new image. Experiments show promising results of our approach.

### 7879-23, Session 6

#### **Image categorization for marketing purposes**

M. I. Almishari, H. Lee, N. Gnanasambandam, Xerox Corp. (United States)

Images meant for marketing and promotional purposes (e.g. coupons) represent a basic component in incentivizing customers to visit shopping outlets and purchase discounted commodities. They also help department stores in attracting more customers and potentially, speeding up their cashflow. While coupons are available from various sources - print, web, etc. there is still a gap in terms of a centralized aggregator of these images. Aggregation of these images helps dispensing these coupons in an on-demand fashion from a centralized repository. But aggregating and gathering of meta-data relating to the promotional material is not easily achieved. Firstly the creation of such marketing aids is a distributed and artistic endeavor. Coupons contain both images and text. Designs are arbitrary (structure is hidden) and most often don't conform to any specification. On the contrary text ads in the web domain are well structured and follow strict word limitations. Further responding to changing conditions in the market or inventories are harder with coupons than with, say, text ads as with coupons the artists or creators of the marketing material have to be involved in redesign.

In our work, we aim for a mechanism that decouples the design process and promotion activity.

### 7879-24, Session 6

#### **Text extraction from web images**

C. Liu, C. Yang, X. Ding, Tsinghua Univ. (China)

Images play a key-role in web pages and they are an important part of web document analysis and understanding. Statistics show that many web images containing text carry important information about both the layout and the content of the page document. Detecting and recognizing text embedded in web images becomes potentially essential for applications such as effective indexing and searching. Since web images have special characteristics that distinguish them from conventional complex background images, most text segmentation and recognition algorithms with good performance in other fields fail to recognize web images text. In this paper we present a survey of the methods and principles that have been proposed to handle segmentation and recognition of text in web images. And the purpose of this paper is to classify and review these algorithms, discuss benchmark data and performance evaluation, and to point out promising directions for future research.

### 7879-25, Session 6

#### **Web image annotation using two-step filtering on social tags**

S. Cho, J. Cha, H. Byun, Yonsei Univ. (Korea, Republic of)

Web image annotation has become an important issue with exploding web images and the necessity of effective image search. The social tags have recently utilized at image annotation because they can reflect the user's tagging tendency, and reduce the semantic gap. However, an effective filtering procedure is required to extract the relevant tags since the user's subjectivity and noise of tags. In this paper, we propose a two-step filtering on social tags for image annotation. This method conducts the filtering and verification tasks by analyzing the distribution of visual features and the relation between tags on visual neighbor images. Our method consists of the following three steps: 1) the tag candidate set is founded by searching the visual neighbor images, 2) from a given tag candidate set, coarse filtering is conducted by tag grouping and voting technique, 3) the dense filtering is conducted by using similarity verification for coarse filtered candidate tag set. To evaluate the performance of our approach, we conduct the experiments on a social-tagged image database obtained from Flickr. We compare the accuracy between the voting technique and our proposed technique. Our experimental results show that our method has a significant improvement in image annotation.

# Conference 7880: Media Watermarking, Security, and Forensics XIII

Monday-Wednesday 24-26 January 2011

Part of Proceedings of SPIE Vol. 7880 Media Watermarking, Security, and Forensics III

## 7880-01, Session 1

### Signal rich art: enabling the vision of ubiquitous computing

B. Davis, Digimarc Corp. (United States)

Advances in networking and mobile computing are converging with digital watermarking technology to realize the vision of Ubiquitous Computing, wherein mobile devices can sense, understand, and interact with their environments. Watermarking is the primary technology for embedding signals in the media, objects, and art constituting our everyday surroundings, and so it is a key component in achieving Signal Rich Art: art that communicates its identity to context-aware devices. However, significant obstacles to integrating watermarking and art remain, specifically questions of incorporating watermarking into the process of creating art. This paper identifies numerous possibilities for research in this arena.

## 7880-02, Session 2

### Comparison of three solutions to correct erroneous blocks to extract an image of a multiplicative homomorphic cryptosystem

N. Islam, W. Puech, R. Brouzet, LIRMM (France)

Multiplicative homomorphic properties of a cryptosystem can be used in various applications requiring security, protection and authentication e.g. digital ngerprinting, electronic voting, online betting etc. Secret sharing between two or more parties exploiting multiplicative homomorphic property of RSA results into erroneous blocks while extracting the message. The generation of these erroneous blocks limits the capabilities of homomorphic properties of RSA to be used in its full extend. This paper provides three dierent approaches as solutions to the problem of erroneous blocks in image. These solutions are, mean value approach, shortest distance approach and image preprocessing approach. It has been observed that shortest distance approach results into good PSNR but it is computationally expensive. The best approach with high PSNR is image preprocessing approach before sharing process, which results into no erroneous blocks in the extracted image, thus no extra extraction techniques are required.

## 7880-08, Session 4

### Using feature point-based extraction for STDM 3D-mesh watermarking that withstands the cropping attack

M. Montañola Sales, I. R. M. Darazi, J. Giard, Univ. Catholique de Louvain (Belgium); P. Rondao Alface, Alcatel-Lucent Bell Labs. (Belgium); B. M. Macq, Univ. Catholique de Louvain (Belgium)

State-of-the-art blind and robust 3D watermarking schemes already withstand combinations of a wide variety of attacks (e.g. Noise addition, simplification, smoothing, etc) except cropping.

Spread Spectrum Dithering Modulation (STDM) method is an extension of Quantization Index Modulation (QIM). Besides the simplicity and the trade-off between high capacity and robustness provided by QIM methods, it is also resistant against re-quantization.

This paper focuses on two state-of-the-art techniques which offer different and complementary advantages, respectively QIM-based 3D watermarking and feature point-based watermarking synchronization. The idea is to combine both in such a way that the new scheme would benefit from the advantages of both techniques and compensate for their respective fragilities.

We show that robustness against cropping and other common attacks is achieved provided that at least one feature point as well as its corresponding local neighborhood is retrieved.

## 7880-12, Session 5

### A curiosity regarding steganographic capacity of pathologically nonstationary sources

A. D. Ker, Univ. of Oxford (United Kingdom)

No abstract available

## 7880-13, Session 5

### Design of adaptive steganographic schemes for digital images in spatial domain

T. Filler, J. Fridrich, Binghamton Univ. (United States)

Most steganographic schemes for real digital media embed messages by minimizing a suitably defined distortion function. In practice, this is often realized by syndrome codes which offer near-optimal rate--distorton performance. However, the distortion functions are designed heuristically and the resulting steganographic algorithms are thus suboptimal. In this paper, we present a practical framework for optimizing the parameters of additive distortion functions to minimize the statistical detectability. We first introduce a rich parametric model which assigns a cost of making a change at every pixel based on its neighborhood. Then, we present a practical method for optimizing the parameters with respect to a chosen detection metric and feature space used in blind steganalysis. A computationally appealing choice for a detection measure is the so-called Maximum Mean Discrepancy (MMD) also used in steganalysis. Unfortunately, the parameters optimized w.r.t. MMD do not lead to more secure stegosystems. We explain this behavior by recalling the direct connection between MMD and Bayes risk of Parzen window classifiers. A tighter connection to practice is obtained using a more theoretically founded detection metric -- the size of the margin between support vectors in soft-margin SVMs. We show that model parameters with smaller margin lead to more secure stegosystems. Optimal parameters obtained by the Nelder-Mead simplex algorithm are presented and embedding methods are tested by blind steganalyzers utilizing various feature sets. Experimental results show that as few as 50 images are sufficient for obtaining optimal parameters of the cost model allowing us to speed up the parameter search.

## 7880-14, Session 5

### Feature restoration and distortion metrics

V. Chonev, A. D. Ker, Univ. of Oxford (United Kingdom)

No abstract available.

## 7880-15, Session 6

### Image and video manipulation: Past, present, and future

No abstract available



## 7880-16, Session 7

### Lossless image data embedding in plain areas

M. Fallahpour, D. Megias, Univ. Oberta de Catalunya (Spain); Y. Q. Shi, New Jersey Institute of Technology (United States)

This letter presents a lossless data hiding scheme for digital images which uses an edge detector to locate plain areas for embedding. The proposed method takes advantage of the well-known gradient adjacent prediction utilized in image coding. In the suggested scheme, prediction errors and edge values are first computed and then, excluding the edge pixels, prediction error values are slightly modified through shifting the prediction errors to embed data. The aim of proposed scheme is to decrease the amount of modified pixels to improve transparency by keeping edge pixel values of the image. The experimental results have demonstrated that the proposed method is capable of hiding more secret data than the known techniques at the same PSNR, thus proving that using edge detector to locate plain areas for lossless data embedding can enhance the performance in terms of data embedding rate versus the PSNR of marked images with respect to original image.

## 7880-17, Session 7

### Re-synchronizing audio watermarking after nonlinear time stretching

M. Steinebach, S. Zmudzinski, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

No abstract available.

## 7880-19, Session 8

### On locating steganographic payload using residuals

T. Quach, Sandia National Labs. (United States)

Locating steganographic payload using Weighted Stego-image (WS) residuals has been proven successful provided a large number of stego images are available. In this paper, we revisit this topic with two goals. First, we argue it is a promising approach to locate payload by showing in the ideal scenario where the cover images are available, the expected number of stego images needed to perfectly locate all load-carrying pixels is the logarithm of the payload size. Second, we generalize cover estimation to a maximum likelihood decoding problem and demonstrate that a second order statistical cover model can be used to compute residuals to locate payload.

## 7880-20, Session 8

### Steganalysis using logistic regression

I. Lubenko, A. D. Ker, Univ. of Oxford (United Kingdom)

No abstract available.

## 7880-21, Session 8

### Steganalysis in high dimensions: fusing classifiers built on random subspaces

J. Kodovsky, J. Fridrich, Binghamton Univ. (United States)

Modern steganographic methods achieve steganographic security by minimizing an appropriately defined distortion function in a feature space of a very high dimension. On the other hand, steganalysis, as implemented today using classifiers, does not scale as easily --

working with very high dimensional features leads to problems with the lack of training data, infeasible complexity of training, degradation of generalization abilities, lack of robustness to cover source, and saturation of performance below its potential. To address these problems collectively known as the curse of dimensionality, we propose a new clean approach in which we strive to minimize the role of human design and put emphasis on automatization and generality of the entire process. Based on the character of the media being analyzed, the steganalyst first puts together a high-dimensional set of "prefeatures" selected to capture dependencies among individual cover elements. Then, a family of weak classifiers is built on random subspaces of the prefeature space. The final classifier is constructed by fusing the decisions of individual classifiers. The advantage of this approach is its universality, low complexity, simplicity, and improved performance when compared to classifiers trained on the entire prefeature set. Experiments with the steganographic algorithm nsF5 demonstrate the usefulness of this approach in comparison with feature sets built using heuristic "by hand."

## 7880-22, Session 9

### Private content identification based on soft fingerprinting

S. V. Voloshynovskiy, T. S. Holotyak, O. J. Koval, F. P. Beekhof, Univ. of Geneva (Switzerland)

Content identification systems are widely used in various emerging applications ranging from identification of physical objects and humans to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on binary digital fingerprinting. A digital fingerprint represents a short, robust and distinctive content description allowing fast and privacy-preserving operations. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data.

Digital content identification based on digital fingerprinting become a de facto standard in biometrics applications, physical object identification based on physical unclonable functions (PUFs) as well as various multimedia and security and management applications. During last years, certain important practical and theoretical achievements were reported. The main efforts on the side of practical algorithms have been concentrated on robust feature selection and fast indexing techniques mostly borrowed from content-based retrieval applications [1,2]. The information-theoretic limits of content identification under infinite length and ergodic assumptions have been investigated by Willems et. al. [3] using the jointly typical decoder. The detection-theoretic limits have been first studied in [4] under geometrical desynchronization distortions and a further extension of this framework was proposed in [5] for the case of finite-length fingerprinting and null hypothesis. The used decision rule is based on minimum Hamming distance decoder with a fidelity constraint under binary symmetric channel model. Since this decision rule requires the computation of likelihoods/distances between the query and all database entries, the complexity of the considered identification is exponential with the input length. Due to the additional fact that identification services are often outsourced to third parties and state authorities, the privacy of data owners is an important issue and remains largely unexplored.

Privacy issues have been mainly studied in the authentication applications due to the public sharing of helper data and extended to various practical implementations based on Slepian-Wolf and Wyner-Ziv distributed coding. In our previous work [6] we have considered the rate-privacy-complexity trade-off for identification applications based on a framework presented in [7]. This approach is based on a global privacy amplification where all bits of stored fingerprint are randomized with the same probability disregarding their reliabilities. The main benefit from the presented framework of bit reliability (a.k.a. soft fingerprinting) was demonstrated for the reduction of identification complexity based on bounded distance decoder. Obviously such a construction does not fully benefit from the fact that the information about the reliable bits can be present at the encoder and decoder that can be used not only for the efficient decoding but also for the enhanced privacy amplification.

Therefore, in this paper we introduce an information-theoretic framework for the analysis of private content identification based on finite length fingerprinting with bit reliability side information. Contrary to previous works, we propose a privacy amplification mechanism, which is adaptive to the bit reliability and demonstrate its advantages over the state-of-the-art privacy amplification. We present and analyze a privacy-preserving technique which asymptotically achieves the theoretical performance limits in terms of identification rate. The proposed technique is based on Forney's type of erasure/list decoding [8] implemented in the form of bounded distance decoder. The analysis is performed for the case of perfect match between the side information shared between the encoder and decoder as well as for the case of imperfect side information. We analyze the optimal trade-off between the achievable rate and privacy as the solution to a constraint optimization problem. Finally, we will show that content identification is closely related to the problem of erasure and list decoding [8] and further investigation of this connection might reveal many interesting insights to the analysis and design of future identification systems based on soft fingerprinting.

- [1] J. Haitisma, T. Kalker, and J. Oostveen, "Robust audio hashing for content identification," in International Workshop on Content-Based Multimedia Indexing, Brescia, Italy, September 2001, pp. 117-125.
- [2] F. Lefebvre and B. Macq, "Rash : RAdon Soft Hash algorithm," in Proceedings of EUSIPCO - European Signal Processing Conference, Toulouse, France, 2002.
- [3] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in Proc. 2003 IEEE Int. Symp. Inform. Theory, Yokohama, Japan, June 29 - July 4 2003, p. 82.
- [4] S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Robust perceptual hashing as classification problem: decision-theoretic and practical considerations," in Proceedings of the IEEE 2007 International Workshop on Multimedia Signal Processing, Chania, Crete, Greece, October 1-3 2007.
- [5] A. L. Varna, A. Swaminathan, and M. Wu, "A decision theoretic framework for analyzing hash-based content identification systems," in ACM Digital Rights Management Workshop, Oct. 2008, pp. 67-76.
- [6] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, T. Holotyak, Information-Theoretical Analysis of Private Content Identification, IEEE Information Theory Workshop, ITW 2010, Dublin, Ireland, August 30-September 3, 2010.
- [7] S. Voloshynovskiy, F. Beekhof, O. Koval, and T. Holotyak, "On privacy preserving search in large scale distributed systems: a signal processing view on searchable encryption," in Proceedings of the International Workshop on Signal Processing in the Encrypted Domain, Lausanne, Switzerland, 2009.
- [8] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," IEEE Trans. Inf. Theory, vol. 14, pp. 206-220, March 1968.

## 7880-23, Session 9

### **Geometrically robust perceptual fingerprinting: an asymmetric case**

O. J. Koval, S. V. Voloshynovskiy, Univ. of Geneva (Switzerland)

In this paper we consider the problem of geometrically resilient multimedia object identification. We analyze performance limits attainable in this application under a certain parametric class of geometrical distortions. In the analysis we assume that the query and database entries are distorted and desynchronized versions of the same ideal/original multimedia data. We present conditions on the geometrical desynchronization parameter set cardinality to ensure reliable communications.

## 7880-24, Session 9

### **Trade-offing privacy-complexity of identification problem**

T. S. Holotyak, S. V. Voloshynovskiy, O. J. Koval, F. P. Beekhof, Univ. of Geneva (Switzerland)

In this paper we advocate the extension of techniques for the fast identification of multimedia content. Solving performance - complexity - privacy preserving optimization problem we propose an approach, where performance - complexity trade-off is analyzed for the predefined levels of the private information disclosure. The proposed identification method is based on a soft fingerprinting and consists of two stages: at the first stage, the list of possible candidates is estimated based on the most reliable bits of soft fingerprint, and, at the second stage, the traditional maximum likelihood decoding is applied to the obtained list to find a single the best match. The complexity-performance trade-off is investigated by considering different distortions introduced during image acquisition. The soft fingerprint is computed based on random projections with sign-magnitude decomposition of projected coefficients. The estimate of bit reliability is deduced directly from the observed coefficients. We investigate different decoding strategies to estimate the list of candidates, which minimize the probability of missing the right index on the list. The attained complexity-performance trade-off demonstrates superiority of the proposed technique over certain state-of-the-art methods including local sensitivity hashing.

## 7880-25, Session 10

### **A context model for microphone forensics and its application in evaluations**

C. Krätzer, K. Qian, M. Schott, J. Dittmann, Otto-von-Guericke- Univ. Magdeburg (Germany)

In this paper we first design a suitable context model for microphone recordings, formalizing and describing the involved signal processing pipeline and the corresponding influence factors. As a second contribution we apply the context model to devise empirical investigations about: a) the identification of suitable classification algorithms for statistical pattern recognition based microphone forensics, evaluating 74 supervised classification techniques and 8 clusterers; b) the determination of suitable features for the pattern recognition (with very good results for second order derivative MFCC based features), showing that a reduction to the 20 best features has no negative influence to the classification accuracy, but increases the processing speed by factor 30; c) the determination of the influence of changes in the microphone orientation and mounting on the classification performance, showing that the first has no detectable influence, while the latter shows a strong impact under certain circumstances; d) the performance achieved in using the statistical pattern recognition based microphone forensics approach for the detection of audio signal compositions.

## 7880-26, Session 10

### **Double H.264/AVC compression detection using quantized nonzero AC coefficients**

D. Liao, R. Yang, H. Liu, J. Li, J. Huang, Sun Yat-Sen Univ. (China)

Developments of video processing technology make it much easier to tamper with video. In some situation, such as in a lawsuit, it is necessary to prove videos are not tampered. This contradiction poses challenges to ascertain integrity of digital videos. Most of tamperings occur in pixel domain. However, nowadays videos are usually stored in compressed format, such as H.264/AVC. For attackers it is necessary to decompress original video bitstreams and recompress it into compressed domain. As a result, by detecting double compression,

we can authenticate integrity of digital video. In this paper, we propose an efficient method to detect whether or not a digital video has been double compressed. Specifically, we use probability distribution of quantized nonzero AC coefficients as features to distinguish double compressed video from those original one compressed video. If a smaller QP is used in the second compression, the original distribution law will be violated, which can be used as the evidence of tampering.

#### 7880-27, Session 10

### Forensic printer detection using intrinsic signatures

A. K. Mikkilineni, N. Khanna, E. J. Delp III, Purdue Univ. (United States)

The ability to intrinsically characterize a printer leads to questions about privacy and anonymity. There are many instances where existence of the intrinsic signature in the printed document is undesirable. This is useful, for example, in protecting the anonymity of people distributing printed documents during peaceful protest. On the other hand, some groups may want to hide the intrinsic signature for illegal purposes, such as distribution of counterfeit currency. We have shown results that indicate the addition of masking signals and/or noise to the document before printing does not prevent estimation of the intrinsic signature. The intrinsic signature could still be obtained simply by extending the feature set in order to maintain the same performance over those attacks.

In this work we follow up on those results and design a distance based metric for use in printer identification based on the intrinsic signature. This will provide a solution to the printer detection problem for printed documents.

As we found in our earlier work, using both texture and banding features to characterize printed regions of the page captures the intrinsic signature irrespective of any modifications performed on the document before printing.

We will present new results showing that this set of features can be used for forensic detection of printers.

#### 7880-28, Session 11

### Non-destructive forensic latent fingerprint acquisition with chromatic white light sensors

M. Leich, S. Kiltz, J. Dittmann, C. Vielhauer, Otto-von-Guericke- Univ. Magdeburg (Germany)

Latent fingerprints are of vital importance for modern crime scene investigation. The most frequently used methods to secure these fingerprints (i.e. dusting with powder) destroy the original evidence irreversibly and thus make it unavailable for additional verification or further analysis like tests for substance abuse and age estimation.

In this paper a series of tests is performed to investigate the overall suitability of a high resolution off-the-shelf chromatic white light sensor for the contact-less and non-destructive acquisition. In particular 3D height field and reflection data of 10 different latent fingerprints on four different types of surfaces (hard disk platter, painted car body, brushed aluminum, veneered plywood) are experimentally studied. Standard algorithms as well as customized methods for the visual enhancement of the acquired fingerprints are assessed.

While the quality of the acquired data is highly dependent on surface structure, the quality of the fingerprint and the processing, preliminary results for scans on ideal surfaces are very detailed and enable even the use of level three fingerprint features (pores) for matching. Under these circumstances error rates are currently expected to be below 0.005 (False Acceptance Rate) and 0.05 (False Rejection Rate).

#### 7880-29, Session 11

### Detecting messages of unknown length

T. Pevny, Czech Technical Univ. in Prague (Czech Republic)

No abstract available.

#### 7880-30, Session 11

### A new paradigm for steganalysis via clustering

A. D. Ker, Univ. of Oxford (United Kingdom); T. Pevny, Czech Technical Univ. in Prague (Czech Republic)

No abstract available.

#### 7880-31, Session 12

### Collusion-secure patchwork embedding for transaction watermarking

W. Berchtold, S. Zmudzinski, M. Schäfer, M. Steinebach, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

No abstract available.

#### 7880-32, Session 12

### Probabilistic fingerprinting codes used to detect traitor zero-bit watermark

M. Desoubeaux, G. Le Guelvouit, Orange Labs. (France); W. Puech, LIRMM (France)

Traitor tracing aims at preventing unauthorized redistribution of multimedia content by embedding each authorized copy with individual sequences of bits having robustness against collusion attacks. Collusion is the process used by dishonest users to forge an untraceable content with their copies. In this paper we present a new method dedicated to video content distribution. It is based on a probabilistic traitor tracing code and an orthogonal zero-bit informed watermark. We use the well known Tardos fingerprinting tracing function to reduce the search space of suspicious users. Their guiltiness is then confirmed by detecting the presence of their personal watermark embedded with a personal key. To prevent watermarking keys storage for the distributor we use a part of the user fingerprinting sequence as a personal embedding key. This method permits to reduce the code length in function of the false alarm probability of the watermarking method. Therefore it ensures a global smallest false alarm probability compared to original probabilistic codes. However efficiency of this process depends strongly on the number of colluders at the watermarking side. Indeed we assume that colluders attempt to erase their personal watermark with average attacks. To increase robustness against such attacks we propose an additive correlation method based on successive watermarked images. We present the robustness of such a method for different sizes of collusion. We finally study the false alarm probability of this additive correlation method and present tracing results for short length of Tardos code.

7880-33, Session 12

**Rihamark: perceptual image hash  
benchmarking**

M. Steinebach, Fraunhofer-Institut für Sichere Informations-  
Technologie (Germany); H. Eckehard, C. Zauner, FH OÖ  
Studienbetriebs GmbH (Austria)

No abstract available.

7880-34, Session 13

**A spatio-temporal framework based on  
eigenvectors for improved face recognition**

M. Ouaret, J. E. Dugelay, EURECOM (France)

Most of state-of-the-art accurate face recognition systems require heavy processing and face normalization. In this paper, we introduce a real time hybrid face recognition system, combining spatial and temporal video information while maintaining a fair processing complexity. The introduced hybrid solution combines spatial (eigenfaces) with temporal (tomofaces) eigenvectors in a two layers fusion scheme for improved face recognition. Initially, eigenfaces and tomofaces are simultaneously applied to a non-normalized input video sequence. Then, several fusion methods (first fusion layer) are applied to the resulting scores from both, eigenfaces and tomofaces. The proposed system generates the final result by majority voting of all the fusion modules from the first layer of both biometric traits. The proposed technique shows improvements of around 13% and 8% in correct identification rate over standalone tomofaces and eigenfaces, respectively. Thus, the proposed scheme achieves a good performance under realistic conditions with low complexity (PCA and Canny edge detector) and without heavy pre-processing (image normalization).

### 7881A-01, Session 1

#### **Towards a multimedia remote viewer for mobile thin clients**

M. P. Mitrea, B. Joveski, L. Gardenghi, TELECOM & Management SudParis (France); P. Simoens, IBBT (Belgium); J. Marshall, Prologue (France); B. Vankeirsbilck, IBBT (Belgium); F. J. Prêteux, TELECOM & Management SudParis (France); B. Dhoedt, IBBT (Belgium)

Be there a mobile user wanting to connect to a multimedia server. In order to allow him/her to enjoy the same user experience (play, interact, edit, store and share capabilities) as in a fixed LAN environment, several dead-locks are to be dealt with: (1) a heavy and heterogeneous content should be sent through a bandwidth constraint network; (2) the displayed content should be of good quality; (3) user interaction should be processed in real-time and (4) the complexity of the practical solution should not exceed the features of the mobile client in terms of CPU, memory and battery.

The present paper demonstrates that the MPEG-4 scene technologies (BiFS and LAsEr) can provide for all the needs of a remote mobile thin viewer.

First, a BiFS/LAsEr-based architecture for remote viewer is advanced. In order to ensure backward compatibility with the legacy applications, this architecture takes as input the traditional X11 graphical content and it converts it into MPEG-4 BiFS/LAsEr. Once converted, the content is streamed live to a thin client device, where the user can watch and interact with it.

The second part of the paper is devoted to an objective assessment of this architecture. By considering three types of content (simple graphics, a text editor and www browsing) the two MPEG technologies (BiFS and LAsEr) are compared to their competitors. The overall results demonstrate that MPEG technologies are more efficient for mobile thin clients than the direct extension of the traditional solutions.

The final part of the paper presents the perspectives opened by this proof of concept.

### 7881A-02, Session 1

#### **Multimodal sensing-based camera applications**

M. Bordallo Lopez, J. Hannuksela, O. J. Silvén, Univ. of Oulu (Finland); M. Vehviläinen, Nokia Research Ctr. (Finland)

The increased sensing and computing capabilities of mobile devices can provide enhanced mobile user experience. Integrating the data from different sensors offers a way to improve application performance in camera based applications.

A key advantage of using cameras as an input modality is that it enables recognizing the context. Therefore, computer vision has been traditionally utilized in user interfaces to look at people and automatically detecting the user actions. The imaging applications can also make use of various sensors for improving user interaction and robustness of the system.

In this context, two applications fusing the sensor data with the results obtained from video analysis have been implemented on a Nokia Nseries mobile device.

The first application is a real-time panorama builder that uses the mobile device's accelerometers to improve the overall quality, providing also instructions during the capture. The second solution is a real-time user interface that can be used for browsing large images. The solution enables the display to be controlled by the motion of the user's hand using the built-in sensors as complementary information.

The experiments show that fusing the sensor data greatly improves camera based applications especially when the conditions are not optimal for approaches using only cameras.

### 7881A-03, Session 1

#### **Mobile text messaging solutions for obesity prevention**

D. Akopian, V. Jayaram, L. Aaleswara, M. Esfahanian, The Univ. of Texas at San Antonio (United States); C. Mojica, D. Parra-Medina, The Univ. of Texas Health Science Ctr. at San Antonio (United States); S. Kaghyan, Yerevan State Univ. (Armenia)

This paper provides an overview of the state-of-the-art in mobile phone technologies which can be used for health promotion interventions and related data collection. It also describes a proposed system architecture customized for an obesity prevention program.

Recently several healthcare projects have been reported that integrate cell phones into the data communication chain. They use cell-phone technologies to various extents. While it is tempting to incorporate phones as a healthcare instrument broadly, the cell phone market is very diverse with many technologies available for application development. This paper summarizes market data for general and smart phone sales, systematizes software development layers from a portability point of view, and analyzes existing development tools.

As a case study, a messaging system is proposed for a health-promotion research study to prevent obesity and obesity-related health disparities among low-income Latino adolescent girls. Messaging and polling mechanisms are used to communicate and automatically process response data from the target constituency. The aim of the project is to incorporate low-cost, mobile technology to promote health and connect youth to community resources, design an intervention to increase moderate to vigorous physical activity, and to ascertain the feasibility of the approach and effects among Latino adolescent girls.

### 7881A-04, Session 2

#### **Quality and noise measurements in mobile phone video capture**

D. Petrescu, J. Pincenti, Motorola, Inc. (United States)

The quality of videos captured with mobile phones has become increasingly important particularly since resolutions and formats have reached a level that rivals the capabilities available in the digital camcorder market, and since many mobile phones now allow direct playback on large HDTVs. The video quality is determined by the combined quality of the individual parts of the imaging system including the image sensor, the digital color processing and the video compression, each of which has been studied independently. In this work, we study the combined effect of these elements on the overall video quality. We do this by evaluating the capture under various lighting, color processing, and video compression conditions. First, we measure full reference quality metrics between encoder input and the reconstructed sequence, where the encoder input changes with light and color processing modifications. Second, we introduce a system model which includes all elements that affect video quality, including a low light additive noise model, ISP color processing, as well as the video encoder. Our experiments show that in low light conditions and for certain choices of color processing the system level visual quality may not improve when the encoder becomes more capable or the compression ratio is reduced.

### 7881A-05, Session 2

#### **3D scene reconstruction based on multiview distributed video coding in the Zernike domain for mobile applications**

V. Palma, M. Carli, A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this paper a Multi-view Distributed Video Coding (DVC) scheme for mobile applications is presented. Specifically a new fusion technique between temporal and spatial side information in Zernike Moments (ZM) space is proposed.

As well known DVC introduces a flexible architecture that enables the design of very low complex video encoders compared to its traditional counterparts. The main goal of our work is to generate at the decoder the side information that optimally blends temporal and interview data. Multi-view DVC performance strongly depends on the side information quality built at the decoder. At this aim to improve its quality a spatial view compensation/prediction in Zernike moments' domain is applied. More in detail, we first apply state of the art key point extraction and matching algorithms to estimate the parameters characterizing the effects of the geometrical transformations among different views in the image planes. Then, to handle rotations, we partition each view in blocks and for each of them we compute the Zernike moments as a projection of the function defining the Region Of Interest onto a set of orthonormal functions within circles whose radii are selected according to the previously estimated zoom factors. Spatial and temporal motion activity will be fused together to obtain the overall side-information. The proposed method will be evaluated by rate-distortion performances for different inter-view and temporal estimation quality conditions.

### 7881A-06, Session 2

#### **Psycho-physiological effects of head-mounted displays in ubiquitous use: a comparison of see-through and non-see-through, binocular, and monocular conditions**

T. Kawai, Waseda Univ. (Japan); J. P. Häkkinen, University of Helsinki (Finland) and Nokia Research Center (Finland) and Aalto University (Finland); K. Oshima, Waseda University (Japan); H. Saito, T. Yamazoe, Waseda Univ. (Japan); H. Morikawa, Waseda University (Japan); G. S. Nyman, University of Helsinki (Finland)

Multimedia devices are now found everywhere in many different forms, including devices that can be worn like clothing. The head mounted-displays (HMDs) are wearable devices to present information as the user views the surrounding environment. HMDs can be classified into following four types: binocular see-through, binocular non-see-through, monocular see-through, monocular non-see-through. These types may be different in terms of user's visual information processing, and consequently the user experience. In practical use, the workload may also differ by the types.

In this paper, the authors carried out an experiment to examine the workload by the use of three of above-mentioned four types HMDs: binocular and monocular see-through, and monocular non-see-through. As the tasks to be performed, the ten subjects were asked short walking through a University building while wearing three types HMDs and perceiving visual stimulation. The total distance walked was approximately 600 meters. News video was provided to the subjects as the audio-visual stimulation. The subjective indexes were measured using a task load index (NASA-TLX) after each trial. As objective indexes, the heart rate was measured during each trial. The results showed common tendencies in both subjective and objective indexes.

### 7881A-07, Session 2

#### **Progressive imagery with scalable vector graphics**

G. A. Fuchs, H. Schumann, Univ. Rostock (Germany); R. U. Rosenbaum, Sr., Univ. of California, Davis (United States)

Vector graphics can be scaled without loss of quality, making them suitable for mobile image communication where a given graphics must be typically represented in high quality for a wide range of screen resolutions. One problem is that file size increases rapidly as content becomes more detailed, which can degrade response times and

efficiency in mobile settings. Similar challenges have been addressed for raster images using progressive refinement schemes, however taking advantage of compliant progressive vector graphics in common image communication is still an open research question.

Therefore in this publication we show how to provide progressive refinement schemes based on the extensible Scalable Vector Graphics (SVG) standard. We propose two strategies: decomposition of the original SVG and incremental transmission using (1) several linked files and (2) element-wise streaming of a single file. The first strategy exploits SVG's ability to reference external resources, the second uses a transcoder to sequentially stream individual geometric primitives and for client-side reassembly of the SVG file.

The publication further discusses how both strategies are employed in mobile image communication scenarios where the user can interactively define Rols for prioritized image communication. Our contribution closes with results we obtained from a prototypically implemented client/server setup.

### 7881A-08, Session 3

#### **Mobile 3D quality of experience evaluation: a hybrid data collection and analysis approach**

A. P. Gotchev, S. Jumisko-Pyykkö, A. R. Boev, T. Utrainen, J. Häyrynen, M. Mikkola, Tampere Univ. of Technology (Finland); M. Hannuksela, Nokia Research Ctr. (Finland)

The paper presents a hybrid approach to study the user's experienced quality of 3D visual content on mobile auto-stereoscopic displays. It combines extensive subjective tests with collection and objective analysis of eye-tracked data. 3D cues which are significant for mobiles are simulated in the generated 3D test content. The methodology for conducting subjective quality evaluation includes hybrid data-collection of quantitative quality preferences, qualitative impressions, and binocular eye-tracking. We present early results of the subjective tests along with gaze fixation maps obtained from raw eye-tracked data after statistical analysis. The study contributes to the question what is important to be visualized on portable auto-stereoscopic displays and how to maintain and visually enhance the quality of 3D content for such displays.

### 7881A-09, Session 3

#### **Overcome the shortcoming in mobile stereoscopy**

K. Lee, S. Kim, Korea Institute of Science and Technology (Korea, Republic of)

In stereoscopic camera system, representative two types such as a parallel and a convergence have been reported. Recently a diverging type is introduced by Dr. Son. It has a similar stereoscopic property to the converging type. Diverging stereo camera alignment may be to be a suitable to a mobile stereoscopy because the mobile device has the lacked space for configuring cameras to make either ortho- or hyperstereoscopic condition with a small size of display, there is only a hypostereoscopic condition giving a card-board effect. This matter means that mobile stereoscopy cannot provide a presence with a good depth sense to an observer. For this reason, we focused on the depth sense control method with a switchable stereo camera alignment. In converging type, the fusible stereo area becomes wider compared to a parallel type when the same focal length was used in both types. This matter means that the stereo fusible area formed by converging type to be equal to the parallel type with a shorten focal length. Therefore there is a kind of the zoom-out effect at the reconstructed depth sense, because the disparity obtained by the converging type to be equal to the disparity by the parallel type having a shorten focal length than before the comparison. In diverging type, the fusible stereo area becomes narrower than the parallel. As the same way, the diverging type guarantees the same characteristic of that an increased focal length is considered in parallel type. Therefore there is a zoom-in effect existing. We considered the permitted disparity

about 2.5mm at the mobile display for taking suitable stereo fusion. As the result, The satisfied both converging and diverging angles are taken by the theoretical consideration such as 4 and 1.5 degrees under the condition of the narrowed inter camera distance (5mm) with a wide FOV (common mobile phone camera). Additionally, the zoom-in effect becomes rapidly changed by the increased angle but zoom-out becomes retarded relatively.

### 7881A-10, Session 3

#### Comparative study of autostereoscopic displays for mobile devices

A. R. Boev, J. Häyrynen, A. P. Gotchev, Tampere Univ. of Technology (Finland)

We present a comparative study of several portable auto-stereoscopic displays. We overview the technologies, they are based on and present parameters which are used for their evaluation, such as crosstalk, 3D contrast, optimal viewing zone, optimal disparity range, and frequency domain throughput. We present a simply yet precise methodology for their measurement and summarize and discuss the results of the measurements.

### 7881A-11, Session 3

#### Subjective evaluation of mobile 3D content: depth range versus compression artifacts

T. Haustola, S. Jumisko-Pyykkö, A. R. Boev, A. P. Gotchev, Tampere Univ. of Technology (Finland)

The theories about human visual perception state that the views formed by the eyes are used to form a fused central, so-called 'cyclopean', view and to form stereopsis. Correspondingly, the perceived quality of a 3D scene is a combination of two components - "2D" quality (image details), and its "3D" quality (quality of the binocular depth cues). However, the very presence of stereoscopic depth changes the way image details are perceived and equivalent loss of image fidelity might cause different quality experience of 2D and 3D video.

In this study, we aim at quantifying the impact of varying binocular depth in 3D video on the overall perceptual quality in the presence of varying compression artefacts. We have designed and conducted a set of subjective experiments, where the binocular depth and the image quality of a scene are independently varied on denser grids. Four real-world multiview videos with different characteristics of scene dynamics are used in the tests. From each multiview sequence, a number of stereoscopic videos were created using different camera pairs, thus achieving varying camera baseline for the same scene. More specifically, the depth was varied between 2D, HD-optical baseline and mobile-optimal baseline. Each stereoscopic video was compressed with varying quality, i.e. with five different QPs. In the conducted tests, the participants were asked to grade the quality of each compressed video in a single stimulus setting. The results obtained for a solid base for designing an objective metric for 3D video quality evaluation.

### 7881A-12, Session 3

#### Development of a 3D mobile receiver for stereoscopic video and data service in T-DMB

G. Lee, H. Lee, K. Jung, N. Hur, S. Lee, Electronics and Telecommunications Research Institute (Korea, Republic of)

3D Mobile broadcasting that delivers 3D contents via mobile broadcasting network is believed to be a very attractive service because single-user environment at the terminal side is suitable for glasses-free 3D viewing and a variety of multimedia services are applicable on 3D mobile devices. With the help of a high-quality auto-stereoscopic technologies, a 3DTV service over T-DMB (Terrestrial-

digital Multimedia Broadcasting), what we call a 3D DMB service, has recently been introduced to provide further realistic mobile broadcasting services. As a killer application of the 3D DMB, 3D data service has been also developed based on MPEG-4 system technologies, where stereoscopic data contents with formats of JPEG, PNG and MNG are delivered and rendered on the 3D mobile terminal supporting T-DMB. It is believed that such 3D service technology will be useful for many 3D broadcasting applications such as advertisement, education, sports, movie, drama, and so on. Generally, backward compatibility is a crucial requirement for successful launch of new services in the digital broadcasting.

### 7881A-13, Session 3

#### A right scaled depth sense formed by using a distorted objective space based on CG stereoscopy

K. Lee, Korea Institute of Science and Technology (Korea, Republic of) and Konkuk Univ. (Korea, Republic of); D. Kim, Korea Institute of Science and Technology (Korea, Republic of); G. Um, E. Chang, G. Bang, N. Hur, Electronics and Telecommunications Research Institute (Korea, Republic of); S. Kim, Korea Institute of Science and Technology (Korea, Republic of)

In stereoscopy, depth distortion is a serious problem to provide the correct depth sense of the object when the reconstructed depth image was displayed and perceived. Uncorrected disparity is a main cause inducing the problem and it is depended on the stereoscopic circumstances such as a stereoscopic system configured a stereo camera and display and an observation related a viewing condition. The numerous studies have been reported to solve the problem but they did not give us a general solution because the causes inducing problem are crossed linking between the stereoscopic system and observation. In this paper, we suggested a new way based on computer graphics (CG) to overcome the aforementioned shortcoming of a common stereoscopy. In terms of the way, let the objective space transform as the distorted space to make a correct perceived depth sense as if we are seeing the scaled object volume which is well adjusted to user's stereoscopic circumstance. Since the distorted depth sense strong relates to the fixed objective space. In detail, all parameters which related the depth distortion such as a focal length, an inter-camera distance, an inner angle between camera's axes, a size of display, a viewing distance and an eye distance can be altered to the amount of inversed distortion in the transformed objective space by the linear relationship between the reconstructed image space and the objective space. Therefore, the depth distortion will be removed after image reconstruction process with a distorted objective space. As the result, we prepared a stereo image having a right scaled depth from -50mm to +200mm with an interval as 50mm in an official stereoscopic circumstance and showed it to 5 subjects. All subjects recognized and indicated the designed depths. Consequently, the adoption of distorted objective space is more the powerful way to present a right scaled depth sense without the depth distortion than the existing ways in CG based stereoscopy.

### 7881A-14, Session 4

#### Smart travel guide: from internet image database to intelligent system

G. Chareyron, J. Da Rugna, Pôle Univ. Léonard de Vinci (France)

From many years, tourism is a primordial matter to region economy. To help the tourist to discover a city, a region or a park, many options are provided by public tourism travel centers, by free online guides or by dedicated book guides. Nonetheless, these guides provide only mainstream information which are not conform to a particular tourist behavior. On the other hand, we may find several online image databases allowing users to upload their images and to localize each image on a map. A recent work has demonstrated that these websites

are representative of tourism practices and constitute a proxy to analyze tourism flows. This work intends to answer this question: knowing what I have visited and what other people have visited, where I should go now? This process needs to profile users, sites and photos. Our paper presents the acquired data and relationship between photographers, sites and photos and introduces the Bayesian model designed to correctly estimate the site interest of each tourism point. The third part shows an application of our schema: a smart travel guide on geolocated mobile devices. This application permits the travel guide to match the user wishes

#### 7881A-15, Session 4

### Revised benchmarking of contact-less fingerprint scanners for forensic fingerprint detection: challenges and results for chromatic white light scanners (CWL)

S. Kiltz, C. Kraetzer, J. Dittmann, C. Vielhauer, Otto-von-Guericke-Univ. Magdeburg (Germany)

Mobile contact-less fingerprint scanners can be very important tools for the forensic investigation of crime scenes. To be admissible in court, data and the collection process must adhere to rules w.r.t. technology and procedures of acquisition, processing and the conclusions drawn from that evidence. Currently, no overall accepted benchmarking methodology is used to support some of the rules regarding the localisation, acquisition and pre-processing using contact-less fingerprint scanners. Benchmarking is seen essential to rate those devices according to their usefulness for investigating crime scenes.

Our main contribution is a revised version of our extensible framework for methodological benchmarking of contact-less fingerprint scanners using a collection of extensible categories and items. The suggested main categories describing a contact-less fingerprint scanner are properties of forensic country-specific legal requirements, technical properties, application-related aspects, input sensory technology, pre-processing algorithm, tested object and materials. Using those it is possible to benchmark fingerprint scanners and describe the setup and the resulting data. Additionally, benchmarking profiles for different usage scenarios are defined. First results for all suggested benchmarking properties, which will be presented in detail in the final paper, were gained using an industrial device (FRT Microprof 200) and conducting 18 tests on 10 different materials.

#### 7881A-16, Poster Session

### A new mobile service: automatic lottery winning identification system

F. Tan, The Hong Kong Polytechnic Univ. (Hong Kong, China); Q. Huang, Univ. of Missouri-Kansas City (United States)

The number-based lottery tickets are increasingly popular all around the world and most of people will buy more than one combination for one time. When the count of tickets increases, the winning identification will become a trouble and time consuming. In this paper, a new service will be presented as a total solution for the automatic lottery winning identification. This service based on mobile device and uses image processing, special optical character recognition (OCR) in its analysis. Initially, the cell-phone camera will be used to capture a photo of the lottery tickets. Then the target numbers combinations will be extracted automatically. Furthermore, the application will automatically seek the latest winning numbers combination from the lottery agency's online server by the cell-phone. Lastly, it will show the identification result of whether winning or not, if winning, which grade will be. The application was developed on an android-based mobile device and used the Hong Kong Mark Six tickets as the training and test targets, good performance and usability were obtained.

#### 7881A-17, Poster Session

### Development of testbeds for AGPS mobile applications

D. Akopian, G. K. Ramachandran, A. Soghoian, G. V. S. Raju, The Univ. of Texas at San Antonio (United States)

During recent years location technologies have emerged as a research area. This is essentially driven by the success of US Global Positioning System (GPS) and the development of other Global Navigation Satellite Systems (GNSS).

This paper studies testbed design options based on GNU-Radio open source and Labview development platforms. GNU radio native environment allows easy incorporation of custom C/C++ units which accelerates the system faster and make it suitable for processing real GPS signals in real time. The benefits of using GNU radio SDK is that it provides already an interface between the software and USRP hardware. The paper provides details on the enhancements of the conventional AGPS receivers implemented on GNU radio platform including advanced novel acquisition and tracking units.

Labview is data flow software which provides flexible and facilitated design environment. Different from other similar software the Labview has been built to conveniently interface with data acquisition hardware for real-time signal processing including RF signals. In addition, Labview native environment also allows easy incorporation of custom C/C++ units which is important for real-time performance evaluations. Also the availability of NI GPS simulation toolkit provides facilitated options to simulate scenarios and test receiver operations.

The implementation and analysis of software GPS receiver in GNU Radio platform and using Labview simulator toolkit empowered with several algorithmic improvements in the GPS receiver architecture itself is achieved.

#### 7881A-18, Poster Session

### Integrity monitoring and mobile platform implementation for WLAN positioning

D. Akopian, S. Yalamanchili, A. Melkonyan, The Univ. of Texas at San Antonio (United States)

Global Positioning System (GPS) products help to navigate while driving, hiking, boating, and flying. GPS uses a combination of orbiting satellites to determine position coordinates. This works great in most outdoor areas, but the satellite signals are not strong enough to penetrate inside most indoor environments. As a result, a new strain of indoor positioning technologies that make use of 802.11 wireless LANs (WLAN) is beginning to appear on the market. In WLAN positioning the system either monitors propagation delays between wireless access points and wireless device users to apply trilateration techniques or it maintains the database of location-specific signal fingerprints which is used to identify the most likely match of incoming signal data with those preliminary surveyed and saved in the database. In this paper we investigate the issue of deploying WLAN positioning software on mobile platforms with typically limited computational resources. We suggest a novel received signal strength rank order based location estimation system to reduce computational loads with a robust performance. The proposed system performance is compared to conventional approaches.

#### 7881A-19, Poster Session

### Optimizing bandwidth and storage requirements for mobile images using perceptual-based JPEG recompression

D. Gill, T. Shoham, S. Carmel, ICVT Ltd. (Israel)

The increasing quality and resolution of cellular phone cameras is creating a significant burden on the device storage and the cellular network bandwidth. In this paper we propose a novel method for



recompressing digital photos, which significantly reduces their file size, without affecting their perceptual quality. The method is applied by iteratively recompressing the input image by different amounts, and computing the value of a robust, perceptual image quality measure, which consists of a pixel value difference, a blockiness detector and a texture distortion detector. The iterative process ensures that the maximum amount of compression, which still yields a perceptually identical image, is applied to each input image.

In subjective testing we have found that using our proposed method, the file size of high resolution photos may be reduced by a factor of 3X to 4X (66% - 75% reduction) on average without affecting their perceptual visual quality. We have implemented the algorithm in a mobile application for the iPhone 3Gs device, which recompresses the photos captured by the device's 3 Megapixel camera by 2.5X (60%) on average. The algorithm has also been implemented as a command-line application in Windows, Linux and MacOS.

## 7881A-21, Poster Session

### mQIM principles for MPEG-4 AVC watermarking

M. P. Mitrea, M. Hasnaoui, M. Belhaj, F. J. Prêteux, TELECOM & Management SudParis (France)

Watermarking imposed itself as an efficient yet flexible solution to digital video protection: by persistently (robustly) and imperceptibly (transparently) inserting some extra data (a mark) into a video excerpt, illegal copies can be tracked down to the last legal distribution.

The present paper takes the challenge of video watermarking for mobile (thin) terminals where video is subject to highly performing compression schemes. One of the author previous papers established the proof of concepts for compressed domain watermarking: the first transparent method, robust against transcoding and geometric attacks has been designed. However, the main limitation was connected to the data payload.

In order to solve this problem, a new watermarking scheme inserting the mark in the MPEG-4 AVC stream is proposed. The main novelty consists in considering multiple symbol Quantisation Index Modulation (mQIM) watermarking techniques instead of binary QIM. In this respect, the embedding/detection rules are first mathematically demonstrated and then objectively assessed in industrial partnership (under the framework of the MEDIEVALS French national project); with respect to previous studies, this method allows the size of the inserted mark (the data payload) to be increased by an factor  $\log_2(m)$ , while keeping the same good transparency (objectively and subjectively evaluated) and robustness (against transcoding and geometric attacks).

## 7881A-22, Poster Session

### Generalized Phi number system and its application for image decomposition and encryption

S. Agaian, Stanford Uni. (United States); Y. Zhou, Tufts Univ. (United States)

In this paper, we introduce a new generalized Phi number system (GPNS). By selecting appropriate parameters, The GPNS can be specified to the traditional PNS, the binary number system (base-2), and other integer base number systems. We investigate the applications of the new GPNS in image processing. We introduce a new parameter bit-plane decomposition method using the new GPNS. Integrating this new decomposition method with the chaotic logistic map, a new image encryption algorithm is introduced. Experimental results are given to demonstrate that the GPNS shows excellent performance in image decomposition and encryption.

## 7881A-23, Poster Session

### Local polynomial approximation and local binary pattern based face classification

R. Mehta, J. Yuan, K. Egiazarian, Tampere Univ. of Technology (Finland)

Face Classification is widely studied topic in field of Computer Vision and Multimedia Information Processing. One of the most fundamental parts of this process is an efficient face representation. The face should be represented in such a way that the feature vector is robust to illumination changes and the pose variation of the subject. In literature of face classification many methods have been proposed where features are extracted at multiple scales for robust classification. These methods however are not able to completely capture the information from different directions of the face image. Face features are aligned in specific directions, e.g. eyes, eyebrows and lips are aligned in horizontal direction while nose and face contour are aligned in vertical direction. By capturing the information in specific directions at different scales we can represent the face image in a way which is better suited for the classification purpose. In this paper we have proposed a novel method which utilizes Local Polynomial Approximation (LPA) techniques to capture the directional information of the face image at different scales. Since a face image is spatially varied and classification works better when local descriptors are used, we have incorporated Local Binary Pattern (LBP) operator in order to obtain LPA-LBP map. A blockwise operation is performed on the LPA-LBP map to extract the face image descriptor. The dimensionality of the final feature vector is quite high due to the blockwise operation. In order to further enhance the classification accuracy and to reduce the complexity, the dimensionality of the feature vector is reduced by taking into account those features which vary across the classes while remaining relatively constant within a class.

In this method first of all the directional estimates of face images (called LPA Directional Faces) are obtained using LPA filters for a specific number of directions and scales. After extracting the directional information LBP operator is applied on LPA Directional Faces to obtain LPA-LBP map which completely capture the texture information from them. The LPA-LBP map is a holistic representation of the face. In order to have a final local descriptor LPA-LBP maps are divided into blocks and histogram is evaluated for each block. Then all the histograms are concatenated to form the LPA-LBP Histogram Sequence (LPA-LBPHS). Since the dimensionality of this feature vector is high, Linear Discriminant Analysis (LDA) is used to reduce the dimensionality of LPA-LBPHS. This reduced LPA-LBPHS feature vector is used for face representation. Finally Support Vector Machine (SVM) classifier is learned in the reduced LPA-LBPHS feature space for face classification. Experiments done on standard datasets demonstrate that the proposed method has higher classification accuracy than previously proposed state-of-the-art methods.

## 7881A-24, Poster Session

### Anisotropic multiscale Lucas Kanade pyramid

J. Yuan, K. Egiazarian, Tampere Univ. of Technology (Finland)

The Lucas Kanade (LK) algorithm provides a smart iterative parameter-update rule for efficient image alignment, and it has become one of the most widely used techniques in computer vision. Applications range from optical flow and tracking to layered motion, mosaic construction, and face coding. The LK algorithm has been proved to be effective under small-noise conditions. But in real world applications, especially for face recognition, object tracking and video recognition, the images to be aligned are always captured by surveillances meaning that they could be quite noisy and might be taken from various angles. In such cases, the LK algorithm could not handle: the accuracy severely degrades when the image quality is poor; and the LK algorithm may not even converge when the angle between template and the captured image is too large. At this point, a novel concept of Lucas Kanade pyramid emerges in 2009. By extracting image pyramids from the original images and iteratively implementing LK algorithm at each level,

the LK pyramid gained better robustness and accuracy.

Yet the result of LK pyramid still suffers from heavy noise conditions and severe image distortions, this can be mainly attributed to the disability of correct image gradient calculation from the distorted image. Thus, on the basis of LK pyramid, this paper proposes a novel Anisotropic Multi-Scale Lucas Kanade Pyramid (AMSLKP) method. Instead of calculating gradients in single direction with fixed scale sizes, this paper introduces anisotropic local polynomial approximation (LPA) and intersection of confidence intervals (ICI) method to the LK pyramid. The proposed AMSLKP method first calculates the directional estimates and gradients with multiple scales; then for each direction, it adaptively selects the optimal scale for each pixel in the image using ICI rule; at last, the estimate and gradients of the distorted image is computed by fusing the directional results together.

The proposed method is evaluated in different noise conditions with various distortion levels. Experiment results show that the AMSLKP method improves the accuracy by more than ten percent compared to LK pyramid; moreover, the convergence process is accelerated.

## 7881A-25, Poster Session

### **iPhone forensics with Mac OS X based open source tools**

R. Creutzburg, T. Höne, Fachhochschule Brandenburg (Germany)

The aim of this article is to show the usefulness of Mac OS X based Open Source Tools for forensic investigation of modern iPhones.

It is demonstrated how important data stored in the iPhone are investigated.

Three different scenarios of investigations are presented that are well-suited for a forensics lab work.

## 7881A-26, Poster Session

### **Forensic investigation of certain types of mobile devices**

R. Creutzburg, S. Luttenberger, Fachhochschule Brandenburg (Germany)

The aim of this article is to show the usefulness of Windows based Open Source Tools for forensic investigation of modern mobile devices.

It is demonstrated how important data stored in the mobile device are investigated.

Different scenarios of investigations are presented that are well-suited for a forensics lab work.

# Conference 7881B: Multimedia Content Access: Algorithms and Systems V

Tuesday-Wednesday 25-26 January 2011

Part of Proceedings of SPIE Vol. 7881B Multimedia Content Access: Algorithms and Systems V

## 7881B-52, Poster Session

### No-reference blur estimation based on the average cone ratio in the wavelet domain

L. Platiša, A. Pizurica, E. Vansteenkiste, W. R. Philips, Univ. Gent (Belgium)

With extensive technological advancements in electronic imaging today, high image quality is becoming an imperative necessity in the modern imaging systems. An important part of quality assurance are techniques for measuring the level of image distortion. Recently, we proposed a wavelet based metric of blurriness in the digital images named CogACR. The metric is highly robust to noise and able to distinguish between a great range of blurriness. Also, it can be used either when the reference degradation-free image is available or when it is unknown. However, the metric is content sensitive and thus in a no-reference scenario it was not fully automated. In this paper, we further investigate this problem. First, we propose a method to classify images based on edge content similarity. Next, we use this method to automate the CogACR estimation of blur in a no-reference scenario. Our results indicate high accuracy of the method for a range of natural scene images distorted with the out-of-focus blur. Within the considered range of blur radius of 0 to 10 pixels, varied in steps of 0.25 pixels, the proposed method estimates the blur radius with an absolute error of up to 1 pixel in 80 to 90% of the images.

## 7881B-53, Poster Session

### Texture based Markovian modelling for image retrieval

D. Benboudjema, Ecole Nationale Supérieure de l'Electronique et de ses Applications (France)

Texture is one of the main features with color, shape, edges ... by which human being perceives image. It can be viewed as a set of pixels within an image whose local statistics or local properties (e.g. periodicity, frequency) are constant or slightly varying. In this paper we address from the statistical standpoint, the image indexing problem for image retrieval. Two new Markov model based approaches allowing texture feature extraction will be proposed and a comparison to the texture features based on Gabor filters will be presented. The three methods have been tested for image retrieval task using SVM classifier with a Gaussian kernel on texture-oriented database, Brodatz, and on another texture database. The experimental results show for the proposed scheme promising results.

## 7881B-54, Poster Session

### Non-supervised macro segmentation of the large-scale TV videos

H. Bai, Y. Dong, France Telecom R&D Beijing (China)

In this paper, a novel non-supervised macro segmentation algorithm is presented by detecting duplicate sequences of large-scale TV videos. Motivated by the fact that "Inter-Programs" are repeatedly inserted into the TV videos, so the macro structure of the videos can be effectively and automatically generated by identifying the special sequences. There are four sections in the algorithm, namely, keyframe extraction, discrete cosine transform-based feature generation (a fixed-size \$64D\$ signature), Locality-Sensitive Hashing (LSH)-based frame retrieval and macro segmentation through the duplicated sequence detection and the dynamic programming. The main contributions are: (1) supply one effective and efficient algorithm for the macro segmentation in the large-scale TV videos, (2) LSH can quickly query the similar frames, and

(3) the non-supervised learned duplicate sequence models are used to find the lost duplicate sequences by the dynamic programming. The algorithm has been tested in 15-day different-type TV streams. The F-measure of the system is greater than 96%. The result shows that the algorithm is efficient and effective for the macro segmentation.

## 7881B-39, Session 5

### Material classification and automatic content enrichment of images using supervised learning and knowledge bases

G. Knapp, S. A. Mallepudi, R. A. Calix, Louisiana State Univ. (United States)

In recent years there has been a rapid increase in the size of video and image databases. Effective searching and retrieving of images from these databases is a significant current research area. In particular, there is a growing interest in query capabilities based on semantic image features such as objects, locations, and materials, known as content-based image retrieval. This study investigated mechanisms for identifying materials present in an image. These capabilities provide additional information impacting conditional probabilities about images (e.g. objects made of steel are more likely to be buildings). These capabilities are useful in Building Information Modeling (BIM) and in automatic enrichment of images. I2T methodologies are a way to enrich an image by generating text descriptions based on image analysis. In this work, a learning model is trained to detect certain materials in images. To train the model, an image dataset was constructed containing single material images of bricks, cloth, grass, sand, stones, and wood. For generalization purposes, an additional set of 50 images containing multiple materials (some not used in training) was constructed. Two different supervised learning classification models were investigated: a single multi-class SVM classifier, and multiple binary SVM classifiers (one per material). Image features included Gabor filter parameters for texture, and color histogram data for RGB components. All classification accuracy scores using the SVM-based method were above 85%. The second model helped in gathering more information from the images since it assigned multiple classes to the images. A framework for the I2T methodology is presented.

## 7881B-40, Session 5

### Personal photo album summarization for global and local photo annotation

M. Broilo, F. G. B. De Natale, Univ. degli Studi di Trento (Italy)

Although content-based media retrieval tools are continuously improving, personalized image annotation is still one of the most reliable ways to index large image archives. Unfortunately, it is also a time consuming and repetitive operation. Using content to facilitate the user in media annotation may lead to reduced effort and more accurate results. In this paper we propose a content-based interactive tool that supports a user in annotating his personal photo albums. The system provides two main functionalities: to summarize a photo collection in salient moments, and to annotate pictures in a semi-supervised way based on their global and local content. The summarization is based on a bottom-up unsupervised hierarchical clustering that exploits two different matrices of visual distances, while the local tagging uses an object retrieval method based on local image features. Experiments on personal photo collections show that the proposed technique produces good results in terms of organization and access to data.

## 7881B-41, Session 5

### Event-driven people re-identification across photo collections

L. Lo Presti, M. Morana, M. La Cascia, Univ. degli Studi di Palermo (Italy)

Person re-identification across personal photo albums enables the development of new automated tools to support the user in browsing and managing his own collection.

In this paper, a new system is presented to support the user during the annotation task. The system automatically detects persons in each photo and tries to infer hypothetical correspondences among person across the photo album. In this way, the system is able to find the groups of photos where a person was detected taking advantage from the fact that each person can appear at most once in each photo.

We propose to model the problem of people re-identification in photos as a data association problem driven by temporal events.

Noting that appearance information such as clothing descriptors are more reliable within short temporal window while facial features are more reliable across wider temporal window, we propose to perform the re-identification process in two steps: first identifying persons within the same event using both facial and clothing descriptors, then inferring association among persons identified across temporal events using facial information. Our method is fully automated and does not require any initialization neither a priori knowledge of the number of persons that are in the photo collection.

Experiments were performed on a publicly available dataset and results are compared with those obtained by using standard clustering methods.

## 7881B-42, Session 6

### Web-scale multimedia processing

M. Slaney, Yahoo! Inc. (United States)

No abstract available

## 7881B-43, Session 7

### Spatially organized visualization of image query results

G. Ciocca, C. Cusano, Univ. degli Studi di Milano-Bicocca (Italy); S. Santini, Univ. Autónoma de Madrid (Spain); R. Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

In this work we present a system which visualizes the results obtained from image search engines in such a way that users can conveniently browse the retrieved images. The way in which search results are presented allows the user to grasp the composition of the set of images "at a glance". To do so, images are grouped and positioned according to their distribution in a prosemantic feature space which encodes information about their content at an abstraction level that can be placed between visual and semantic information. The compactness of the feature space allows a fast analysis of the image distribution so that all the computation can be performed in real time.

## 7881B-45, Session 7

### Image retrieval considering people co-occurrence relations using relevance feedback

K. Shimizu, N. Nitta, N. Babaguchi, Osaka Univ. (Japan)

The recent popularity of digital cameras allows us to take images easily. The large numbers of images taken by consumers are stored in

personal computers, on web servers, etc., and there is an increasing need for efficiently and accurately retrieving images containing specific objects from such image collections. One of the popular approaches for realizing such image retrieval is example-based retrieval, where the user presents an example image containing a specific object as a query image to retrieve other images containing the same object from image collections. Visual features are extracted from object regions in the query image and in every image in the image collection, and the images are presented to the user in the order of similarities between these features. Especially for retrieving images containing specific persons, the features are usually extracted from their face regions in images. Furthermore, the features extracted from other regions in images are used to improve the retrieval or recognition performance recently. We focus on the fact that some people such as family or friends are more likely to appear in the same images than others and use visual features of not only the queried person but also people who have strong co-occurrence relations with the queried person to improve the retrieval performance.

In this paper, people who appear with the queried person in the same images are called co-occurring persons. Our proposed system retrieves images of specific persons by considering the co-occurrence relations between the queried persons and co-occurring persons. Given a query image, the system calculates the expectation degrees which indicate how high the queried person is expected to be in images and presents the images with high expectation degrees to the user. The relevance feedback is used to learn the co-occurring relations between the queried and co-occurring persons. After the user specifies if each presented image contains the queried person, the system learns various co-occurring persons who are not in the query image, various facial appearances of the queried and co-occurring persons, and the strength of the co-occurrence relations between each co-occurring person and the queried person. Then, the retrieval results can be improved by considering the learned people co-occurrence relations.

In order to evaluate the proposed system, we collected images taken by consumers from 6 persons, all of which are his/her images with family or friends. The number of persons appearing in an image ranges from 1 to 6 but most frequently 3. From the persons contained in these images, we manually selected 20 persons who often appear with family or friends as the queried persons. We evaluated the retrieval results with the three measures Mean Average Precision (MAP), recall rate, and average rank of correctly retrieved images. After five feedback iterations, the recall rate was improved from 34 %, which is the result obtained when considering only the queried person, to 53 % by considering the people co-occurrence relations after learning various co-occurring persons by using only the queried person. The results have verified the effectiveness of learning the people co-occurrence relations for retrieving images of a specific person.

## 7881B-47, Session 9

### Face detection and recognition in Facebook

R. Yan, J. Yang, Facebook (United States)

No abstract available

## 7881B-48, Session 9

### Applications of consumer photo content understanding

D. Tretter, H. Chao, Hewlett-Packard Labs. (United States); Y. Gao, Hewlett-Packard Co. (United States); N. Lyons, F. Tang, C. Willis, P. Wu, J. Xiao, T. Zhang, X. Zhang, Q. Lin, Hewlett-Packard Labs. (United States)

No abstract available

7881B-49, Session 9

**Know your data: understanding implicit usage versus explicit action in video content classification**

J. Yew, D. A. Shamma, Yahoo (United States)

In this paper, we present a method for video category classification using only social metadata from websites like YouTube. In place of content analysis, we utilize communicative and social contexts surrounding videos as a means to determine a categorical genre, e.g. Comedy, Music. We hypothesize that video clips belonging to different genre categories would have distinct signatures and patterns that are redirected in their collected metadata. In particular, we define and describe social metadata as usage or action to aid in classification. We trained a Naive Bayes classifier to predict categories from a sample of 1,740 YouTube videos representing the top video genre categories. Using just a small number of the available metadata features, we compare the classifications produced by our Naive Bayes classifier with those provided by the uploader of that particular video. Compared to random predictions with the YouTube data (21% accurate), our classifier attained a mediocre 33% accuracy in predicting video genres. However, we found that the accuracy of our classifier significantly improves by nominal factoring of the explicit data features. By factoring the ratings of the videos in the dataset, the classifier was able to accurately predict the genres of 75% of the videos. We argue that the patterns of social activity found in the metadata are not just meaningful in their own right, but are indicative of the meaning of the shared video content. The results presented by this project represent a first step in investigating the potential meaning and significance of social metadata and its relation to the media experience.

7881B-50, Session 9

**Multimedia information retrieval at FX Palo Alto Laboratory**

M. L. Cooper, J. Adcock, A. Girgensohn, J. Pickens, L. D. Wilcox, FX Palo Alto Lab. (United States)

This paper describes research activities at FX Palo Alto Laboratory (FXPAL) in the area of multimedia browsing, search, and retrieval. Overviews of relevant systems are presented and references with additional details are provided. We first consider interfaces for organization and management of personal photo collections. We then survey our work on interactive video search and retrieval. Throughout we discuss the evolution of both the research challenges in these areas and the proposed solutions.

7881B-51, Session 9

**Image and video content analysis: challenges of scale**

J. Yagnik, Google Inc. (United States)

No abstract available

# Conference 7882: Visual Information Processing and Communication II

Tuesday-Wednesday 25-26 January 2011

Part of Proceedings of SPIE Vol. 7882 Visual Information Processing and Communication II

## 7882-12, Session 1

### Visual search: a tutorial overview

R. Grzeszczuk, Nokia Research Ctr. (United States)

No abstract available

## 7882-02, Session 2

### A hybrid video codec based on extended block sizes, recursive integer transforms, improved interpolation, and flexible motion representation

M. Karczewicz, P. Chen, R. Joshi, X. Wang, W. Chien, R. Panchal, M. Coban, I. S. Chong, Qualcomm (United States); Y. A. Reznik, Qualcomm Inc. (United States)

No abstract available

## 7882-03, Session 2

### Achieving H.264/AVC performance using distributed video coding combined with super-resolution

R. Klepko, D. Wang, G. Huchet, Communications Research Ctr. Canada (Canada)

Distributed Video Coding (DVC) is an emerging video coding paradigm for the systems that require low complexity encoders supported by high complexity decoders as would be required for, say, real-time video streaming from one mobile phone to another. Under the assumption of an error-free transmission channel, the coding efficiency of current DVC systems is still below that of the latest conventional video codecs, such as H.264/AVC. To increase coding efficiency we propose in this paper that either every second Key frame or every Wyner-Ziv frame is downsampled by a factor of two in both dimensions prior to coding. However, this in turn would require upsampling coupled with interpolation at the decoder. Simple interpolation (e.g., bicubic or FIR filter) would not suffice since high-frequency (HF) spatial image content would be missing. Instead, we propose the incorporation of a super-resolution (SR) technique, specifically based upon an example-based scene-specific method, to allow this HF content to be recovered. The SR technique will add computational complexity to the decoder side of the DVC system, which is allowable within the DVC framework. Rate-distortion curves will show that this novel combination of SR with DVC improves the system performance by up to several decibels as measured by the PSNR, and can actually exceed the performance of an H.264/AVC codec using GOP=IP.

## 7882-04, Session 2

### Real-time priority-aware transfer of SVC encoded video over MIMO communications system

D. Radakovic, Y. Yao, R. Ansari, Univ. of Illinois at Chicago (United States)

A novel cross-layer method is proposed for real-time transmission of standard compliant scalable video over a power-limited multiple-input

multiple-output (MIMO) system with channel state feedback. In the MIMO system, adaptive power allocation and antenna selection are utilized for creation of unequal bit error rate (BER) sub-channels. BER across all the sub-channels can be improved by reducing the channel throughput. In the proposed method, the scalable video is first divided into multiple video sub-streams of unequal importance by content-based partitioning and sorting of video layers. A novel technique is utilized to select the sub-stream data to be sent over the available MIMO sub-channels as to match the importance of the video data to both the channel BER and data transmission delay. Video packets that are delayed excessively are discarded at the transmitter. A trade-off exists between the losses in video peak signal-to-noise ratio (PSNR) resulting from discarded video packets at the transmitter, and gains in video PSNR due to lower channel BER. Simulation results show that the proposed method results in significantly improved performance compared with video transmission over constant BER channels with throughput equal to the video bit rate.

## 7882-05, Session 2

### Optimal power allocation and joint source-channel coding for wireless DS-CDMA visual sensor networks

K. Pandremmenou, L. P. Kondi, K. E. Parsopoulos, Univ. of Ioannina (Greece)

No abstract available

## 7882-06, Session 3

### A device and an algorithm for the separation of visible and near infrared signals in a monolithic silicon sensor

G. Langfelder, Politecnico di Milano (Italy); T. Malzbender, Hewlett-Packard Labs. (United States); A. F. Longoni, F. Zaraga, Politecnico di Milano (Italy)

No abstract available

## 7882-07, Session 3

### Localization of buildings with a gable roof in very-high-resolution aerial images

L. Hazelhoff, CycloMedia Technology B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands); P. H. N. de With, Technische Univ. Eindhoven (Netherlands) and CycloMedia Technology B.V. (Netherlands)

This study aims at the robust automatic detection of buildings with a gable roof in varying rural areas from very-high-resolution aerial images. The originality of our approach resides in a custom-made design extracting key features close to modeling, such as e.g. roof ridges and gutters. In this way, we allow a large freedom in roof appearances. The proposed method is based on a combination of two hypotheses. First, it exploits the physical properties of gable roofs and detects straight line-segments within non-vegetated and non-farmland areas, as possibilities of occurring roof-ridges. Second, for each of these candidate roof-ridges, the likely roof-gutter positions are estimated for both sides of the line segment, resulting in a set of possible roof configurations. These hypotheses are validated based on the analysis of size, shadow, color and edge information, where for each roof-ridge

candidate the optimal configuration is selected. Roof configurations with unlikely properties are rejected and afterwards ridges with overlapping configurations are fused. Experiments conducted on a set of 200 images covering various rural regions, with a large variation in both building appearance and surroundings, show that the algorithm is able to detect 75% of the buildings with a precision of 69.4%. We consider this as a reasonably good result, since the computing is fully unconstrained, numerous buildings were occluded by trees and because there is a significant appearance difference between the considered test images.

### 7882-08, Session 3

#### **Impact of near-lossless and lossy coding on information extraction from hyperspectral data**

A. C. Miguel, Seattle Univ. (United States)

In this paper, we evaluate the ability to extract meaningful information from the decompressed imaging spectrometer data. We compare the results of image processing performed on hyperspectral data compressed with a near-lossless bit plane coder with the results obtained on images coded with a lossy JPEG2000-based algorithm. Our study is extensive: we investigate a wide range of bit rates, use all scenes of the AVIRIS 224-band Cuprite, Jasper Ridge, and Moffett Fields radiance images, and employ several measures of rate-distortion and information extraction performance including PSNR, MAD, spectral angle mapper (SAM), and the true positive and negative rates of whole- and mixed-pixel analysis, and anomaly detection. Our results show that restricting the compression error to be uniform over the whole image does not improve the results of image processing, when compared to an efficient lossy coding technique. At higher bit rates, the results of post-processing performed on data compressed with the near-lossless coder are comparable to those obtained with an efficient lossy coder. At lower bit rates, the lossy coder outperforms the near-lossless coder. We conclude that lossy compression algorithms are preferable for hyperspectral data compression and processing applications over near-lossless methods.

### 7882-09, Session 3

#### **Motion adaptive Kalman filter for superresolution**

M. Richter, F. Nasse, H. Schroeder, Technische Univ. Dortmund (Germany)

Superresolution is a strategy to enhance image quality of both low and high resolution video. Especially recursive superresolution algorithms can fulfill these quality aspects because they control the video output using a feed-back loop and adapt the result in the next iteration. A very promising approach is the utilization of Kalman filters as proposed by Farsiu et al. Reliable motion estimation is essential for superresolution and therefore robust global motion models are preferred over local models, thereby limiting the application of superresolution.

The contribution of our paper is an investigation how the Kalman filter can be extended to allow improved handling of sequences with complex motion. Motion adaptive variance estimation and a segmentation to apply superresolution only in case of global motion are key features to reach this goal. Experiments confirm the potential of our proposal for ideal and real video sequences in comparison to state-of-the-art methods like trainable filters. The final paper will give a detailed explanation of the algorithm with our proposed improvements. Moreover, results of a detailed evaluation will be presented, e.g. investigation of required motion estimation accuracy to reach a quality higher than spatial processing, algorithm convergence speed and quality gain reached by our improvements.

### 7882-10, Session 3

#### **Hyper-cube watermarking scheme**

M. Chaumont, D. Goudia, W. Puech, LIRMM (France)

In 2007, Li and Cox showed that their scheme called Perceptual-QIM (P-QIM) was one of the solutions the most successful (even the best) in order to watermark multi-bits in an image by a quantization approach. Our research led us to take some of their ideas and brought new proposals. Thus, this paper presents a new scheme which we call Hyper-Cube. In addition to re-express the mechanisms of watermarking from a different angle, we propose three improvements: the use of the JPEG quantization table in setting the size of lattices, the calculation of modified Watson slacks on a neighborhood, and the use of a correcting code more efficient than the simple repetition code. Given the obtained results, we can conclude that the Hyper-Cube watermarking scheme is currently one of the most successful technique when one wants to watermark an image using quantization-based approaches.

### 7882-11, Session 3

#### **A joint JPEG2000 compression and watermarking system using a TCQ-based quantization scheme**

D. Goudia, M. Chaumont, W. Puech, LIRMM (France); N. Hadj Said, Univ. Mohamed Boudiaf Des Sciences Et De La Technologie d'Oran (Algeria)

In this paper, we describe a Trellis Coded Quantization (TCQ)-based quantization and watermarking technique in the framework of JPEG2000 still image compression. Furthermore, we investigate the design of a novel joint compression and watermarking scheme based on a hybrid TCQ module which can perform at the same time quantization and watermark embedding. The watermark extraction process can be achieved both during and after image decompression. Another advantage is the lower complexity of the system because the quantization stage is used for both compression and watermarking purposes. Experimental results have demonstrated that the proposed joint scheme successfully survives JPEG2000 compression with minimal degradation of the image quality. We also study the robustness of the scheme against gaussian filtering attack and valumetric attack.

### 7882-23, Poster Session

#### **Depth map coding based on color motion information**

B. T. Oh, Samsung Advanced Institute of Technology (Korea, Republic of) and The Univ. of Southern California (United States); H. Wey, D. Park, Samsung Advanced Institute of Technology (Korea, Republic of)

This paper presents an efficient depth map coding method based on color motion information in multi-view plus depth (MVD) system. As compared to the conventional depth map coding in which the depth video is separately coded, the proposed scheme involves the color information for depth coding. In details, the proposed algorithm subsamples the input depth data along temporal direction to reduce the bit-rate, and non-encoded depth frames are fully recovered at the decoder side guided by the motion information extracted from the decoded color video. The simulation results shows the high coding efficiency of the proposed scheme, and it also shows that the recovered depth data is not much different from the reconstructed one, and it even provides the temporally consistent depth map which results in better subjective quality for view-interpolation.

7882-25, Session 4

**Joint design of optics and image processing for application-specific sensors: overthrowing old optical design principles in the new era of electro-optics**

D. G. Stork, Ricoh Innovations, Inc. (United States)

No abstract available

7882-13, Session 5

**Robust HOSVD-based multi-camera motion trajectory indexing and retrieval**

Q. Li, X. Shi, D. Schonfeld, Univ. of Illinois at Chicago (United States)

We present a novel method for robust indexing and retrieval of multiple motion trajectories obtained from a multi-camera system. Motion trajectories describe the motion information by recording the objects' coordinates in the video sequence. We generate a four-dimensional tensor representation of multiple motion trajectories from multiple cameras. We subsequently rely on high-order singular value decomposition (HOSVD) for compact representation and dimensionality reduction of the tensor. We show that HOSVD-based representation provides a robust framework that can be used for a unified representation of the HOSVD of all subtensors. We thus demonstrate analytically and experimentally that the proposed HOSVD-based representation can handle flexible query structure consisting of an arbitrary number of objects and cameras. Simulation results are finally used to illustrate the superior performance of the proposed approach to multiple trajectory indexing and retrieval from multi-camera systems compared to the use of a single camera.

7882-14, Session 5

**Particle filtering with missing frames and its application to video tracking over lossy networks**

J. Huang, D. Schonfeld, Univ. of Illinois at Chicago (United States)

Many practical scenarios such as video tracking in lossy environment require a robust accurate tracking algorithm with dropped frames. A novel robust approach is proposed for visual tracking in the first part of this paper in the presence of frame loss with the Bayesian Importance Sampling framework based on first-order hidden Markov model (HMM). The graphical methods are firstly used to provide an exact solution for estimation using first-order hidden Markov model (HMM) with dropped frames. We subsequently rely on Sequential Importance Sampling to derive the first-order particle filtering algorithm with missing frames. In the second part of the paper, we promote this result and present that graphical methods can also be used to provide an exact solution to particle filtering with missing frames for an  $m$ th-order hidden Markov model (HMM) and cycle-free graphs. The resulting algorithm requires a small number of particles for efficient tracking. Experimental results demonstrate the superiority and robustness of the proposed approach to the standard methods, yet the additional computational time required is negligible.

7882-15, Session 5

**Affine image registration with curve mapping**

Y. Li, R. L. Stevenson, Univ. of Notre Dame (United States)

No abstract available

7882-16, Session 5

**Optimal optical flow based disparity map estimation for lossless stereo image coding**

A. Kumar K. C., I. R. M. Darazi, B. M. Macq, Univ. Catholique de Louvain (Belgium)

In dependent stereo image compression, the aim is to minimize the bitrate of disparity map and that of residual image. Traditionally, focus has been paid on either disparity map or residual image. In this paper, we compute an optimal disparity map (in terms of bitrates) by jointly exploiting the trade-off between disparity map and residual image. Firstly, the dense disparity map is obtained using existing optical flow technique. Secondly, the dense disparity map is quantized. Consequently, the bitrate for disparity map decreases significantly at the cost of slight increase in residual image. As a result, the overall bitrate attains minimum value. The proposed scheme is compatible and can be integrated in JPEG2000 framework.

7882-17, Session 6

**Background subtraction using pixel-wise adaptive learning rate for object tracking initialization**

K. K. Ng, E. J. Delp III, Purdue Univ. (United States)

In this paper we present a new method for object tracking initialization using background subtraction.

We propose an effective scheme for updating a background model adaptively in dynamic scenes. Unlike the traditional method that uses the same "learning rate" for the entire frame or sequence, our method assigns a learning rate for each pixel according to two parameters. The first parameter depends on the difference between the pixel intensities of the background model and the current frame. The second parameter depends on the duration of the pixel being classified as a background pixel.

We also introduce a method to detect and compensate for sudden illumination change.

Experimental results show significant improvements in moving object detection in dynamic scenes such as waving tree leaves and sudden illumination change, and it has much lower computational cost compared to Gaussian mixture model.

7882-18, Session 6

**Background estimation and update in cluttered surveillance video via the radon transform**

N. Conci, Univ. degli Studi di Trento (Italy); E. Izquierdo, Queen Mary, Univ. of London (United Kingdom)

In this paper we propose a background estimation and update algorithm for cluttered video surveillance sequences in indoor scenarios. The implementation relies on the integration of the Radon transform in the processing chain, applied on a block-by-block basis. The Radon transform is applied in this context to extract the meaningful information in terms of edges and texture, providing a signature for each portion of the image plane. The algorithm is validated in typical surveillance contexts and presented in this paper using a PETS video sequence.



7882-19, Session 6

**People re-identification in camera networks based on probabilistic color histograms**

A. D'Angelo, J. E. Dugelay, EURECOM (France)

No abstract available

7882-20, Session 6

**Estimating the number of people in crowded scenes**

M. Kim, W. Kim, C. Kim, KAIST (Korea, Republic of)

In this paper, we propose a novel method for estimating the number of people in an indirect manner. To this end, we basically employ statistical information of space-time interest points to analyze crowd behavior. Although crowd motion can be easily obtained by salient corners in the space-time domain, the number of those points tends to be highly variable due to difference of crowd motions. To cope with such variations between consecutive frames, we combine foreground information obtained from the Gaussian mixture model (GMM) with extracted salient corners. Based on this combination, we define our features, which are the number of space-time interest points, the number and size of foreground regions, and finally feed them into the multiple regression to precisely estimate the number of people in crowd scenes. To justify the efficiency and robustness of our approach, the experiments are conducted on PETS2009 datasets.

7882-21, Session 6

**MD/PNC with feedback for heterogeneous video multicast in lossy networks**

A. K. Ramasubramonian, J. W. Woods, Rensselaer Polytechnic Institute (United States)

We provide heterogeneous users in a multicast with video based on their available resources. Most existing solutions require knowledge of the network structure, which can be impractical in large networks. We present a scheme that combines multiple description coding and practical network coding (MD/PNC) to provide heterogeneous video multicast in lossy networks. The parity provided by random network codes not only helps in countering channel losses, but also provides different source rates to receivers based on their max-flow bandwidths. This method only requires knowledge of the max-flow bandwidths of the receivers, and not the network structure. The users' feedback information is used to compute the average loss rate they experience, and this is then used by the video server to optimize the source rate allocation in the descriptions. Simulation of multicast of the well-known Foreman test sequence on a random network shows a 1.3 - 1.5 dB improvement in the average PSNR of the receivers while using MD/PNC when compared to unirate multicast using network coding.

7882-22, Session 6

**Object-adaptive depth compensated inter-prediction for depth video coding in 3D video system**

M. Kang, Gwangju Institute of Science and Technology (Korea, Republic of); J. Lee, I. Lim, Samsung Advanced Institute of Technology (Korea, Republic of); Y. Ho, Gwangju Institute of Science and Technology (Korea, Republic of)

Nowadays, the 3D video system including MVD (multi-view video plus depth) is being actively studied. The system has many advantages with respect to virtual view synthesis such as an auto-stereoscopic functionality, but compression of huge input data remains a problem. Therefore, efficient 3D data compression is extremely important in the system, and low temporal consistency and viewpoint correlation problems should be resolved for efficient depth video coding. In this paper, we propose an object-adaptive depth compensated inter prediction method to resolve the problems. To achieve this, a mean-depth difference between a current block to be coded and a reference block is compensated during inter prediction, and unique properties of depth video are exploited to reduce side information required for signaling of the depth difference. To evaluate the coding performance, we have implemented the proposed method into MVC (multiview video coding) reference software, JMVM 6.0. Experimental results have demonstrated that our proposed method is especially efficient for the test depth videos generated by DERS (depth estimation reference software) of MPEG 3DVC group. The coding gain was up to 11.5% bit-saving, and subjective quality of synthesized view was noticeably improved by supporting better performance of inter prediction around object boundaries.