2013
Electronic Imaging

SCIENCE AND TECHNOLOGY

—

Technical Summaries

Conferences & Courses
3–7 February 2013

Location
Hyatt Regency San Francisco Airport Hotel
Burlingame, California, USA

www.electronicimaging.org

IS&T
imaging.org

SPIE®

IS&T/SPIE
Electronic Imaging
SCIENCE AND TECHNOLOGY

3–7 February 2013
Hyatt Regency San Francisco Airport Hotel
Burlingame, California, USA

Symposium Chair
Gaurav Sharma
University of Rochester

Symposium Co-Chair
Sergio Goma
Qualcomm Inc.

Short Course Chair
Choon-Woo Kim
Inha University

Symposium Steering Committee

Gaurav Sharma, Symposium Chair, Univ. of Rochester (United States)

Sergio R. Goma, Symposium Co-Chair, Qualcomm Inc. (United States)

Choon-Woo Kim, Short Course Chair, Inha Univ. (Korea, Republic of)

Andrew J. Woods, Technical Advisor, Curtin Univ. (Australia)

Majid Rabbani, Past Symposium Chair, Eastman Kodak Co. (United States)

Suzanne E. Grinnan, IS&T Executive Director

Rob Whitner, SPIE Event Manager

Contents

Conference 8648: Stereoscopic Displays and Applications XXIV

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8648 Stereoscopic Displays and Applications XXIV

8648-1, Session 1

Simulator sickness analysis of 3D video viewing on passive 3DTV

Kjell E. Brunnström, Kun Wang, Acreo AB (Sweden) and Mid Sweden Univ. (Sweden); Börje Andrén, Acreo AB (Sweden)

No Abstract Available

8648-2, Session 1

Investigation of visual fatigue/discomfort generated by S3D video using eye-tracking data

Iana Iatsun, Mohamed-Chaker Larabi, Christine Fernandez-Maloigne, XLIM-SIC (France)

During the last years expansion of 3D video grown significantly. Stereoscopic imaging is distributed in movie production, advertising, interactive games, and others. Among the advantages of 3D are the strong involvement of viewers and the increased feeling of presence. However, one of the main disadvantages is possibility to affect human health.

In this work, we propose to study the link between pupil diameter reflecting the activity of the automotic nervous system and eye movements representing a characteristic of the visual motor system especially the ciliary muscle, and the visual fatigue.

8648-80, Session K1

History of polarized image stereoscopic display *(Keynote Presentation)*

Vivian K. Walworth, StereoJet, Inc. (United States)

This presentation will recount the historic work of many investigators who contributed to the development of stereoscopic display through utilization of light polarization. The talk will include personal recollections of experiences with key inventors and colleagues, as well as early adventures in exploitation of the technology. Early cinema 3D demonstrations utilized polarizing filters over paired projectors, and viewers wore complementary polarizing glasses. Millions saw the 3D Chrysler film at the New York World's Fair in 1938 and 1939. 3D cinema has also utilized this technology. The Vectograph process, which provided back-to-back image pairs having oppositely oriented polarization, facilitated single-pass printing of stereoscopic images. Black-and-white Vectograph stereoscopic images were used extensively during World War II for both aerial surveillance and military instruction manuals. For many years binocular vision testing and training kits have included black-and-white Vectograph target images. In the postwar years excellent color Vectographs were made in the Polaroid laboratories, using dye transfer technology. Polaroid also collaborated with Technicolor in 3D cinema tests using that company's "imbibition" transfer technology. The introduction of inkjet printing technology in the 1980s gave impetus to exploration of single-pass 3D color printing at Polaroid. In the 1990s the Rowland Institute for Science produced excellent full-color StereoJet 3D Images on Vectograph substrate, which is no longer manufactured. StereoJet, Inc. is presently developing circularly polarized full-color StereoJet image systems using conventional polarizer substrate.

8648-4, Session 2

Optical modeling of lenticular array for autostereoscopic displays

Sung-Min Jung, Jin-Hyuk Jang, Hong-Yun Kang, Keong-Jin Lee, Ji-Na Kang, Seung-Chul Lee, Kyoung-Moon Lim, Sang-Deog Yeo, LG Display (Korea, Republic of)

In this study, we suggest mathematical model describing optical phenomena at the surfaces of lenticular array for autostereoscopic displays. Relations between incident and refracted lights are derived from Snell's law in vector form. By simulating the mathematical model with the actual design condition of autostereoscopic displays, we obtain angular distribution of light and compare the results with experiment. From the results, it is clarified that our simulation model coincide well with the experimental results. We believe that our simulation model is very useful for design and optimization of lenticular type autostereoscopic displays.

8648-6, Session 2

Simple measurement of lenticular lens quality for autostereoscopic displays

Stuart Gray, Robert A. Boudreau, Corning Incorporated (United States)

Autostereoscopic 3D displays utilizing lenticular lenses require precise alignment between the display pixels and the lens array to create a high quality viewing experience. This paper describes a simple technique for estimating the alignment errors between the lens array and pixels across the entire display.

When a multi-view autosteroscopic display is observed inside the optimum viewing distance then portions of several views are observed in one eye. Different views can be identified by creating a test pattern where each view is a single color with different colors in adjacent views. A well aligned lens array will show uniform and parallel bands of the different colors. Deviations in the width or spacing of the bands represent lenticular misalignment.

This was demonstrated on a 9-view, 55-inch display with a test pattern showing red in views 1,4 and 7, green in views 2,5 and 8 and blue in views 3,6 and 9. The pattern was viewed 1.7 meters from the screen rather than the specified optimum of 4.8 meters. The image showed diagonal color bands with deviations from parallel corresponding to misalignment of up to 400 um in some locations.

8648-8, Session 2

Visibility of crosstalk for high-dynamic range displays

Marc Lambooij, Martin Hammer, TP Vision (Netherlands)

Crosstalk is one of the main display-related perceptual factors degrading image quality and causing visual discomfort. In this experiment we investigate the impact of high display contrast and high display luminance on the perception of crosstalk.

A high-dynamic range LCD in combination with a Wheatstone viewer is used. To independently vary luminance, contrast, and binocular crosstalk (defined as (BW-BB)/(WW-BB))m, the display opto-electrical response was characterized and modeled. Luminance varied from moderate to very high (125, 500, and 1,500 cd/m^2) and contrast from low to high (1:100, 1:1,000, and 1:2,500).

Results show that increase in luminance leads to a reduced crosstalk visibility threshold resulting in a minimal visibility threshold of 0.20%

at 1,500 cd/m^2. The visibility threshold was independent of the tested contrast levels, indicating that contrast levels above 1:100 do not affect visibility thresholds.

In conclusion, with the introduction of OLEDs the system crosstalk will rise, but the increase in visibility of stereoscopic crosstalk is absent. If peak luminance is considerably higher than existing TVs, such as in high-dynamic range displays, the crosstalk visibility threshold will be lower. As the absolute threshold levels are very low (at or below 0.3%) this will result in little perceptual effect.

8648-10, Session 2

Method to test and quantify 3D active shutter glasses

Kurt Hoffmeister, Mechdyne Corp. (United States)

Liquid Crystal Shutter (LCS) 3D glasses (also known as 3D shutter glasses, active glasses, or just shutter glasses) are the selection device commonly used to view 3D stereoscopic content on time-sequential 3D displays. There are many available models of LCS glasses to choose from; although, unfortunately, there can be wide performance differences between these glasses. Ultimately the perceived quality of 3D image viewing depends on the display viewed as well as the performance of the shutter glasses used. The objective of this paper is to define key glasses performance parameters and to present a testing method to help quantify glasses and select between models available. Differences in shutter opening speed and open transparency (shuttering performance) are shown between tested models using the method presented here. Additional differences are noted for other performance parameters and features illustrating that there can be many product differences to consider that do not directly affect shuttering performance.

8648-12, Session 2

Towards standardized 3DTV QoE assessment: cross-lab study on display technology and viewing environment parameters

Marcus Barkowsky, Jing Li, Polytech' Nantes (France); Taehwan Han, Sungwook Youn, Jiheon Ok, Chulhee Lee, Yonsei Univ. (Korea, Republic of); Christer Hedberg, Inirajith V. Ananth, Acreo AB (Sweden); Kun Wang, Kjell E. Brunnström, Acreo AB (Sweden) and Mid Sweden Univ. (Sweden); Patrick Le Callet, Polytech' Nantes (France)

The perception of degradations measured in subjective assessments of images and video sequences is influenced by the viewing conditions. In stereoscopic 3DTV, the influence of the viewing environment, such as illumination, viewing distance, voting interface, observer screening, training and introduction to the experiment, is expected to differ significantly from the influence that was perceived when 2D recommendations, such as ITU-R BT.500, were established. The subjective measurement of Quality of Experience (QoE) in 3DTV extends over several psychophysical dimensions such as picture quality, depth sensation, and visual comfort which may be combined to higher level indications such as naturalness, presence and visual experience. This study focuses on the influence of display technology and viewing conditions on the subjective scores obtained in assessing degradations of 3D picture quality introduced by video coding and spatial resolution reduction.Three laboratories in Sweden, South Korea and France performed subjective experiments using the Absolute Category Rating with Hidden Reference (ACR-HR) method on the same video sequences. Differences in viewing distance, display technology (active or passive), and illumination (dark or ITU conforming background illumination) have been included on purpose. Additional observer data such as 3D viewing experience and screening results were obtained. The votes obtained in the studies are analyzed and compared using parametric and non-parametric statistical methods such as Anova, Pearson Correlation, RMSE, and Wilcoxon and Student tests to obtain information about the discrimination power for the video conditions.

8648-3, Session 5

Optical axes misalignment compensation method for stereo camera zooming

Jinmo Kang, Junyong Lee, Changseob Park, Korean Broadcasting System (Korea, Republic of)

Zooming of stereo camera suffers from optical axes misalignment problem. As the optical axis of a zoom lens is rarely aligned with the center of camera's image sensor, the position of the optical axis on the image sensor weaves with a circular motion as the zoom is pulled.

To correct this optical axes error for stereo camera, we need to calibrate the stereo rig for every zoom value using near and far chart. Usually this is done manually and it takes long time to calibrate the entire zoom range.

In this paper, we propose a new method which compensates optical axes misalignment for stereo camera zooming automatically. Optical axes misalignment is modeled as image sensor center translation from the lens center. To match the image point of left and right image, right camera translation and rotation equation is devised. By using these equations we can easily get the rig calibration value. Then, we can compensate optical axes misalignment by saving and applying this rig calibration information when we change the zoom of stereo camera.

The explanation of the proposed method and devised equation will be provided. And experimental verification results using stereo camera rig will be presented in the paper.

8648-5, Session 5

Stereo rendering of rain in real-time

Syed A. Hussain, David F. McAllister, North Carolina State Univ. (United States)

The rendering of photorealistic rain has been previously studied for monoscopic viewing. We extend the monoscopic statistical rain models to simulate the behavior and distribution of falling rain for stereo viewing. Our goal is to be able to render in real-time frame rates. In this investigation we ignore the complex issues of scene illumination and concentrate on the parameters that produce a realistic rain distribution. Using the concept of retinal persistence we render a visible falling raindrop as a linear streak. To speed rendering we use pre-computed images of such rain streaks. Rain streak positions for the left and right eye views are created by generating random numbers that depend on the view volume of the scene. We permit interactive but controlled modification of rain parameters such as density and wind gusts. We compare our approach to the use of existing 2D-3D conversion methods. The results demonstrate that using commercial 2D-3D converters are not sufficient in producing realistic stereo rain effects. Future research will concentrate on including complex lighting interactions.

8648-7, Session 5

A new 3D video format for multiview/ freeview depth image based rendering

Lin Du, Yan Xu, Gang Cheng, Technicolor Beijing (China)

This paper describes a novel 3D multiview/freeview video format, 2D plus Delta and Embed Occlusion and Transparency information (2D+DEOT), which embeds the additional occlusion information into a frame-compatible 2D+delta format using digital watermarking techniques. The multiple occlusion and transparency (OT) layers are weighted, encoded and multiplexed using spread spectrum watermarking and then embedded into a frame-compatible 2D+depth/ delta base layer using least significant bits (LSB) watermarking. The multiview/freeview 3D video can be rendered at the player using depth image based rendering methods after the occlusion information is retrieved as watermarks and the original 2D+delta base layer is recovered by subtracting the re-produced watermark from the LSB of each frame. Therefore, this video format can provide OT information

for multiview/freeview 3D video synthesis while only introduces minor noise at the least significant bits for conventional 2D+delta 3D video rendering systems. Several encoding and embedding schemes of this 3D video format are tested using the Breakdancing and Ballet sequences from Microsoft Research. The test results show that using this format can both generates good quality multiview/freeview 3D video and keeps the compatibility to current 2D encoding/decoding and transmission systems.

8648-9, Session 5

Subjective evaluation of an edge-based depth image compression scheme

Yun Li, Mårten Sjöström, Ulf Jennehag, Roger Olsson, Sylvain Tourancheau, Mid Sweden Univ. (Sweden)

We have previously proposed a depth image coding scheme that preserves significant edges and encodes smooth areas between these. An objective evaluation considering the structural similarity (SSIM) index for synthesized views demonstrated an advantage to the proposed scheme over the high efficiency video coding (HEVC) intra mode in certain cases. However, some discrepancies exist between the objective evaluation and our visual inspection, which motivated this study of subjective tests. The test was conducted according to ITU-R BT.500-13 recommendation with Stimulus-comparison method. The synthesized images of the test sequences were paired such that they used encoded depth map of similar bit rates produced from the proposed scheme and HEVC respectively. The results from the subjective test showed that the proposed scheme performs slightly better than HEVC with statistical significance at almost all tested bit rates for the given contents. The paper reveals that preserving significant depth continuities in a depth image can result in an overall better quality of experience for synthesized views. It also illustrates that the objective metric SSIM fails to predict the importance of preserved edges for perceived quality in a synthesized view.

8648-11, Session 5

Cooperative ARQ for 3D video transmission

Yan Xu, Lin Du, Technicolor (China) Technology Co., Ltd. (China)

Error control is important for 3D video streaming especially over wireless networks which have highly variable nature. Various schemes exist that mitigate the effects of errors during transmission of 2D video data. Existing error control methods can be applied to some 3D format, such as frame compatible format, in which the two views are multiplexed into a single frame. However, for 2D plus metadata format being transmitted in different channels, there has no error control method to deal with it. Automatic Repeat Query (ARQ) is widely used to mitigate the effects of errors, which can be tailored to 3D error control in consideration of 3D transmission characteristics. We propose an error control method for transmitting 2D plus metadata in two different channels. It applies selective repeat ARQ with consideration of two factors. One is the interdependent relationship between 2D video and its metadata information; the other is time constraints for continuous video playback. Simulation results show that it allows the sender to control ARQ strength adaptively on a per-packet basis. And it makes a good tradeoff between playback delay and packet loss rate, which provides a guideline of error control system design for 3D video streaming.

8648-14, Session 3

Natural 3D Content on Glasses-free Light-field 3D Cinema

Tibor Balogh, Zsolt Nagy, Péter Tamás Kovács, Vamsi Kiran Adhikarla, Holografika Kft. (Hungary)

This paper presents a hardware-software system to visualize synthetic or natural 3D light-field videos to a larger audience on a cinema-sized screen size then what was previously possible with autostereoscopic displays. We present the hardware system including optical and mechanical design considerations, the electrical system and rendering cluster, and the software driving the whole system, and feeding it with 3D data. The display is the first front-projected light-field 3D display HoloVizio system, controlling 63 MPixels, has all the advantages of previous light-field displays, but allows for a more flexible arrangement, larger screen size, and simpler set-up. The software system makes it possible to show 3D applications in real-time as well as natural content captured with dense light-field camera arrangements as well as content captured with wide baseline few-camera rigs. The software system is able to visualize pre-recorded Multiview Plus Depth (MVD) videos on this light-field glasses-free cinema system, interpolating and extrapolating missing information not captured by the cameras.

8648-15, Session 3

Large-scale autostereoscopic outdoor display

Joerg Reitterer, Franz Fidler, Ferdinand Saint Julien-Wallsee, TriLite Technologies GmbH (Austria); Gerhard Schmid, Wolfgang Gartner, Walter Leeb, Ulrich Schmid, Technische Univ. Wien (Austria)

State-of-the-art autostereoscopic displays are often limited in size, effective brightness, number of 3D viewing zones, and maximum 3D viewing distances, all of which are mandatory requirements for applying autostereoscopic 3D to large-scale outdoor displays. Conventional autostereoscopic indoor concepts like lenticular lenses or parallax barriers cannot simply be adapted for these screens due to the inherent loss of effective resolution and brightness, which would have a significantly disadvantageous impact on both image quality and sunlight readability. We have developed a modular autostereoscopic multi-view laser display with sunlight readable effective brightness, several thousand of 3D viewing zones, and a maximum 3D viewing distance of up to 50 meters. For proof of concept purposes a prototype display with 4x5 pixels is realized. Due to various manufacturing tolerances each individual pixel of the display has slightly different optical properties and, hence, the perceived 3D image quality of the display has to be described stochastically. In this paper we present the corresponding stochastic model and evaluate the simulation and measurement results of the prototype display.

8648-20, Session 3

Multi-view display module using MEMS projectors for an ultra-large screen autostereoscopic display

Kenji Hirabayashi, Tokyo Univ. of Agriculture and Technology (Japan); Hiromitsu Takenaka, Osamu Konuma, Yasuhiro Morimoto, Samsung Yokohama Research Institute Co., Ltd. (Japan); Yasuhiro Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

A multi-view display module using MEMS projectors is proposed to realize an ultra-large screen autostereoscopic display. The module consists of an array of MEMS projectors, a vertical diffuser, and a lenticular lens. All MEMS projectors having different horizontal positions project images that are superimposed on the vertical diffuser. Each cylindrical lens constituting the lenticular lens generates multiple 3D pixels at different horizontal positions near its focal plane. Because the 3D pixel is an image of a micro-mirror of the MEMS projector, the number of 3D pixels in each lens is equal to the number of projectors. Therefore, the horizontal resolution of the module can be increased using more projectors. By properly modulating lasers in the MEMS projector, the horizontal positions of dots constituting a projected image can be altered at different horizontal scan lines. When several scan lines correspond to one 3D pixel, a number of views can be increased. Because the module has a frameless screen, the number of modules can be arranged two-dimensionally to obtain a large screen. The prototype module was constructed using four MEMS projectors. The module had a 3D resolution of 160 x 120, and provided 64 views. The screen size was 14.4 inches.

Conference 8648

8648-13, Session 6

Real-time handling of existing content sources on a multi-layer display

Darryl Singh, Jung Shin, PureDepth Inc. (New Zealand)

Multi-Layer Displays (MLD) consists of two or more imaging planes separated by physical depth where the depth is a key component in creating a glasses-free 3D effect. It's core benefits include being viewable from multiple angles, having full panel resolution for 3D effects with no side effects of nausea or eye-strain, however, typically content must be designed for it's optical configuration in foreground and background image pairs. A process was designed to give a consistent 3D effect in a 2-layer MLD from existing stereo video content in real-time. Optimization of the stereo matching algorithms to generate depth maps in real-time specifically tailored for the optical characteristics and image processing algorithms of a MLD was required. The end-to-end process included improvements to the Hierarchical Belief Propagation HBP stereo matching algorithm, improvements to optical flow and temporal consistency. Imaging algorithms designed for the optical characteristics of a MLD provided some visual compensation for depth map inaccuracies. The result can be demonstrated in a PC environment, displayed on a 22" MLD, used in the casino slot market, with 8mm of panel seperation. Prior to this development, stereo content had not been used to achieve a depth-based 3D effect on a MLD in real-time.

8648-16, Session 6

Switching dual layer display with dynamic LCD mask

Quinn Y. J. Smithwick, Mark Reichow, Walt Disney Imagineering (United States)

Dual layer displays are simple effective means of creating dimensional images. A common implementation is a stacked spaced front layer transparent LCD panel and back layer monitor. Foreground content doesn't appear over dark backgrounds, and appears low contrast and semi-transparent over bright backgrounds. The front screen prevents access to the image volume when used in interactive gestural, haptic, or touch applications.

The Switching Dual Layer Display with a Dynamic Mask produces high-contrast floating front content with opacity independent of the background. The display consists of optically stacked front transparent 120Hz LCD panel and back 120Hz monitor, and a relay mirror. In the first mode, the back display shows foreground content and the front display is clear. The mirror reimages the back display through the transparent front display, so the foreground floats in front of the display. In the second mode, the back display shows white with a mask of the foreground, reflects off the mirror, backlighting the front display showing background content. Rapidly alternating between the modes, the viewer perceives opaque high-contrast foreground content floating above the background.

The Switched Dual Layer Display is a volumetric display, exhibiting coupled accommodation/vergence cues, occlusion, and real imagery.

8648-18, Session 6

Design of time-multiplexed autostereoscopic displays based on virtual stacking of multi-layer panels

Hironobu Gotoda, National Institute of Informatics (Japan)

The term, "slice-stacking displays", refers to a particular kind of volumetric displays, which present a large number of volumetric elements or voxels to viewers. Voxels are often packed in a 2D array, called "slice" or "depth plane", which sequentially fills a 3D volume. Recently, "multi-layer displays" have been proposed and implemented by several researchers. Unlike slice-stacking displays, multi-layer displays can present multiple slices at the same time. Moreover, slices in multi-layer displays always remain stationary, i.e., there are no moving or rotating slices, which can be typically found in slice-stacking displays.

In this paper, we consider a mixture of slice-stacking and multi-layer displays. More specifically, we present a new design of multi-layer displays, where slices can move back and forth, changing periodically the relative distance from the viewers. By allowing the slices to move around, we can, in effect, increase the number of layers in the displays. This results in finer resolution of the light field observable from the display, and improvement in the precision of the accommodation cues. Two configurations, implementing the concept of "moving slices", are presented using either a moving mirror or a varifocal lens, and compared with each other.

8648-19, Session 6

A new method to enlarge a range of continuously perceived depth in DFD (Depth-fused 3D) display

Atsuhiro Tsunakawa, Tomoki Soumiya, Hirotsugu Yamamoto, Shiro Suyama, Univ. of Tokushima (Japan)

In DFD (Depth-fused 3-D) display which can present a continuous 3-D image by using only layered two transparent displays, we clarify that the perceived depth dependence on luminance ratio of front and rear images is widely changed as the depth difference of two planes is increased. When the depth difference of two planes is small, the perceived depth is changed almost linearly as the luminance ratio is changed. On the other hand, when the depth difference is large, the perceived depth is around front plane or around rear plane at luminance ratio of about 0-40% or of about 60-100%, resulting in non-linear change of perceived depth. This indicates that this large depth difference results in out of range in continuously perceived depth. In order to enlarge this range of continuously perceived depth, we changed the spatial-frequency distributions of front and rear images. Even when the depth difference of two planes is large enough, almost linear change of perceived depth can be successfully obtained by increasing low spatial-frequency part of two images, resulting in deeper 3-D image in DFD display.

8648-22, Session 4

Subjective and objective measurements of visual fatigue induced by excessive disparities in stereoscopic images

Yong Ju Jung, Dongchan Kim, Sohn Hosik, Seong-il Lee, Hyun Wook Park, Yong Man Ro, KAIST (Korea, Republic of)

As stereoscopic displays have spread, it is important to find evidences of what really causes fatigue and discomfort and what happens in the visual system under an uncomfortable stereoscopic viewing experience. To the best of our knowledge, no single subjective or objective measurement is a thorough method to find these evidences, since visual fatigue and discomfort are related to many aspects of the visual system. Few studies have been done for the combined subjective and objective measurements of visual fatigue and discomfort.

In this paper, we employ both subjective and objective measurements for visual fatigue in stereoscopic displays. In particular, we use functional magnetic resonance imaging (fMRI) as an objective measurement of visual fatigue in stereoscopic displays. Our experiment consisted of a two-pass measurement of visual fatigue: a subjective measurement and an objective measurement. Using the subjective measurement results, we selected comfortable videos and uncomfortable videos in our dataset. Then, fMRI experiments were conducted to observe what happened in the visual system under an uncomfortable stereoscopic viewing experience. From the double stimulus comparison analysis, the results indicated that there were statistical differences in brain activation regions between comfortable and uncomfortable video stimuli with different amounts of disparity.

8648-23, Session 4

Depth distortion in color-interlaced stereoscopic 3D displays

Joohwan Kim, Paul V. Johnson, Martin S. Banks, Univ. of California, Berkeley (United States)

We measured whether color-interlaced stereoscopic 3D (S3D) displays can reduce the depth distortions that arise in conventional temporally interlaced S3D displays. We used a haploscope to emulate color-interlaced S3D displays and conventional temporally interlaced S3D displays. Our experimental results show that the human visual system calculates disparity using primarily luminance information rather than color information. This suggests that we can reduce depth distortion in temporally interlaced S3D displays by arranging the order of color presentation to keep overall luminance temporally constant. In the recently proposed color-interlaced S3D display method, we present green to the left eye and red and blue to the right eye at one moment in time while we exchange color channels in the next moment in time. Assuming that the luminance signal is the primary source of information for disparity calculation, this is the optimal ordering to reduce the depth distortion in 3-chip projectors.

8648-26, Session 4

Depth perception from stationary and moving stereoscopic three-dimensional images

Yu-Chi Tai, Sowjanya Gowrisankaran, Shunnan Yang, James E. Sheedy, John Hayes, Pacific Univ. (United States); Audrey C Younkin, Philip J Corriveaub, User Experience Research Group, Intel Corporation (United States)

The study evaluated the accuracy of depth perception afforded by static and dynamic stereoscopic three-dimensional (S3D) images with proportional (scaled to disparity) and constant size cues. Adult participants with good binocular vision participated in the study. For static S3D trials, participants were asked to indicate the depth of stationary S3D images rendered with different magnitudes of crossed disparity, and with either proportional or a constant size. For dynamic S3D trials, participants were asked to indicate the time when moving S3D images matched the depth of a reference image closer to the viewer. Results show that viewers perceived S3D images as being closer than would be predicted by the magnitude of image disparity, and correspondingly they overestimated the motion speed in depth. The resultant depth perception and estimate of motion speed were more accurate for certain conditions with proportional and larger image size, slower motion-in-depth and larger image disparity. These findings possibly explain why effects such as looming are over stimulating in S3D viewing. To increase the accuracy of depth perception, S3D content should match image size to its disparity level, utilize larger depth separation (without inducing excessive discomfort) and render slower motion in depth.

8648-27, Session 4

Immersion, tangibility, and realism: explaining the qualitative experience of stereopsis

Dhanraj Vishwanath, Univ. of St. Andrews (United Kingdom)

The fundamental property that drives 3D technology is the compelling qualitative experience of tangible solid objects, immersive space and realism that is lacking in conventional 2D displays. This qualitative perceptual phenomenon, referred to as 'stereopsis', is widely assumed to be a by-product of binocular vision or parallax. However, its underlying cause, variation and functional role remain largely unexplained. I will present evidence that suggests that stereopsis is not simply a product of binocular disparities or parallax but is a more basic visual property. I present an alternative theory that links the impression of stereopsis to the precision of absolute (scaled) depth. I present perceptual evidence (qualitative and quantitative) in support of the theory and discuss the implications for some important problems in stereoscopic display technology.

8648-21, Session 7

Aerial 3D LED display by use of retroreflective sheeting

Hirotsugu Yamamoto, Shiro Suyama, Univ. of Tokushima (Japan)

We propose a new optical configuration to form aerial 3D LEDs by use of retroreflective sheeting. The proposed configuration is composed of LEDs, a half mirror, and retroreflective sheeting. A half of the LED lights are reflected by the half mirror and impinge on the retroreflective sheeting. The retroreflective sheeting reflects the lights back to their sources. On the way to the sources, a half of the lights transmit through the half mirror and form the real images of LEDs. Although less than 25 % of the output lights are contributed to the aerial image, recent LED panels have enough luminance to enjoy the aerial image with a quarter of the luminance. We have made a prototype of the proposed aerial LED display. An aerial image of the LED panel has been successfully formed in free space. Its viewing angle was significantly improved compared to the aerial display by use of crossed mirrors, which limit the viewing angle by aperture size and height of mirror walls. The viewing angle in the proposed configuration is mainly limited by the size of the retroreflective sheeting. Furthermore, by using LEDs in different depths, we realized an aerial 3D display in free space.

8648-24, Session 7

Implementation of shading effect for reconstruction of smooth layer-based 3D holographic images

Jhen-Si Chen, Univ of Cambridge (United Kingdom); Quinn Y. J. Smithwick, Walt Disney Imagineering (United States); Daping Chu, Univ of Cambridge (United Kingdom)

A 3D image can be composed by a number of 2D sliced images. The holograms for each of them can be calculated very quickly and then combined together for reconstructing the 3D image. Meanwhile, a 3D image can be separated into views from different angles using a so-called angular tiling approach with each view being generated by a sub-hologram. The algorithm developed in this work is based on both these two methods, that each view is calculated by a layer-based decomposition while the whole 3D image is composed of different views.

The individuality of angular tiled views allows the introduction of shading to a layer-based image because it can be varied from view to view. The angular tiling can also reduces the depth gaps caused by layer slicing due to the superposition of layers from different views. In this way, to replay a smooth 3D image with correct shadings will require less memory in computation than using conventional method, and the whole calculation will be fast and easily adaptable for parallel computations.

In conclusion, a practical algorithm to compute holograms for 3D images is presented. It can incorporate all the necessary diffractive information for image reconstruction at high speed.

8648-28, Session 7

Hologram synthesis using integral imaging camera with synthetic aperture technique

Sung-Keun Lee, Yong-Soo Kim, Sung-In Hong, Jae-Hyeung Park, Chungbuk National Univ. (Korea, Republic of)

In this paper, we propose a novel incoherent hologram capture method. In the proposed method, the integral imaging capture

system is used to capture an array of different perspective images of the three-dimensional (3D) object. The captured perspective array is then transformed to an array of the images refocused at various distances, forming the focal image stack. This refocusing process at arbitrary distance can be performed using a computational integral imaging reconstruction (CIIR) technique which is well-known in integral imaging literatures. Using the generated focal image stack, the hologram is synthesized by considering a point spread function of the integral imaging system. Since the focal image stack is not captured sequentially but generated from the perspective array which is captured simultaneously, the proposed method can capture the hologram in real-time. In integral imaging, the synthesis of the focal image stack preserves high resolution, enabling high resolution hologram synthesis. In the presentation, we will explain the principle with experimental results.

8648-81, Session K2

Coverage of the London 2012 Olympic Games in 3D *(Keynote Presentation)*

Jim DeFilippis, Consultant (United States)

The London 2012 Olympics provided the opportunity to broadcast a full-time 3D channel. In the past, 3D coverage was only available on a closed circuit basis of limited events. During the 2012 Olympics, the host broadcaster (Olympic Broadcast Services) operated a 3D channel consisting of the opening and closing ceremonies, multiple sports, both live and near live coverage, and produced over 275 hours of 3D stereoscopic programming. The Olympic 3D channel was available to all the rights holding broadcasters including: BBC, Eurovision, SkyItalia, Nine Network Australia, Japan Pool, NBC and others. The core of the 3D coverage was provided from three mobile production units and six single camera field production units. A variety of stereoscopic rigs were used in each of four venues alongside the Panasonic ENG/EFP P2 3D Camcorder. Some special stereo cameras were also used including: pole cameras, rail cameras, RF cameras and underwater cameras. I will present the unique challenges in providing 3D coverage, from organizing the 3D channel as well as the technical challenge of covering sports in 3D while accommodating the full-up 2D production with an emphasis on what worked and what did not.

8648-29, Session 8

The psychology of the 3D experience

Sophie H. Janicke, Andrew Ellis, Florida State Univ. (United States)

With the swift arrival of 3D technology in our living rooms, content creators are pressured to keep up with increased demand by consumers. Experience suggests to us that simply applying 2D storytelling conventions to 3D is rarely the best option to meet this demand. So, developing new forms and approaches to storytelling within a 3D environment is a must. But how do we go about this? Our study will contribute to the exploration of this question from the perspective of media psychology. Our project investigates several variables underlying a positive 3D entertainment experience, as well the process through which entertainment narratives can impact a viewer's attitudes about particular issues (i.e., narrative persuasion). We specifically look at the process of transportation into the narrative world, an experience that has been shown to underlie both of effects. We contend that 3D theoretically lends itself to the creation of a greater feeling of transportation for the viewer, which ultimately should impact their enjoyment and narrative persuasion. Our project involves two experimental studies, using a variety of 2D and 3D content. Our work takes a first step in better understanding the 3D entertainment experience from a media-psychological perspective, offering a model of how academic research can assist in the development of this industry.

8648-30, Session 8

A semi-automatic 2D to stereoscopic 3D image and video conversion system in a semi-automated segmentation perspective

Raymond Phan, Dimitrios Androutsos, Ryerson Univ. (Canada)

We create a system for semi-automatically converting unconstrained 2D images and videos into stereoscopic 3D. Current efforts are done automatically or manually by rotoscopers. The former prohibits user intervention, or error correction, while the latter is time consuming, requiring a large staff. Semi-automatic mixes the two, allowing for faster and accurate conversion, while decreasing time to release 3D content. User-defined strokes for the image, or over several keyframes, corresponding to a rough estimate of the scene depths are defined. After, the rest of the depths are found, creating depth maps to generate stereoscopic 3D content, and Depth Image Based Rendering is employed to generate the artificial views. Here, depth map estimation can be considered as a multi-label segmentation problem, where each class is a depth value. Optionally, for video, only the first frame can be labelled, and the strokes are propagated over all frames through a modified robust tracking algorithm. Our work combines the merits of two respected segmentation algorithms: Graph Cuts and Random Walks. The diffusion of depths from Random Walks, combined with the edge preserving properties from Graph Cuts is employed to create the best results possible. Results demonstrate good quality stereoscopic images and videos with minimal effort.

8648-31, Session 8

Temporally consistent disparity estimation using PCA dual-cross-bilateral grid

Jin Zhu, Technische Univ. Eindhoven (Netherlands); Patrick Vandewalle, Philips Research (Netherlands); Gerard de Haan, Philips Research Nederland B.V. (Netherlands)

Disparity estimation has been extensively investigated in recent years. Though several algorithms have been reported to achieve excellent performance on a common reference data set, few of them reach a satisfying balance between accuracy and efficiency for real-time applications, and few of them consider the problem of temporal coherence. In this paper, we introduce a novel disparity estimation approach, which improves the accuracy for static images and the temporal coherence for videos. For static images, the proposed approach is inspired by the adaptive support weight method proposed by Yoon et al. and the dual cross-bilateral grid introduced by Richardt et al. Principal component analysis (PCA) is used to reduce the color dimensionality in the cost aggregation step. This simple, but efficient technique helps the proposed method to be comparable to the best local algorithms on the reference data set, and it is proven to have the potential to be implemented in real-time. A computationally efficient method for temporally consistent behavior is also proposed. Moreover, in the user evaluation experiment, the proposed temporal approach achieves the best overall user experience among the selected comparison algorithms.

8648-32, Session 8

Automatic depth grading tool to successfully adapt stereoscopic 3D content to digital cinema and home viewing environments

Cédric Thébault, Didier Doyen, Technicolor S.A. (France); Pierre Routhier, Technicolor S.A. (Canada); Thierry Borel, Technicolor S.A. (France)

After the enthusiasm initially generated by the digital cinema stereo experience, we are clearly now in a second phase of the adoption of stereoscopic 3D (S3D) content. The spectator is now not only waiting for a novel experience but for one that will bring him something more. To ensure looking at S3D content will be a good experience, an

adaptation of the content should be applied according to the targeted display and its environment. We have presented an automatic tool to adapt S3D content to digital cinema and home viewing environments. The stereo point of interest in each shot is automatically defined and then the appropriate parallax shift is applied to ensure proper format. Based on disparity estimation, this tool also helps to automatically smooth transition between shots according to the targeted application. The development is not yet finalized since temporal variation of the processing has to be developed to better optimize the adaptation all over a sequence.

8648-33, Session 8

Disparity remapping to ameliorate visual comfort of stereoscopic video

Sohn Hosik, Yong Ju Jung, Seong-il Lee, KAIST (Korea, Republic of); Filippo Speranza, Communications Research Ctr. Canada (Canada); Yong Man Ro, KAIST (Korea, Republic of)

Disparity remapping is one of the solutions proposed to improve visual comfort of stereoscopic 3D content. One common approach is a linear mapping of disparity that retargets an entire scene disparity range into a comfortable range. However, retargeting the overall scene disparity range might not always achieve the desired level of visual comfort. In fact, visual discomfort could still be induced by fast changing disparities, e.g., disparities resulting from objects moving in depth. More importantly, such discomfort might persist even when the overall scene disparity range is within a nominally comfortable range.

In this paper, we propose a novel disparity remapping method aimed at reducing the visual discomfort induced by fast changes in disparity. The proposed remapping approach selectively scales the disparities of the discomfort regions where fast spatial and temporal changes in disparity occur. In this approach, it is also assumed that the overall degree of visual discomfort is dominantly influenced by visually important regions in the scene. The results of the subjective assessment of visual comfort show that the proposed local remapping approach is capable of reducing the visual discomfort induced by fast changes of disparity in stereoscopic 3D video.

8648-34, Session 9

Disparity analysis of 3D movies and emotional representations

Takashi Kawai, Masahiro Hirahara, Yuya Tomiyama, Daiki Atsuta, Waseda Univ. (Japan); Jukka Häkkinen, Aalto Univ. (Finland)

In the field of stereoscopic (3D) images, industry associations for hardware and software, consisting mainly of display manufacturers, have been established to undertake initiatives with the aim of promoting 3D. Recently there have been many initiatives from the viewpoint of 3D content creators. It has been pointed out that the establishment of know-how for creating 3D content in Hollywood is forward compared with the other countries. Here "know-how" means producing a depth sensation mainly by imposing a binocular disparity. The authors have carried out multifaceted and continuous analysis of the disparity in famous 3D movies that were created in Hollywood and that have been highly evaluated throughout the world. The objective was to provide useful knowledge for creating 3D content in future, by quantitatively determining the characteristics and trends in the disparity in these movies.

This paper introduces examples of the results for analysis of movies. In addition, the detailed analysis is described for feature scenes that are considered to express emotion within the movies.

8648-35, Session 9

Methodology for stereoscopic motion-picture quality assessment

Alexander Voronov, Dmitriy Vatolin, Denis Sumin, Vyacheslav Napadovsky, Alexey Borisov, Moscow State Univ. (Russian Federation)

Creating and processing stereoscopic video places additional quality requirements relating to view synchronization. In this work we propose a set of algorithms for detecting typical stereoscopic-video problems, which appear owing to imprecise setup of capture equipment or incorrect postprocessing. We developed a methodology for analyzing the quality of S3D motion pictures and for revealing their most problematic scenes. We then processed 10 modern stereo films, including Avatar, Resident Evil:

Afterlife, and Hugo, and analyzed changes in S3D-film quality over the years. This work presents real examples of common artifacts (color and sharpness mismatch, vertical parallax, and excessive horizontal parallax) in the processed motion pictures, as well as possible solutions for each problem. Our results enable improved quality assessment during the filming and postproduction stages.

8648-36, Session 9

Critical alignment methods for stereoscopic production and post-production image registration

Christopher A. Mayhew, Craig M. Mayhew, Vision III Imaging, Inc. (United States)

The quality of the three-dimensional perception of a stereoscopic production is directly dependent upon the precision of its left/right image registration. Tests show that pixel-precise registration improves three-dimensional perception and viewing comfort. Current software tools that rely on techniques like 50/50 views for image alignment have difficulty achieving pixel-accurate registration.

The Authors have developed a "Critical Alignment" (CA) stereoscopic software plug-in toolset that allows for accurate registration of camera systems and left/right image pairs. The tools simplify and increase the precision of the registration process. The CA toolset is compatible with standard video editing software and can be viewed in the field on normal unaided 2D displays.

CA takes advantage of the human eye's high comparative capabilities. Humans evoke an autostereoscopic (automatic three-dimensional perception) response to parallax imagery presented alternately at between 3 to 6 Hz. Transformations made on alternating imagery allow a high level of registration in all vectors at a sub-pixel level.

Applying CA to the camera system alignment during production and to left/right image registration in post-production can significantly improve stereoscopic perception and reduce viewer fatigue. Further, tests have shown that the CA toolset also benefits scene-to-scene transitions and edits by precise manipulation of convergence points.

8648-37, Session 9

Towards a metric of antialiasing sufficiency for stereoscopic displays

Charles J. Lloyd, Visual Performance, LLC (United States)

This paper describes the development, measurement, computation, and validation of a metric of antialiasing sufficiency for stereoscopic display systems. A summary is provided of two previous evaluations that demonstrated stereoscopic disparity thresholds in the range of 3 to 10 arcsec are attainable using electronic displays with a pixel pitch as coarse as 2.5 arcmin, however, only if sufficient antialiasing is performed. An equation is provided that describes the critical level of antialiasing required as a function of pixel pitch. The proposed

metric uses a radial test pattern that can be photographed from the user eyepoint using a hand held consumer color camera. Several candidate unitary metrics that quantify the spatial sampling noise in the measured test pattern are described. The (R-squared) correlation obtained between the best candidate metric and stereoscopic disparity thresholds was 0.98. The variance introduced due to making repeated measurements with a hand held camera was less than 1% of the variance due to the antialiasing function. The proposed method is display technology independent and requires no knowledge of how the antialiasing is accomplished.

8648-38, Session 10

Stereoscopic game design and evaluation

Nicolas S. Holliman, Durham Univ. (United Kingdom)

During the last ten years at the Durham Visualisation Laboratory we have designed a number of games to investigate whether high scores can be improved by playing in S3D compared to 2D. The results have varied from no improvement in a 3D archery style game, to a small improvement of 8% in a 3D selection task, to 28% improvement in a 3D tetris-style task, to 68% improvement in a 3D asteroids game. Clearly the type of game and the use of depth cues in the game make a significant difference to the potential benefit of S3D.

Following this experience we were challenged to create a game that would be impossible to play on a 2D display. We designed a game where the user must successfully fly through a series of hoops to score points, this looked 3D but varying the hoop size made the perspective cue an unreliable guide to object depth. We found that, while not impossible, the player's high scores were on average 300% higher when the game was played in S3D compared to 2D. We conclude there are games where high scores will significantly benefit from S3D, and bring real benefit to competitive games players.

8648-39, Session 10

Impact of floating windows on the accuracy of depth perception in games

Brodie M. Stanfield, Chris R. Zerebecki, Andrew Hogue, Bill Kapralos, Univ. of Ontario Institute of Technology (Canada); Karen Collins, Univ. of Waterloo (Canada)

With the recent growth of stereoscopic 3D within the entertainment industry it has created the impetus for stereoscopic 3D in video games. With the mass consumers in the video game industry being relatively new to the use of stereoscopic 3D it has created the need for guidelines and standards for creation and use of stereoscopic 3D in games. The quality of these stereoscopic 3D parameters will determine the level of engagement and performance that a user will receive should a developer choose to add or omit such guidelines and standards. In this paper we use the Floating Window Technique to determine its effect on how accurately an individual can perceive in screen and out of screen depth of an object in a virtual environment. Through this technique we hope to determine how the Floating Window can be used to affect a user's depth perception. Thus, allowing for game designers to know if and how the Floating Window Technique can be effectively used for gaming to enhance the stereoscopic 3D experience.

8648-41, Session 10

The effects of 5.1 sound presentations on the perception of stereoscopic imagery in video games

Brian Cullen, Univ. of Ontario Institute of Technology (Canada); Daniel Galperin, Karen Collins, Univ. of Waterloo (Canada); Andrew Hogue, Bill Kapralos, Univ. of Ontario Institute of Technology (Canada)

Stereoscopic 3D (S3D) content in games, film and other audio-visual media has been increasing over the past few years. However, before S3D becomes more widespread, there are still open, fundamental questions regarding its implementation, particularly as it relates to a multi-modal experience that involves sound. Research has shown that sound has considerable impact on our perception of 2D phenomena, but very little research has considered how sound may influence S3D.

Here we present the results of an experiment that examined the effects of 5.1 surround sound and stereo loudspeaker setups on depth perception in relation to S3D imagery within a video game environment. Our aim was to answer the question: "can 5.1 surround sound enhance the participant's perception of depth in the stereoscopic field when compared to stereo sound presentations?" In addition, our experiment examined how the presence or absence of Doppler shift and frequency fall-off audio effects can influence depth judgement under these conditions. Results suggest that 5.1 surround sound presentations enhance the apparent depth of S3D imagery when compared to stereo presentations. Results also suggest that the addition of audio effects. such as Doppler shift and frequency fall-off filters, can also influence the apparent depth of S3D objects.

8648-42, Session 10

Depth perception of audio sources in stereo 3D environments

David Corrigan, Marcin Gorzel, John Squires, Frank Boland, Trinity College Dublin (Ireland)

In this paper we undertake perceptual experiments to determine the allowed differences in depth between audio and visual stimuli in stereo-3d environments while being perceived as congruent. We also investigate whether the nature of the environment and stimuli affects the perception of congruence. This was achieved by creating an audio-visual environment consisting of a photorealistic visual environment captured by a camera under orthostereoscopic conditions and a virtual audio environment generated by measuring the acoustic properties of the real environment. The visual environment consisted of a room with a loudspeaker and person forming the visual stimulus and was presented to the viewer using a passive stereoscopic display. Both pink noise samples and voiced phrases were used as audio stimuli. The 3D audio stimuli are presented over headphones using binaural recordings. The stimuli were generated at different depths from the viewer and the viewer was asked to determine whether the audio stimulus was nearer, further away or at the same depth as the visual stimulus. From our experiments it is shown that there is a significant range of depth differences for which audio and visual stimuli are perceived as congruent. Furthermore, this range increases as the depth of the visual stimulus increases.

8648-73, Session 10

An interactive in-game approach to user adjustment of stereoscopic 3D settings

Mina Tawadrous, Andrew Hogue, Bill Kapralos, Univ. of Ontario Institute of Technology (Canada); Karen Collins, Univ of Waterloo (Canada)

Since movie theatres have propagated 3D technology, content developers have been working hard at creating customizable Stereoscopic 3D (S3D) experiences for the user to take home. S3D game developers often offer a 'white box' approach in which too many controls and settings are exposed to the average consumer. This can lead to users being uncomfortable or unimpressed with their own user-defined S3D settings. Our goal is to investigate interactive approaches to setting these parameters and compare the final stereoscopic settings with the methods provided by the content developers and determine the effects this strategy has on engagement. Inspired by standard testing methodologies experienced at the optometrist, we've created a split-screen game with the same gameplay running in both screens, but with different interaxial distances. We then compare those results with the same test run on a single screen and allowing the user to move a slider changing the interaxial distance until they are most

comfortable. We expect that the interactive nature of the calibration will impact the final game engagement scores significantly providing us with an indication of whether in-game, interactive, S3D parameter calibration is a mechanism that game developers should adopt.

8648-44, Session 11

Accommodation responses to horizontal-parallax-only super multiview display

Junya Nakamura, Kosuke Tanaka, Yasuhiro Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

Super multi-view (SMV) displays have been developed to solve the vergence-accommodation conflict that causes visual fatigue with conventional 3D displays. An ideal SMV display should generate dense viewpoints in both the horizontal and vertical directions. However, the SMV displays developed so far generate dense viewpoints only in the horizontal direction because of the system complexity required for full-parallax systems. Therefore, the accommodation responses to SMV displays were measured mostly for 3D images that contain distinctive vertical edges. In this study, we investigated the influences of the edge properties contained in 3D images upon the accommodation responses evoked by horizontal-parallax-only (HPO) SMV displays. We used the recently developed reduced-view SMV display, whose interval of viewpoints was 2.6 mm, for the accommodation measurements. Two test images were used: a "Maltese cross," which contains several sharp edges, and a standard test image of "Lenna," which contains various types of edges. We found that the HPO–SMV display still evoked the accommodation responses when Lenna was displayed. There were two types of accommodation responses for Lenna; the amounts of the evoked accommodation were smaller than or similar to those for the Maltese cross, depending on where the viewers gazed.

8648-45, Session 11

Lightweight spatial-multiplexed dual focal-plane near to eye display using two freeform prisms

Dewen Cheng, Yongtian Wang, Beijing Institute of Technology (China); Guofan Jin, Tsinghua Univ. (China)

Accommodation and convergence play important roles in the natural viewing of 3-D scenes and need to be accurately matched to avoid visual fatigue. However, conventional stereoscopic displays lack the ability to adjust accommodation cues because they only have a single, fixed image plane, but the 3-D virtual objects generated by two stereoscopic images are displayed at different depths, either in front or behind the focal plane. The eyes are therefore forced to converge on those objects while maintaining accommodation fixed on the image plane to view objects clearly. The contribution of the paper is to using freeform optical surfaces in the design of a lightweight and wearable spatial-multiplexed dual focal-plane near-to-eye display that capable of adjusting the accommodation cue in accordance with the convergence cue and of generating the retinal blur cue. The system has great potential applications in both scientific research and commercial market.

8648-47, Session 11

An optical see-through, multi-focal-plane stereoscopic display prototype enabling nearly-correct focus cues

Xinda Hu, Hong Hua, College of Optical Sciences, The Univ. of Arizona (United States)

Conventional stereoscopic displays render 3D scenes from a single pair of 2D images at a fixed distance to the viewer. Therefore conventional stereoscopic displays force an unnatural decoupling of the accommodation and convergence cues, which may contribute to various visual artifacts and have significant effects on depth perception accuracy. In this paper, we present the design and implementation of an optical see-through depth-fused multi-focal-plane stereoscopic display. The prototype is capable of rendering nearly-correct focus cues for a large volume of 3D objects extending into a depth range from 0 to 3 diopters and a diagonal field of view of 40 degrees at a flicker-free speed. With the optimized optical system including a freeform prism eyepiece and a see-through compensator, the prototype system demonstrates high image quality while having minimal degradation to the see-through view. Our multi-focal-plane display potentially provides higher depth perception accuracy, higher stereoacuity and lower fatigue; it can also be a necessary tool to investigate the effects of focus cues on human depth perception and visual fatigue, therefore it has great potentials for both 3D display applications as well as basic vision research.

8648-43, Session 13

S3D depth-axis interaction for video games: performance and engagement

Chris R. Zerebecki, Brodie Stanfield, Andrew Hogue, Bill Kapralos, Univ. of Ontario Institute of Technology (Canada); Karen Collins, Univ. of Waterloo (Canada)

The film industry has embraced the Stereoscopic 3D (S3D) format, and experimented with how the inclusion might modify or enhance their products. The games industry has also included the format into their products, but rarely have interactive game mechanics been developed to take advantage and enhance the stereoscopic 3D experience. In games, interactions with the player are more important than only the visual experience. Stereoscopic displays allow players to perceive objects at different depth, and it is our hypothesis that designing a core mechanic around this would create compelling content. In this study we develop a game where the core mechanic requires the player to interact directly along the stereoscopic depth axis. A user-based experiment is then used to investigate the performance, perception and enjoyment of this game in stereoscopic 3D vs. traditional 2D viewing.

8648-48, Session 13

Discernible difference and change in object depth afforded by stereoscopic three-dimensional content

Shunnan Yang, Sowjanya Gowrisankaran, Pacific Univ. (United States); Audrey C Younkin, Philip J Corriveau, User Experience Research Group, Intel Corporation (United States); James E. Sheedy, John Hayes, Pacific Univ. (United States)

Stereoscopic three-dimensional (S3D) viewing enhances depth perception with differential image disparity but excessive disparity incurs significant viewing discomfort. The present study measured the necessary difference and change in image disparity to perceive different depth. Sixty young adults between age 18 to 40 years and with good binocular vision were recruited to detect depth difference, and their vergence eye movements were measured when performing the tasks. They was asked to detect an object located at a depth different from other identical objects at the same depth and to discern the direction this target object was subsequently shifted in depth. Smaller baseline disparity (< 2.48% screen depth) resulted in higher accuracy and shorter response time in reporting the target object, which was accompanied by more accurate convergence angle. Change in disparity was more accurately discerned when the original disparity difference was larger (>= 4 pixels) and when the change increased crossed disparity. These suggest that in constructing S3D content, depth difference is better perceived with a smaller overall disparity and a larger disparity difference between objects. Depth change is best rendered with increased rather than decreased crossed disparity.

8648-64, Session 13

Comfortable stereo viewing on mobile devices

Takashi Shibata, Tokyo Univ. of Social Welfare (Japan) and Waseda Univ. (Japan); Fumiya Muneyuki, Keisuke Oshima, Junki Yoshitake, Takashi Kawai, Waseda Univ. (Japan)

The situation of viewing stereo 3D imagery is becoming diverse as 3D technology spreads. Nowadays, 3D images are viewed on television and mobile devices as well as in cinema. The viewing of mobile devices especially has a unique feature that allows observers to view images at a short distance. We conducted an experiment that examined a relationship between visual comfort and individual observer's near phoria and interpupillary distance. The results showed that observers with large exophoria tended to prefer stereo images behind the screen. On the other hand, observers with small exophoria or esophoria tended to prefer stereo images in front of the screen. In addition, the results also showed that observers with larger interpupillary distance tended to prefer stereo images behind the screen. These findings suggest that users can adjust the depth of stereo images in advance based on their phoria or interpupillary distance for a comfortable stereo viewing on mobile devices.

8648-50, Session 12

Is the use of 55" LCD 3D screen practicable in large seminar to lecture-hall size audiences?

Justus F. Ilgner M.D., Ingo Sparrer M.D., Martin Westhofen M.D., Univ. Hospital Aachen (Germany)

Background: The presentation of surgical contents to undergraduate medical students can be challenging, as the surgical approach is often different from the anatomist's perspective that is reproduced in textbooks. Although there are many options to record endoscopic, microscopic as well as "open" surgical procedures, presentation of contents still can be costly and entail a loss in picture quality including depth impression.

Material and Methods:

We presented seven stereoscopic clips of 30 seconds to minute and 20 seconds each to 64 medical students (43 female / 21 male) as part of the "sensory organs" course block in 4th year; using one 55" LCD 3D screen with line-alternating, circular polarization. Students were asked for their subjective viewing impression and about their opinion on the usefulness of 3D presentations in medical lectures.

Results:

63% of students returned their questionnaires completed. The main results (multiple answers allowed) were: 70% noted that 3D presentations made complex anatomy easier to comprehend from an unknown perspective, 48% would feel better motivated to learn surgical procedures, 38% would generally prefer a 3D lecture to a 2D lecture, while 23% would not see any advantage of 3D presentations whatsoever.

Discussion:

While the screen size compared to audience size was far from ideal, it gave medical students, who had not been exposed to surgical procedures in the operating theatre yet, an impression of general approach to microsurgery and how the choice of surgical approach in relation of vital structures can minimize trauma and unwanted effects to the patient. The availability of larger screens, however, may necessitate changes in production of 3D material from the microscope camera onward.

8648-52, Session 12

Immersive stereoscopic panoramas

John S. Toeppen, HoloGraphics (United States); Jason Buchheim, Odyssey Expeditions (United States)

Immersive Stereoscopic Panoramas 192 words

CONTEXT; Immersive stereoscopic imaging requires sharp wide field images, special software, and high resolution displays.

OBJECTIVE; Examples of some successful mage capture, splicing, viewing, hosting, and posting techniques used in digital stereoscopic panoramic photography are given.

METHODS

Image capture uses camera movements that approximate natural eye positions reasonably well by using manual or motorized gimbal mounted systems designed for the purpose. Processing requires seamlessly stitching dozens or hundreds of images into stereo pairs. Creating stereoscopic images over 50 mega pixels benefits from programmable motorized equipment. The 2 gig limit of TIFFs is often exceeded and requires the use of GigaPan.org viewing and hosting technologies. Gigapixel stereoscopic images are viewed as a single whole while small files are quickly uploaded to improve the sharpness of the areas viewed.

RESULTS

Immersive stereo content, active scrolling and deep zoom capabilities are presently web accessible and formatted for different displays.

CONCLUSSION

Immersive content, active scrolling and deep zoom capabilities take stereoscopic photography from snapshots into the realm of immersive virtual presence when combined with modern display technologies. Scientific, artistic, and commercial applications can make effective use of existing stereoscopic displays systems by using these extended capabilities.

http://gigapan.com/galleries/7454/gigapans

8648-55, Session 12

Experimental verification of conditions for a bubble-projection three-dimensional display

Takuma Uchida, Hitoshi Ozaki, Hiroshi Kawakami, Jippei Suzuki, Mie Univ. (Japan); Koichi Nakayama, Saga Univ. (Japan)

Our goal is to develop a bubble-projection three-dimensional display (BP3DD). BP3DD does not cause a contradiction between the binocular convergence and the focal accommodation of our eyes. A 3D image on a BP3DD can be seen from all directions without requiring the use of 3D eyeglasses.

When an infrared laser beam is condensed underwater, water boils locally and bubbles are generated. Each bubble corresponds to a pixel of an underwater image. By controlling the position at which the laser beam is condensed, an arbitrary-shaped 3D screen can be generated with many bubbles. We can see a colorful 3D underwater image when visible light is projected on this 3D screen.

We created bubbles under various conditions in order to determine the conditions that were most suitable for BP3DD. When the water pressure was lower, the energy required decreased linearly. This result showed that the energy required to generate bubbles decreased when the water pressure in BP3DD came close to the water vapor pressure.

Because the output energy of a laser is limited, a high-resolution image cannot be generated. We plan to create high-resolution motion video by developing a method for decreasing the energy required by the laser.

8648-63, Session 12

Extended depth-of-field in Integral Imaging by depth-dependent deconvolution

Conference 8648

Hector Navarro, Genaro Saavedra-Tortosa, Manuel Martinez-Corral, Univ. de València (Spain); Mårten Sjöström, Roger Olsson, Mid Sweden Univ. (Sweden)

OBJECTIVE:

The aim of this work is designing a new method to significantly extend the depth of field in the reconstruction stage.

METHOD:

The idea is based in a two-step method. We first propose obtaining the depth map for the different captured perspectives. Now, the depth information of the 3D scene is available and we can filter each elemental image selecting the pixels that are associated with a given distance to the camera objective. The axial range spanning the scene is divided in intervals. It is possible to define in good approximation an effective PSF over each interval of distances. Only the pixels belonging to a certain interval are deconvolved with the effective PSF calculated for that distance. The final image is the sum of the pixels of every interval after being filtered and deconvolved. We have used an iterative algorithm based in the Richardson-Lucy deconvolution. To show the feasibility of our method, we have conducted optical experiments to obtain an integral image of a 3D scene. Over an optical table, we prepared a scene composed by three resolution charts located at three different axial positions from the in-focus plane. As a capture system, instead of a microlens array, we used the so called Synthetic Aperture Integral Imaging method, where a digital camera is translated mechanically to the positions where the microlenses would be located over the microlens array. This method let us having a large amount of pixels in every captured image and hence obtaining an accurate depth map. With this configuration we got a set of 10x10 different perspectives of the scene, focusing the camera in a plane located 650 mm away from the camera. The three tests were located 200, 370 and 500 mm away from the camera. The camera focal length was set to 35 mm and the f#=9. The above-proposed method to extend the depth of field has been applied to each captured perspective.

RESULTS:

Computational reconstructions of the 3D image have been performed using back projection algorithms in order to compare the results obtained with the original captured integral image and those obtained after applying our method. The analysis of these reconstructions in the planes where each one of the resolution charts were located shows that, with the proposed method, it is possible to recover all the frequencies that were lost due to the limited depth of field of the camera, even for the resolution chart which was farthest from the in-focus plane.

CONCLUSIONS:

We have proposed a new method to significantly extend the depth of field for 3D image reconstruction. We have illustrated our method with an experiment in which we have recovered the high-frequency information of a 3D scene, even for the most defocused planes. The proposed method has broad applications in 3D display, and 3D recognition.

8648-49, Session 14

Depth enhanced multi-layered display using polarization dependent double reflection

Na-Young Jo, Hee-Seung Kim, Chungbuk National Univ. (Korea, Republic of); Hong-Gi Lim, Chungbuk National University (Korea, Republic of); Jae-Hyeung Park, Chungbuk National Univ. (Korea, Republic of)

Recently, multi-layered displays attract growing attention. Multi-layered displays effectively compose light ray field by using several liquid-crystal display (LCD) panels. It can display high quality three-dimensional (3D) images with horizontal and vertical parallax in a volume around the display system. In multi-layered displays, the depth range wherein 3D images can be displayed depends on the number of layers. The more layers can have more information and thus increase the depth range. However, there is a practical limitation on the number of the layers since each LCD has a limited transmittance.

In this paper, we propose a depth enhancement of the multi-layered display using polarization dependent double reflection. The proposed system consists of a multi-layered display, a polarization control panel, two pieces of quarter-wave-plates, a half-mirror and a reflective polarizer. The rays from the multi-layered display pass through the polarization control panel which alternates the polarization between two orthogonal states. In one polarization state, the light rays pass through two quarter-wave-plates, a half mirror, and a reflective polarizer directly. In another polarization state, the light rays undergo additional reflection between the reflective polarizer and the half mirror before exiting from the reflective polarizer, increasing the apparent distance of the multi-layered display from the observer. By alternating these two states, it is possible to double the effective number of the layers of the conventional multi-layered display, achieving the enhancement of the depth range. In the presentation, we will explain the principle with experimental verifications.

8648-51, Session 14

Polarization analysis of the light emitted by liquid-crystal-based 3D displays

Pierre M. Boher, Thierry Leroux, Thibault Bignon, Véronique Collomb-Patton, ELDIM (France)

Liquid crystal technology is presently more or less used for all types of commercial applications involving displays and in particular 3D displays. For auto-stereoscopic 3D displays, liquid crystals are even used to build driven lenses to control the angular emission. Liquid crystal cell is essentially a polarization state modulator and the polarization state of the light is of course a tool of choice to measure precisely the performances of such devices. ELDIM has developed viewing angle and imaging instrument capable to measure the polarization state of the light in addition to standard values such as luminance, color or radiance. In the proposed paper, we use these instruments to characterize different auto-stereoscopic, passive and active glass 3D displays and we show how their can help understanding better the limit and the defects of each type of technology.

8648-53, Session 14

accurate depth estimation using spatiotemporal consistency in arbitrary camera array

Woo-Seok Jang, Yo-Sung Ho, Gwangju Institute of Science and Technology (Korea, Republic of)

Depth estimation is essential task for natural 3D image generation. In this paper, we propose an accurate depth image sequence using stereo video captured by an arbitrary camera array. In general, stereo-video-based depth estimation is carried out with two input videos that are obtained from the parallel camera array. Recently, the arc camera array has been actively used to create immersive 3D videos. In the arc camera arrays, it is difficult to adopt image rectification methods for correspondence point matching due to serious image distortion. In this work, we estimate depth data without image rectification. Consecutively, we define the potential energy function for mismatching depth detection based on spatial consistency. Energy optimization confirms mismatching depth pixels. Reasonable depth values are assigned to the obtained mismatching pixels using distances and intensity differences between the mismatched pixel and its neighbors. Finally, we pursue improvement of temporal consistency to reduce visual discomfort to viewer's eyes. Experimental results show that the proposed method provides more stable results than the image rectification-based method.

8648-56, Session 14

Stereoscopic 3D video coding quality evaluation with 2D objective metrics

Kun Wang, Kjell E. Brunnström, Acreo AB (Sweden) and Mid

Sweden Univ. (Sweden); Marcus Barkowsky, Univ. de Nantes (France); Matthieu Urvoy , University of Nantes (France); Mårten Sjöström, Mid Sweden Univ. (Sweden); Patrick Le Callet, Univ. de Nantes (France); Sylvain Tourancheau, Mid Sweden Univ. (Sweden); Börje Andrén, Acreo AB (Sweden)

The aim of this work is to compare and evaluate the quality of Stereoscopic 3D videos with existing 2D quality metrics, and examining the metrics performance by comparing it with conducted subjective measurement results on the same processed video set. The performance of the 2D objective metrics is assessed according to evaluation criteria from VQEG HDTV phase 1 procedure and the S3D general quality is evaluated by the correlation coefficient of the objective scores and the subjective DMOS. Various pooling functions are investigated to apply on the stereo 2D image quality and depth quality scores. Preliminary results show the VIFP objective metrics have a high correlation with subjective DMOS.

8648-57, Session 15

Case study: The Avengers 3D: cinematic techniques and digitally created 3D

Graham D. Clark, Stereo D (United States)

In the digital creation of a 3D film from a 2D image capture, recommendations to the filmmakers cinematic techniques are offered at each step from pre-production onwards, through set, into post.

As the footage arrives at our facility we respond in depth to the cinematic qualities of the imagery in context of the edit and story, with the guidance of the Directors and Studio, creating stereoscopic imagery.

These cinematic considerations that we recommend and follow, in the digital creation of 3D imagery, include the traditional standard Visual Depth Cues that have long been in use in the stereoscopic field.

Traditional Cinematic Techniques have been further developed that support and utilize the various Visual Depth Cues, in most cases they are accentuated by Stereoscopic 3D imagery.

Our involvement in The Avengers was early, after reading the script we had the opportunity and honor to meet the Director Joss Whedon, and DP Seamus McGarvey on set.

We presented what could only seem obvious to such great filmmakers in the ways of cinematic techniques as they related to the standard depth cues, in particular Stereoscopic.

Our hope was any cinematic habits that supported better 3D would be emphasized. And they were!

8648-25, Session IPI

An autostereoscopic display system with four viewpoints in full resolution using active anaglyph parallax barrier

Qu Zhang, Hideki Kakeya, Univ. of Tsukuba (Japan)

Parallax barrier is one of the most widely acknowledged autostereoscopy solutions. However, it suffers from the limited number of viewers at the same time. We apply anaglyph into active parallax barrier and create a display pattern to show four viewpoints in full resolution so that two viewers can enjoy at the same time. In one frame, stripes with a "magenta-black-green-black" pattern are used as the parallax barrier, while image stripes from four different views (A, B, C and D) are shown in a "AG/CM-BG/DM-CG/AM-DG/BM" pattern, where AG/CM means the green component of view A overlapped with the magenta component of view C. With parallax barrier properly positioned, four different views with half resolution and half anaglyph component at each pixel can be shown in one frame. By changing the patterns of barrier and images every frame, four full-color views in full resolution can be achieved from four viewpoints every four frame. We made a system based on this method with two 120 Hz displays. Although it should have required 240 Hz displays to show stable

stereoscopy for this system, the result showed that anaglyph with green and magenta helped to reduce flickers and four viewpoints were successfully created.

8648-58, Session IPI

Wide viewing angle three-dimensional display using curved HOE lens array

Yoshiaki Oshima, Hideya Takahashi, Osaka City Univ. (Japan); Kenji Yamada, Osaka Univ. (Japan)

We propose an omnidirectional 3D display system. This is a tool for communication around a 3D image between a small number of people. This 3D display consists of multiple basic 3D display units. Each basic 3D display unit consists of a projector and a curved HOE lens array sheets. The projector is located on the center of the curvature of the curved HOE lens array, and it projects light rays on the curved HOE lens array with ascending vertical angle 45-degree. Projected light rays are reflected and reconstruct the 3D image over the center of the curvature by the curved HOE lens array. In this method, even though the viewing angle of each HOE lens does not increase, the viewing angle of the 3D image by the curved HOE lens array can be increased. To verify the effectiveness of the proposed method, we constructed the prototype basic 3D unit. The viewing angle of a 3D image was 44-degree. Therefore, the viewing angle was increased by the proposed method compared to 17-degree by the conventional flat HOE lens array. This paper describes the principle of proposed 3D display system, and also describes the experimental results.

8648-59, Session IPI

Volumetric display based on optical scanning of an inclined image plane by an image rotator and imaging by a dihedral corner reflector array

Yuki Maeda, Diasuke Miyazaki, Takaaki Mukai, Osaka City Univ. (Japan); Satoshi Maekawa, National Institute of Information and Communications Technology (Japan)

We propose to use a rotational optical scanning method for a volumetric three-dimensional(3D) display based on optical scanning of an inclined image plane. The image plane is scanned by an image rotator and formed by a dihedral corner reflector array, which forms a real image at a plane-symmetrical point. Stack of the scanned and formed image plane creates displayable space of a 3D image, which satisfies all the criteria of stereoscopic vision and can be seen by the naked eyes.

The image rotator used in this study consists of two prism sheets and a planar mirror. The image plane is scanned rotationally by rotating the proposed image rotator along an axis parallel to the mirror plane. Since just extending a distance between the image plane and the mirror makes the displayable space large, this scanning method is effective to display a large 3D image easily.

We made a prototype display and observed scanned image plane to measure the displayable space. The size of the space was 1500 cubic centimeters that is approximately ten times as large as our previous display that uses a translational scanning method.

8648-60, Session IPI

Resolution-enhanced display of three-dimensional objects by using the dynamic mask pattern in a depth-priority integral-imaging system

Suk-Pyo Hong, Eun-Seong Kim, Kwangwoon Univ. (Korea, Republic of); Donghak Shin, Dongseo Univ. (Korea, Republic of); Eun-Soo Kim, Kwangwoon Univ. (Korea, Republic of)

In this paper, we propose a novel approach for resolution-enhanced display of three-dimensional (3-D) objects by employing a dynamic mask pattern (DMP) in the depth-priority integral-imaging (DPII) system. Basically, the maximum resolution of a DPII system might be limited by the number of picked-up elemental images (EIs), so that several approaches including the moving array lenslet technique (MALT) and the intermediate-view reconstruction technique (IVRT) have been suggested to improve the resolution of the DPII system. But, these methods require a mechanical movement of the lenslet array or suffer from incorrect image formation. To overcome these drawbacks, we propose a novel DMP-based DPII system, in which the location of the mask pattern displayed on a LCD panel can be electronically controlled and the mask pattern expects to be moved fast enough to make an after-image effect. Accordingly, in the proposed system, high-resolution EIs can be picked up just by time-multiplexed movement of the electronic mask pattern displayed on a LCD panel, so that the upper resolution limit imposed by the Nyquist sampling theorem could be overcome, which results in a remarkable improvement of the resolution of the conventional DPII system. To confirm the feasibility of the proposed method, experiments are carried out and the results are discussed.

8648-61, Session IPI

A method of reducing number of pixels on display device for super-multiview display

Yuki Adachi, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan)

A super multi-view display provides smooth motion parallax without special glasses, and it is expected the observer is free from the visual fatigue caused by the accommodation-vergence conflict. However, there is a problem a huge number of pixels are required on a display device such as liquid crystal display panel because high-density rays are required for good quality images and each ray needs corresponding pixel. We proposed a new three-dimensional display based on lenticular method to reduce the required number of pixels. The rays are shot out to only around observer's pupils. To do this, the lenticular lens of which viewing zone angle is narrowed is used and the lenticular lens is illuminated by parallel light made by cylindrical lenses and LEDs to suppress side lobes. The direction of the parallel light is changed to follow the observer's eye. We designed a display using these components and confirmed the availability of the proposed method by computer simulation. As a result, the parallax pitch and the viewing zone nearly equaled to designed values when the display was observed from the front, but these values were increased with the viewing angle. After analysis, the reason why the parallax pitch and the viewing zone were expanded is thought as the curvature of field of the lenticular lens. Thus, Future work includes proposing the solution of the effect of this aberration.

8648-62, Session IPI

Development of super multiview head-up display and evaluation of motion parallax smoothness

Hiroyuki Nishio, Yasuhiro Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

A super multi-view head-up display (SMV-HUD) was developed. The smooth motion parallax provided by the SMV technique enables a precise superposition of 3D images on real objects. The developed SMV-HUD was used to explore display conditions to provide smooth motion parallax. It had three configurations that display 3D images in short-, middle-, and long-distance ranges, assuming the usage environments of PC monitors, TVs, and advertisement displays, respectively. The subjective evaluation was performed by changing the depth of 3D images and the interval of viewing points. The interval of viewing points was changed by displaying identical parallax images to succeeding viewing points. We found that the ratio of the image shift between adjacent parallax images to the pixel pitch of 3D images dominated the perception of the discontinuous motion parallax. When the ratio was smaller than 0.2, the discontinuity was not perceived.

When the ratio was larger than 1, the discontinuity was always perceived and the 3D resolution decreased two times at transition points of viewing points. When the crosstalk between viewing points was relatively large, the discontinuity was not perceived even when the ratio was one or two, although the resolution decreased two or three times throughout the viewing region.

8648-65, Session IPI

Subjective assessment of visual discomfort induced by binocular disparity and stimulus width in stereoscopic image

Seong-il Lee, Yong Ju Jung, Sohn Hosik, Yong Man Ro, KAIST (Korea, Republic of)

Stereopsis is such a complex mechanism that stereoscopic depth perception can be influenced by not only disparity magnitude, but several other factors. For instance, a number of psychophysical evidences have been reported that binocular fusion limit could be affected by characteristics of visual stimulus, such as disparity gradient, stimulus size, spatial frequency, motion, etc. Subjective sensation of visual discomfort could also be different depending on characteristics of visual stimulus. In our previous study, we have observed a clue that the horizontal size of objects could affect the perceived visual comfort. However, no thorough subjective assessment has been conducted to investigate the relation among horizontal stimulus size, disparity magnitude, and visual comfort. To the best of our knowledge, further study is necessary to investigate the effect of horizontal stimulus size and disparity magnitude on visual comfort of stereoscopic image. In this paper, we investigate the influence of stimulus size on visual comfort of stereoscopic image. To that end, we conduct extensive subjective assessments using visual stimuli with various horizontal sizes and disparity magnitudes. The statistical analysis results show that the perceived degree of visual discomfort is significantly increased as the horizontal size of visual stimulus become smaller than a given disparity value.

8648-66, Session IPI

Evaluation of the no-disparity realistic image from a sense of presence and low fatigue

Hisaki Nate, Nobutaka Natui, Naoki Hayashi, Kazuo Ishikawa, Tokyo Polytechnic Univ. (Japan); Toyohiko Hatada, Tokyo Optometric College (Japan); Yutaka Ichihara, Nobuyuki Miyake, Yoshijiro Ushio, Nikon Corp. (Japan)

We evaluated an observer's fatigue and a sense of presence, in observing the no disparity realistic image (NDR image). NDR image is consisted of two images (right and left image). Right image is created by shifting all pixels in left image same amount. Consequently, there are no disparities in all area of NDR image. NDR image which is reconfigured the contents that contain image with large disparity might have a possibility that it suppress an observer's fatigue and let him feel high presence.

Subjects observed three condition's videos (stereoscopic video, NDR video and 2D video). Subjects observed two videos (Each video was about 30 minutes.) in each condition. There were parts with large disparity (more than 5 degrees) in videos.

Subjects responded SSQ (Simulator Sickness Questionnaire) and VAST (Visual Analogue Scale of Tiredness). After observing the first video, they responded VAST. After observing second video, they responded SSQ and VAST and answered the questionnaire about a sense of presence of the displayed image.

Results showed that NDR image let observer feel high presence and an observer's fatigue was low. NDR image is effective, even if contents which contain large disparity are converted into NDR image.

IS&T / SPIE
Electronic Imaging
SCIENCE AND TECHNOLOGY

Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs

Jing Li, Marcus Barkowsky, Patrick Le Callet, Polytech' Nantes (France)

No Abstract Available

Stereoscopic cameras for the real-time acquisition of panoramic 3D images and videos

Luis E. Gurrieri, Eric Dubois, Univ. of Ottawa (Canada)

There are different panoramic techniques to produce outstanding omni-stereoscopic images of static scenes. However, a camera configuration capable to sample the scene in real-time to produce stereoscopic panoramas is still a subject of research. In this paper, two different multiple-camera configurations capable to produce high-quality stereoscopic panoramas in real-time are presented. Unlike existing methods, the proposed multiple-camera systems acquire all the information necessary to render stereoscopic panoramas at once. The first configuration exploits micro-stereopsis arising from a narrow baseline to produce omni-stereoscopic images. The second configuration uses an extended baseline to produce poly-centric panoramas and additional depth information, e.g. disparity and occlusion maps, which are used to synthesize stereoscopic views at arbitrary directions. The results of simulating both camera configurations and the pros and cons of each set-up are presented.

Applications of 2D to 3D conversion for educational purposes

Yoshihisa Koido, Hiroyuki Morikawa, Saki Shiraishi, Wataru Maruyama, Soya Takeuchi, Waseda Univ. (Japan); Toshio Nakagori, Tokai Univ. (Japan); Masataka Hirakata, Hirohisa Shinkai , Toppan Printing (Japan); Takashi Kawai, Waseda Univ. (Japan)

There are three main approaches creating stereoscopic S3D content: stereo filming using two cameras, stereo rendering of 3D computer graphics, and 2D to S3D conversion by adding binocular information to 2D material images. Although manual "off-line" conversion can control the amount of parallax flexibly, 2D material images are converted according to monocular information in most cases, and the flexibility of 2D to S3D conversion has not been exploited. If the depth is expressed flexibly, comprehensions and interests from converted S3D contents are anticipated to be differed from those from 2D. Therefore, in this study we created new S3D content for education by applying 2D to S3D conversion. For surgical education, we created S3D surgical operation content under a surgeon using a partial 2D to S3D conversion technique which was expected to concentrate viewers' attention on significant areas. And for art education, we converted Ukiyoe prints; traditional Japanese artworks made from a woodcut. The conversion of this content, which has little depth information, into S3D, is expected to produce different cognitive processes from those evoked by 2D content, e.g., the excitation of interest, and the understanding of spatial information. In addition, the effects of the representation of these contents were investigated.

A stereoscopic archive for the heritage of industrial modernization and the evaluation

Hayoung Yoon, Kohei Ando, Waseda Univ. (Japan); Takashi Kawai, Hiroyuki Morikawa, Waseda University (Japan); Keiji Ohta, Quality eXperience Design Co., Ltd. (Japan)

In recent years, record and retention methods of cultural assets by stereoscopic images, stereoscopic archive have attracted public attentions, and various researches on the archive building methods have been carried out. However, cultural assets without tangible properties at present cannot be pictured and so recording and retaining them as stereoscopic images is not possible. To overcome the limitation, the use of the 2Dto3D conversion technology, which converts 2D images to 3D images by applying binocular information, has been emphasized. In this study produced a stereoscopic archive for the heritage of industrial modernization (the JVC KENWOOD Yokohama Factory façade), which was dismantled in March 2011, by converting the 2D images to 3D, and examined the impressions. Moreover, the drawing images of the cultural asset were also converted to 3D images in order to evaluation the expression as a stereoscopic archive.

For the experiment, head mounted display and 42-inch 3D TV were used. 30 students in twenties years of age. with normal binocular vision. participated through questionnaires and interviews to compare the impressions, between the conditions.

A wavelet-based image quality metric for the assessment of 3D synthesized views

Emilie Bosc, Institut National des Sciences Appliquées de Rennes (France); Federica Battisti, Marco Carli, Univ. degli Studi Roma Tre (Italy); Patrick Le Callet, Polytech' Nantes (France)

In this paper we present a novel image quality assessment technique for evaluating virtual synthesized views in the context of multi-view video. In particular, Free Viewpoint Videos (FVV) are generated from uncompressed color views and their compressed associated depth maps by means of the View Synthesis Reference Software (VSRS), provided by MPEG. Prior to the synthesis step, the original depth maps are encoded with different coding algorithms thus leading to the creation of additional artifacts in the synthesized views. The core of proposed wavelet-based metric is in the registration procedure performed to align the synthesized view and the original one, and in the skin detection that has been applied considering that the same distortion is more annoying if visible on human subjects rather than on other parts of the scene. The effectiveness of the metric is evaluated by analyzing the correlation of the scores obtained with the proposed metric with Mean Opinion Scores collected by means of subjective tests. These results are also compared against those of well known objective quality metrics (i.e. PSNR, SSIM, MSSIM). In more details, 27 observers assessed 42 FVV sequences based on six different sequences, each of them processed with 7 depth map coding algorithms.

Stereoscopic display gray-to-gray crosstalk measurement

Yi-Heng Chou, Fu Hao Chen, Kuo-Chung Huang, Kuen Lee, Ching-Chiu Liao, Yi-hen Chen, Industrial Technology Research Institute (Taiwan)

There are several studies on estimating crosstalk of 3D displays. In previous study?gray to gray crosstalk model has been modified. In this paper, we use six commercial stereoscopic displays including passive polarized glasses and active shutter glasses, both left and right eyes were measured by a luminance meter to verify the gray to

gray crosstalk model and analyze the stability of the measurement data using our standard operating procedure (SOP). According to the SOP, we use simple statistical method to verify the repeatability of data. These results, can be used as an important parameters in stereoscopic display visual quality, also can be used as a design feedback for engineer.

8648-77, Session IPI

Compressing stereo images in discrete Fourier transform domain

Chee Sun Won, Dongguk Univ. (Korea, Republic of); Shahram Shirani, McMaster Univ. (Canada)

Frequency features of stereo images are investigated in the DFT (Discrete Fourier Transform) domain by characterizing phase and magnitude properties originated from the horizontal parallax of the stereo images. Also, the well-known DFT properties including the conjugate symmetry property are utilized to identify essential frequency components of stereo images. Our investigation reveals that the DFT of the stereo images has useful properties that can prioritize the DFT coefficients for compact representations and compressions.

8648-78, Session IPI

Depth inloop resampling using dilation filter for free viewpoint video system

Seok Lee, Seungsin Lee, Hocheon Wey, Jaejoon Lee, Dusik Park, Samsung Advanced Institute of Technology (Korea, Republic of)

A depth dilation filter is proposed for free viewpoint video system based on mixed resolution multi-view video plus depth (MVD). By applying gray scale dilation filter to depth images, foreground regions are extended to background region, and synthesis artifacts occur out of boundary edge. Thus, objective and subjective quality of view synthesis result is improved. A depth dilation filter is applied to inloop resampling part in encoding/decoding, and post processing part after decoding. Accurate view synthesis is important in virtual view generation for autostereoscopic display, moreover there are many coding tools which use view synthesis to reduce interview redundancy in 3D video coding such as view synthesis prediction (VSP) and depth based motion vector prediction (DMVP), and compression efficiency can be improved by accurate view synthesis. Coding and synthesis experiments are performed for performance evaluation of a dilation filter with MPEG test sequences. Dilation filter was implemented on the top of the MPEG reference software for AVC based 3D video coding. By applying a depth dilation filter, BD-rate gains of 0.5% and 6.0% in terms of PSNR of decoded views and synthesized views, respectively.

8648-79, Session IPI

Adaptive hole filling for 3D warping-based virtual view synthesis

Tatsuro Mori, Nagoya Univ. (Japan); Mehrdad Panahpour Tehrani, National Institute of Information and Communications Technology (Japan); Toshiaki Fujii, Nagoya Univ. (Japan); Masayuki Tanimoto, Nagoya Industrial Science Research Institute (Japan)

As one of methods to realize stereoscopic image synthesis, especially multi-viewpoint generation, 3D warping is usually employed with View + Depth format. This process includes the problem that holes appear in the virtual image. In conventional method, the holes were dealt with by median filter. There are different reasons why holes appear. So, treating them collectively is improper. We analyze the factors, and recognize that two ones exist, boundary between foreground and background, and decline of resolution.

In this paper, we propose a new hole filling method considering these factors. In the first step, we classify nearby pixels into boundary or same object area according to the gradient of depth value. For boundary case, we hold them and refer to other two real cameras. For another case of same object area, we set up sub-pixels between nearby pixels and warp them if the depth area has layer or an object gets more close to camera after 3D warping, which probably causes decline of resolution.

We implement these methods in simulation. As the result, we confirm that proposed method prevent boundary in virtual image from being ambiguous, and can fill holes in object area mostly though holes remain in conventional method.

Conference 8649: The Engineering Reality of Virtual Reality 2013

Monday - Tuesday 4 –5 February 2013

Part of Proceedings of SPIE Vol. 8649 The Engineering Reality of Virtual Reality 2013

8649-1, Session 1

CalVR: an advanced open-source virtual reality software framework

Jurgen Schulze, Andrew Prudhomme, Philip Weber, Thomas A. DeFanti, Univ. of California, San Diego (United States)

Despite a variety of existing virtual reality frameworks, we decided to develop our own, because none of the existing systems offer the features we were looking for. Our new software system is called CalVR, and is an open source project, which we offer to anyone in the academic community free of charge. Key features of our system, which are not present in existing ones, are: rendering support for autostereoscopic displays, multi-user and multi-hand input, multi-viewer output, a clear separation between GUI functionality and graphical implementation, asynchronous input device handling with a message queue, and programming objects to simplify the creation of applications with multiple, independent objects on the screen. Like other systems, we support VR displays of any shape and size, driven by PC clusters.

8649-2, Session 1

CAVE2: a Hybrid Reality Environment for Immersive Simulation and Information Analysis

Alessandro Febretti, Arthur Nishimoto, Terrance Thigpen, Jonas Talandis, Lance Long, J. D. Pirtle, Univ. of Illinois at Chicago (United States); Tom Peterka, Electronic Visualization Laboratory (United States); Alan Verlo, Maxine Brown, Dana Plepys, Daniel J. Sandin, Luc Renambot, Andrew Johnson, Jason Leigh, Univ. of Illinois at Chicago (United States)

No Abstract Available

8649-3, Session 2

MASCARET : creating virtual learning environments from system modelling

Ronan Querrec, Paola Vallejo, Cédric Buche, Ecole Nationale d'Ingénieurs de Brest (France)

The design process for a VLE such as that put forward in the SIFORAS project (SImulation FOR training and ASsistance) means that system specifications can be differentiated from pedagogical specifications. Equally, the system specifications can be obtained directly from the specialists' expertise; that is to say directly from the Product Lifecycle Management (PLM) tools. To do this, the system model needs to be considered as a piece of VLE data. We propose a meta-model which can be used to represent these system models : MASCARET. In order to ensure that the meta-model is capable of describing, representing and simulating such systems, MASCARET is based on a standard defined by OMG : SYSML .

To design a VLE from PLM specifications, we need 4 kinds of information:

• the 3D geometry (Collada files),

• the system structure (Blocks, properties, compositions in SYSML),

• the behaviour of the entities within the system (StateMachines, Event, Signals in SYSML),

• the system's usage or maintenance procedures (Activities, partitions, actions in SYSML).

8649-4, Session 2

Employing WebGL to develop interactive stereoscopic 3D content for use in biomedical visualization

Semay Johnston, Univ. of Illinois at Chicago (United States) and Biomedical and Health Information Sciences (United States); Luc Renambot, Univ. of Illinois at Chicago (United States); Daniel Sauter, Univ. of Illinois at Chicago (United States) and School of Art and Design (United States)

Web Graphics Library (WebGL), the forthcoming web standard for rendering native 3D graphics in a browser, represents an important addition to the biomedical visualization toolset. This project defines a WebGL design methodology for a target audience of biomedical artists with a basic understanding of web languages and 3D graphics. The methodology was informed by a literature review of 3D web technologies and their value in biomedical education, a survey of current WebGL resources and frameworks, and the development of an interactive web application depicting the anatomy and various pathologies of the human eye. The application supports several modes of stereoscopic displays for a better understanding of 3D anatomical structures. Implemented by most modern browsers, the WebGL specification has begun changing the landscape of the web with native hardware accelerated 3D graphics—otherwise achievable only through browser plugins like Adobe Flash and the Unity3D Web Player. Though not yet universally supported, WebGL has steadily gained importance due to shrinking support for third-party plugins and is projected to become a mainstream method of delivering 3D online content. Finally, the availability of stereoscopic displays is steadily increasing (3D TV, desktop, and mobile), allowing the delivery of a VR experience to web users. The value of WebGL for education has been apparent since the technology's inception. Google's Body Browser in of 2009 showcased the potential of WebGL to bring a rich 3D interactive learning environment to the browser. A number of developers have explored the use of WebGL not only for creating interactive anatomical models, but also for chemical and molecular visualization and web-based volume rendering. Several obstacles have hindered widespread adoption of this technology in biomedical education. The WebGL specification is currently in draft state and has not yet been implemented by all browsers. Additionally, because WebGL utilizes hardware acceleration, older generation machines may lack the computational power to meet the higher requirements for rendering 3D graphics. However, WebGL will eventually become a web standard, and newer generation computers and mobile devices with data access and embedded 3D graphics accelerators will replace older equipment. Learning WebGL can prove difficult for practitioners in biomedical visualization due a lack of instructional literature. Programming directly in WebGL is inherently challenging due to its low-level nature. Fortunately, open-source developers have created high-level frameworks. These frameworks vary in quality and purpose. Despite the challenges, the task of learning to use a WebGL framework is very achievable. Many professionals in biomedical visualization already have transferable skills that will aid in the learning process (3D graphics animation, illustration and stereoscopic 3D). The pivotal role of web and mobile learning in biomedical education demands that practitioners in biomedical visualization adapt to the shifting trajectory of the web. By embracing new standards, they will gain the ability to deliver quality interactive VR content in a format that promises widespread accessibility and longevity. This project addresses a major obstacle hindering broad adoption of WebGL technology in biomedical visualization by exploring the question: How can biomedical artists with a basic understanding of web languages begin using WebGL to create 3D interactive web applications? An example workflow for creating a WebGL application is provided. The demonstration, named Eye3D, is an interactive 3D eye model based on the UIC VRMedLab's VR project, The Virtual

Eye, which remains largely inaccessible due to the need for expensive equipment with which to display it. Bringing this content to the web fulfils three objectives of 1) making available a valuable educational tool, 2) providing a platform for learning HTML5 and WebGL technologies and 3) delivering a VR experience through the web.

8649-5, Session 2

FreeVR: honoring the past, looking to the future

William R. Sherman, Indiana Univ. (United States); Daniel Coming, Google (United States); Simon Su, Ball Aerospace & Technologies Corp. (United States)

Ten years of experience in designing and implementing a VR integration library have produced a wealth of lessons upon which we can further build and improve our capabilities of writing worthwhile virtual reality applications. The VRX virtual library is a mature library, yet continues to progress and benefit from the insights and requests produced by application development. We compare VRX with the standard provisions of virtual reality integration libraries, and provide an in-depth look at VRX itself. We examine what design decisions worked, and which fell short. In particular, we look at how the features of VRX serve to restore applications of the past into working condition and aid in providing longevity to newly developed applications.

8649-6, Session 2

An industrial approach to design compelling VR and AR experience

Simon Richir, Ecole Nationale Supérieure d'Arts et Métiers (France); Philippe Fuchs, Ecole des Mines ParisTech (France); Domitile Lourdeaux, UTC (France); Ronan Querrec, Cedric Buche, ENIB (France)

The convergence of technologies currently observed in the field of VR, AR, robotics and consumer electronic reinforces the trend of new applications appearing every day. But when transferring knowledge acquired from research to businesses, research laboratories are often at a loss because of a lack of knowledge of the design and integration processes in creating an industrial scale product. In fact, the innovation approaches that take a good idea from the laboratory to a successful industrial product are often little known to researchers.

The objective of this paper is to present the results of the work of several research teams that have finalised a working method for researchers and manufacturers that allow them to design virtual or augmented reality systems and enable their users to enjoy "a compelling VR experience". That approach, called "the I2I method", present 11 phases from "Establishing technological and competitive intelligence and industrial property" to "Improvements" through the "Definition of the Behavourial Interface, Virtual Environment and Behavioural Software Assistance". As a result of the experience gained by various research teams, this design approach benefits from contributions from current VR & AR research. Our objective is to validate and continuously move such multidisciplinary design team methods forward.

8649-7, Session 2

3D interactive augmented reality-enhanced digital learning systems for mobile devices

Kai-Ten Feng, National Chiao Tung Univ. (Taiwan); Po-Hsuan Tseng, National Taipei Univ. of Technology (Taiwan); Pei-Shuan Chiu, Jia-Lin Yang, Chun-Jie Chiu, National Chiao Tung Univ. (Taiwan)

With enhanced processing capability of mobile platforms, augmented reality (AR) has been considered a promising technology for achieving enhanced user experiences (UX). Augmented reality is to impose virtual information, e.g., videos and images, onto a live-view digital display. UX on real-world environment via the display can be effectively enhanced with the adoption of interactive AR technology. Enhancement on UX can be beneficial for digital learning systems. There are existing research work based on AR targeting for the design of e-learning systems. However, none of these work focuses on providing three-dimensional (3-D) object modeling for enhanced UX based on interactive AR techniques. In this paper, the 3-D interactive augmented reality-enhanced learning (IARL) systems will be proposed to provide enhanced UX for digital learning. The proposed IARL systems consist of two major components, including the markerless pattern recognition (MPR) for 3-D models and velocity-based object tracking (VOT) algorithms. Realistic implementation of proposed IARL system is conducted on Android-based mobile platforms. UX on digital learning can be greatly improved with the adoption of proposed IARL systems.

8649-8, Session 2

User contexts: identifying specificity of user needs

Diane Gromala, Simon Fraser Univ. (Canada)

User Contexts: Identifying Specificity of User Needs

This paper explores the specific needs of users on two levels. FIrst, it reviews the basic needs of users, and extends this basic approach by identifying needs specific to female users. Second, it identifies specific needs of users who experience acute and chronic pain. Users who have chronic pain, a newly identified disease that is estimated to affect 1 in 5 people in industrialized nations, present several established needs that range from physical and perceptual sensitivities and impairments. Although these are known in Pain Medicine, very little has been published in VR research outside of acute pain contexts. This papers presents user needs identified in evidence-based medical research, several of which have been recently been verified in immersive VR contexts in the Pain Studies Lab at Simon Fraser University.

Finally, we present a new framework for approaching VR users in general, based on the lessons learned from these two groups, and from VR research from the 1980s to the present.

8649-10, Session 3

Using the computer-driven VR environment to promote experiences of natural world immersion

Lisa A. Frank, Beckman Institute for Advanced Science and Technology (United States) and The Wisconsin Institute for Discovery (United States)

In December, 2011, nearly 800 people experienced the exhibit, <1>:"der"//pattern for a virtual environment, created for the fully immersive CAVE™ at the UW-Madison. This exhibition took my 2D nature-based photographic work and reinterpreted it for VR.Varied responses such as: "It's like a moment of joy," or "I had to see it twice," or "I'm still thinking about it weeks later" were common. Although an implied goal of my 2D artwork is to create a connection that makes viewers more aware of what it means to be a part of the natural world, these six VR environments opened up an unexpected area of inquiry that my 2D work has not. Even as the experience was mediated by machines, there was a softening at the interface between technology and human sensibility. Somehow, for some people, through the unlikely auspices of a computer-driven environment, the project spoke to a human essence that they connected with in a way that went beyond all expectations and felt completely out of my hands. Other interesting behaviors were noted: in some scenarios some spoke of intense anxiety, acrophobia, claustrophobia–even fear of death when the scene took them underground. These environments were believable enough to cause extreme responses and disorientation for some people; were fun, pleasant and wonder-filled for most; and were liberating, poetic and meditative for many others. The exhibition seemed to promote imaginative skills, creativity, emotional insight, and environmental sensitivity. It also revealed the CAVE™ to be a powerful

tool that can encourage uniquely productive experiences. Quite by accident, I watched as these nature-based environments revealed and articulated an essential relationship between the human spirit and the physical world. The CAVE™ is certainly not a natural space, but there is clear potential to explore virtual environments as a path to better and deeper connections between people and nature. We've long associated contact with nature as restorative, but those poetic reflections of Thoreau and others are now confirmed by research. Studies are showing that contact with nature can produce faster, greater recovery from stress and other illnesses, reduction in anger, and an increased sense of well-being. Additionally, I discovered that the novelty of a virtual reality experience can bring new focus and fresh attention to elements of our world that we have grown immune to. Possibly, the boletus edulis in one scene seemed to have been made more remarkable and mysterious in VR than if it was seen in the backyard. A VR environment can be used to create opportunities to experience being in the world differently. Here they can be inside of an egg that is inside of a nest that is held by tree branches over a creek bed in a floating landscape where a light spring snow is falling. We are liberated from the worldly limitations of our body. The question is this: in an anti-natural environment, can immersants in a CAVE™ become more ecologically sympathetic and spiritually connected? Although the exhibit was not put through any form of testing as of yet, my observations amount to a remarkable vision of what VR might provide for us as an instrument to expand consciousness and promote wellness. Creating exceptional, transformative experiences may seem like a lofty goal for VR but that purpose is at the heart of any art making process.

8649-11, Session 3

A virtual reality environment of oriental residences

Catherine Chi, Indiana Univ. (United States)

The presentation is a virtual reality environment built with 3D models and synthetic textures. By navigating through the space, the viewers can feel themselves personally immersed in the world, and interact with the building structures. In the project I want to express the relationship between architectural style and cultural difference, and how the emotional condition or characteristics of the residents are affected by their residencies. The community we live in is a symbol for self-reflection. The differences between the geographical conditions and cultural influences change the appearance of residences and city landscapes.

I was inspired by the building structure in my hometown-Taiwan, which is really different with the architecture style in the United States. In the US the buildings often spread out horizontally, but in Taiwan the buildings try to extend vertically and densely as much as possible due to the limitation of space. The growth and the slowly disappearance of the buildings represent the progress of human's existence. The piece demonstrates that a virtual reality environment enhances the immersive experience, which makes the viewers present in the space and at that moment.

8649-12, Session 3

Mrs. Squandertime

Josephine Anstey, Dave Pape, Univ. at Buffalo (United States)

In this paper we discuss Mrs. Squandertime, a real-time, persistent simulation of a virtual character, her living room, and the view from her window, designed to be a wall-size, projected art installation. Through her large picture window, the eponymous Mrs. Squandertime watches the sea: boats, clouds, gulls, the tide going in and out, people on the sea wall. The hundreds of images that compose the view are drawn from historical printed sources. The program that assembles and animates these images is driven by weather, time, and tide data constantly updated from a real physical location. The character herself is rendered photographically in a series of slowly dissolving stills which correspond to the character's current behavior. We intend Mrs. Squnadertime to be a quietly compelling, virtual, experience for the

viewer, to stimulate the out-of-clock-time, regenerative, contemplative sense of well-being that a real land or seascape offers. At a time when groups as diverse as eco-spiritualists and earnest social scientists suggest that the mental and physical health of much of the world's increasingly urban population is damaged by lack of contact with nature, can virtual wild space be part of a solution?

8649-13, Session 4

There's an app for that shirt!: evaluation of augmented reality tracking methods on deformable surfaces for fashion design

Silvia Ruzanka, Rensselaer Polytechnic Institute (United States); Katherine Behar, Baruch College (United States); Ben Chang, Rensselaer Polytechnic Institute (United States)

In this presentation we will discuss appAREL, a creative research project at the intersection of augmented reality, fashion, and performance art. AppAREL is a mobile augmented reality application that transforms otherwise ordinary garments with 3D animations and modifications. With appAREL, entire fashion collections can be uploaded in a smartphone application, and "new looks" can be downloaded in a software update. The project will culminate in a performance art fashion show, scheduled for March 2013.

appAREL includes textile designs incorporating fiducial markers, garment designs that incorporate multiple markers with the human body, and iOS and Android apps that apply different augments, or "looks", to a garment.

We will demonstrate appAREL; discuss our philosophy for combining computer-generated and physical objects; and share the challenges we encountered in applying fiduciary markers to the 3D curvatures of the human body.

8649-14, Session 4

Augmented reality: past, present, and future

Laura Inzerillo, Palermo Univ. (Italy)

A great opportunity has permitted to carry out a cultural, historical, architectural and social research with great impact factor on the international cultural interest. We are talking about the realization of a museum whose the main theme is the visit and the discovery of a monument of great prestige: the monumental complex of Steri in Palermo. The museum runs is divided into sub themes including the one above all, that has aroused the international interest so much that it was forward the instance to include the museum like cultural heritage of UNESCO. It is the realization of a museum path that regards the cells of the Inquisition, which are located just inside of some buildings of the monumental complex. The project, as a whole, is faced in a synergy view between the various competences implicated: historical, chemical, architectural, topographical, representative informatics. The birth of the museum will be a sums of the results of all these disciplines involved.

Methodology, implementation, fruition, virtual museum, parameters, potentials, setbacks on the cultural heritage and landscape knowledge, that's and more than that has been studied and development during all the research phases.

8649-15, Session 4

Vroom: Designing an augmented environment for remote collaboration in digital cinema production

Todd Margolis, Tracy Cornish, Univ. of California San Diego (United States)

As media technologies become increasingly affordable, compact and inherently networked, new generations of tele-collaborative platforms continue to arise which capitalize on these new affordances. Virtual

reality has been primarily concerned with creating simulations of environments that can transport participants to real or imagined spaces that replace the "real world". Meanwhile Augmented Reality systems have evolved to embed elements from Virtual Reality environments into the physical landscape. Perhaps now there is a new class of systems that reverse this precept to capitalize on dynamic media landscapes and physical display environments to enable data exploration through collaboration.

Vroom (Virtual Room) is a next-generation reconfigurable tiled display environment in development at the California Institute for Telecommunications and Information Technology at the University of California, San Diego. Vroom enables freely scalable digital collaboratories, connecting distributed, high-resolution visualization resources for collaborative work in the sciences, engineering and the arts. Vroom transforms a physical space into a completely immersive mixed media environment with large format interactive display surfaces, video teleconferencing and spatialized audio based on high-speed optical networks.

The system utilizes narrow bezel display walls, directional audio and advanced interaction devices to share ultra high resolution data. Vroom builds on previous research from the NSF funded OptiPuter project by integrating modular OptiPortables (portable tiled display systems) with open-source middleware for local and remote collaboration.

Vroom enables group collaboration for local and remote participants to share knowledge and experiences. Possible applications include: remote learning, brainstorming/storyboarding, post-production editorial review, high resolution video playback, tele-immersive 3D interaction, easy-to-use screencasting and simple image, video and text file sharing. To support these various scenarios, Vroom features support for multiple user interfaces (optical tracking, touch UI, gesture interface, etc.), support for directional and spatialized audio, giga-pixel image interactivity, 4K video streaming, 3D visualization and telematic production.

8649-16, Session 4

New perspectives and limitations in the use of virtual reality in the rehabilitation of motor disorders

Alessandro De Mauro, Aitor Ardanza, Vicomtech-IK4 (Spain); Esther Monge, Ctr. de Terapia Integral del Niño, Momo (Spain); Francisco Molina Rueda, Univ. Rey Juan Carlos (Spain)

Several studies have shown that both virtual and augmented reality are technologies suitable for rehabilitation therapy due to the inherent ability of simulating real daily life activities while improving patient motivation.

In this paper we will first present the state of the art in the use of virtual and augmented reality applications for rehabilitation of motor disorders and second we will focus on the analysis of the results of our project. In particular, requirements of patients with cerebrovascular accidents, spinal cord injuries and cerebral palsy to the use of virtual and augmented reality systems will be detailed.

8649-17, Session P1

Art, science, and immersion: data-driven experiences

Ruth G. West, Univ. of California San Diego (United States); Laura Monroe, Los Alamos National Lab. (United States); Jacquelyn Ford Morie, The Univ. of Southern California (United States); Julieta Cristina Aguilera, Adler Planetarium & Astronomy Museum (United States)

This panel and dialog-paper explores the potentials at the intersection of art, science, immersion and highly dimensional, "big" data to create new forms of engagement, insight and cultural forms. We will address questions such as: "What kinds of research questions can be addressed at the intersection of art + science + immersive environments that can't be addressed otherwise?" "How is

art+science+immersion distinct from state-of-the art visualization?" "What does working with immersive environments and visualization offer that other approaches don't or can't?" "Where does immersion fall short?" We will also explore current trends in the application of immersion for gaming, scientific data, entertainment, simulation, social media and other new forms of big data. We ask what expressive, arts-based approaches can contribute to these forms in the broad cultural landscape of immersive technologies.

8649-18, Session 5

Nomad devices for interactions in immersive virtual environments

Paul George, Renault S.A. (France) and Arts et Métiers ParisTech, Institut Image, LE2I – CNRS (France) and Univ. de Technologie Compiègne (France); Andras Kemeny, Renault Technocentre (France) and Arts et Métiers ParisTech, Institut Image, LE2I – CNRS (France); Frédéric Merienne, Arts et Métiers ParisTech, Institut Image, LE2I – CNRS (France); Jean-Rémy Chardonnet, Ecole Nationale Supérieure d'Arts et Métiers (France); Indira Thouvenin, Univ. de Technologie Compiègne (France); Javier Posselt, Emmanuel Icart, Renault S.A. (France)

Renault is currently setting up a new CAVE, a 5 rear-projected wall virtual reality room with a combined resolution of ~140 Mpixels, distributed over sixteen 4k projectors and two 2k projector as well as an additional 3D HD collaborative powerwall.

Renault's CAVE aims at answering needs of different steps of conception of a vehicle. Starting from vehicle architecture through the subsequent phases of design, ergonomic and perceived quality control to production, Renault has built up a list of use-cases and defined the specifications of the French national collaborative project VARI3 (Virtual & Augmented Reality Intuitive and Interactive Interface) in collaboration with ON-X, Theoris, CEA LIST and Institute Image.

One of the goals of this project is to study interactions in a CAVE, especially with nomad devices such as IPhone or IPad to explore the possibilities of manipulation and observation of objects. Inspired by current uses of nomad devices (multi-touch gestures, IPhone UI look'n'feel and AR applications), we have implemented an early feature set taking advantage of these popular input devices. In this paper, we present its performance through measurement data collected in our test platform, a 4-sided homemade low-cost virtual reality room, powered by ultra-short-range and standard HD home projectors.

8649-19, Session 5

Analysis of tactors for wearable simulator feedback: a tactile vest architecture

David Prater, Stephen Gilbert, Eliot Winer, Iowa State Univ. (United States)

Current training simulators for police officers and soldiers lack two critical qualities for establishing a compelling sense of immersion within a virtual environment: a strong disincentive to getting shot, and accurate feedback about the bodily location of a shot. This research addresses these issues with a hardware architecture for a Tactical Tactile Training Vest (T3V). In this study, we have evaluated the design space of impact tactors and present a T3V prototype that can be viscerally felt.

This research focuses on determining the optimal design parameters for creating maximum tactor hitting energy. The energy transferred to the projectile directly relates to the quality of the disincentive. The complete T3V design will include an array of these tactors on front and back of the body to offer accurate spatial feedback.

The impact tactor created and tested for this research is an electromagnetic projectile launcher, similar to a solenoid, but lower profile and higher energy. Our best tactor produced the efficiency at just above 0.1% with projectile energy of approximately 0.08 Joules. Users in an informal pilot study described the feeling as "surprising," "irritating," and "startling," suggesting that this level of force is approaching our target level of disincentive.

8649-20, Session 5

Use of virtual reality to promote hand therapy post-stroke

Daria Tsoupikova, Univ. of Illinois at Chicago (United States); Nikolay Stoykov, Rehabilitation Institute of Chicago (United States); Randy M. Vick, The Art Institute of Chicago (United States); Yu Li, Derek Kamper, Molly Listenberger, Rehabilitation Institute of Chicago (RIC) (United States)

A novel, art-based virtual rehabilitation system was developed and tested for use as a rehabilitation protocol for upper extremity therapy following stroke. The system was developed by an interdisciplinary team of engineers, art therapists, occupational therapists, and VR artists to improve patients' motivation and engagement. Here we describe system design, development, and user testing for efficiency, subject's satisfaction and clinical feasibility. We report results of the completed qualitative, pre-clinical pilot study of the system effectiveness for therapy. We are currently conducting a longitudinal intervention study over 6 weeks in stroke survivors with chronic hemiparesis. Initial results following use of the system on the first subjects demonstrate that the system is operational and can facilitate therapy for post stroke patients with upper extremity impairment.

8649-21, Session 5

Collaborative imaging of urban forest dynamics: augmenting re-photography to visualize changes over time

Ruth G. West, Univ. of California San Diego (United States); Jarlath O'Neil-Dunne, Univ. of Vermont (United States); Robert B. Pless, Washington Univ. in St. Louis (United States)

Ecological measurement is a challenge because there is a need measure subtle changes in data captured over large areas and over varied time-scales. Furthermore, while changes happen over large regions, they are measured based on the growth/health of individual organisms, and the metrics of individuals are not always accurately reflected in large-scale measurement. The challenge is thus to measure patterns of slow, subtle change occurring along multiple spatial and temporal scales, and then to visualize those changes in a way that makes important variations visceral to the observer. This research envisions a community-based and participatory approach based around augmented re-photography of ecosystems. We show a case study for the purpose of monitoring the urban tree canopy. The goal is to explore, for a set of urban locations, the integration of ground level re-photography with available LIDAR data, and to create a dynamic view of the urban forest, and its changes across various spatio-temporal scales. This case study gives the opportunity to explore various augments to improve the ground level image capture process, protocols to support 3D inference from the contributed photography, and both in-situ and web based visualizations of the temporal change over time.

8649-22, Session IPI

Integral virtual display for long distance view

Kosuke Shimazu, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan)

In driving a vehicle, if the information that is needed to drive can be displayed like Augmented Reality (AR), a driver operates the vehicle effectively. For example, it is efficient to display the images that indicate an intersection where the vehicle should turn next. Unlike conventional AR system such as using see through Head Mounted Display, the image needs to be displayed to long distance at environment of the vehicle. On the other hand, Head Up Display (HUD) is currently used on vehicle for displaying speed meter, tachometer and so on. But it is difficult that the HUD is applied to on-vehicle AR because viewing field is insufficient. To realize this system, it needs a HUD that displays the optical virtual images for long distance view and covers large viewing field. Therefore, we propose a display which is divided up many small optical systems. Basic idea is using a convex lens array and elemental images similarly to Integral Photography (IP). Each lens has corresponding one elemental image in front, and each lens generates the virtual image. In this paper, on the proposed method, a theoretical formula of the position relation among the elemental images to create the continuous virtual images is solved. And we simulated the system with ray tracing method.

Conference 8650: 3D Image Processing (3DIP) and Applications 2013

Wednesday - Thursday 6 –7 February 2013

Part of Proceedings of SPIE Vol. 8650 Three-Dimensional Image Processing (3DIP) and Applications 2013

8650-1, Session 1

Depth image completion using a modified robust PCA

Hyunjung Shim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

In recent years, depth cameras have been widely adopted into various applications due to its ease of accessibility and an affordable price. However, the depth image acquired by those cameras often includes invalid or missing measurements due to the limit of its sensing principle. These eventually prevent from robust depth sensing, yielding a substantial amount of missing data in measurements.

In this paper, we introduce a novel approach to completing the depth image with missing data using a modified robust principle component analysis (PCA). The original formulation of robust PCA considers the entries of data sample are statistically independent. Our data sample is a 3D position and three entries are solely dependent on depth value. It is because the 3D point should lies on a line, defined by a center of camera and its pixel coordinate. Hence, we propose a modified robust PCA so to enforce the ray constraint on depth completion.

The basic intuition of proposed algorithm is that the depth can be represented by a combination of clusters and each cluster holds the regularity in PCA subspace. By imposing such constraints, we demonstrate that the proposed algorithm is effective to handle the missing data on repeatable, uneven surfaces.

8650-2, Session 1

Depth image post-processing method by diffusion

Yun Li, Mårten Sjöström, Ulf Jennehag, Roger Olsson, Mid Sweden Univ. (Sweden)

Multi-view three-dimensional television relies on view synthesis to reduce the number of views being transmitted. Depth images from stereo pairs or range cameras may contain erroneous values, which entail artifacts in a rendered view. Post-processing of the data may then be utilized to enhance the depth image with the purpose to reach a better quality of synthesized views utilizing the data.

We propose a partial differential equation PDE-based interpolation method for the reconstruction of the smooth areas in depth images, while preserving significant edges. We modeled and evaluated the depth image and the view synthesis quality by adjusting thresholds for (Canny) edge detection and a uniform sparse sampling factor followed by the second order PDE interpolation. The edge detector identifies significant edges and the uniform sparse samples controls the agreement between the reconstructed depth image and the original one, while the interpolation enforces the smoothness constraint for the areas between significant edges. The objective results show that a depth image processed by the proposed scheme can achieve a better quality of synthesized views than the original depth image. Visual inspection confirmed the results.

8650-3, Session 1

Beta-function B-spline smoothing on triangulations

Lubomir T. Dechevsky, Peter Zanaty, Narvik Univ. College (Norway)

In this work we investigate a novel family of smooth rational basis functions on triangulations for fitting, smoothing, denoising and compressing geometric data.

In a parallel communication [Dechevsky, L. T., Zanaty, P., Asymptotically optimal K-functional Tikhonov regularization of multivariate non-parametric regression problems with non-uniformly sampled deterministic design, Computational Imaging XI (2013)] we explore the performance of a general smooth construction over a very general class of simply connected partitions or covers of (possibly, not simply) connected multidimensional domains when using a specific approach to the so-called K-functional Tikhonov regularization using a smoothness penalty. Here we carry through an upgraded version of this research program for the important special case of simplicialization (in 2D, triangulation) partitions and star-1 neighborhood covers for which we discuss two new constructions: a semi-smooth upgrade of the classical piece-wise affine construction and a smooth upgrade simultaneously of the general smooth construction and the semi-smooth construction on simplicializations.

We perform comparative analysis and graphical comparison of the results obtained for both simplicialized constructions with the analogous results obtained for the general construction in the special case of simplicializations.

Based on this comparison, we derive a set of conclusions and recommendations about possibly adaptive, context-dependent choice of one or more of the considered constructions and data-dependent tuning of respective parameters of the selected construction(s).

8650-4, Session 2

Evaluation of efficient high quality depth upsampling methods for 3DTV

Luc P. J. Vosters, Technische Univ. Eindhoven (Netherlands); Chris Varekamp, TP Vision Netherlands (Netherlands); Gerard de Haan, Technische Univ. Eindhoven (Netherlands)

Depth maps play a vital role for the generation and transmission of 3D-TV content. High quality 3D content generation requires high quality depth maps. In practice, depth maps generated by stereo-matching, time-of-flight cameras, or decoders, inherently have a low resolution and suffer from unreliable estimates and noise. Depth post-processing is therefore typically required to enhance the result.

State-of-the-art methods for depth-map enhancement perform up-sampling, edge-alignment and de-noising in a single step. Depth up-sampling methods can be roughly classified into optimization and filter based depth up-sampling. While the performance of optimization based depth up-samplers is high, their computational complexity and memory usage is currently prohibitive for real-time implementations. Conversely, filter based techniques avoid the high computational cost and memory requirements and allow for scan-line implementations. Therefore, filter based methods are potential candidates for implementation in real-time 3D-TV systems.

Despite the importance for high quality 3D video, there has never been a fair attempt to quantitatively assess the performance of the state-of-the-art filter-based, depth up-sampling methods on a large dataset for multiple upscale factors and different levels of noise. Therefore, in this paper we provide a thorough benchmarking comparing each method's depth accuracy and interpolation quality on a large number of images in the Middlebury stereo dataset. To get a fair comparison we first compute the optimal parameter setting for each method by searching the whole parameter space and minimizing the number of bad pixels over all test images. Additionally, we analyze each method's computational complexity with the big O notation and measure the runtime of a GPU implementation we built for each method.

8650-5, Session 2

Multiview ToF sensor fusion technique for high-quality depth map

Deuk Hyeon Kim, Jinwook Choi, Kwanghoon Sohn, Yonsei Univ. (Korea, Republic of)

In this paper, we proposed a novel method to fuse and upsample multiview depth maps obtained by Time-of-Flight (ToF) based depth sensors. The proposed method can be robust to the camera calibration error and effectively applied to the Multiview Video plus Depth (MVD) format based on multiple depth sensors. ToF based depth sensor is widely used in computer vision fields. However, there are some problems in a sense that it provides a low resolution and unreliable z-depth around object boundaries and highly reflective regions. We address these problems by using multiview ToF sensor fusion technique. It is essential for high quality multiview 3D contents. The proposed method is based on the depth calibration which adjusts the distribution of z-depth between two depth sensors. It provides a coherent z-depth for the corresponding points between depth maps. Confidence map based multiview depth fusion makes it possible to restore the acquisition error and be aligned well with the color image by using only reliable z-depth in the MVD format. Experimental results show that the proposed method based on multiview depth maps is superior to the conventional method which means a 2D-plus-depth format using one color image and one depth map

8650-6, Session 2

Time-of-flight depth image enhancement using variable integration time

Sun Kwon Kim, Ouk Choi, Byong Min Kang, James D. K. Kim, Chang-Yeong Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

Time-of-Flight (ToF) cameras are used for a variety of applications because it delivers depth information at a high frame rate. These cameras, however, suffer from challenging problems such as noise and motion artifacts. To increase signal-to-noise ratio (SNR), the camera should calculate a distance based on a large amount of infrared light, which needs to be integrated over a long time. On the other hand, the integration time should be short enough to suppress motion artifacts. We propose a ToF depth imaging method to combine advantages of short and long integration times exploiting an image fusion scheme proposed for color imaging. To calibrate depth differences due to the change of integration times, a depth transfer function is estimated by analyzing the joint histogram of depths in the two images of different integration times. The depth images are then transformed into wavelet domains and fused into a depth image with suppressed noise and low motion artifacts. To evaluate the proposed method, we captured a moving bar of a metronome with different integration times. The experiment shows the proposed method could effectively remove the motion artifacts while preserving high SNR comparable to the depth images acquired during long integration times.

8650-7, Session 2

Pseudo-random modulation for multiple 3D time-of-flight camera operation

Dong-Ki Min, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Ilia Ovsiannikov, SAMSUNG Electronics Co., Ltd. (United States); Yohwan Noh, Wanghyun Kim, SAMSUNG Electronics Semiconductor (Korea, Republic of); Sunhwa Jung, Joonho Lee, Deokha Shin, Hyekyung Jung, Lawrence Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Gregory Waligorski, Lilong Shi, Samsung Semiconductor, Inc. (United States); Yoondong Park, Chilhee Chung, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

3D time-of-flight depth cameras utilize modulated light sources to detect the distance to objects as phase information. A serious limitation may exist in cases when multiple depth time-of-flight cameras are imaging the same scene simultaneously. The interference caused by the multiple modulated light sources can severely distort captured depth images. To prevent this problem and enable concurrent 3D multi-camera imaging, we propose modulating the camera light source and demodulating the received signal using sequences of pulses, where the phase of each sequence is varied in a pseudo-random fashion. The proposed algorithm is mathematically derived and proved by experiment.

8650-8, Session 2

Lossy contour-coding in segmentation-based intra depth map coding

Jan Hanca, Adrian Munteanu, Peter Schelkens, Vrije Univ. Brussel (Belgium) and Interdisciplinary Institute for Broadband Technology (Belgium)

Efficient depth map compression is mandatory in order to fulfill the strenuous technical demands of various 3D video applications such as 3DTV and Free Viewpoint Video. In this paper we present an intra-depth map coding system employing optimized segmentation procedures for depth maps and lossy or lossless coding of the segment contours. Lossless contour coding makes use of chain coding techniques while lossy contour coding is performed using contour approximation and predictive coding mechanisms. Both systems employ high order arithmetic entropy coding. Depth values in each segment are encoded using predictive coding and DCT-based encoding of the residuals. The experimental results show that the lossless contour codec outperforms the conventional JBIG2 codec employed in other segmentation-based depth map coding approaches in the literature. The effect of lossy contour coding on the synthesized views is analyzed: the experiments show that lossy contour coding outperforms the lossless coding method for various input data. Finally, the experimental comparison against JPEG-2000 shows that the proposed system is a viable alternative for light intra coding of depth maps.

8650-9, Session 3

Estimation of spreading fire geometrical characteristics using near infrared stereovision

Lucile L. Rossi, Tom Toulouse, Univ. di Corsica Pasquale Paoli (France); Moulay Akhloufi, Ctr. of Robotics and Vision (Canada); Antoine Pieri, Yves Tison, Univ. di Corsica Pasquale Paoli (France)

In fire research and forest firefighting, there is a need of robust metrological systems able to estimate the geometrical characteristics of outdoor spreading fires. In recent years, we assist to an increased interest in wildfire research to develop non destructive techniques based on computer vision. This paper presents a new approach for the estimation of fire geometrical characteristics using near infrared stereovision. Spreading fire information like position, rate of spread, height and surface, are estimated from the computed 3D fire points. The proposed system permits to follow fire spreading on a ground area of 5mx10m.

8650-10, Session 3

Uniform grid upsampling of 3D lidar point cloud data

Prudhvi Gurram, Shuowen Hu, John F. Dammann, U.S. Army Research Lab. (United States)

Airborne laser scanning light detection and ranging (LIDAR) systems are used for remotely sensing topology and bathymetry. The most frequently used data collection technique used in LIDAR systems employs a linear mode scanning. The resulting scanning data is a non-uniformly sampled 3D point cloud. To interpret the 3D point cloud

data and further process the data, it is most commonly converted to digital elevation model (DEM). In order to obtain DEMs in a uniform and upsampled raster format, the elevation information from the available non-uniform 3D point cloud data is mapped onto the uniform grid points. After mapping is done, the grid points where the elevation information is not available are filled by using interpolation techniques. In this paper, partial differential equations (PDE) based approach has been proposed to perform the interpolation and upsample the 3D point cloud onto a uniform grid. By employing higher order PDEs, smoothness is maintained over homogeneous regions and sharp edge information in the scene is retained. The proposed algorithm will reduce the draping effects near the edges of objects like buildings in the scene, which are caused by existing algorithms. Simulation results will be presented to illustrate the advantages of the new algorithm.

8650-11, Session 3

Improvements on an MMI-based method for automatic texture mapping of 3D dense models

Pasquale Ferrara, Istituto Nazionale di Ottica (Italy); Francesca Uccheddu, Univ. degli Studi di Firenze (Italy); Anna Pelagotti, Istituto Nazionale di Ottica (Italy)

Not all range devices acquire, along with 3D data, the object's texture. Moreover, not always the desired texture is the visible light image.

We propose to use depth maps from several hierarchical viewpoints to match the homologue points between the 3D model and the given texture image. Every depth map keeps ineherently the correspondence between the 3D points and its pixels. By finding the depth map that better overlap with the given texture we find the better correspondence between the 3D vertices and the texture image. The Maximization of Mutual Information (MMI) methods proved to be highly reliable as similarity measure, however, MI is strongly dependent on the number of samples on which it is computed. In order to correctly compare estimates of MI, it is therefore advisable to perform the analysis on the same number of pixels.

We propose a weighted estimation of MI depending on how many pixels are considered. The new developed algorithm involves Laplace's rule of succession to estimate MI. In this way, an a priori information is introduced when low amount of pixels are involved in the MI estimation. We also present a comparison of the results in case the MMI is applied straightforwardly, with a bounding box and with the employ of Laplace's rule of succession.

8650-27, Session IPI

Passive stereoscopic panomorph system

Anne-Sophie Poulin-Girard, Simon Thibault, Denis Laurendeau, Univ. Laval (Canada)

In the last decade, wide-angle stereoscopic systems using fish-eye lenses have been proposed but the compromise made for having a large field of view is low resolution and high distortion resulting in imprecise depth estimation of objects in a 3D scene. High and non-uniform distortion, especially in the azimuthal direction, is often considered as a weakness of panoramic lenses because it is sometimes difficult to compensate for by image processing.

The aim of this paper is to present an alternative to existing stereoscopic panoramic systems by taking advantage of non-uniform distortion and anamorphosis in Panomorph lenses. There are many challenges related to this project like the calibration of the system and the creation of a 3D depth estimation algorithm that suits the resolution of the different areas in the images.

This paper presents the steps leading to the choice of two specific Panomorph lenses that are used in a Passive Stereoscopic Panomorph System. It also lists different configurations and their related applications. Then, we present preliminary results for the calibration of one selected configuration. Finally, for this configuration, 3D reconstruction of an object in a scene located at the ROI is presented.

8650-28, Session IPI

3D hand localization by low-cost webcams

Cheng-Yuan Ko, Chung-Te Li, Chen-Han Chung, Liang-Gee Chen, National Taiwan Univ. (Taiwan)

In recent years, Human Computer Interaction (HCI) is an active research region. The most intuitive way for user to interactive with device is using user's hand. Thus, the most important information for system to interactive with user is the location of user's hand. Because of our goal is create a new user interface and widely used in any consumer electronics, such as interactive 3DTV or interactive Tablet Personal Computer, so the cost is an important factor. We use two low cost webcams to build the 3D hand localization and use stereo matching algorithm to get the depth map. However, the quality of the depth map is so poor due to the noise in captured image pair, so unlike to common segmentation method in depth map which sets the threshold and then segments the user's hand easily, we have more things to do. In this paper, unlike traditional 2D hand localization which are based on skin-color or feature extraction, or RGB-D cameras use both skin color detection and depth, we only use two low cost commodity webcams to build a stereo system and get the depth map by stereo matching. According to the experiment results, it is robust to any hand shapes.

8650-29, Session IPI

3D shape extraction of internal and external surfaces of glass objects

Alban Bajard, Olivier Aubreton, Frédéric Truchetet, Univ. de Bourgogne (France)

Three-dimensional (3D) digitization of manufactured objects has been investigated for several years and consequently, many techniques have been proposed. Even if some techniques have been successfully commercialized, most of them assume a diffuse or near diffuse reflectance of the object's surface, and difficulties remain for the acquisition of "optically non cooperative" surfaces, such as transparent or specular ones. To address such surfaces, we propose a non conventional technique, called "Scanning from Heating" (SfH). In contrast to classical active triangulation techniques that acquire the reflection of visible light, we measure the thermal emission of the heated surface. The aim of this paper is to demonstrate, by using the experimental setup designed for specular (transparent or not) objects, how this method allows reconstruction both of internal and external surfaces of glass objects from a unique measure.

8650-30, Session IPI

Real-time self-calibration of depth sensing systems based on structured-light 3D

Vikas Ramachandra, Kalin Atanassov, James Nash, Sergio R. Goma, Qualcomm Inc. (United States)

A structured-light system for depth estimation is a type of 3D active sensor that consists of a structured-light projector, that projects an illumination pattern on the scene (e.g. mask with vertical stripes) and camera, which captures the illuminated scene. Based on the received patterns, depths of different regions in the scene can be inferred. For this setup to work optimally, the camera and projector must be aligned such that the projection image plane and image capture plane are parallel, i.e. free of any relative rotations (yaw, pitch and roll).

In reality, due to mechanical placement inaccuracy, the projector-camera pair will not be aligned. In this paper we present a calibration process which measures the misalignment. We also estimate a scale factor to account for differences in focal lengths of projector and camera.

The 3 angles of rotation can be found by introducing a plane in the field of view of the camera and illuminating it with projected light patterns. This plane image is captured and processed to obtain the relative pitch, yaw and roll angles, as well as scale through an iterative process. This algorithm uses the effects of rotations on the depth map of the plane image.

8650-31, Session IPI

Discovering unexpected information using a building energy visualization tool

Benoit Lange, William Puech, Nancy Rodriguez, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France)

Building energy consumption is an important problem in construction field, old buildings are gap of energy and they need to be refactored. Energy footprint of buildings needs to be reduced. New buildings are designed to be suitable with energy efficiency paradigm. To improve energy efficiency, Building Management Systems (BMS) are used: BMS are IT (Information Technology) systems composed by a rules engine and a database connected to sensors. Unfortunately, BMS are only monitoring systems: they cannot predict and mine efficiently building information. RIDER project has emerged from this observation. This project is conducted by several French companies and universities, IBM at Montpellier, France, leads the project. The main goal of this project is to create a smart and scalable BMS. This new kind of BMS will be able to dig into data and predict events. This IT system is based on component paradigm and the core can be extended with external components. Some of them are developed during the project: data mining, building generation model and visualization. All of these components will provide new features to improve rules used by the core. In this paper, we will focus on the visualization component. This visualization use a volume rendering method based on sensors data interpolation and a correlation method to create new views. We will present the visualization method used and which rules can be provided by this component.

8650-32, Session IPI

An efficient anaglyph stereo video compression pipeline

Adhatus Solichah A., National Taiwan Univ of Science and Technology (Taiwan); Guan-Ming Su, Dolby Labs., Inc. (United States); Kai-Lung Hua, Yu-Chi Lai, National Taiwan Univ of Science and Technology (Taiwan)

In this paper, we consider the end-to-end stereo video distribution pipeline consisting of singe-sensor digital camera pairs, the legacy consumer-grade video decoder, and anaglyph displays. As we observe that only part of the color information is used in the final reconstructed anaglyph video, we propose an efficient video compression and processing pipeline to significantly improve the final reconstructed anaglyph video quality with lower computation complexity. Unlike the traditional pipeline which demosiackes the color filter array (CFA) raw pixels directly after the camera capturing stage at the stereo video encoder side, the proposed pipeline packs the required color information used in the final anaglyph image from the raw CFA stereo image and compresses the packed pixels using the legacy video codec. The demosaicking is conducted at the stereo video decoder side after the pixels are unpacked. With this new proposed pipeline, we can eliminate the color transform and chroma reformatting operations. Besides, the stereo video compression efficiency can be further improved. The simulation results show that the proposed framework outperforms the traditional one at all bit rates and can have up to 4.66dB PSNR gain, which demonstrates the superior performance over the traditional distribution pipeline.

8650-33, Session IPI

3D/2D image registration by image transformation descriptors (ITDs) for thoracic aorta imaging

Pawel J. Lubniewski, Laurent Sarry, Univ. d'Auvergne Clermont-Ferrand I (France); Bruno Miguel M.D., CHU de Clermont-Ferrand (France); Christophe Lohou, Univ. d'Auvergne Clermont-Ferrand I (France)

In this article, we present a novel image registration technique. Unlike most state of the art methods, our approach allows us to compute directly the relationship between images. The proposed registration framework, built in a modular way, can be adjusted to particular problems. Tests on sample image database prove that our method is fast and robust and could be successfully used for many cases.

The proposed registration method, unlike most current techniques, does not use an iterative framework. The transformation parameters are obtained via certain quantities issued from images - Image Transformation Descriptors (ITD). An ITD estimates single transformation parameter for every image independently and the transformation is calculated in a single pass. Such approach is effective for registering multiple images. It is easier to estimate the time of data processing than for iterative algorithms.

The presented technique based on ITD is interesting alternative for classic algorithms. Its unique properties can be advantageous for many image alignment problems. The possibility of using different descriptors, adapted for particular cases, makes our approach very flexible. Fast time of computing is an important feature and motivates to use our technique even as an initialization step before execution of other algorithms which could be more precise, but slow and sensitive to initialization of the parameters.

8650-34, Session IPI

Stereo matching with partial information

Yusuf Cem Subakan, Omer Can Gurol, Cagatay Dikici, Bogaziçi Üniv. (Turkey)

In this paper, we address the stereo matching problem in stereo videos using partially informed Markov Random Fields (MRFs) using the motion as a side information. We first estimate the motion within one of the videos and then regularize the stereo matching using the estimated motion field from one of the videos. The proposed scheme is applied to both dense and sparse images such as edge maps.

8650-12, Session 4

Analysis of weighting of normals for spherical harmonic cross-correlation

Robert L. Larkins, Michael J. Cree, Adrian A. Dorrington, The Univ. of Waikato (New Zealand)

The registration of overlapping point clouds has recently received increased attention due to the growing availability of 3D sensors. Spherical harmonic cross-correlation is a robust registration algorithm that uses the normals of two overlapping point clouds to bring them into coarse rotational alignment. However, not all of these normals are beneficial, as there are often normals that contribute noise to the cross-correlation, detrimentally affecting registration. By applying a weighting scheme to the normals, the accuracy of the cross-correlation can be improved. In this paper we perform an in-depth analysis investigating how different weighting schemes impact the registration accuracy of the cross-correlation. It is found that the initially tested scheme, in which each normal is given a binary classification, increased the likelihood of correct registration when contrasted against the base approach of equally weighting every normal. Schemes with greater complexity are shown to further improve registration alignment, even with decreased overlap between point clouds. The primary goal of this analysis is to identify which schemes for weighting normals provides the best aid for point cloud registration when used with spherical harmonic cross-correlation.

8650-13, Session 4

Edge-aided virtual view rendering for multiview video plus depth

Suryanarayana M. Muddala, Mårten Sjöström, Roger Olsson, Sylvain Tourancheau, Mid Sweden Univ. (Sweden)

Depth-Image-Based Rendering (DIBR) of virtual views is an essential method in three dimensional 3-D video applications to produce additional perspectives from texture and depth information, such as the multi-view-plus-depth (MVD) format. However, there are still artifacts in virtual views as a consequence of rendering using existing DIBR methods. In this paper, we propose an alternative DIBR method for MVD. The proposed method introduces an edge aided pixel and interpolates pixel values in the virtual view using the actual projected coordinates from two adjacent views, by which cracks and disocclusions are automatically filled. In particular, we propose a method to merge pixel information from two adjacent views in the virtual view before the interpolation; a weighted averaging of projected pixels is applied within the range of one pixel in the virtual view. We compared virtual view images rendered by the proposed method to the corresponding view images rendered by state-of-the-art methods. Objective metrics demonstrated an advantage of the proposed method for most investigated media contents. Subjective tests gave no significant difference to state-of-the-art methods, which we find encouraging as the proposed method omits specific processing steps to remove different artifacts.

8650-14, Session 4

Smarter compositing with the Kinect

Alex S. Karantza, Roxanne L. Canosa, Rochester Institute of Technology (United States)

In this work, we apply the concept of a deferred shading pipeline to the data gathered by a depth sensor. The depth image corresponds to a depth buffer, from which the position of a pixel in 3D space can be calculated, assuming knowledge of the camera's field of view. Another buffer containing surface normals is represented as an XYZ triplet representing the direction of each surface element relative to the camera. This information is found from the geometry of the scene and is usually combined with a texture map that represents fine surface detail. We propose using the depth image directly, instead of a texture map, to approximate the surface orientation. Once the buffers are populated with real data, the synthetic scene is rendered by combining the data into a single set of buffers and a mask is used to overlay the depth information. This set of buffers is used to re-shade the final composite, incorporating effects from conventional computer graphics to add realism to the final scene.

8650-15, Session 4

Real time 3D object model reconstruction using self-imaging

Arezoo Movaghar, Reza Safabakhsh, Khosro Madanipour, Amirkabir Univ. of Technology (Iran, Islamic Republic of)

Among the current methods of 3D imaging and visualization, optical topography is one of the most popular one. It is simple, inexpensive and a nondestructive procedure which has no contact with the object. In this paper we are using fringe projection in a modern way in order to acquire highly-detailed models of 3D objects. A collimated laser beam is projected on a grid and produces self-images of the grid in different distances. A complete sinusoidal pattern is projected on the object's surface and the deformed pattern was used for analyzing three dimensional profile of the object.

8650-17, Session 5

Novel calibration procedure for SVBRDF estimation of 3D objects using directional illumination and structured light projection

Jakub F. Krzeslowski, Robert Sitnik, Grzegorz Mączkowski, Warsaw Univ. of Technology (Poland)

Estimation of geometry and reflectance of 3D objects requires that surface geometry is registered together with photometric data. We present a method which combines geometrical camera calibration and photometric calibration into a single procedure utilizing only one calibration target. Using structured light projection and directional illumination, the surface of a 3D object can be registered with an integrated measuring device. To estimate spatial distribution of reflectance parameters, a Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) model is used. We also show a 3D image processing method to estimate SVBRDF parameters using an arbitrary defined array of illuminators and algorithms to reconstruct this surface using specialized visualization software. This approach allows for effective measurement of geometry and visual properties of 3D objects represented by a dense point cloud model. It can become a valuable tool for documentation of digital heritage and in industrial computer vision applications.

8650-18, Session 5

Wide range time-of-flight camera: design, analysis, and simulation

Ouk Choi, Samsung Advanced Institute of Technology (Korea, Republic of)

Time-of-flight cameras measure the distances to scene points by emitting and detecting a modulated infrared light signal. The modulation frequency of the signal determines a certain maximum range within which the measured distance is unambiguous. If the actual distance to a scene point is longer than the maximum range, the measured distance suffers from phase wrapping, which makes the measured distance to be shorter than its actual distance by an unknown multiple of the maximum range. This paper proposes a time-of-flight camera that is capable of restoring the actual distance by simultaneously emitting light signals of different modulation frequencies and detecting them separately in different regions of the sensor. We analyze the noise characteristic of the camera, and acquire simulated depth maps using a commercially available time-of-flight camera, reflecting the increased amount of noise due to the use of dual-frequency signals. We finally propose a phase unwrapping method for restoring the actual distances from such a dual-frequency depth map. Through experiments, we demonstrate that the proposed method is capable of extending the maximum range to at least twice, with high success rates.

8650-19, Session 5

Efficient intensity-based camera pose estimation in presence of depth

Maha El Choubassi, American Univ. of Beirut (Lebanon); Oscar Nestares, Yi Wu, Igor Kozintsev, Horst Haussecker, Intel Corp. (United States)

The widespread success of Kinect enables users to acquire both image and depth information with satisfying accuracy at relatively low cost. We leverage the Kinect output to efficiently and accurately estimate the camera pose in presence of rotation, translation, or both. The applications of our algorithm are vast ranging from camera tracking, to 3D points clouds registration, and video stabilization.

The state-of-the-art approach uses point correspondences for estimating the pose. More explicitly, it extracts point features from images, e.g., SURF or SIFT, and builds their descriptors, and matches features from different images to obtain point correspondences. However, while features-based approaches are widely used, they perform poorly in scenes lacking texture due to scarcity of features or in scenes with repetitive structure due to false correspondences. Our algorithm is intensity-based and requires neither point features' extraction, nor descriptors' generation /matching. Due to absence of depth, the intensity-based approach alone cannot handle camera translation. With Kinect capturing both image and depth frames, we extend the intensity-based algorithm to estimate the camera pose in case of both 3D rotation and translation. The results are quite promising.

Conference 8650

8650-20, Session 5

Depth correction in ToF camera

Byong Min Kang, Keechang Lee, James D. K. Kim, Chang-Yeong Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

A captured depth using a Time-of-Flight (ToF) camera is sometimes distorted when ToF camera captures a depth of some objects closely. In this case, the ToF camera captures depth of objects differently, although objects are located at same distance from camera. This problem causes errors or low quality in several applications using ToF camera such as game (gesture recognition), 3D reconstruction and 3D display (multi-view generation). There are several reasons for this such as saturation, non-linearity and asymmetry of capacitances in pixel. In this paper, we will take saturation into account and propose a depth correction method for distortion.

8650-21, Session 6

Karate moves recognition from skeletal motion

Simone Bianco, Francesco Tisato, Univ. degli Studi di Milano-Bicocca (Italy)

The proposed works aims at automatically recognizing sequences of complex karate movements and giving a measure of the quality of the movements performed. Since this is a problem which intrinsically needs a 3D model, in this work we propose a solution taking as input sequences of skeletal motions that can derive from both motion capture hardware or consumer-level, off the shelf, depth sensing systems.

The proposed system is constituted by four different modules: skeleton representation, pose classification, temporal alignment, and scoring.

The proposed system is tested on a set of different punch, kick and defense karate moves executed starting from the simplest case, i.e. fixed static stances (heiko dachi) up to sequences in which the starting stances is different from the ending one.

The dataset has been recorded using a single Microsoft Kinect. The dataset includes the recordings of both male and female athletes with different skill levels, ranging from novices to masters.

The exemplar sequences were recorded from athletes of the World Champion Italian Karate National team. The dataset will be made freely available after the conference

8650-22, Session 6

Analyzing the relevance of shape descriptors in automated recognition of facial gestures in 3D images

Julian S. Rodriguez, Flavio Prieto, Univ. Nacional de Colombia (Colombia)

The present article shows the description and results emerging from analyzing the behavior of two shape descriptors (DESIRE and Spherical Spin Image) at recognizing facial gestures in 3D images. DESIRE is made of depth images, silhouettes and extended rays from the polygonal mesh; in contrast, Spherical Spin Image (SSI) is associated to a point in the polygonal mesh and it is a 2D histogram built from neighboring points using position information that captures the features in a local way. The database used contains images with 6 facial expressions out of which 3 were recognized in great percentage by the first descriptor implementing a neuronal mesh as classifier whereas the second descriptor only one of the facial expressions is recognized.

8650-23, Session 6

3D segmentation of the true and false lumens on CT aortic dissection images

Nawel Fetnaci, Pawel J. Lubniewski, Univ. d'Auvergne Clermont-Ferrand I (France); Bruno Miguel M.D., Univ. d'Auvergne Clermont-Ferrand I (France) and CHU Gabriel Montpied, Service de Chirurgie Cardio-Vasculaire, Pôle Cardiologie (France); Christophe Lohou, Univ. d'Auvergne Clermont-Ferrand I (France)

Our works are related on aortic dissections which are a medical emergency and can quickly lead to death. In this paper, we want to retrieve the false lumen and the true lumen which are aortic dissection features. Our aim is to provide a 3D view of the lumens (augmented reality) that we can't obtain directly by volume rendering or by another visualization tool which gives only the outer contour of the aorta. We chose to segment these data by using a deformable model based on the fast marching method.

In the classical fast marching approach, a speed function is used to control the front propagation of the deforming curve. The speed function is only based on the image gradient. In our CT images, due to the low resolution, the front propagates from a lumen to the other; therefore, the gradient data is insufficient to have accurate segmentation results.

In the paper, we have modified the fast marching method to segment the two lumens separately, by adapting the speed function.

We give our segmentation results [on the uploaded document] on an aortic dissection image with the classical fast marching method and with the adapted one with our new speed function.

8650-24, Session 7

Adaptive quality assurance of the product development process of additive manufacturing with modern 3D data evaluation methods

Julia W. Kroll, Sabine Botta, Jannis Breuninger, Alexander Verl, Fraunhofer-Institut für Produktionstechnik und Automatisierung (Germany)

In this abstract, the possibilities of modern 3D data evaluation for metrology and quality assurance are presented for the special application of the plastic laser sinter process, especially the Additive Manufacturing process. We use the advantages of the computer tomography and of the 3D focus variation to all stages of a production process for an increased quality of the resulting products. With the CT and the 3D focus variation the modern quality assurance and metrology have up to date instruments that allow non-destructive, complete and accurate measuring of parts. Therefore, these metrological methods can be used in many stages of the product development process for non-destructive quality control. In this work, studies and evaluation of 3D –data and the conclusions for relevant quality criteria are presented. Additionally, new developments and implementations for adapting the evaluation results for quality prediction, comparison and for correction are described to show how an adequate process control can be achieved with the help of modern 3D metrology techniques. The focus is on the optimization of laser sintering components with regard to their quality requirements so that the functionality during production can be guaranteed and quantified.

8650-25, Session 7

Quality assessment of adaptive 3D video streaming

Samira Tavakoli, Jesús Gutiérrez Sánchez, Univ. Politécnica de Madrid (Spain); Narciso García Santos , Univ. Politécnica de Madrid (Spain)

HTTP Adaptive Streaming (HAS) represents a great complement for many applications. Yet, current studies generally focus on technical enhancement of technology which can improve the robustness of IP video delivery over an unmanaged network. However, considering the user as the ultimate receiver, study about the Quality of Experience (QoE) of this technology is a crucial issue which has not been done extensively. To this aim, we discuss the impact of possible HAS client's behavior while adapting to the network capacity on end-user. This has been done through a subjective experiment of testing the end-user response to various clients' behavior for increasing, decreasing, and oscillation of quality in 3D video. Furthermore, the end-user's QoE of some of the HAS typical impairments during the adaptation has been evaluated. The results have shown us that the majority of the proposed adaption scenarios have efficiently provided a good perceptual quality to the end–user. The experimental conclusions have made good insight into the user's response to different adaption scenarios and visual impairments which can be used to develop the adaptive streaming algorithm to improve the QoE.

8650-26, Session 7

Evaluation methodology for structured light 3D based depth maps

Tom Osborne, Vikas Ramachandra, Kalin Atanassov, Sergio R. Goma, Qualcomm Inc. (United States)

A structured-light system for depth estimation is a type of 3D active sensor that consists of a structured-light projector, that projects an illumination pattern on the scene, and camera, which captures the illuminated scene. Based on the received patterns, depths of different scene regions can be inferred.

Here, we propose an evaluation framework for the depth maps obtained using different structured-light systems based on certain criteria described below.

a. Resolution: This test the smallest object size that can be detected at a given distance from the system

b. Object contour fidelity: This criterion measures the shape fidelity, especially of objects with thin contours, like human fingers.

c. Depth granularity: Measures the smallest detectable change in depth when an object moves.

d. Flatness of planes: When a planar object is placed in the field of view and parallel to the system, its depth map should appear flat. This may not be the case in reality, and is closely linked to the fact that the projector/ camera pair might not be well calibrated.

e. Sensitivity to human body factors including human body part detection.

f. Temporal consistency artifacts due to speed and post processing.

Conference 8651: Human Vision and Electronic Imaging XVIII

Monday - Thursday 4 –7 February 2013

Part of Proceedings of SPIE Vol. 8651 Human Vision and Electronic Imaging XVIII

8651-2, Session K1

Predicting visual memorability *(Keynote Presentation)*

Aude Oliva, Massachusetts Institute of Technology (United States)

When glancing at a magazine or browsing the Internet, we are continuously exposed to images. Despite this overflow of visual information, humans are extremely good at remembering thousands of pictures along with their visual details. But not all images are created equal. Whereas some images stick in our minds, others are ignored or quickly forgotten. What makes an image memorable? Our recent work shows that one can predict image memorability, opening a new domain of investigation at the interface between human cognition and computer vision.

8651-3, Session K1

World, environment, umwelt, and innerworld: a biological perspective on visual awareness *(Keynote Presentation)*

Jan J. Koenderink, Technische Univ. Delft (Netherlands)

The "world" is all physical reality (Higgs bosons, and so forth), the "environment" is a geographical locality (your city, …), the "Umwelt" is the totality of possible actions of the environment on the sensitive body surface of an agent (you, your dog, …) and the possible actions of the agent on the environment (mechanical, chemical, …), whereas the "innerworld" is what it is for the agent to be, that is awareness. Awareness is pre-personal, proto-conscious, and (perhaps) proto-rational. The various "worlds" described above are on distinct ontological levels. The world, and the environment are studied in the exact sciences, the Umwelt is studied by physiology and ethology. Ethology is like behavioristic psychology, with the difference that it applies to all animals. It skips the innerworld, e.g., it considers speech to be a movement of air molecules.The innerworld can only be known through first person reports, thus is intrinsically subjective. It can only be approached through "experimental phenomenology", which is based on intersubjectivity among humans. In this setting speech may mean something in addition to the movements of molecules. These views lead to a model of vision as an "optical user interface". It has consequences for many applications.

8651-4, Session K1

Does evolution favor true perceptions? *(Keynote Presentation)*

Donald D Hoffman, Univ. of California, Irvine (United States); Manish Singh, Rutgers University (United States); Justin Mark, Univ. of California, Irvine (United States)

Does natural selection favor veridical perceptions, those that more accurately depict the objective environment? Students of perception often claim that it does. But this claim, though influential, has not been adequately tested. In this talk I formalizethe claim and a few alternatives. To test them, I introduce"interface games," a class of evolutionary games in which perceptual strategies compete. I present the results of Monte Carlo simulations of some simpler games that assume frequency-dependent selection and complete mixing in infinite populations. These simulations show that veridical perceptions can bedriven to extinction by non-veridical strategies that are tuned to utility rather than objective reality. This suggests that naturalselection need not favor veridical perceptions, and that the effects of selection on sensory perception deserve further study.

8651-5, Session 1

Lightness perception in imaging and art *(Invited Paper)*

Alan L. Gilchrist, Rutgers, The State Univ. of New Jersey (United States)

The high dynamic range of most natural images, due to variations in both reflectance and illumination level, poses a challenge both for the visual system and for the artist. The visual system must assign a limited range (about 30:1) of reflectances (white: 90%, to black: 3%) to a range of image luminances that often far exceeds 1000:1. The painter must represent that large range of intensities using the same (30:1) low range of reflectances. While the visual system organizes these high range images into regions of high and low illumination, each comprising a limited range (< 30:1) the artist must compress the high range image into the limited range of pigment. When the painting is viewed, the observer's visual system then expands the compressed range. Measurements of perceived reflectance range have shown expansion in paintings and low-range abstract Mondrian patterns, but dramatic compression in very high range (>5000:1) Mondrians, revealing a tendency to normalize the perceived range towards a canonical white-to-black range. When a 3D canvas is used, as in a Patrick Hughes-type Reverspective, directional illumination can be exploited to extend the luminance range, and the observer can have a genuine experience of light-emitting self-luminosity.

8651-6, Session 1

Human lightness perception is guided by simple assumptions about shape and reflectance

Richard F. Murray, York Univ. (Canada)

Lightness constancy is the remarkable ability of human observers to perceive surface reflectance accurately despite variations in illumination and context. Two successful approaches to understanding lightness perception that have developed along independent paths are anchoring theory and Bayesian theories. Anchoring theory is a set of rules that predict lightness percepts under a wide range of conditions (Gilchrist, 2006). Some of these rules are counterintuitive, e.g., a rule that large surfaces tend to look lighter than small surfaces. Bayesian theories are formulated as probabilistic assumptions about lights and objects, and they model percepts as rational inferences from sensory data (e.g., Adelson, 2000). Here I reconcile these two seemingly divergent approaches by showing that many rules of anchoring theory follow from simple probabilistic assumptions about lighting and reflectance. I describe a simple Bayesian model that makes maximum a posteriori interpretations of luminance images, and I show that this model predicts many of the phenomena described by anchoring theory, e.g., anchoring to white, scale normalization, and articulation effects. Thus anchoring theory can be naturally formulated in a Bayesian framework, and this approach shows that many seemingly idiosyncratic properties of human lightness perception are actually rational consequences of simple assumptions about lighting and reflectance.

8651-7, Session 1

Spatial imaging in color and HDR: Prometheus unchained *(Invited Paper)*

John J. McCann, McCann Imaging (United States)

The Human Vision and Electronic Imaging Conference at EI has brought together research in the fundamentals of both vision and digital technology. This conference has included many color disciplines that have contributed to the theory and practice of today's imaging: theory of human vision, color constancy, models of vision, digital output, high dynamic range imaging, and the understanding of perceptual mechanisms. Before digital imaging, silver halide color was a pixel based mechanism. Color films are closely tied to colorimetry, the science of matching pixels in a black surround. The quanta catch of the sensitized silver salts determines the amount of colored dyes in the final print. The rapid expansion of digital imaging over the past 25 years has eliminated the limitations of using single pixels in forming images. Spatial interactions can now generate images more like vision. Since the 1950's, neurophysiology has shown that post-receptor neural processing is based on spatial interactions. These results reinforced the findings of 19th century experimental psychology. This paper reviews the role of HVEI in color, emphasizing the interaction of research on vision and the new algorithms and processes made possible by electronic imaging.

8651-8, Session 2

Presenting visual stimuli: past and present *(Invited Paper)*

Gerald Westheimer, Univ. of California, Berkeley (United States)

Candle, light bulb, Xenon arc, LED, computer monitor; Nicol prism and Polaroid; cobalt glass, interference filter and monochromatic laser light. In his 60 years as an experimental vision researcher the speaker has used them all, except the very first. But even that is covered in this talk on the influence of the state of technological development on the aspects of vision studied at any one time and the way it is studied, from Maxwell and Mach to the present.

8651-9, Session 2

Emerging technologies: 25 years *(Invited Paper)*

Hawley K. Rising III, Consultant (United States)

This paper will talk about the technologies that have been emerging over the 25 years since the Human Vision and Electronic Imaging conference began that the conference has been a part of, and that have been a part of the conference, and will look at those technologies that are emerging today, such as social networks, haptic technologies, and still emerging imaging technologies, and what we might look at for the future.

Twenty-five years is a long time, and it is not without difficulty that we remember what was emerging in the late 1980s. Yet to be developed: The first commercial digital still camera was not yet on the market, although there were hand held electronic cameras. Personal computers were not displaying standardized images, and image quality was not something that could be talked about in a standardized fashion, if only because image compression algorithms were not standardized yet for several years hence. Even further away were any standards for movie compression standards, there was no personal computer even on the horizon which could display them. What became an emergent technology and filled many sessions later, image comparison and search, was not possible, nor the current emerging technology of social networks -- the world wide web was still several years away. Printer technology was still devising dithers and image size manipulations which would consume many years, as would scanning technology, and image quality for both was a major issue for dithers and Fourier noise.

From these humble beginnings to the current moves that are changing

computing and the meaning of both electronic devices and human interaction with them, we will see a course through the changing technology that holds some features constant for many years, while others come and go.

8651-10, Session 2

What brain imaging technology can tell us about perception and consciousness *(Invited Paper)*

Lora T. Likova, The Smith-Kettlewell Eye Research Institute (United States)

The ultimate goal of electronic imaging is to provide effective presentation of visual information, and the HVEI Conference is the unique interface between electronic media and human perception/cognition. The inclusion of brain imaging is the latest manifestation of this linkage, which has bloomed into a theme for well-attended Special Sessions in the last several years. The topics have ranged from detailed mapping of primary visual cortex and subsequent cortical hierarchy to the neural circuitry underlying the perception of dynamic images, including those giving rise to the 3D percepts of depth and stereomotion, as well as to the issue of visual attention and visual consciousness. These fields have even extended to cross-modal brain plasticity and awareness in not only image encoding and retrieval, but in sensorimotor brain processing of motor control and human image reproduction (i.e., drawing). The progressive incorporation of the brain imaging field into our interdisciplinary community will be crucial for breakthrough advance to the next generation of electronic media.

8651-11, Session 2

Perceptual approaches to finding features in data *(Invited Paper)*

Bernice E. Rogowitz, Visual Perspectives Consulting (United States)

Electronic imaging applications hinge on the ability to discover features in data. For example, doctors examine diagnostic images for tumors, broken bones and changes in metabolic activity. Financial analysts explore visualizations of market data to find correlations, outliers and interaction effects. Seismologists look for signatures in geological data to tell them where to drill or where an earthquake may begin. These data are very diverse (images, numbers, graphs, 3-D graphics, text, etc.) and, mainly because of automatic data collection technologies such as sensors and digital imaging, are growing exponentially. This paper explores how the art and science of finding features in data have evolved over the past 25 years, and how research in human vision and cognition have driven this progress. We will discuss key contributions from the Conference on Human Vision and Electronic Imaging, covering key topics such as shape perception, interactive visualization, virtual reality and haptic interfaces, and will explore research trends that will drive the next generation of pattern-recognition technologies.

8651-12, Session 3

Is image quality a function of contrast perception? *(Invited Paper)*

Andrew M. Haun, Eli Peli, Schepens Eye Research Institute (United States)

In this retrospective we trace the development of image quality measures based on models of the early stages of the human visual system (HVS). Luminance contrast is the fundamental attribute of natural images, and understanding its natural and psychophysical properties has been important in developing numerous digital image technologies. While the SPIE Human Vision and Electronic Imaging meeting has strived to find points of contact between the study of human psychophysics and the development of computer vision and

image quality algorithms, progress in the field has not always been made on these terms. It was assumed that by understanding the psychophysical properties of contrast perception, and by modeling these more and more precisely, models of image quality could be designed. Encoded images could be presented to a simulated HVS, which would then return a verdict on quality. Instead, the most successful image quality metrics today tend to include only highly abstracted HVS components. We conclude with a puzzle: why do different quality metrics with such disparate structures as the VSNR, VIF, and SSIM to name just a few, all work so well? What do they have in common, and what do they have to do with human spatial vision?

8651-13, Session 3

Visible contrast energy metrics for detection and discrimination *(Invited Paper)*

Albert J. Ahumada, NASA Ames Research Ctr. (United States)

Contrast energy was proposed by Watson, Robson, & Barlow (Science, 1982) as a useful metric for representing luminance contrast target stimuli because it represents the detectability of the stimulus in photon noise for an ideal observer. Like the eye, the ear is a complex transducer system, but relatively simple sound level meters are used to characterize sounds. These meters provide a range of frequency sensitivity functions and integration times depending on the intended use. We propose here the use of a range of contrast energy measures with different spatial frequency contrast sensitivity weightings, eccentricity sensitivity weightings, and temporal integration times. When detection threshold are plotting using such measures, the results show "what the eye sees best" when these variables are taken into account in a standard way. The suggested weighting functions revise the Standard Spatial Observer (Watson & Ahumada, J. Vision, 2005) for luminance contrast detection and extend it into the near periphery. Under the assumption that the detection is limited only by internal noise, discrimination performance can be predicted by metrics based on the visible energy of the difference images (Watson & Ahumada, J. Vision, 2008).

8651-14, Session 4

Visual image quality: coding *(Invited Paper)*

Sheila S. Hemami, Cornell Univ. (United States)

No Abstract Available

8651-15, Session 4

Initial spatio-temporal domain expansion of the Modelfest database

Thom Carney, Sahar Mozaffari, Sean Sun, Ryan Johnson, Sharona Shirvastava, Priscilla Shen, Emma Ly, Univ. of California, Berkeley (United States)

The first Modelfest group publication appeared in the SPIE Human Vision and Electronic Imaging conference proceedings in 1999. "One of the group's goals was to develop a public database of test images with threshold data from multiple laboratories for designing and testing Human Vision Models" The group collected and made public detection threshold data for static images to be used to test human vision models. In the same vein we have begun extending the database into the temporal domain with an expanded stimulus set of 41 video clips, partially overlapping with the original stimuli but spanning a broad range of temporal frequencies. Average detection thresholds for seven subjects are presented along with historical values from the literature. Like the original Modelfest dataset, the participation of multiple laboratories is needed ensure the reliability of the results and to compare methods across laboratories. We look forward to other groups joining us in this endeavor.

8651-16, Session 4

A database of local masking thresholds in natural images

Md. Mushfiqul Alam, Kedarnath P. Vilankar, Damon M. Chandler, Oklahoma State Univ. (United States)

A database of local contrast detection thresholds of fifteen natural images was presented. Currently, we are extending the database by experimenting on additional fifteen natural images. Through quantitative measures, we proved the intra-subject and inter-subject consistency of our experimental procedure. We also analyzed the relationship between some local low-level mask features and corresponding thresholds. We found that, except sharpness, individual features have weak relationship with contrast thresholds. Our next step is to study, whether the combined low-level features have strong relationship with the detection thresholds. We also analyzed the performance of Watson and Solomon's neural masking model and two image-quality assessment algorithms - MAD and MS-SSIM in predicting the local contrast detection thresholds. Among the three models, Watson and Solomon's neural model performed the best in predicting the thresholds. Nevertheless, the predictions are not quite close to the ground-truth thresholds. These results suggest a wide-scope of improving the neural masking models in predicting local contrast detection thresholds in natural images. Our database provides the primary data for this modeling. Currently, we are designing a neural masking model for natural images with the consideration of multiple low-level mask features. We intend to report the outcomes of our study in the full-paper.

8651-17, Session 4

Interplay between image coding and quality estimation

Guilherme O. Pinto, Sheila S. Hemami, Cornell Univ. (United States)

In image processing, image quality models are useful in several areas, such as in image quality estimation and in image coding. In the former, the image models are used to estimate the perceived quality of images that have undergone different types of distortions, including but not limited to white noise, transmission errors, lossy-compression and blur. In the latter, the image models are used to achieve a perceptually optimal bit-allocation across the different subbands and code-blocks. In the area of image quality estimation, several algorithms, such as SSIM, VIF, MAD and VSNR, have been proposed in order to overcome the known limitations of the mean squared error (MSE) distortion metric. In addition, a recent work in image quality and utility estimation has also indicated that the MSE quality estimator presents a poor performance in the Low-Quality Regime (LQR), while other quality and utility estimators are able to perform well. Although there have been some significant developments in the field of quality estimation, the JPEG-2000 standard still uses the MSE distortion metric. One possible reason for this lies in the fact that many of the state-of-the-art quality and utility estimation algorithms were developed after the development of JPEG-2000 image coding standard. This because many of the state-of-the-art quality estimation algorithms were developed after 2000, which corresponds to the year of development of JPEG-2000. The final paper will study the use of different image quality and utility estimators, such as MS-SSIM, VIF and MS-NICE, in image coding applications. % For a given coding pass and estimator, it is possible to compute the resulting distortion in two different ways. The distortion can be computed for a given code-block and subband in the wavelet domain, or the same estimator can be used to compute the resulting distortion in the coded image, in the pixel domain. Considering different quality and utility estimators, the final paper will compare the performances of these two different approaches. The preliminary results given in this abstract focus on the former approach.

Conference 8651

8651-18, Session 5

From image quality to atmosphere experience: how evolutions in technology impact experience assessment *(Invited Paper)*

Ingrid Heynderickx, Philips Research Nederland B.V. (Netherlands) and Technische Univ. Delft (Netherlands); Huib de Ridder, Technische Univ. Delft (Netherlands)

Image quality is a concept that for long very well served to optimize display performance and signal quality. New technological developments, however, forced the community to look into higher level concepts to capture the full experience. Terms as naturalness and viewing experience were used to optimize the full experience of 3D-displays and Ambilight TV. These higher level concepts capture differences in image quality and differences in perceived depth, or perceived viewing field. With the introduction of solid state lighting, further enhancing the multimedia experience, yet more advanced quality evaluation concepts will be needed in the future to optimize the overall experience.

8651-19, Session 5

Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance

Scott J Daly, Dolby Labs Inc (United States); Timo Kunkel, Xing Sun, Suzanne Farrell, Poppy Crum, Dolby Labs., Inc. (United States)

No Abstract Available

8651-20, Session 5

Quantifying image quality in graphics: perspective on subjective, objective metrics, and their performance in computer graphics applications

Rafal Mantiuk, Bangor Univ. (United Kingdom)

We explore three problems related to quality assessment in computer graphics: the design of efficient user studies; the scene-referred metrics for comparing high-dynamic-range images; and the comparison of metric performance for the database of computer graphics distortions. This paper summarizes the most important observations from investigation of these problems and gives a high level perspective on the problem of quality assessment in graphics.

8651-21, Session 5

Visualizing lighting with images: converging between the predictive value of renderings and photographs

Ulrich Engelke, Mariska G. M. Stokkermans, Philips Research (Netherlands); Michael J. Murdoch, Philips Research Nederland B.V. (Netherlands)

Lighting systems are usually designed based on real-world prototypes, which are expensive and time consuming to build. Computer-based renderings are typically used to circumvent these shortcomings. The success of a virtual prototype depends on its ability to predict human perception of the real-world. Previously we showed that renderings predict a real lighting system reasonably well. Still, a margin of prediction error remained even after identifying the most suitable rendering pipeline. This could be partly due to presenting a 3D real-world scene on a 2D display, thus, losing one dimension that might be instrumental in judging lighting perception.

This work aims at better understanding the upper limit of prediction accuracy that can be achieved with an image compared to the real-world. We consider photographs as a ground truth because of their geometrical accuracy. We performed psychophysical experiments in which people rated the lighting of a real office room as well as of photographs and renderings of that room. The outcomes reveal that an error margin remains also between the photographs and the real-world. However, the photographs generally predict lighting perception superior to the renderings. The results are utilized to further improve the renderings to converge towards the predictive value of photographs.

8651-22, Session 6

A survey on 3D quality of experience and 3D quality assessment

Anush K. Moorthy, Texas Instruments Inc. (United States); Alan C. Bovik, The Univ. of Texas at Austin (United States)

This article summarizes recent advances in algorithmically evaluating the quality of stereoscopic presentations. We undertake a brief review of research in the area of 3D quality of experience which encompasses (1) ge- ometry, visual discomfort etc., (2) display issues such as cross-talk and (3) quality assessment as previously defined. We describe algorithms and databases that have been proposed in the literature and analyze the merits and demerits of the proposed approaches. We begin by summarizing what has been done in the area of visual discomfort and geometry when viewing 3D presentations. We explain the state-of-the-art of 3D quality assess- ment, although the field is still nascent and the topic so diverse that our survey is by no means comprehensive. Rather we provide a birds-eye view of current prevailing thought on the topic. We summarize databases that have recently been proposed for 3D quality assessment. Finally, we describe a recent, extremely useful tool for researchers in the area of 3D image quality assessment (IQA) – the LIVE 3D IQA database. The LIVE 3D IQA database is the only database of its kind that not only incorporates reference and distorted images, but also true depth information captured using a high-precision laser range scanner. The LIVE 3D IQA database is freely available to researchers to enable rapid progress in 3D IQA research.

8651-23, Session 6

Visual quality beyond artifact visibility

Judith A. Redi, Technische Univ. Delft (Netherlands)

In the last three decades a lot of effort has been devoted to the development of technologies that can predict the visual quality of images and videos, as a basis for the delivery of optimal visual quality to the user. These systems have been based for the most part on a visibility-centric approach, assuming that high artifact visibility corresponds to high annoyance and, therefore, to low visual quality. Despite the remarkable results achieved with this approach, recently a number of studies suggested that the visibility-centric approach to visual quality might have limitations, and that other factors might influence the overall quality impression of an image or video, depending on cognitive and affective mechanisms that work on top of perception. In particular, interest in the visual content, engagement and aesthetic appeal have been found to positively impact on the overall quality impression of the image/video. In this paper, we review these studies and explore the impact that affective and cognitive processes have on the visual quality. In addition, as a case study, we will present the results of an experiment investigating on the impact of aesthetic appeal on visual quality.

8651-24, Session 6

Subjective matters: from image quality to image psychology *(Invited Paper)*

Elena A. Fedorovskaya, RadixNova (United States); Huib de Ridder, Technische Univ. Delft (Netherlands)

Conference 8651

From the advent of digital imaging through several decades of studies, human vision research community systematically focused on perceived image quality and digital artifacts due to resolution, compression, gamma, dynamic range, capture and reproduction noise, blur, etc., to help overcome existing technological challenges and shortcomings. Technological advances made digital images and digital multimedia nearly flawless in quality, and ubiquitous and pervasive in usage, providing us with the exciting possibility and at the same time demanding to turn to the domain of human experience including higher psychological functions, such as cognition, emotion, awareness, social interaction, consciousness and Self.

In the paper we will outline the evolution of human centered multidisciplinary studies related to imaging as we see it and propose steps and potential foci of future research.

8651-25, Session 7

The rough side of texture: texture analysis through the lens of HVEI *(Invited Paper)*

Thrasyvoulos N. Pappas, Northwestern Univ. (United States)

No Abstract Available

8651-26, Session 7

Adapting environments to observers *(Invited Paper)*

Michael A. Webster, Univ. of Nevada, Reno (United States)

Visual coding continually adapts to match the current environment. Yet the extent to which this adaptation affects perception and performance remains uncertain, in part because adaptation is typically studied only over brief timescales. We explored potential consequences of very long-term adaptation to color in different environments, by adjusting mages to simulate their appearance under theoretically complete adaptation. The simulation is based on plausible properties of color mechanisms and how they would adapt to natural color distributions, with images rendered based on the adapted responses. The simulated images allow us to explore a number of questions about the consequences and functions of adaptation that are difficult to study by instead adapting the observer. These include how much color appearance is predicted to vary when the same observer is placed in different environments, or when different observers are exposed to a common environment. Measures of visual search also reveal the extent to which complete adaptation could enhance sensitivity and salience. Pre-adapting images could especially enhance information in uncommon environments. The method thus provides a novel theoretical approach to understanding adaptation at long timescales and a novel technique for processing images to facilitate how well observers can interact with them.

8651-27, Session 8

Efficient image representations and features *(Invited Paper)*

Erhardt Barth, Univ. zu Lübeck (Germany); Michael Dorr, Harvard Medical School (United States); Eleonora Vig, Harvard Univ. (United States)

Interdisciplinary research in human vision and electronic imaging has

greatly contributed to the current state of the art in imaging technologies. Image compression and image quality are prominent examples and the progress made in these areas relies on a better understanding of what natural images are and how they are perceived by the human visual system. A key research question has been: given the (statistical) properties of natural images, what are the most efficient and perceptually relevant image representations, what are the most prominent and descriptive features of images and videos?

We give an overview of how these topics have evolved over the 25

years of HVEI conferences and how they have influenced the current state of the art. There are a number of striking parallels between human vision and electronic imaging. The retina does lateral inhibition, one of the early coders was using a Laplacian pyramid; primary visual cortical areas have orientation- and frequency-selective neurons, the current JPEG standard defines similar wavelet transforms; the brain uses a sparse code, engineers are currently excited about sparse coding and compressed sensing. Some of this has indeed happened at the HVEI conferences and we would like to distill that.

8651-28, Session 8

Highly overcomplete sparse coding

Bruno A. Olshausen, Helen Wills, Univ. of California, Berkeley (United States)

No Abstract Available

8651-29, Session 8

Blind image quality assessment without training on human opinion scores

Anish Mittal, Rajiv Soundararajan, Gautam S. Muralidhar, Alan C. Bovik, Joydeep Ghosh, The Univ. of Texas at Austin (United States)

We propose a family of image quality assessment (IQA) models based on natural scene statistics (NSS), that can predict the subjective quality of a distorted image without reference to a corresponding distortionless image, and without any training results on human opinion scores of distorted images. These 'completely blind' models compete well with standard non-blind image quality indices in terms of subjective predictive performance when tested on the large publicly available 'LIVE' Image Quality database

8651-46, Session IPI

Picture perception and visual field

Andrea J. van Doorn, Huib de Ridder, Technische Univ. Delft (Netherlands); Jan J. Koenderink, Technische Univ. Delft (Netherlands) and Katholieke Univ. Leuven (Belgium)

Looking at a picture fills part of the visual field. In the case of straight photographs there is a notion of the "Field of View" of the camera at the time of exposure. Is there a corresponding notion for the perception of the picture? In most cases the part of the visual field (as measured in degrees) filled by the picture will be quite different from the field of view of the camera. The case of works of arts is even more complicated, there need not even exist a well defined central view point. With several examples we show that there is essentially no notion of a corresponding "field of view" in pictorial perception. This is even the case for drawings in conventional linear perspective. Apparently the "mental eye" of the viewer is often unrelated to the geometry of the camera (or perspective center used in drawing). Observers often substitute templates instead of attempting an analysis of perspective.

8651-47, Session IPI

Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation

luminance

Kil Joong Kim, Seoul National Univ. (Korea, Republic of); Rafal Mantiuk, Bangor Univ. (United Kingdom); Kyoung Ho Lee, Seoul National Univ. (Korea, Republic of) and Bundang Hospital (Korea, Republic of)

Introduction: The contrast sensitivity functions (CSFs) are one of the

main components of visual models and metrics. Most models of CSFs were developed based on the data from numerous historical psychophysical exeperiments, which often differed in their methods, stimuli and viewing conditions. Also, the such data is usually available for very limited range of spatial frequencies and luminance adaptation levels. Very few measurements were done for the chromatic CSFs and those measurements were measured for the sensitivity to equiluminant color contrast rather than luminance contrast. For those reasons, we conducted a series of experiments to measure achromatic and chromatic CSFs at different levels of adaptation luminance, ranging from 0.02 to 200 cd/m2.

Experiment setting: The experiment design was inspired by the ModelFest and ColorFest data sets. The stimuli were shown on a 24" LCD display with a 10-bit panel and RGB LED backlight. Stimuli were observed from a fixed distance of 93 cm, which gave an angular resolution of 60 pixels per visual degree. The display was calibrated using a spectro-radiometer. The display white point was fixed at D65. The stimuli were generated for different color directions, frequencies, and levels of adaptation luminance.

Experiment procedure: Five observers participated. For each stimulus, each observer found the threshold using the QUEST procedure with 25 trials. The experiment involved a 4-alternative-forced-choice method in which an observer was asked to choose one of the four stimuli, of which only one contained the pattern. The stimuli were shown side-by-side on the same screen and the presentation time was not limited.

Stimuli: The stimuli consisted of vertical sine-gratings attenuated by the Gaussian envelope with sigma of 1.5 deg. The stimuli included three color directions, black-white, red-green, and blue-yellow. The frequency range of the sine gratings varied from 0.125 to 16 cycles per degree (cpd). The background luminance varied from 0.02 to 200 cd/m2. The luminance levels below 10 cd/m2 were achieved by wearing modified welding goggles in which the protective glass was replaced with neutral density filters. Stimuli were observed with a natural pupil.

Results: The results were tabulated in the figures below. The data will be discussed in the full paper if it is accepted.

8651-48, Session IPI

Viewer preferences for adaptive playout

Sachin G. Deshpande, Sharp Labs. of America, Inc. (United States)

Adaptive media playout techniques are used to avoid buffer underflow in a dynamic streaming environment where the available bandwidth may be fluctuating. In this paper we report human perceptions from audio quality studies that we performed on speech and music samples for adaptive audio playout. Test methods from ITU-R BS. 1534-1 recommendation were used. Studies were conducted for both slow playout and fast playout. Two scales - a coarse scale and a finer scale was used for the slow and fast audio playout factors. Results from our study can be used to determine acceptable slow and fast playout factors for speech and music content. An adaptive media playout algorithm could use knowledge of these upper and lower bounds to decide its playback schedule.

8651-49, Session IPI

The effect of familiarity on perceived interestingness of images

Elena A. Fedorovskaya, RadixNova (United States); Sharon Lynn Chu, Virginia Polytechnic Institute and State Univ. (United States); Jeffrey Snyder, Kodak Research Labs. (United States); Francis Quek, Virginia Polytechnic Institute and State Univ. (United States)

This paper addresses the individualized content adaptation of media for image experience and consumption. We investigate factors by which images/media may be personalized with respect to interestingness, considering how two dimensions of familiarity: facial familiarity and familiarity with image context or setting, relates to interestingness. Through a sequence of experiments, we looked at 1. How manipulation of facial similarity through morphing may influence the sense of familiarity with an image; 2. Whether varying the source face in the morphing procedure towards a target face affects perceived visual similarity of the morph result with the target face; and, 3. How the viewer's familiarity with the person and context in photographs can influence their interest in the photo. We posited that by creating personal meaning, familiarity can cause images to be perceived as interesting. Four levels of person familiarity varying in degree of person knowledge, and two levels of context familiarity varying in frequency of exposure, were considered: Self, Friend, Celebrity, and Stranger in Familiar and Unfamiliar contexts. Person familiarity was obtained through face morphing. Experimental results showed interesting significant main effects of context and person familiarity. Our findings deepen our understanding of the critical element of familiarity and its relation to interestingness in the context of images, and can impact the design of image-related systems. We discuss how our findings may be applied in media adaptations to influence interestingness to cultural groups and individuals.

8651-51, Session IPI

Quantifying patterns of dynamics in eye movement to measure goodness in organization of design elements in interior architecture

Hasti Mirkia, Arash Sangari, Mark Nelson, Amir H. Assadi, Univ. of Wisconsin-Madison (United States)

Architecture brings together diverse elements to enhance the observer's measure of esthetics and the convenience of functionality. Architects often conceptualize synthesis of design elements to invoke the observer's sense of harmony and positive affect. How does an observer's brain respond to harmony of design in interior spaces? One implicit consideration by architects is the role of guided visual attention by observers while navigating indoors. Prior visual experience of natural scenes provides the perceptual basis for Gestalt of design elements. In contrast, Gestalt of organization in design varies according to the architect's decision. We outline a quantitative theory to measure the success in utilizing the observer's psychological factors to achieve the desired positive affect. We outline a unified framework for perception of geometry and motion in interior spaces, which integrates affective and cognitive aspects of human vision in the context of anthropocentric interior design. The affective criteria are derived from contemporary theories of interior design. Our contribution is to demonstrate that the neural computations in an observer's eye movement could be used to elucidate harmony in perception of form, space and motion, thus a measure of goodness of interior design. Through mathematical modeling, we argue the plausibility of the relevant hypotheses.

8651-52, Session IPI

Development of a human vision simulation camera and its application: implementation of specific color perception

Hiroshi Okumura, Shoichiro Takubo, Shoichi Ozaki, Takeru Kawasaki, Indra N. Abdullah, Kohei Arai, Saga Univ. (Japan); Osamu Fukuda, National Institute of Advanced Industrial Science and Technology (Japan)

Human eye has a lot of photoreceptor cells in its retina. Human photoreceptor cells consist of "cone" type and "rod" type.

Cone cells that are responsible for color vision are concentrated at the macula and rod cells are concentrated arround the macula and are used in peripheral vision. More sensitive than cone cells, rod cells are almost entirely responsible for scotopic vision. Purkinje effect for mesopic and scotopic vision and adaptation are caused by uneven distribution and sensitivity of photoreceptor cells. In this study, HuVisCam, a human vision simulation camera, that can simulate not only Purkinje effect for mesopic and scotopic vision but also dark and light adaptation, abnormal miosis and abnormal mydriasis caused by the influence of mydriasis medicine or nerve agent is developed. In this article, improvement of HuVisCam for specific color perception is discussed.

For persons with normal color perception, simulation function of various types of specific color perception is provided.

In addition, for persons with specific color perception, color information picker is also provided.

8651-53, Session IPI

IMF-based chaotic characterization of AP and ML visually-driven postural responses

Hanif Azhar, Guillaume Giraudet, Jocelyn Faubert, Univ. de Montréal (Canada)

The objective was to analyze visually driven postural responses and characterize any nonlinear behaviour. We recorded physiological responses for two adults, 260 trials each. The subjects maintained quite stance while fixating for four seconds within an immersive room, EON Icube™ [9], where the reference to the visual stimuli, i.e., the virtual platform, randomly oscillated in Gaussian orientation 900 and 2700 for antero-posterior (AP), and, 00 and 1800 for medio-lateral (ML) at three different frequencies (0.125, 0.25, and 0.5 Hz). We accomplished stationary derivatives of posture time series by taking the intrinsic mode functions (IMFs). The phase space plot of IMF shows evidence of the existence of nonlinear attractors in both ML and AP. Correlation integral slope with increasing embedding dimension is similar to random white noise for ML, and similar to nonlinear chaotic series for AP. Next, recurrence plots indicate the existence of more non-linearity for AP than that for ML. The patterns of the dots after 200th time stamp (near onset) appears to be aperiodic in AP. At higher temporal windows, AP entropy tends more toward chaotic series, than that of ML. There are stronger non-linear components in AP than that in ML regardless of the speed conditions.

8651-54, Session IPI

Application of imaging technology for archaeology researches: framework design for connectivity analysis in pieces of Jomon pottery

Kimiyoshi Miyata, National Museum of Japanese History (Japan); Ryota Yajima, Kenichi Kobayashi, Chuo Univ. (Japan)

Jomon pottery is one kind of earthenware produced in Jomon period in Japan. Potteries are found by the excavations in archaeological sites, however their original whole shapes have been dismissed because those are broken and separated into small pieces. In the investigation process, reproduction of the whole shape of the potteries is an important and difficult task because there are a lot of pieces and the number of combinations among the pieces is huge.

In this paper, a framework for the application of the imaging technology is discussed at first, then connectivity analysis among the pieces of Jomon potteries is focused on to reduce the number of the trial and error to find connectable combinations of the pieces. The real pieces are chosen and taken by a digital camera, and each piece in the image is labeled to calculate the statistical information including histograms used as features in the analysis of the connectivity. An index showing the connectivity of the pieces is defined and calculated to indicate the connectivity as a matrix showing possibility of all of combinations among the pieces. Finally, validity of calculated index is confirmed by the archaeologist as an experimental result.

8651-55, Session IPI

Top-down visual search in Wimmelbild

Julia Bergbauer, Technische Univ. München (Germany); Sibel Tari, Middle East Technical Univ. (Turkey)

Wimmelbild which means "teeming figure picture" is a popular genre of visual puzzles. Abundant masses of small figures are brought together in complex arrangements to make one scene in a Wimmelbild. It is

picture hunt game. As such, it has the potential to be an excellent test bed for top-down visual search algorithms.

Figure hunting games starting with early Gestalt have played a noticeable role in visual perception studies. Today, a quick search in internet retrieves plenty of children-oriented samples of the genre of Wimmelbild as well as striking samples of camouflage art, forming a wider genre of camouflaged objects pictures. These pictures vary in style, but hardly any of them seem amenable to purely bottom-up analysis. For some of them, it may be possible that processing starts bottom-up, delivering initial groupings as well as locations of salience. For some deceptively simple black and white drawings, however, it appears to us that the process is highly likely to start top-down.

Here, we address the problem of locating a given form in a Wimmelbild. A key point is that we resort to field models of shape representation to have smooth well-conditioned dependence of the goodness-of-fit to pose and scale parameters.

8651-56, Session IPI

Visual discrimination and adaptation using non-linear unsupervised learning

Sandra Jiménez, Valero V Laparra, Jesus Malo Lopez, Univ. de València (Spain)

Understanding human vision not only involves empirical descriptions of how it works, but also organization principles that explain why it does so (Barlow 1961). Identifying the guiding principles of visual phenomena requires learning algorithms to optimize specific goals. Moreover, these algorithms have to be flexible enough to account for the non-linear and adaptive behavior of the system. For instance, linear redundancy reduction transforms certainly explain a wide range of visual phenomena (Buchsbaum and Gottschalk 1983; Atick et al. 1992, 1993; Olshausen and Field 1996; Bell and Sejnowski 1997; Hoyer and Hyvarinen 2000; Simoncelli and Olshausen 2001; Doi et al. 2003). However, the generality of this organization principle is still in question (Barlow 2001): it is not only that and additional constraints such as energy cost may be relevant as well (Laughlin 2004), but also, statistical independence may not be the better solution to make optimal inferences in squared error terms (MacLeod 2003; Simoncelli 2009; Laparra et al. 2012a). Moreover, linear methods cannot account for the non-uniform discrimination in different regions of the image and color space: linear learning methods necessarily disregard the non-linear nature of the system. Therefore, in order to account for the non-linear behavior, principled approaches commonly apply the trick of using (already non-linear) parametric expressions taken from empirical models (Schwartz and Simoncelli 2001; Kayser et al. 2003; Lyu 2011). Therefore these approaches are not actually explaining the non-linear behavior, but just fitting it to image statistics. In summary, a proper explanation of the behavior of the system requires flexible unsupervised learning algorithms that (1) are tunable to different, perceptually meaningful, goals; and (2) make no assumption on the non-linearity. Over the last years we have worked on these kind of learning algorithms based on nonlinear ICA (Malo and Guti´errez 2006), Gaussianization (Laparra et al. 2011), and principal curves (Laparra et al. 2012a,b). In this work we stress the fact that these methods can be tuned to optimize different design strategies, namely statistical independence, error minimization under quantization, and error minimization under truncation. Then, we show (1) how to apply these techniques to explain a number of visual phenomena, and (2) suggest the underlying organization principle in each case.

8651-57, Session IPI

Chromatic induction and contrast masking: similar models, different goals?

Sandra Jiménez, Univ. de València (Spain); Xavier Otazu, Univ. de Barcelona (Spain); Valero V Laparra, Jesus Malo Lopez, Univ. de València (Spain)

Normalization of signals coming from linear sensors is an ubiquitous mechanism of neural adaptation (Carandini and Heeger 2012).

Local interaction between sensors tuned to a particular feature at certain spatial position and neighbor sensors explains a wide range of psychophysical facts including (1) masking of spatial patterns (Watson and Solomon 1997), (2) non-linearities of motion sensors (Simoncelli and Heeger 1998), (3) adaptation of color perception (Hillis and Brainard 2005), (4) brightness and chromatic induction (Otazu et al. 2008, 2010), and (5) image quality assessment (Laparra et al. 2010). However, the formal and the qualitative similarities of the local interaction in the above models do not necessarily mean that the mechanisms involved are pursuing the same statistical goal. For instance, in the case of chromatic mechanisms (disregarding spatial information), different parameters in the normalization give rise to optimal discrimination or adaptation (Abrams et al. 2007), and different non-linearities may give rise to error minimization or component independence (Laparra et al. 2012). In the case of spatial sensors (disregarding color information), a number of studies have pointed out the benefits of masking in statistical independence terms (Schwartz and Simoncelli 2001; Malo and Guti´errez 2006; Malo and Laparra 2010; Lyu 2011). However, such statistical analysis has not been performed for spatio-chromatic induction models where chromatic perception depends on spatial configuration. In this work we investigate whether a successful spatio-chromatic induction models (Otazu et al. 2010), increases component independence as previously reported for masking models. Mutual information analysis suggests that despite the formal similarity of the models, their statistical role may be different.

8651-58, Session IPI

Aesthetics and entropy II: a critical examination

Melville R Sahyun, Consultant (United States)

No Abstract Available

8651-30, Session 10

The evolution of attention research: exploring the interplay of bottom-up and top-down processes *(Invited Paper)*

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

No Abstract Available

8651-31, Session 10

Saliency identified by absence of background structure

Fred W. Stentiford, Univ. College London (United Kingdom)

Visual attention is commonly modelled by attempting to characterise objects using features that make them special or in some way distinctive in a scene. These approaches have the disadvantage that it is never certain what features will be relevant in an object that has not been seen before. This paper provides a brief outline of the approaches to modeling human visual attention together with some of the problems that they face. A graphical representation for image similarity is described that relies on the size of maximally associative structures (cliques) that are found to be reflected in pairs of images. While comparing an image with itself, the similarity mechanism is shown to model pop-out effects when constraints are placed on the physical separation of pixels that correspond to nodes in the maximal cliques. Background regions are found to contain structure in common that is not present in the salient regions which are thereby identified by its absence. The approach is illustrated with figures that exemplify asymmetry in pop-out, the conjunction of features, orientation disturbances and the application to natural images.

8651-32, Session 10

Investigation of eye-catching colors using eye tracking

Mokryun Baik, Hyeon-Jeong Suk, Jeongmin Lee, Kyungah Choi, KAIST (Korea, Republic of)

An eye tracking experiment was carried out in order to investigate the relationship between eye gazing movements and the characteristics of color attributes. The experiment focused on identifying the color attributes that are decisive for eye-catching efficiency. The icon array used in the experiment to design stimuli was adopted with consideration to the context of recent communication media, such as smart phones or smart TVs. Each stimulus set was composed of 25 color square patches arrayed in the format of a 5 by 5 grid. The experiment was divided into three parts to focus on a specific attribute of a color one at a time, while controlling its other attributes: In Part I, hue difference was examined, and thus each stimulus set contained 25 hues while the tone remained controlled. It was revealed that subjects were more attentive to warm colors than to cool colors, particularly when warm colors were arranged along the horizontal and vertical axes; In Part II, the experiment dealt with tone difference, and thus 25 tone variations within red or blue were provided as a stimulus set. However, the result indicated that changes in tone does not have a significant influence on subjects' initial attention; Lastly, in Part III, whether color combinations have an influence on participants' attention in a manner different from that of single colors was examined. In the stimulus set, 25 icons were arranged. Among them, icons with complementary color combinations gained the greatest attention. In the experiment, fifteen college students were recruited to participate in all three parts. Throughout the experiments, either black or white were applied as background, but the contrast effect between foreground and background color did not have a noticeable influence on one's attention. Based on the empirical results, this paper provides a guideline for an eye-catching color composition and discusses its applications.

8651-33, Session 10

Can skill be determined from a photographic portfolio?

Abhishek Agrawal, Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore); Rajesh Somavarapu, The Univ. of Texas at Dallas (United States)

We plan to study and understand the role that photographic skill has to play in deciding image composition and aesthetics. Few scientific studies have been carried out in the past in this domain and even fewer to address the supposition that skill is apparent in photography, even without considering the quality of the instrument used to take the image. We conduct and obtain results from an experiment in which 33 subjects used identical equipment to photograph 7 pre-determined commonplace scenes such as a portrait, zebra crossing, building corner etc. We explore and demonstrate the use of Amazon Mechanical Turk platform for the collection of scientific data and obtain ratings on our images from a large number (bet. approx. 150-200) of judges. We also use this platform to discern and categorize the judges by obtaining information on their skill, knowledge and prior experience in photography. We improve the experimental design used in the previous studies and compare our findings. Our results demonstrate the role that photographic skill has to play in deciding composition in perception based imaging, and how it relates to judges' ratings of the images to a statistically significant level.

8651-34, Session 11

Binocular eye movements in health and disease *(Invited Paper)*

Christopher W. Tyler, The Smith-Kettlewell Eye Research Institute (United States)

Recording binocular eye movements for stereoscopic targets can be challenging because the range of useful disparities is only a few degrees and the range within stereograms is even narrower. This issue will be addressed in the context of the recent papers at SPIE on binocular control and stereoscopic image quality in 3D displays. With careful calibration it is possible to record binocular (vergence) eye movements to an accuracy of a few arc minutes. I here report the results of the first quantitative survey of the dynamics of vergence eye movements in a full-spectrum non-academic population to purely stereoscopic, purely accommodative and full-cue stimulation. The results show a wide variety of vergence behaviors that need to be taken into account in the design of stereoscopic displays. Although optimal vergence movements may be accomplished in as little as 300 ms, a majority of the population have slow vergence movements in at least one direction of vergence, taking over 500 ms to reach the target disparity (usually for divergence), or other deviations from optimal vergence capabilities. For larger vergence movements, the eyes may take as much as 2 s to reach the target disparity, even under full cue conditions.

A substantial proportion of the population has suffered from an incident of diffuse traumatic brain injury (dTBI; an impact to the head resulting in loss of consciousness but no evidence of tissue damage), which in turn can often result in impairment of binocular vergence control. At up to 2 million incidents per year in the US, dTBI may affect as many as 20% of the population, of which 50% or more may show vergence control problems. We find that many dTBI sufferers exhibit no vergence at all to disparity change stimuli, despite having normal stereopsis.

It is important to take this range of vergence capabilities into account in the design of stereoscopic 3D displays and film products.

8651-35, Session 11

Reflexive and voluntary control of smooth eye movements

Jeffrey B. Mulligan, NASA Ames Research Ctr. (United States); Scott B. Stevenson, Univ. of Houston (United States)

An understanding of visually evoked smooth eye movements is required to predict the visibility and legibility of moving displays, such as might be encountered in vehicles like aircraft and automobiles. We have studied the response of the oculomotor system to various classes of visual stimuli, and analyzed the results separately for horizontal and vertical version (in which the two eyes move together), and horizontal and vertical vergence (where they move in opposite directions). Of the four types of motion, only vertical vergence cannot be performed under voluntary control, and we found that certain stimuli (all having relatively long latencies) were incapable of evoking it. In another experiment, we instructed observers to track one of two targets, and measured weak but reliable responses to the unattended target, in which the long-latency component of the response is abolished. Our results are consistent with a system containing two distinct processes, a fast reflexive process which responds to a restricted class of stimuli, and a slower voluntary process capable of following anything that can be seen, but incapable of controlling vertical vergence.

8651-36, Session 11

Simple gaze-contingent cues guide eye movements in a realistic driving simulator

Laura Pomarjanschi, Univ. zu Lübeck (Germany); Michael Dorr, Peter J. Bex, Schepens Eye Research Institute (United States); Erhardt Barth, Univ. zu Lübeck (Germany)

Looking at the right place at the right time is a critical component of driving skill. Therefore, gaze guidance has the potential to become a valuable driving assistance system. In previous work, we have already shown that complex

gaze-contingent stimuli can guide attention and reduce the number of accidents in a simple driving simulator. We here set out to investigate whether cues that are simple enough to be implemented in a real car can also capture gaze during a more realistic driving task in a high-fidelity driving simulator. This immediately raises another question, namely how such cues would interfere with the driving task itself.

We used a state-of-the-art, wide-field-of-view driving simulator with an integrated eye tracker. Gaze-contingent warnings were implemented using two arrays of light-emitting diodes horizontally fitted below and above the simulated windshield. Twelve volunteers drove along predetermined routes in the simulated environment populated with autonomous traffic. Warnings were triggered during the approach to half of the intersections, cueing either towards the right or to the left. The remaining intersections were not cued, and served as controls. A preliminary analysis shows that gaze-contingent cues led to a significant shift in gaze position towards the highlighted direction.

8651-37, Session 11

Designing an obstacle display for helicopter operations in degraded visual environment

Patrizia M. Knabl, Niklas Peinecke, Deutsches Zentrum für Luft- und Raumfahrt e.V. (Germany)

Flying in degraded visual environment is an extremely challenging task for a helicopter pilot. The loss of the outside visual reference causes impaired situation awareness, high workload and spatial disorientation leading to incidents like obstacle or ground hits. DLR is working on identifying ways to reduce this problem by providing the pilot with additional information from fused sensor data. Therefore, different display design solutions are being developed.

In a first study, the design focused on the use of a synthetic head-down display, considering different representations for obstacles, color coding and terrain features. Results show a subjective preference for the most detailed obstacle display, while objective results reveal better performance for the display with little lower detail. Due to the fact that attention is limited, a very complex representation might actually hinder fast and accurate information processing in a high workload situation respectively when divided attention is required. In a second study, symbology for a helmet-mounted display is being designed and will be evaluated in autumn 2012. Design considerations will focus on different levels of terrain transparency, shape, size and complexity of obstacles and associated visual and perceptual aspects (e.g. attentional tunneling).

8651-38, Session 12

Visual storytelling in 2D and stereoscopic 3D video: effect of blur on visual attention

Quan Huynh-Thu, Cyril Vienne, Laurent Blondé, Technicolor S.A. (France)

Visual attention is an inherent mechanism that plays an important role in the human visual perception. Visual composition of scenes in stereoscopic 3D content creation today still mostly follows the same techniques than those used in 2D. In particular, out-of-focus blur is often used in 2D motion pictures and photography to drive the viewer's attention towards a sharp area of the image. However, in stereoscopic 3D content viewing, a conflict may appear with defocused foreground objects as our vision is naturally drawn both by sharp areas in the image and by objects appearing closest to us. In this paper, we study the impact of defocused foreground objects on visual attention deployment in stereoscopic 3D content. For that purpose, we conducted an eye-tracking experiment using a controlled production of stereoscopic 3D video content. 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 presented with the following test conditions: presence or absence of foreground object, focus or defocus on this foreground object, magnitude of disparity difference between foreground/background object planes. We discuss our results in the context of stereoscopic content creation, sense of immersion and visual comfort.

8651-39, Session 12

Using natural versus artificial stimuli to perform calibration for 3D gaze tracking

Christophe Maggia, Nathalie Guyader, Anne Guérin-Dugué, Gipsa-Lab (France)

The presented study tests which type of image, natural or artificial, is more adapted to perform efficient and reliable calibration to track 3D gaze in space using classical 2D eye tracker. Running an experiment with four observers, we compared natural scene images (indoor and outdoor) and grid of squares on 3D gaze calibration. Two types of models, linear and non-linear, were developed to predict real disparity from the disparity recorded (difference of the x-coordinates of the two eyes). Models were computed for different spatial positions and then merged over the whole space. We compared the type of images and the models using Root Mean Square errors between predicted and real disparities. We did not obtain any difference for two subjects whereas for the two others natural images provided better results. This suggests that it might be important to test which kind of stimulus is best suitable for particular observer to perform 3D calibration on the chosen stimulus before running an experiment, for example, on 3D natural scene images to study visual attention on stereo images.

8651-40, Session 12

Study of center-bias in the viewing of stereoscopic image and a framework for extending 2D visual attention models to 3D

Junle Wang, Matthieu Perreira Da Silva, Patrick Le Callet, Vincent Ricordel, Univ. de Nantes (France)

Compared to the good performance that can be achieved by many 2D visual attention models, predicting salient regions of a 3D scene is still challenging because of the additional depth information led by binocular disparity. An efficient way to achieve this can be to exploit existing studies dedicated for 2D content. Since it has been demonstrated the influence of 2D visual features (e.g. color, contrast, orientation, and center of the screen) in stereoscopic 3D viewing condition.

In this paper, we quantitatively evaluate the degree of center-bias during the viewing of stereoscopic 3D images of natural content. We also propose a simple computational model of 3D visual attention which can easily take advantage of center-bias and existing 2D models. For validation, an eye-tracking experiment is conducted, in which 35 subjects are involved. We thus create a database containing eye-movement data recorded during the viewing of eighteen 3D images and their corresponding 2D version. Our results indicate a clear difference between center-bias in 2D and 3D viewing conditions. By integrating center-bias in the proposed model, an added value of center-bias has been demonstrated in the prediction of saliency maps for 3D images.

8651-41, Session 12

How visual attention is modified by disparities and textures changes?

Darya Khaustova, Jérôme Fournier, Emmanuel Wyckens, France Telecom R&D (France); Olivier Le Meur, IRISA / INRIA Rennes (France)

3D image/video quality of experience is a multidimensional concept that depends on 2D image quality, depth quantity and visual comfort. The relationship between these parameters is not clearly defined yet. With this perspective, we try to understand how texture, depth quantity and visual comfort influence the way people observe 3D content in comparison to 2D. Five scenes with different structural parameters were generated using Blender software. For these five scenes the following parameters were modified: texture, camera baseline and convergence distance creating then different contents for different presentations to each observer. Our study was conducted using an eye-tracker and a 3DTV display. During psychovisual experiment each observer examined freely each scene with different textures, camera baselines and convergence distances. To avoid memory bias, we made sure that every observer had seen scene content only once. Collected fixation data are used to build saliency maps and to analyze differences between 2D and 3D condition. If the regions of interest (RoI) are coincided in both conditions, it was interesting to track the gaze in 3D. The discomfort introduced by the background or the foreground might be neglected in case when observers pay attention to RoI. From this analysis, we can answer the question if 3D visual attention is content dependent and whether quality of experience influence the way we observe stereoscopic pictures.

8651-43, Session 13

Copy-paste in depth: quantifying the pictorial space of paintings

Maarten W. A. Wijntjes, Technische Univ. Delft (Netherlands)

Whereas pictorial space plays an important role in art historic discussions, there is little research on the quantitative structure of pictorial spaces. Recently, a number of methods have been developed, one of which relies on size constancy: two spheres are rendered in the image while the observers adjusts the relative sizes such that they appear of equal size in pictorial space. This method is based on pair-wise comparisons, resulting in n(n-1)/2 trials for n samples. Furthermore, it renders a probe in the image that does not conform to the style of the painting: it mixes computer graphics with a painting.

The method proposed here uses probes that are already in the scene, not violating the paintings' style. An object is copied from the original painting and shown in a different location. The observer can adjust the scaling such that the two objects (one originally in the painting, and the other copy-pasted) appear to have equal sizes in pictorial space. Since the original object serves as a reference, the number of trials is merely n-1, an experimental reduction of 1/n trials with respect to the original method.

We measured the pictorial spaces of two paintings using our method, one Canaletto and one Breughel. We found that observers typically agreed will with respect to each other, coefficients of determination as high as 0.9 were found when the probe was a human, while other probes scored somewhat (but significantly) lower. These initial findings appear very promising for the study of pictorial space.

8651-44, Session 13

Drawing accuracy measured using polygons

Linda C. Carson, Univ. of Waterloo (Canada); Matthew J. H. Millard, Stanford Univ. (United States); Nadine Quehl, James Danckert, Univ. of Waterloo (Canada)

The study of drawing, for its own sake and as a probe into human visual perception, generally depends on ratings by human critics and self-reported expertise of the drawers. To complement those approaches, we have developed an objective continuous performance-based measure of drawing accuracy. This measure is based on representing drawings as sets of landmark points and studying features of particular research interest by comparing polygons of those features' landmark points with their counterpart polygons in a ground truth image. Any element of a drawing—an object, a plane, a shadow, a space—could be represented by the polygon formed by its landmark points. This approach produces both local accuracy measures (for each polygon) and a global accuracy measure (the mean across several polygons) for analysis. Furthermore, there are four distinct properties of a polygon to measure: its size, its position, its orientation and the proportionality of its shape. We can decompose error into four components and investigate how each contributes to drawing performance. We briefly describe the method and its potential applications to research in drawing education and visual perception, then report on its application to a specific research question: Are we more accurate when drawing in the so-called "positive space" (or figure)?

8651-45, Session 13

Fractals in art and nature: why do we like them?

Branka Spehar, The Univ. of New South Wales (Australia); Richard P. Taylor, Univ. of Oregon (United States)

Fractals have experienced considerable success in quantifying the visual complexity exhibited by many natural patterns, and continue to capture the imagination of scientists and artists alike. Fractal patterns have also been noted for their aesthetic appeal, a suggestion further reinforced by the discovery that the poured patterns of the American abstract painter Jackson Pollock are also fractal, together with the findings that many forms of art resemble natural scenes in showing scale-invariant, fractal-like properties. While some have suggested that fractal-like patterns are inherently pleasing because they resemble natural patterns and scenes, the relation between the visual characteristics of fractals and their aesthetic appeal remains unclear. Motivated by our previous findings that humans display a consistent preference for a certain range of fractal dimension across fractal images of various types we turn to scale-specific processing of visual information to understand this relationship. We extend our investigations to the visual preference for random noise images varying in terms of the slope of the rotationally averaged Fourier amplitude spectrum. Furthermore, we make a direct comparison between visual preference to 1/f grayscale images and a comparison set of thresholded (black and white) images. We found no significant differences in preferences between gray-scale images and binary comparison images obtained by simply thresholding the original gray-scale images. For both set of images, the visual preference peaked for images with the amplitude spectrum slopes from 1.25 to 1.5, thus confirming and extending the previously observed relationship between fractal characteristics of images and visual preference.

Conference 8652: Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8652 Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications

8652-1, Session 1

A spherical perceptual color model *(Invited Paper)*

Tieling Chen, Univ. of South Carolina (United States); Zhongmin Deng, Wuhan Textile Univ. (China); Jun Ma, Wuhan Textile University (China)

The paper introduces a transformed spherical model to represent the color space. The circular cone with a spherical top tightly circumscribing the color cube is equipped with a spherical coordinate system, with the origin at the black vertex and the diagonal as the vertical axis. Every point in the color cube is then represented by three spherical coordinates, with the radius Rho measuring the distance to the origin, indicating the brightness attribute of the color, the azimuthal angle Theta measuring the angle on the horizontal plane, indicating the hue attribute of the color, and the polar angle Phi measuring the opening of the circular cone with the vertical axis as its center, indicating the saturation attribute of the color. Similar to the HSV and HSL models, the spherical model specifies colors by describing the color attributes recognized by human vision. The conversion formulas are much mathematically neater than that of the HSV model and HSL model, and the interpretation of the model is more intuitive too. Most importantly, the spherical color model does not have the ray phenomenon that occurs in the HSV and HSL models. Applications of the spherical model including color comparisons are studied in the paper too.

8652-2, Session 1

Chroma-preserved luma controlling technique using YCbCr color space

Sooyeon Lee, Youngshin Kwak, Ulsan National Institute of Science and Technology (Korea, Republic of); Youn Jin Kim, Samsung Electronics Co., Ltd. (Korea, Republic of)

YCbCr color space composed of luma and chrominance components is preferred for its ease of image processing. However the non-uniformity and non-orthogonality of YCbCr between YCbCr components induce change of perceived lightness, perceived chroma and perceived hue as controlling luma, Cb and Cr values. In this study, a new method was designed for the chroma compensation generated by luma change. The 6 YCC_hue angles were selected to design this. For each YCC_hue angle, data points named 'Original data' generated with uniformly distributed various luma and YCC_chroma value. Then the weight values were applied to luma value of 'Original data' as 'Test data'. And new YCC_chroma having minimized CIECAM02 ?C between original and test data was calculated. The new was used to design this model. This model implemented for image processing as luma controlling algorithm having constant perceived chroma. The performance was tested numerically and perceptually. For 1324 'test data' by comparison with 'Original data' the CIECAM02 ?C has been decreased with 53.31%. When the new method was applied to luma changed image, it has 67.3% improvement of CIECAM02 ?C for 130% increment of luma

8652-3, Session 1

Analysis of a color space conversion engine implemented using dynamic partial reconfiguration

Ryan Toukatly, Dorin Patru, Eli Saber, Eric Peskin, Rochester Institute of Technology (United States); Gene Roylance, Brad Larson, Hewlett-Packard Co. (United States)

Dynamic Partial Reconfiguration allows parts of a Field Programmable Gate Array to be reconfigured, while the rest of the system continues uninterrupted operation. A Color Space Conversion Engine is a digital image-processing pipeline, which requires frequent reconfiguration of some, but not all of its stages. Therefore, it is a digital signal processing system that presumably can take advantage of dynamic partial reconfiguration. This paper describes the necessary design changes, testing, and performance analysis of a color space conversion engine implemented onto a field programmable gate array using dynamic partial reconfiguration. The analysis provides insight into the operational scenarios in which dynamic partial reconfiguration is advantageous or not.

8652-4, Session 2

Color reproductivity improvement with additional virtual color filters for WRGB image sensor

Shun Kawada, Rihito Kuroda, Shigetoshi Sugawa, Tohoku Univ. (Japan)

We have developed a high accuracy color reproduction method based on an estimated spectral reflectivity of objects using additional virtual color filters for a wide dynamic range CMOS image sensor with WRGB color filter which we have developed. The WRGB image sensor does not need to switch the way of signal processing depending on the light intensity even where a large sensitivity difference is exist between the color pixels. That is because the four signals are available to the point of saturation light intensity. The virtual color filters are created by multiplying the spectral sensitivity of W pixel by normal distribution functions, and the virtual sensor outputs of those virtual filters are estimated from the four output signals of the WRGB image sensor. The accuracy of color reproduction was evaluated with the Macbeth Color Checker. The averaged value of the color difference delta-Eab of 24 colors is 2.10 by using the minimum norm method based on six signals which are a combination of the three real RGB signals and the three virtual signals.

8652-5, Session 2

Glare and shadow reduction for desktop digital camera capture systems

Thanh H. Ha, Chyuan-Tyng Wu, Purdue Univ. (United States); Peter Majewicz, Flextronics International (United States); Kurt R. Bengtson, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

The quality of images of objects with significant 3D structure, captured at close range under a flash, may be substantially degraded by glare and shadow regions. In this paper, we introduce an imaging system

and corresponding algorithm to address this situation. The imaging system captures three frames of the stationary scene using a single camera in a fixed position, but an illumination source in three different positions, one for each frame. The algorithm includes two processes: shadow detection and image fusion. Through shadow detection, we can locate the area of shadows. After getting the shadow maps, we generate a more complete final image by image fusion. Our experimental results show that in most cases, the shadow and glare are markedly reduced.

8652-6, Session 2

Reducing flicker due to ambient illumination in camera captured images

Minwoong Kim, Purdue Univ. (United States); Kurt R. Bengtson, Lisa Li, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

No Abstract Available

8652-8, Session 3

Binary image compression using conditional entropy-based dictionary design and indexing

Yandong Guo, Purdue Univ. (United States); Dejan Depalov, Peter Bauer, Brent Bradburn, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States); Charles A Bouman, Purdue University (United States)

Details withheld pending submission of a patent application that is in preparation, and which will cover the technology to be described in this paper.

8652-9, Session 3

Segmentation for better rendering of mixed-content pages

Yi-Ting Chen, Purdue Univ. (United States); Di-Yuan Tzeng, Terry Nelson, Mark Q. Shaw, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

We propose a method for improving rendering of mixed-content pages by segmenting and classifying the regions of these pages according to the nature of the content. Our algorithm uses edge operators and connected components applied to both the continuous-tone page image and an object map generated during ripping of the page from page description format to raster format. The rendering is done using an object-based halftoning strategy that was reported at this conference in 2012.

8652-11, Session 3

YACCD2: yet another color constancy database updated

Alessandro Rizzi, Cristian Bonanomi, Davide Gadia, Giuseppe Riopi, Univ. degli Studi di Milano (Italy)

In 2003, at this conference, we have presented an image database (IDB) to test color constancy and other kinds of algorithms. The motivation to add another IDB to the many already present on the web is due to the fact that each IDB is characterized by choices that can fit or not with the algorithms to test. E.g. a database of images containing a white area is suitable for algorithms based on the White Patch approach, on the contrary, the complete absence of white areas can advantage algorithms with alternative approaches.

YACCD has the following characteristics:
- different backgrounds with a wide frequency range, containing white but with an average reflectance close to middle gray
- a set of natural, artificial and extreme illuminants
- images with and without casted shadows
We decided to redo YACCD in order to add the following characteristics:
- multiple exposures for HDR imaging
- RAW and jpg formats
- reflectance data
In this way YACCD2 can be suitable to test a wider variety of algorithms like e.g. computational color constancy, human vision models, HDR tone rendering, intrinsic images and other computer vision algorithms.

8652-42, Session 3

An efficient flicker noise reduction method for single images

Pan Pan, Yuan He, Shufu Xie, Jun Sun, Satoshi Naoi, Fujitsu Research and Development Center Co., Ltd. (China)

In this paper, we present a novel efficient flicker noise reduction method for single images scanned by overhead line sensors. The flicker noise here is perceived as horizontal bands which are not necessarily periodic. We view the flicker pattern as the noise of row cumulative histogram along the vertical direction, and propose two novel cumulative histogram filtering approaches to smooth the artifact, including using different Gaussian variance and padding the image. The proposed algorithm is then used to reduce the flicker noise in our scanned color images. The computational complexity of the proposed algorithm is further analyzed. The algorithm operates on singe images, and does not rely on the frequency of alternative currency, nor requires the horizontal bands are periodic. Experimental results show the superior performance of the proposed method in comparison to other existing methods.

8652-12, Session 4

Gray-world-assumption-based illuminant color estimation using color gamuts with high and low chroma

Harumi Kawamura, Nippon Telegraph and Telephone Corp. (Japan); Shunichi Yonemura, Shibaura Institute of Technology (Japan); Jun Ohya, Waseda Univ. (Japan); Akira Kojima, Nippon Telegraph and Telephone Corp. (Japan)

This paper proposes a new approach for estimating illuminant color based on the gray world assumption combined with color gamut method. This assumption hypothesizes the average color of all the objects in the scene is achromatic, therefore, it is difficult to estimate an illuminant color correctly if the colors of the objects in a scene are dominated by certain colors. Our previous method extends the images to which this hypothesis can be applied by using several but not all colors. With this method, however, illuminant colors cannot be estimated in cases where there are several colors in the image and they are localized in one part of the color space. To solve this problem, our approach uses two kinds of color gamuts, a high-chroma one and a low-chroma one. The high chroma gamut is used for adding the colors to the image to makes it possible to select colors to meet the assumption and the low chroma one is used to estimate the candidates for illuminant colors. Experimental results using color images show that the illuminant color estimation obtained with our approach is statistically closer to the correct one than that obtained with the conventional method.

8652-13, Session 4

Estimation of reflectance based on properties of selective spectrum with adaptive Wiener estimation

Ji-hoon Yoo, Wang-Jun Kyung, Ho-Gun Ha, Yeong-Ho Ha, Kyungpook National Univ. (Korea, Republic of)

In this paper, we propose an advanced estimation of reflectance by updating the autocorrelation matrix of the reflectance in Wiener estimation and using the selective wavelength of the reflectance from a spectral analysis. The key point of Wiener estimation is to find a transform matrix whose role converts the acquired images into the estimated reflectance. The transform matrix is defined as minimizing a mean square error between the measured and the estimated reflectance. Also, this matrix consists of an autocorrelation matrix of a reflectance, an autocorrelation matrix of a noise, and a matrix of a spectral sensitivity which involves an illuminant, a filter transmittance, and camera sensitivity. In proposed method, the autocorrelation matrix of the reflectance is updated by calculating the similarity between the measured and the estimated reflectance. The training samples is used as 1485 color patches of Munsell and the test sample is used as 24 color patches of Macbeth Color Checker. Also, the selective wavelength of the reflectance is used to estimate accurate reflectance. As a result, the proposed method showed more accurate estimation of reflectance than the conventional Wiener estimation in experiments.

8652-14, Session 4

Metal-dielectric object classification by combining polarization property and surface spectral reflectance

Shoji Tominaga, Hideki Kadoi, Keita Hirai, Takahiko Horiuchi, Chiba Univ. (Japan)

Most object surfaces in natural scenes can be classified into two material classes: metal and dielectric. In conventional methods for material classification, there are two approaches; a color-based approach and a polarization-based approach. This paper proposes a method for classifying multiple object surfaces into the two material classes by combining surface spectral reflectance and polarization property. We utilize polarization property in order to classify the objects into metal and dielectric, and surface-spectral reflectance in order to segment the scene image into different object surface regions. An imaging system is developed using a liquid crystal tunable filter for capturing both polarization and spectral images simultaneously. Our classification algorithm consists of three stages; (1) highlight detection based on luminance threshold, (2) material classification based on the spatial distribution of the degree of polarization at the highlight area, and (3) image segmentation based on illuminant-invariant representation of the spectral reflectance. The feasibility of the proposed method is examined in detail in experiments using real-world objects.

8652-15, Session 5

An experiment on the color rendering of different light sources

Simonetta Fumagalli, ENEA (Italy); Cristian Bonanomi, Alessandro Rizzi, Univ. degli Studi di Milano (Italy)

The color rendering index (CRI) attempts to measure how much the color appearance of objects is preserved when they are illuminated by a given light source. This problem is of great importance for various industrial and scientific fields, such as lighting architecture, design, ergonomics, etc.

A light source can be specified through the Correlated color temperature or CCT. However two (or more) light sources with the same CCT but different spectral power distribution can exist. Therefore color samples viewed under two light sources with equal CCTs can appear different. Hence, the need for a method to assess the quality of a given illuminant.

Recently CRI has had a renewed interest because of the new LED-based lighting systems. They usually, have a color rendering index rather low, but good preservation of color appearance and a pleasant visual appearance (visual appeal). Various attempts to develop a

new color rendering index have been done so far, but still research is working for a better one.

This article describes an experiment performed by human observers concerning the appearance preservation of some light sources, comparing it with a range of available color rendering indices.

8652-16, Session 5

Color universal design: analysis of color category dependency on color vision type (4)

Tomohiro Ikeda, Kogakuin Univ. (Japan); Yasuyo G. Ichihara, Kogakuin Univ. (Japan) and NPO Color Universal Design Organization (CUDO) (Japan); Natsuki Kojima, Kogakuin Univ (Japan); Hisaya Tanaka, Kogakuin Univ. (Japan); Kei Ito, The Univ. of Tokyo (Japan) and NPO Color Universal Design Organization (CUDO) (Japan)

Color Universal Design is an easy-to-understand system that was created to convey color-coded information accurately to most people. In this study, we conducted two experiments to develop Color Universal Design. In the first experiment, the confusion locus was verified. We researched the inconsistency of the confusion locus that we found in a previous study using an experimental color chart based on the CIELAB uniform color space. The subjects had P-type and D-type color vision. In this experiment, a practical confusion locus was determined and verified. The second experiment involved color classification using the 100 Hue Test. We investigated a color classification system that does not depend on the color names from C-type, P-type, and D-type color vision. In this experiment, we could identify unique color categories that did not depend on the color names from each type of color vision.

8652-17, Session 5

Color naming 65,274,705,768 pixels

Nathan Moroney, Hewlett-Packard Labs. (United States); Giordano Beretta, Hewlett-Packard Co. (United States)

Starting with a snapshot of a quarter of a million Wikipedia images, roughly two hundred thousand JPEG files were analyzed using a number of algorithms, including a machine color naming module. The processing techniques and corresponding results of color naming 65,274,705,768 pixels is both challenging and informative. The specific images analyzed are based on the ImageCLEF 2010 wikipedia retrieval task and basic image processing operations, such as decompression, made use of open source imaging code. Custom analysis, such as machine color naming, was implemented as a separate module in which a listing of input images was processed to derive descriptive metrics per image. The scale of the data and uniqueness of the resulting metrics was also such that it was more efficient to create custom HTML5 2D canvas visualizations of the results. This paper provides a more detailed description of the processing steps and the corresponding results. This is includes an introductory discussion of data-intensive research for imaging and visualization. On a relative scale, the roughly 20 gigabytes of wikipedia image data is both much smaller than the terabyte image databases generated by some projects and also much larger than the half dozen or so images processed and evaluated by other projects. However wikipedia is a widely known web reference and general observations about the properties of this collection of images should be of general interest. After this background section, a number of basic image analysis modules are described and the results reported before the machine color naming is performed. This background analysis provides a better summary of the images and includes metrics such as aspect ratio, median device color, histogram statistics, and inferred white and black points for the images. The trends and distributions for each of these is considered in summary form. Finally, the machine color naming is applied and the resulting distributions are reported. These results are based on analysis that uses a fixed number of pre-defined color terms or color vocabulary. For a roughly 30 color term vocabulary it is easy to identify

the roughly 12% of wikipedia images than only include achromatic color terms. Of the remaining images the most frequently assigned color terms are described and illustrated. The color terms used per image is also considered and a trend for a mostly monotonically increasing number of color terms per image is investigated.

8652-18, Session 5

Analysis of brain activity and response to colour stimuli during learning tasks: an EEG study

Raffaella Folgieri, Claudio Lucchiari, Daniele L. R. Marini, Univ. degli Studi di Milano (Italy)

All we know the attractive power of colours, and many studies have been done about the psychology of colours and their impact on human activities, perception, memory and education.

In this work we analysed data collected from a sample of students involved in a learning process during which they received visual stimuli based on colour variation. The stimuli concerned both the background of the text to learn and the colour of the characters. The experiment indicated some interesting results concerning the use of primary and complementary colours, identified following the definition of the Chevreul's circle. The observed data, collected through the registration of EEG waves (by a BCI device) and through a final evaluation test, and the following analysis indicated some variation depending on the choose of the colours (primary or complementary). The obtained results encouraged us to continue experiments in a future phase extending the research to other, more complex, stimuli.

8652-41, Session 5

Prototypical colors of skin, green plant, and blue sky

Huanzhao Zeng, Qualcomm Inc. (United States); M. Ronnier Luo, Univ. of Leeds (United Kingdom)

Colours of skin, green plant, and blue sky of digital photographic colour images were studied for modelling and detection of these three important memory colour regions. The colour modelling of these three regions in CIELAB and CAM02-UCS was presented, and the properties of these three colour groups were investigated.

8652-19, Session 6

Direct binary search (DBS) algorithm with constraints

Kartheek Chandu, Mikel J. Stanich, Ricoh Production Print Solutions, LLC (United States); Chai Wah Wu, Barry M. Trager, IBM Thomas J. Watson Research Ctr. (United States)

In this paper, we describe adding constraints to the Direct Binary Search (DBS) algorithm, such as requiring the halftone pattern to have only one dot per column and row, which is implemented using more than two toggles during each trial operation. Implementations of the DBS algorithm traditionally limit operations to either one toggle or swap during each trial. In the example case, used to illustrate the DBS algorithm with a constraint in this paper, the algorithm produces a wrap-around pattern with uniformly distributed ON pixels which has a pleasing appearance that contains precisely one ON pixel per each column and row. The algorithm starts with an initial continuous tone image and an initial pattern having only one ON pixel per column and row. The auto correlation function of Human Visual System (HVS) model is determined along with an initial perceived error. Multiple operation pixel error processing during each iteration is used to enforce the one ON pixel per column and row constraint. The constraint of a single ON pixel per column and row is used as an example in this

paper. Further modification of the DBS algorithm for other constraints is possible, based on the details given in the paper. A mathematical framework to extend the algorithm to the more general case of Direct Multi-bit Search (DMS) is presented.

8652-20, Session 6

Improved spectral vector error diffusion by dot gain compensation

Daniel Nyström, Linköping Univ. (Sweden); Ole L. Norberg, Voxvil AB (Sweden)

Spectral Vector Error Diffusion, sVED, is an alternative approach to achieve spectral color reproduction, i.e. reproducing the spectral reflectance of an original, creating a reproduction that will match under any illumination. For each pixel in the spectral image, the colorant combination producing the spectrum closest to the target spectrum is selected, and the spectral error is diffused to surrounding pixels using an error distribution filter. However, since the colorant separation and halftoning is performed in a single step in sVED, the compensation for dot gain cannot be made for each color channel independently, as in a conventional workflow where the colorant separation and halftoning is performed sequentially. In this study, we modify the sVED routine to compensate for the dot gain, applying the Yule-Nielsen n-factor to modify each of the target spectra. A global n-factor, optimal for each print resolution, reduces the spectral reproduction errors by a factor of 4, while an n-factor that is individually optimized for each target spectrum reduces the spectral reproduction error to 7% of that for the unmodified prints. The results clearly illustrate the necessity to account for the dot gain in the printing process in order to make Spectral Vector Error Diffusion a realistic alternative for spectral color reproduction.

8652-21, Session 6

Extending color primary set in spectral vector error diffusion by multilevel halftoning

Ole L. Norberg, Voxvil AB (Sweden); Daniel Nyström, Linköping Univ. (Sweden)

In view of the fact that the human visual system includes three different types of color receptors, three color channels are sufficient to approximate the range of colors seen by the human. However, this tricromatic approach only has the ability to match colors when the illumination for the reproduction agrees with that of the original. By adding additional color channels to the printing system an optional approach, spectral printing, can be applied to reproduce the spectral information of the original color. A tempting approach to spectral printing is Spectral Vector Error Diffusion, sVED, were the transformation from spectral data to multi-binary colorant image is performed in one single process including both color separation and halftoning.

Essential for the performance of the sVED is the set of available colorant combinations. Larger color sets and optimal spectral characteristics are expected to significantly improve the color accuracy of the spectral reproduction.

In this study, sVED has been applied on a ten channel inkjet system. The output resolution has been reduced and the underlying higher print resolution has been used for multi-level halftoning technique and thereby significantly expanded the available colorant combinations from the Neugebauer primaries. Results from this study shows that by combine Neugebauer primaries with lighter combinations the average color difference can be reduced to one tenth of the reproduction by Neugebauer primaries only. The results illustrates the importance of the available colorant combinations and in the prolongation also the spectral characteristics of the process inks used.

8652-22, Session 6

Reducing auto moiré in discrete line juxtaposed halftoning

Vahid Babaei, Roger D. Hersch, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

Discrete line juxtaposed halftoning creates color halftones with discrete lines of various rational thicknesses laid out side by side, forming the different colorant surfaces. Screen elements are made of parallelogram screen tiles incorporating the colorant discrete lines. The repetition of discrete lines from one screen element to the next may create auto moiré artifacts. By decomposing each screen tile into two sub-tiles of rational thicknesses, we ensure that two successive discrete lines have different phases in respect to the underlying pixel grid. This results in a higher screen frequency and in a repetition vector which is different from one discrete line to the next discrete line of the same colorant. The main low-frequency artifact is broken into higher frequency artifacts which are not visible anymore.

8652-23, Session 7

Optimizing CMYK mapping for high speed digital inkjet webpress

Ruzhu Zeng, Liming Vocational Univ. (China); Huanzhao Zeng, Qualcomm Inc. (United States)

The CMYK to CMYK mapping preserving the black channel is a method to fix the problem in standard ICC color management that cannot preserve the K channel in printing CMYK contents. While the method has been successfully used for digital commercial printing, limitations and areas for improvement are found. To address these problems in generating CMYK re-rendering tables, an alternative method is developed. The K usage and total ink usage are optimized in a color separation step. Instead of preserving the K channel globally, it only preserves K-only grays and determines the mapping of other colors by optimizing the print quality.

8652-24, Session 7

Estimating toner usage with laser electrophotographic printers

Lu Wang, Purdue Univ. (United States); Dennis Abramsohn, Thom Ives, Mark Q. Shaw, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

Estimating toner usage is an important problem for manufacturers of laser electrophotographic printers. We have developed a new approach to this problem that combines image analysis techniques with physical measurements. We will present experimental results that demonstrate the efficacy of our method.

8652-25, Session 7

Perceived acceptability of colour matching for changing substrate white point

Kwame F. Baah, Univ. of the Arts London (United Kingdom); Phil Green, Gjøvik Univ. College (Norway); Michael R. Pointer, London College of Communication (United Kingdom)

Production and proofing substrates frequently differ in their white points. Substrate white points frequently differ between reference and sample, for example between proof and print, or between a target paper colour and an actual production paper. To achieve an acceptable visual match between colours specified for the reference substrate, when printed on a different material, printed colours can be adjusted to compensate for observer adaptation to the substrate white point. A widely-used method of adjustment is to convert all measurement data to media-relative measurements, thus scaling the tristimulus values by the ratio of reference to sample white points.

Using this approach evaluations of colour difference of acceptability between reproducible colour and its perceived appearance for sample substrates were conducted using three modes of proof-to-print psychophysical assessments namely, simultaneous, sequential and display.

The psychophysical tests showed that simultaneous and sequential modes had similar results but the display mode produced a higher tolerance. Reference paper thresholds for media type corrections of solid colours produced a colour difference that was larger than retargeting to a similar paper type. For light tints the threshold was the opposite with a threshold of less than 2.5 DEab which is less than what is expected from retargeting similar paper.

8652-26, Session 7

The development of vector based 2.5D print methods for a painting machine

Carinna E. Parraman, Univ. of the West of England (United Kingdom)

Through recent trends in the application of digitally printed decorative finishes to products, CAD, 3D additive manufacturing and research in material perception, there is a growing interest in the accurate rendering of materials and tangible displays. Although current advances in colour management and inkjet printing has meant that users can take for granted high-quality colour and resolution in their printed images, digital methods for transferring a photographic coloured image from screen to paper is constrained by pixel count, file size, colorimetric conversion between colour spaces and the gamut limits of input and output devices. This paper considers new approaches to applying alternative colour palettes by using a vector-based approach through the application of paint mixtures, towards what could be described as a 2.5D printing method. The objective is to not apply an image to a textured surface, but where texture and colour are integral to the mark, that like a brush, delineates the contours in the image. The paper describes the difference between the way inks and paints are mixed and applied. When transcribing the fluid appearance of a brush stroke, there is a difference between a halftone printed mark and a painted mark. The issue of surface quality is significant to subjective qualities when studying the appearance of ink or paint on paper. The paper provides examples of a range of vector marks that are then transcribed into brush stokes by the painting machine.

8652-27, Session 7

Anti-counterfeiting model of printing micro-mirror in CMYK color space

Tang Cheng, Yixin Zhang, Jiangnan Univ. (China)

Optical variable devices(OVD) for anti-counterfeiting applications mainly are realized by micro-grating structure. The diffraction grating should be less than 1 micron and needs expensive hot stamping foil as the carrier of OVD information. The purpose of this paper is to develop a reflective printing micro-mirror structure for the design of optically variable devices(OVD), which can be duplicated onto the security document in normal printing or molding process. Based on the reflective properties of printing micro-mirror and model of CMYK color space, we establish a new printing micro-mirror geometry model. In this model, considering the effects of micro-mirror surface shape and mirror slope angle on the reflected beams, and the utilizing of the Jacobian of micro-mirror surface function and gradient function, we analyze the reflection brightness and stability of CMYK printing micro-mirror. Our results show that Jacobian is approximately equal to 0 and gradient is a function of coordinate values. It is proved that reflected beams of printing micro-mirror surface have sufficient brightness and stability. Therefore, the printing micro-mirror is suitable to create anti-counterfeiting optical variable device, whose intensity of reflected beams will be spatially changed when position or angle of viewing modulate. In this paper, we design an anti-counterfeiting optical variable device which uses some printing micro-mirror of different CMYK values and a simulating security pattern of printing micro-mirror array is provided.

8652-28, Session 8

Unsupervised correction of relative longitudinal aberrations for multispectral imaging using a multiresolution approach

Julie Klein, RWTH Aachen (Germany)

Longitudinal aberrations appear in multispectral cameras featuring a monochrome sensor with several optical filters in front of it. Due to the slightly different optical properties of the filters, the focal lengths are different and the images cannot remain sharp for all the color channels. We seek for an unsupervised correction of these aberrations, relative to a given reference color channel. "Unsupervised" means here that no calibration of the system is needed. We use a multiresolution approach that takes advantage of the high contrast present in the reference channel and that utilizes this information for the other, more blurred channels. The results of this correction are evaluated using the sharpness of the corrected image with respect to the original blurred image and using the color accuracy: an algorithm that would corrupt the spectral information of multispectral images would not be helpful. Moreover, using the original image and the one corrected with the algorithm, we can calculate the point spread function of the longitudinal aberrations. We then compare it to the point spread function obtained with another method which is based on the capture of a noise chart and thus requires calibration.

8652-29, Session 8

Acquisition of multi-spectral flash image using optimization method via weight map

Bong-Seok Choi, Daechul Kim, Kyungpook National Univ. (Korea, Republic of); Oh-Seol Kwon, Changwon National Univ. (Korea, Republic of); Yeong-Ho Ha, Kyungpook National Univ. (Korea, Republic of)

Acquisitioning multi-spectral flash image, In this paper propose optimization method based on weight map to compensate color representation. We generated weight map by applying canny edge operator to Y channel of UV-IR spectrum image. This weight map represent feature of the scene. in order word, Color, detail, and feature information are applied to optimization process. We apply weight map in order to discriminate weighting in optimization process with region and edge. In the optimization process, smooth region to increasing color information weight, decreasing edge information weight. On the contrary to this, edge region to increasing edge information weight, decreasing color information weight. Our method can be enhancing color reproduction and removing artifacts. We evaluate the performance of the proposed algorithm using PSNR, MSSIM, and z-score with long exposure image. Experiments show that better result than previous method.

8652-31, Session 9

Adaptive local backlight dimming algorithm based on local histogram and image characteristics

Ehsan Nadernejad, Nino Burini, Jari Korhonen, Søren O. Forchhammer, Claire Mantel, Technical Univ. of Denmark (Denmark)

Liquid Crystal Displays (LCDs) with LED backlight are today a very popular display technology, used for instance in televisions, monitors and mobile phones. It is possible to increase the energy efficiency of such displays and preserve the image quality through the use of the backlight dimming. There are different backlight dimming algorithms, each one with a different tradeoff between power savings and image quality. In this paper, we present a low complexity and adaptive algorithm for dimming of LED-backlight displays that uses characteristics of the target image. The proposed algorithm uses the local features of the target image such as the local histogram and the average to find the optimal backlight luminance. The proposed method

has been applied on two modeled screens using a large selection of images of different kinds and sources including, for instance, high-contrast and high-luminance images; one screen has a high resolution direct-lit backlight, while the backlight of the other screen is edge-lit with 16 segments placed in two columns and eight rows. Results show that the proposed algorithm can save the power consumption and image quality of the liquid crystal display when comparing against other algorithms.

8652-32, Session 9

Color generation and modulation from transparent electroactive polymer gratings

Eunkyoung Kim, Haejin Shin, Xu Yang, Byeonggwan Kim, Yonsei Univ. (Korea, Republic of)

Nanostructures of the solution-processable electroactive polymers could be efficiently obtained at low temperature via MIMIC an imprinting methods. Subwavelength gratings with electroactive polymers such as poly(3-hexylthiophene) (P3HT) and poly(3,4-propylenedioxythiophene-phenylene) (P(ProDOT-Ph)) controlled the color intensity for various visible colors of diffracted light in a single device. Under the illumination of a white light, at a fixed angle of incidence, the color intensity of the diffracted light was reversibly switched from the maximum value down to 15% (85% decrease) by applying ?2 to 2 V due to electrochemical (EC) reaction. All spectral colors including red, green, and blue were generated by changing the angle of incidence, and the intensity of each color was modulated electrochemically at a single EC device. With electroactive subwavelength gratings (ESWGs) of P3HT, the maximum modulation of the color intensity was observed in the red-yellow quadrant in the CIE color plot, whereas for the ESWGs of P(ProDOT-Ph), the maximum modulation of the color intensity was observed in the yellow-green and green-blue quadrants. Both ESWGs showed a memory effect, keeping their color and intensity even after power was turned off for longer than 40 hours.

In this report, a simple electrochromic (EC) device based on transparent electroactive polymer gratings (EPG) will be presented as a new color generation as well as mdulaton method.

8652-33, Session 9

Optimizing color fidelity for display devices using contour phase predictive coding for text, graphics, and video content

Fritz Lebowsky, STMicroelectronics (France)

High-end monitors and TVs based on LCD technology continue to increase their native display resolution to 4k2k and beyond. Subsequently, uncompressed pixel data transmission becomes costly when transmitting over cable or wireless communication channels. For motion video content, spatial preprocessing from YCbCr 444 to YCbCr 420 is widely accepted. However, due to spatial low pass filtering in horizontal and vertical direction, quality and readability of small text and graphics content is heavily compromised when color contrast is high in chrominance channels. On the other hand, straight forward YCbCr 444 compression based on mathematical error coding schemes quite often lacks optimal adaptation to visually significant image content.

Therefore, we present the idea of detecting synthetic small text fonts and fine graphics and applying contour phase predictive coding for improved text and graphics rendering at the decoder side. Using a predictive parametric text contour model and transmitting correlated phase information across all three color channels combined with foreground/background color vectors of a local color map promises to overcome weaknesses in compression schemes that process luminance and chrominance channels separately. The residual error of the predictive model is more easily being minimized since the decoder is an integral part of the encoder. We also introduce the concept of an auto-adaptive differential color map that improves color quality over time by applying statistical analysis of color loci in text and graphics regions. A comparative analysis based on some competitive solutions

highlights the effectiveness of our approach and identifies its current limitations with regard to high quality color rendering and its remaining visual artifacts.

8652-34, Session 9

Content-dependent contrast enhancement for displays based on cumulative distribution function

Seul-Ki Jang, Yoon-Gyoo Lee, Inha Univ. (Korea, Republic of); Gyoung-Soo Park, Inha Univ (Korea, Republic of); Choon-Woo Kim, Inha Univ. (Korea, Republic of)

Perceived contrast is one of the most important attributes affecting image quality on displays. Three dimensional displays require contrast enhancement techniques to compensate for reduction in luminance levels. Also, mobile displays demand efficient contrast enhancement techniques to improve degree of visibility under outdoor viewing environments and reduce power consumption. This paper presents a new content dependent contrast enhancement technique for displays. Cumulative distribution functions of an image are classified into a number of classes based on the criterion determined by the results of image quality evaluation. Also, a simple tone mapping function is designed for each of class of cumulative distribution functions. The proposed technique is simple enough for online implementation for moving images. In this paper, criterion of classification of cumulative distribution functions is described. Also, construction of tone mapping function is explained.

8652-35, Session 10

Can trichromats really know what dichromats see? *(Invited Paper)*

Michael H. Brill, Datacolor (United States)

Can trichromats really know what dichromats see?

Presumptively the answer would be "yes", judging from several algorithms and software applications that simulate the appearance to a dichromat of any given trichromatic image.1-4 My purpose here is to challenge that presumption.

We know what sets of tristimuli are matches for each kind of dichromat. The confusion loci are parallel line segments through tristimulus space. But that says nothing about how to map the appearance of a dichromatic color on the appearance from a trichromatic space. It is not even necessary for the dichromatic appearance of a light to match the trichromatic appearance of one of the lights on a confusion locus. So how can one make the map?

On one level, the question devolves to the classic philosophical conundrum of my not being able to know if I see the same blue that you do. The situation is saved to some extent by the existence of unilateral dichromats. There the appearance matches between the dichromatic eye and the trichromatic eye promise to be a legitimate "Rosetta Stone". Indeed, unilateral dichromats depose the naïve model of a dichromat's color always having the appearance of one of its confusion aliases in trichromatic vision. But to be trustworthy, color-appearance matches must be made cetera paribus---that is, all other variables being equal. The spatial context of a scene always affects the appearance of a color in that scene, and the contexts themselves cannot be equated between a dichromat and a trichromat. You would have to ask the unilateral dichromat to match all the colors in all the possible scenes in your universe to be sure that you had a good simulation. An additional complication is that unilateral dichromats are so rare that we cannot be too fussy about assuring that the trichromatic eye is really "normal." Finally, the colors dichromats see can be as unstable as Gestalt effects like the Necker cube. If you still want to predict and simulate what dichromats see, you have truly passed…to the Dark Side of the Color.

8652-36, Session 10

Color scales for visualization: traveling though color space *(Invited Paper)*

Bernice E. Rogowitz, Visual Perspectives Consulting (United States)

Color is often used to represent data in visualization and imaging. A colormap, for example, maps numerical values onto a color scale. Each color scale is a trajectory through a three-dimensional color space. But, which trajectories are the best? A popular idea is that any trajectory where equal steps in the data correspond to equal steps in the color space is sufficient to produce a good result. According to this hypothesis, equal steps around the hue circle or equal steps in luminance should produce equally good colormaps. This talk challenges this assumption and provides insight into other perceptual factors at play.

8652-37, Session 10

Color spaces *(Invited Paper)*

Jan J. Koenderink, Technische Univ. Delft (Netherlands)

Color spaces may be divided into "colorimetric" and "eye measure", although there exist some in between cases of indeterminate ontological status. Colorimetric constructions use only judgments of indiscriminability, whereas eye measure methods rely on first person reports. I consider colorimetric approaches that go beyond the linear Hausdorff space. These become possible when additional constraints are introduced. I consider the simplest model of generic object colors, which involves spectrally selective Lambertian surfaces under average daylight illumination. This reflects likely evolutionary pressure to optimize the ability of discriminating surfaces in generic scenes. There exists a unique geometrical representation that maximizes the color gamut captured by the unit cube of the vector basis. In this representation color vision behaves like 3-bin spectroscopy, essentially yielding RGB-coordinates in the range 0-1. In terms of the natural metric, and using the symmetries of the RGB-cube, one obtains a color space that is very similar to the Ostwald space. Although indeed purely colorimetric, there is a strong similarity to eye measure systems. Because this space hardly depends upon arbitrary conventions (only the choice of average daylight is "arbitrary", although reasonable because of evolutionary/ecological considerations), it has many advantages over conventional systems.

8652-38, Session 10

You can't rely on color, yet we all do *(Invited Paper)*

Floris L. van Nes, Technische Univ. Eindhoven (Netherlands)

Color is one of the most salient features of the visual field. Being thrilled by color, man early in his development learned how to use it, as for body paint – which continued in fashion, art, for painting artefacts. In a way, man relied on color, for embellishing and later, when coding was invented, for attaching meaning to objects having a certain color.

However, this magnificent property of all objects around us is also elusive, as soon as you want to really specify it. Notwithstanding perceptual phenomena such as color constancy one has to cope with (i) dependence of all surface colors on the color of illumination; (ii) the fact that it is not easy for one person to communicate the precise color of an object to another person – especially of the opposite sex; (iii) observing that colors from visual displays are often strongly desaturated by surround light. To mention but a few. And all this is rather poorly understood, if at all, by the general public. Sometimes even by people who are reasonably knowledgable in visual matters, such as visual ergonomists or display engineers. After all, it is not a rare occurrence that multi-colored texts, in print or on slides shown by visual scientists are illegible because the text foreground and background colors were ill-chosen.

But color is so important that people can hardly avoid to 'rely on color' – so they do. They rely on its predictability; when they buy clothes via the internet, for instance, or select a holiday destination from an alluring landscape picture on a calendar.

As to professional users: designers of modern electronic visual displays need to master many fields; colorimetry of one kind or another being one of them.

My experience of several decades has taught me that dealing with colors on displays always turned out to tax most writers of ergonomics display standards heavily – as well as the professionals in display manufacturing. Because, the organization responsible for the agreed standards for specification and measurement of all dimensions of color , the Commission Internationale de l'Éclairage, (CIE), has for a long time now created chromaticity diagrams and color spaces that are complex and far from easy to use by 'practitioners' in the color display industry, and the writers of standards for their products.

Of course, quantifying facets of color is a difficult thing to do. Yet, there are many people who would applaud CIE if this illustrious organization would devote a part of its energy to the development of a color specification and measurement system with a simpler yet easily usable color space than, for instance, CIECAM02 – or its successor that may already be waiting in the wings. Perhaps CIE then should resort to the well established ergonomics method of user involvement, the users being color science practitioners. Furthermore, it would be revolutionary but recommendable if an international organization such as the CIE would start an effort to educate the population at large better in color matters. One group that could benefit from such an action are color-defective viewers. Time and again they are the victim of inappropriate color choices for all sorts of products that have been color-coded, sometimes for important information in terms of health or safety, by people who have normal color vision but lack knowledge on its intricasies, and who therefore are naive as to the perception of these choices by those with color vision abnormalities.

8652-39, Session 10

How 'high-level' is human color perception?
(Invited Paper)

Michael E. Rudd, Univ. of Washington (United States)

Color is often used to represent data in visualization and imaging. A colormap, for example, maps numerical values onto a color scale. Each color scale is a trajectory through a three-dimensional color space. But, which trajectories are the best? A popular idea is that any trajectory where equal steps in the data correspond to equal steps in the color space is sufficient to produce a good result. According to this hypothesis, equal steps around the hue circle or equal steps in luminance should produce equally good colormaps. This talk challenges this assumption and provides insight into other perceptual factors at play.

8652-40, Session 10

Complex spatiochromatic interactions in a real world art laboratory

Scott J. Daly, Dolby Labs., Inc. (United States)

Andy Goldsworthy is an environmental artist whose theme for several decades has been the artistic manipulation of in situ natural materials to form impermanent outdoor sculptures. One common thread is the re-organization of objects by their color, turning natural and consequently more random distributions into highly ordered arrangements. In doing so, the colors of the natural materials seem to change as a result. The real-world artwork invokes various physical and perceptual phenomenon, serving as laboratory for color science. In this talk, we will take one these artworks, titled St. Abbs, the Borders, 1985, and qualitatively investigate the physical and perceptual contributors to the intriguing effect. These processes range from atmospheric haze, multiple reflections, spatiochromatic CSF, white-point anchoring, chromatic induction, etc. They will be abstracted in an attempt to create a two-dimensional synthetic version of the essence of Goldsworthy's three-dimensional natural chromatic illusion.

Conference 8653: Image Quality and System Performance X

Tuesday - Thursday 5 –7 February 2013

Part of Proceedings of SPIE Vol. 8653 Image Quality and System Performance X

8653-1, Session 1

A no-reference quality assessment algorithm for JPEG2000-compressed images based on local sharpness

Phong V. Vu, Damon M. Chandler, Oklahoma State Univ. (United States)

In this paper, we present a no-reference quality assessment algorithm for JPEG2000-compressed images called EDIQ (EDge-based Image Quality). The algorithm works based on the assumption that the quality of JPEG2000-compressed images can be evaluated by the sharpness of the edge/near-edge regions and the non-edge regions where no edges are present. EDIQ first separates the input image into edge/near-edge regions and non-edge regions by applying Canny edge detection and edge-pixel dilation. Our previous sharpness algorithm, FISH [1], is used to generate a sharpness map. The part of the sharpness map corresponding to the non-edge regions is collapsed by using root mean square to yield the sharpness index of the non-edge regions. The other part of the sharpness map, which corresponds to the edge/near-edge regions, is weighted by the local root mean square (RMS) contrast and the local slope of magnitude spectrum to yield an enhanced sharpness map, which is then collapsed into the sharpness index of the edge/near-edge regions. These two indices are combined by a geometric mean to yield a quality indicator of the input image. Testing on the JPEG2000 subsets of four different image-quality databases demonstrate that EDIQ is competitive with other no-reference image quality algorithms on JPEG2000-compressed images.

8653-2, Session 2

Measurement and protocol for evaluating video and still stabilization systems

Etienne Cormier, Frédéric Cao, Frédéric Guichard, Clément Viard, DxO Labs (France)

This article presents a system and a protocol to characterize image stabilization systems both for still images and videos. It uses on a six axes platform, three being used for camera rotation and three for camera positioning. The platform is programmable and can reproduce complex motions that have been typically recorded by a gyroscope mounted on different types of cameras in different use cases. The measurement uses a single chart for still image and videos, the texture dead leaves chart. Although the proposed implementation of the protocol uses a motion platform, the measurement itself does not rely on any specific hardware. For still images, a modulation transfer function is measured in different directions and is weighted by a contrast sensitivity function (simulating the human visual system accuracy) to obtain an acutance. The sharpness improvement due to the image stabilization system is a good measurement of performance as recommended by a CIPA standard draft. For video, some markers on the chart are detected with sub-pixel accuracy to determine a homography deformation between the current frame and a reference position. This model describes well the apparent global motion as translations, but also rotations along the optical axis and distortion due to the electronic rolling shutter equipping most CMOS sensors. The protocol is applied to all types of cameras such as DSC, DSLR and smartphones.

8653-3, Session 3

Hyperspectral image quality for unmixing and subpixel detection applications

John P. Kerekes, Daniel S. Goldberg, Rochester Institute of Technology (United States)

The quality of remotely sensed hyperspectral images is not easily assessed visually, as the value of the imagery is primarily inherent in the spectral information embedded in the data. In the context of earth observation or defense applications, hyperspectral images are generally defined as high spatial resolution (1 to 30 meter pixels) imagery collected in dozens to hundreds of contiguous narrow (?/??~100) spectral bands from airborne or satellite platforms. Two applications of interest are unmixing which can be defined as the retrieval of pixel constituent materials (usually called endmembers) and the area fraction represented by each, and subpixel detection, which is the ability to detect spatially unresolved objects. Our approach is a combination of empirical analyses of airborne hyperspectral imagery together with system modeling driven by real input data. Initial results of our study show the dominance of spatial resolution in determining the ability to detect subpixel objects and the necessity of sufficient numbers of spectral bands for unmixing accuracy. While these results are not unexpected, the research helps to quantify these trends for the situations studied. Future work is aimed at generalizing these results and to provide new prediction tools to assist with hyperspectral imaging sensor design and operation.

8653-4, Session 3

Acceptable bit-rates for human face identification from CCTV imagery

Anastasia Tsifouti, Ctr. for Applied Science and Technology (United Kingdom) and Univ. of Westminster (United Kingdom); Sophie Triantaphillidou, Efthimia Bilissi, Univ. of Westminster (United Kingdom); Mohamed-Chaker Larabi, Univ. of Poitiers (France)

The objective of this investigation is to produce recommendations for acceptable bit-rates of CCTV footage of people onboard London buses. The majority of CCTV recorders on buses use a proprietary format based on the H.264/AVC video coding standard, exploiting both spatial and temporal redundancy. Low bit-rates are favored in the CCTV industry but they compromise the image usefulness of the recorded imagery. In this context usefulness is defined by the presence of enough facial information remaining in the compressed image to allow a specialist to identify a person. The investigation includes four steps: 1) Collection of representative video footage. 2) The grouping of video scenes based on content attributes. 3) Psychophysical investigations to identify key scenes, which are most affected by compression. 4) Testing of recording systems using the key scenes and further psychophysical investigations. The results are highly dependent upon scene content. For example, very dark and very bright scenes were the most challenging to compress, requiring higher bit-rates to maintain useful information. The acceptable bit-rates are also found to be dependent upon the specific CCTV system used to compress the footage, presenting challenges in drawing conclusions about universal 'average' bit-rates.

8653-5, Session 3

Visual acuity and task-based video quality in public safety applications

Joel Dumke, Institute for Telecommunication Sciences (United States)

This paper explores the utility of visual acuity as a video quality metric for public safety applications. An experiment has been conducted to track the relationship between visual acuity and the ability to perform a forced-choice object recognition task with digital video or varying

quality. Visual acuity is measured according to the smallest letters reliably recognized on a reduced LogMAR chart.

8653-6, Session 4

Improvement of quality metrics performance using binocular just-noticeable-difference for S3D images

Bilel Sdiri, Mohamed-Chaker Larabi, XLIM-SIC (France); Azeddine Beghdadi, Univ. Paris 13 (France)

In this paper, we consider the BJND for the development of a full reference metric for quality assessment of stereoscopic images based on the well-known metrics: PSNR and SSIM, in which each inter-pixel distortion is weighted by its visibility threshold for a binocular vision. The preliminary validation of such an approach proves that using the Binocular JND adds a perceptual dimension to mathematical metrics.

8653-7, Session 4

Evaluation of biological effect on luminance of stereoscopic displays

Kyosuke Takahashi, Toshiya Nakaguchi, R. Okamoto, I. Shimoyama, Y. Miyake, Chiba Univ. (Japan)

The stereoscopic images prevail in various fields and the biological effect of stereoscopic images is investigated by various researches. However, the evaluation of biological effect on intensity of stereoscopic displays is unknown. Therefore we evaluated biological effect on intensity of stereoscopic displays and difference in indication method by conducting subjective evaluation and objective evaluation. We measured pupillary light reflex and R-R interval in ECG as objective evaluation, and conducted two kinds of double stimuli evaluation and single stimulus evaluation. The significant effect by the intensity change was shown by double stimuli evaluation whereas it was not shown by single stimulus evaluation. Based on this result it was shown that double stimuli evaluation is suited for evaluation of biological effect on intensity of stereoscopic displays than single stimulus evaluation. No significant relationship was noted in the results between the pupillary light reflex and the intensity. Although significant relationship was obtained in the results between the R-R interval in ECG and the elapsed time for 30 min stimulus, no significant relationship was noted in the results between the R-R interval in ECG and the intensity. In addition, we confirmed experimental accuracy and reproducibility by conducting repetitive experiments.

8653-8, Session 4

Three factors that influence the overall quality of the stereoscopic 3D content: image quality, comfort, and realism

Raluca Vlad, Patricia Ladret, Anne Guérin-Dugué, GIPSA-lab (France)

An increasing amount of stereoscopic 3D content is now available, but no standardized objective methodology exists which could indicate whether this content is 'of good quality'. In this case, subjective methodologies could be a preferred compromise for assessing quality. They imply the deployment of important resources, but they illustrate very well human perception. However, no standard quality model that defines the notion of '3D quality' exists to date. Thus, for evaluating the quality of 3D stereoscopic data, exploratory studies are needed beforehand in order to determine which characteristics of the 3D content should be submitted to test.

In this context, we carried out a psycho-visual qualitative exploration of the various impressions that stereoscopic images produce to an observer. The test proved to be rich in results. Among its conclusions, we consider of highest importance the fact that we could thus determine three different factors – quality, comfort and sense of reality

– which are important in studying how the overall quality of 3D data is perceived. The results of our test show that all of these three factors should be included in an independent manner in the assessment of the perceived 3D experience.

8653-9, Session 4

Performance evaluation of HD camcorders: measuring texture distortions using Gabor filters and spatio-velocity CSF

Kongfeng Zhu, Dietmar Saupe, Univ. Konstanz (Germany)

This paper presents a new method of measuring physical texture distortions (PhTD) to evaluate the performance of HD camcorders w.r.t. motion and lossy compression. It is extended to measure perceptual texture distortions (PeTD) by taking into account of the spatio-velocity contrast sensitivity function of human visual system. PhTD illustrates to what extent texture structures are physically distorted by a camcorder, while PeTD measures the perceptual distortion of textures. The dead leaves chart invariant to scaling, translation, rotation, and contrast, was selected as the target. The PhTD/PeTD metrics of the target distorted by camcorders were measured based on a bank of Gabor filters with eight orientations and three scales. Experimental results on six HD camcorders from three vendors showed: 1) the PhTD value decreases monotonically w.r.t. the motion speed, and increases monotonically w.r.t. the lossy compression bitrate; 2) the PeTD value decreases monotonically w.r.t. the motion speed, but stays almost constant w.r.t. the lossy compression bitrate. As a result, the proposed PhTD metric can measure physical textures distortions, and the PeTD metric can measure perceptual texture distortions. The PeTD measurement results showed that lossy compression of tested HD cameras leads to unperceivable texture distortions in our experimental setup.

8653-10, Session 5

Evaluation of differences in quality of experience features for test stimuli of good-only and bad-only overall audiovisual quality

Dominik Strohmeier, Technische Univ. Berlin (Germany); Kristina Kunze, Klemens Göbel, Judith Liebetrau, Ilmenau Univ. of Technology (Germany)

Assessing audiovisual Quality of Experience (QoE) is a key element to ensure quality acceptance of today's multimedia products. The use of descriptive evaluation methods allows evaluating QoE preferences and the underlying QoE features jointly. From our previous evaluations on QoE for mobile 3D video we found that mainly one dimension, video quality, dominates the descriptive models. Large variations of the visual video quality in the tests may be the reason for these findings. A new study was conducted to investigate whether test sets of low QoE are described differently than those of high audiovisual QoE. Reanalysis of previous data sets seems to confirm this hypothesis. Our new study consists of a pre-test and a main test, using the Descriptive Sorted Napping method. Data sets of good-only and bad-only video quality were evaluated separately. The results show that the perception of bad QoE is mainly determined one-dimensionally by visual artifacts, whereas the perception of good quality shows multiple dimensions. Here, mainly semantic-related features of the content and affective descriptors are used by the naïve test participants. The results show that, with increasing QoE of audiovisual systems, content semantics and users' affective involvement will become important for assessing QoE differences.

8653-11, Session 5

Contrast sensitivity and discrimination of complex scenes

Sophie Triantaphillidou, John Jarvis, Gaurav Gupta, Univ. of Westminster (United Kingdom)

Conference 8653

The aim of the research is to specify experimentally and further model spatial frequency response functions, which quantify human sensitivity to spatial information in real complex images. Two visual response functions are measured: the contextual Visual Perception Function (VPF), which describes visual sensitivity to changes in suprathreshold contrast in a given image; the contextual Contrast Sensitivity Function (CSF), which describes the ability of the visual system to detect any spatial signal in an image. In the paper we present our attempts to derive experimentally and further model the VPF and CSF for a range of scene types and conditions.

Progress is presented in the following areas: Apparatus and laboratory set-up; stimulus acquisition and characterization; imaging system characterization; spatial decomposition and contrast manipulation; methodology for subjective tests. Contextual groups of VPFs and CSFs are presented and compared with 'classical' findings that have used simple sine-wave stimuli. An initial survey of models of human contrast sensitivity suggests that the contrast discrimination framework developed by Barten provides a sound starting position for our own modeling studies. We explain our choice of model and its relevance and use in various image quality metrics.

8653-12, Session 5

Grouping strategies to improve the correlation between subjective and objective image quality data

Silvia Corchs, Francesca Gasparini, Raimondo Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

Image Quality Assessment (IQA) methods can be categorized as subjective versus objective ones. The efficiency of studies that involve people judgments is very low compared to a computerized objective study. However to validate automated approaches, subjective IQA is needed. Great effort is devoted to proper correlate objective and subjective data.

In this paper we address this problem focusing on no reference metrics, and evaluating psycho-visual experiments on JPEG corrupted images.

In general NR metrics are not able to measure with the same performance the distortions within their possible full range and with respect to different image contents.

In this paper we emphasize that the subjective data do not show a monotonic behavior either. In fact the crosstalk between content and distortion signals influences the human perception.

To this end we here propose a strategy that groups the images according to their content so that within each subgroup the perceptual scores can be considered monotonic with respect to the distortion level. This grouping strategy is applied to images of LIVE, MICT and of a new dataset properly generated by the authors. We will also show how these results can significantly help in finding better correlations among NR metrics and psycho-visual data.

8653-13, Session 5

Adapting the ISO 20462 softcopy ruler method for online image quality studies

Peter D. Burns, Burns Digital Imaging (United States); Jonathan B. Phillips, Imaging Scientist (United States); Donald R. Williams, Imaging Science Associates (United States)

The ISO 20462 method for subjective image quality evaluation relies on a set of reference images, which are calibrated in terms of known absolute quality differences. This method was then extended so that the scene content could be presented on computer monitors as Digital Reference Stimuli (DRS). Recently, it has become more common to conduct imaging related surveys via Internet-based participation. In this case, images are not viewed under controlled conditions. However, it can be desirable, e.g., when simulating consumer decisions regarding image editing. In addition, such crowdsourcing has the advantage that large numbers of responses can be acquired efficiently. Our

objective was to develop and test a method that uses a commercial online service that allows presentation of images. There were several limitations to be overcome, including image file size, and a static interface rather than one allowing dynamic updating of the reference image. The method that we tested used reference anchor images rather than the slider-selected reference. However, our anchors were drawn from the ISO 20462 set, and therefore were taken as calibrated reference images, albeit viewed under uncontrolled conditions. We describe the verification study that was completed, and compare results with the corresponding Softcopy Ruler data. We conclude that crowdsourcing is useful for this application. When calibrated subjective image quality measures are needed, our adapted method should be considered an efficient alternative to the ISO 20462 standard, provided that common reference images are used.

8653-27, Session IPI

Generation of PDF with vector symbols from scanned document

Ilya V. Kurilin, Ilia V. Safonov, Michael N. Rychagov, Samsung Electronics Co., Ltd. (Russian Federation); Ho Keun Lee, Sang Ho Kim, Don Chul Choi, Samsung Electronics Co., Ltd. (Korea, Republic of)

The paper is devoted to algorithm for generation of PDF with vector symbols from scanned documents. The complex multi-stage technique includes segmentation of the document to text/drawing areas and background, conversion of symbols to lines and Bezier curves, storing compressed background and foreground. In the paper we concentrate on one-colored regions conversion that comprises segmentation of symbol bodies with resolution enhancement, contour tracing and approximation. Our method outperforms existing solutions according to user opinion survey and secures the best compression rate/quality ratio.

8653-28, Session IPI

Minimum image quality assessment based on saliency maps: a human visual approach

João Barreira, Maximino Bessa, Luís G. Magalhães, INESC TEC (Portugal) and Univ. de Trás-os-Montes e Alto Douro (Portugal)

mage quality assessment as perceived by humans is of crucial importance in numerous fields of image processing. Transmission and storage of digital media require efficient methods to reduce the large number of bits to store an image, while maintaining sufficiently high quality compared to the original image. Since subjective evaluations cannot be performed in various scenarios, it is necessary to have objective metrics that predict image quality consistent with human perception. However, objective metrics that considers high levels of the human visual system are still limited. In this paper, we investigate the possibility of automatically predict, based on saliency maps, the minimum image quality threshold from which humans can perceive the elements on a compressed image. We conducted a series of experimental subjective tests where human observers have been exposed to compressed images with decreasing compression rates. To measure the difference between the saliency map of the compressed and the original image it was used the normalized absolute error metric. Our results indicate that the elements on the image are only perceived by most of the human subjects not at a specific compressed image quality level, but depending on a saliency map difference threshold.

8653-29, Session IPI

Qualification process of CR system and quantification of digital image quality

Patrice GARNIER, Laurence HUN, Jeremie Klein, Catherine LEMERLE, CEA-VALDUC (France)

The LCND (Non Destructive Inspection Laboratory) uses several

Conference 8653

X-Ray generators to carry out many inspections: void search, welding expertise, gap measurements, etc. Most of these inspections are carried out on silver based plates. For several years, the LCND has decided to qualify new devices such as digital plates or CCD/flat panel plates. On the one hand, these technological intelligence is to forecast the assumed and eventual disappearance of silver based plates; on the other hand, it is also compulsory the keep our skills mastering up-to-date and to be a driving force behind our sub-contractants.

The main improvement brought by numerical plates is the continuous progress of the measurement accuracy, especially with image data processing. It is now common to measure defects thickness or depth position within a part. In such applications, data image processing is used to obtain complementary information of scanned silver based plates. This scanning procedure is harmful to the measurement for it implies a data corruption of the resolution, the adding of numerical noise and is time expensive. Digital plates enable to suppress the scanning procedure and to increase the resolution. It is nonetheless difficult to define, for digital images, single criteria for the image quality. A procedure is to be defined to judge the quality of the digital data itself; the impact of the scanning device and the configuration parameters are also to be taken into account.

This presentation deals with the detailed qualification process developed by the LCND for digital plates (DUR-NDT) based on the study of quantitative criteria chosen to define a direct numerical image quality that could be compared with scanned silver based pictures and the classical optical density.

The versatility of the X-Ray parameters is also discussed (X-ray tension, intensity, time exposure). The aim is to be able to transfer the year long experience of the LCND with silver-based plates inspection to these new digital plates supports.

8653-30, Session IPI

Challenges of an automated spectral responsivity characterization system

Mark Helmlinger, Greg A. McKee, Christopher Durell, Labsphere, Inc. (United States)

An essential part of imaging system performance measurement and modeling is the determination of spectral responsivity, namely the in-band band-shape and out-of-band response. These complicated measurements have heretofore been difficult to make with consistency. To address this industry-wide problem, Labsphere is developing an automated spectral response measurement station, incorporating several techniques to enhance accuracy. This presentation will cover the physics and considerations behind the scaling of this highly automated system and the experimental methodology it incorporates to assure absolute traceability, as well as detail some of the lessons learned along the way.

8653-14, Session 6

A line-based HDR sensor simulator for motion artifact prediction

Donald J. Baxter, STMicroelectronics Ltd. (United Kingdom)

Modeling only a camera's lens blur, noise and sensitivity is not sufficient to predict image quality. For a fuller prediction, motion blur/artifacts must be included. Automotive applications are particularly challenging for HDR motion artifacts. This paper extends a classic camera noise model to simulate motion artifacts. The motivation is to predict, visualize and evaluate the motion/lighting flicker artifacts for different image sensor readout architectures. The proposed motion artifact HDR simulator has 3 main components; a dynamic image source, a simple lens model and a line based image sensor model. The line based nature of image sensor provides an accurate simulation of how different readout strategies sample movement or flickering lights in a given scene. Two simulation studies illustrating the model's performance are presented. The first simulation compares the motion artifacts of frame sequential and line interleaved HDR readout while the second study compares the motion blur of an 8MP 1.4μm, 5MP

1.75μm and 3MP 2.2μm image sensors under the same illumination level. Good alignment is obtained between the expected and simulated results.

8653-15, Session 6

On the spectral quality of scanner illumination with LEDs

Luke C. Cui, Lexmark International, Inc. (United States)

Document scanner illumination has evolved along with general illumination technologies. LEDs have become more and more popular as the illumination sources for document scanning. LED technologies provide a wide range of choices both in terms of structural design and spectral composition. In this report, we examine some popular LED technologies used for document scanners and their inherent spectral quality with color rendering simulation and some actual examples. It is hoped that the report will help to clarify the impact of different LEDs on the color quality of document scanners.

8653-16, Session 6

Refined measurement of digital image texture loss

Peter D. Burns, Burns Digital Imaging (United States)

Image texture is the term given to the information-bearing fluctuations such as those for skin, grass and fabrics. Since image processing aimed at reducing unwanted fluctuations (noise are other artifacts) can also remove important texture, good product design requires a balance between the two. The texture-loss MTF method, currently under international standards development, is aimed at the evaluation of digital and mobile-telephone for the capture and retention of image texture. These image fields include pseudo-random objects, such as overlapping disks, often referred to as 'dead leaves' targets. We investigated several aspects of this Pseudo-random Object Target (PROT) method that are aimed at improving its practical application and comparison with other imaging performance measures. We indicate how the texture MTF method can be based on noise-power spectra comparable with other noise sources in the imaging system such as signal quantization and print materials. In addition, the current method relies on modeling (knowledge of) the spatial noise characteristics the test target. For evaluation of image processing functions the method can be improved by computing the input signal spectrum directly from the image data. The adapted analysis method can result in improved and less reliance on data smoothing and fitting. This has a particularly advantage at low spatial frequencies, where texture-MTF scaling is performed when comparing various methods.

8653-17, Session 6

F-MAD: a feature-based extension of the most apparent distortion algorithm for image quality assessment

Punit Singh Banga, Damon M. Chandler, Oklahoma State Univ. (United States)

In this paper, we present an improved version of the Most Apparent Distortion (MAD) quality assessment algorithm [Larson and Chandler, SPIE Image Quality and Sys. Perf., 2009] which improves upon its predecessor by using local feature measurements. MAD's appearance-based stage is quite low-level relying only on log-Gabor statistics. Here we show that higher-level factors such as sharpness, clarity of edges, and color fidelity are good predictors of image quality. We use simple image features (sharpness, contrast, edge strength, luminance distance, and color distance) as a measurement of image quality and incorporate them into MAD's appearance-based model. We have titled the resulting algorithm F-MAD (Feature-Based MAD). Testing on several image-quality databases demonstrates that F-MAD yields improved predictive performance compared to MAD and other modern quality-assessment techniques.

8653-18, Session 6

An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes

Yi Zhang, Damon M. Chandler, Oklahoma State Univ. (United States)

In this paper, we present an algorithm for no-reference image quality assessment, called DESIQUE (DErivative Statistics-Based Image QUality Evaluator), which operates by using log-derivative statistics of natural scenes. Log-derivative statistics are preserved across natural scenes, but are changed in the presence of distortions. To capture these changes, DESIQUE extracts log-derivative statistical features at two image scales in both the spatial and frequency domains, from which the same two-stage framework (distortion identification and distortion-specific quality assessment) used in the recent DIIVINE and BRISQUE algorithms are employed to estimate quality. Testing on several image-quality databases shows that DESIQUE is generally better at quality assessment than DIIVINE, BRISQUE, and other no-reference methods. DESIQUE yields SROCC values of 0.918, 0.913, and 0.870, on the CSIQ, TID, and Toyama databases, respectively. As a comparison, the next best competitor yields SROCC values of 0.900, 0.898, and 0.848 on these databases. Furthermore, in terms of computational efficiency, DESIQUE is competitive with the currently fastest performance algorithms. Additional results and analyses can be viewed online at http://vision.okstate.edu/yi/DESIQUE/.

8653-19, Session 7

On the analysis of wavelet-based approaches for print grain artifacts

Ahmed H. Eid, Brian E. Cooper, Edward E. Rippetoe, Lexmark International, Inc. (United States)

Grain is one of several attributes described in ISO/IEC DTS 24790, a draft technical specification for the measurement of image quality for monochrome printed output. It defines grain as aperiodic fluctuations of lightness greater than about 0.4 cycles per millimeter, a definition inherited from the latest official standard on printed image quality, ISO/IEC 13660.

In a previous publication, we introduced a modification to the ISO/IEC 13660 grain measurement algorithm that includes a band-pass, wavelet-based, filtering step to limit the contribution of high-frequency fluctuations. This modification improves the algorithm's correlation with the subjective evaluation of experts who rated the severity of printed grain artifacts.

Seeking to improve upon the grain algorithm in ISO/IEC 13660, the ISO 24790 committee evaluated several graininess metrics. This led to the selection of the above wavelet-based approach as the top candidate algorithm for inclusion in a future ISO/IEC standard. Compared against subjective evaluations conducted on data from the ISO committee, the proposed algorithm offers much better correlation than the ISO/IEC 13660 approach.

In addition, we introduce an alternative approach for measuring grain defects based on spatial frequency analysis of wavelet-filtered images. We establish a link between the spatial-based grain (ISO/IEC DTS 24790) approach and its equivalent frequency-based one in light of Parseval's theorem.

8653-20, Session 7

A general approach for assessment of print quality

Xiaochen Jing, Purdue Univ. (United States); Stephen Astling, Renee Jessome, Eric Maggard, Terry Nelson, Mark Q. Shaw, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

Laser electrophotographic printers are complex systems that can generate prints with a number of possible artifacts that are very different in nature. It is a challenging task to determine a single processing algorithm that can effectively identify such a wide range of print quality defects.

In this paper, we describe an image processing and analysis pipeline that can effectively assess the presence of a wide range of artifacts, as a general approach. In our paper, we will discuss in detail the algorithms that comprise the image processing and analysis pipeline, and will illustrate the efficacy of the pipeline with a number of examples.

8653-21, Session 7

Verification of proposed ISO methods to measure resolution capabilities of printing systems

Milena Cisarova, Gjøvik Univ. College (Norway) and Univ. of Pardubice (Czech Republic); Marius Pedersen, Peter Nussbaum, Gjøvik Univ. College (Norway); Frans Gaykema, Océ PLT (Netherlands)

Printer resolution is an important attribute, and ISO has proposed a standard "ISO/IEC DTS 29112:2011(E): Information Technology – Office equipment – Test charts and Methods for Measuring Monochrome Printer Resolution" including test charts and methods to measure printer resolution. From a customer's point of view quality is what the customers perceive, and therefore it is often a goal for objective measures to reflect and correlate with perceived quality. We investigate if the objective methods proposed by ISO reflect human judgement. Experiments have been conducted in which we evaluated the proposed objective methods against perceptual data.

8653-22, Session 7

Assessment of presence of isolated periodic and aperiodic bands in laser electrophotographic printer output

Jia Zhang, Purdue Univ. (United States); Stephen Astling, Renee Jessome, Eric Maggard, Terry Nelson, Mark Q. Shaw, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

No Abstract Available

8653-23, Session 8

Printing artificial sweat using ink jet printers for the test set generation in forensics: an image quality assessment of the reproducibility of the printing results

Mario Hildebrandt, Otto-von-Guericke-Univ. Magdeburg (Germany); Jennifer Sturm, Fachhochschule Brandenburg (Germany); Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

In order to use scientific expert evidence in court hearing, several criteria must be met. In the US jurisdiction the Daubert decision has defined several criteria that might be assessed if a testimony is challenged. In particular the potential for testing or actual testing, as well as known or potential error rate are two very important criteria.

In order to be able to compare the results with each other, the reproducible creation of evaluation samples is necessary. However, each latent fingerprint is unique due to external influence factors such as sweat composition or pressure during the application of a trace. Hence, Schwarz introduced a method to print latent fingerprints using ink-jet printers equipped with artificial sweat.

In this paper we assess the image quality in terms of reproducibility and

clarity of the printed artificial sweat patterns. For that, we determine the intra-class variance from one printer on the same and on different substrates based on a subjective assessment, as well as the inter-class variance between different printers of the same model using pattern recognition techniques. Our results indicate that the intra-class variance is primarily influenced by the drying behavior of the amino acid. The inter-class is surprisingly large between identical models of one printer. Our evaluation is performed using 100 samples on an overhead foil and 50 samples of 5 different patterns (a line structure, a fingerprint image, two different arrows and a larger area with amino acid) on a compact disk surface acquired with a Keyence VK-X110 laser scanning confocal microscope. The results show a significant difference between the two identical printers allowing for differentiating between them with an accuracy of up to 99%.

8653-25, Session 8

Figure of merit for macrouniformity based on image quality ruler evaluation and machine learning framework

Weibao Wang, Purdue Univ. (United States); Gary Overall, Travis Riggs, Rebecca Silveston-Keith, Julie Whitney, Lexmark International, Inc. (United States); George T. C. Chiu, Jan P. Allebach, Purdue Univ. (United States)

No Abstract Available

8653-26, Session 8

Wavelet-based figure of merit for macro-uniformity

Xing Liu, Purdue Univ. (United States); Gary Overall, Travis Riggs, Rebecca Silveston-Keith, Julie Whitney, Lexmark International, Inc. (United States); George T. C. Chiu, Jan P. Allebach, Purdue Univ. (United States)

Wavelets are a powerful tool that can be applied to problems in image processing and analysis. They provide a multiscale decomposition of an original image into average terms and detail terms that capture the characteristics of the image at different scales. In this project, we develop a figure of merit for macro-uniformity that is based on wavelets. We use the Haar basis to decompose the image of the scanned page into eleven levels, which we then group into four separate frequency bands, each of which overlaps by two levels with its adjacent frequency bands. Each frequency band image consists of the superposition of the detail images within that band. We next compute 1-D horizontal and vertical projections for each frequency band image. For each frequency band image projection, we develop a structural approximation that summarizes the essential visual characteristics of that projection. For the coarsest band comprising levels 8,9,10,11, we use a generalized square-wave approximation. For the next coarsest band comprising levels 6,7,8,9, we use a piecewise linear spline approximation. For the two finest bands comprising levels 4,5,6,7, and 2,3,4,5, respectively, we use a spectral decomposition. For each 1-D approximation signal, we define an appropriate set of scalar-valued features. These features are used to design two predictors – one based on linear regression and the other based on the support vector machine, which are trained with data from our image quality ruler experiments with human subjects.

Conference 8654: Visualization and Data Analysis 2013

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8654 Visualization and Data Analysis 2013

8654-1, Session 1

An interactive visual testbed system for dimension reduction and clustering of large-scale high-dimensional data

Jaegul Choo, Hanseung Lee, Zhicheng Liu, John Stasko, Haesun Park, Georgia Institute of Technology (United States)

Many of the modern data sets such as text and image data can be represented in high-dimensional vector spaces and have benefited from computational methods that utilize advanced techniques from numerical linear algebra. Visual analytics approaches have contributed greatly to data understanding and analysis due to their capability of leveraging humans' ability for quick visual perception. However, visual analytics targeting large-scale data such as text and image data has been challenging due to limited screen space in terms of both the numbers of data points and features to represent. Among various computational technique supporting visual analytics, dimension reduction and clustering have played essential roles by reducing these numbers in an intelligent way to visually manageable sizes. Given numerous dimension reduction and clustering techniques available, however, decision on choice of algorithms and their parameters becomes difficult. In this paper, we present an interactive visual testbed system for dimension reduction and clustering in a large-scale high-dimensional data analysis. The testbed system enables users to apply various dimension reduction and clustering methods with different settings, visually compare the results from different algorithmic methods to obtain rich knowledge for the data and tasks at hand, and eventually choose the most appropriate path for a collection of algorithms and parameters. Using various data sets such as documents, images, and others that are already encoded in vectors, we demonstrate how the testbed system can support these tasks.

8654-2, Session 1

Multi-focus and multi-level techniques for visualization and analysis of networks with thematic data

Michele Cossalter, Ole J. Mengshoel, Ted Selker, Carnegie Mellon Univ. (United States)

Information-rich data sets bring several challenges in the areas of visualization and analysis, even when associated with node-link network visualizations. This paper presents and validates in a user study an interactive approach to making multiple and multi-step comparisons across different parts of such data sets. We describe NetEx, a visualization tool providing multi-focus and multi-level techniques that enable users to simultaneously explore different parts of a network and its underlying data set. NetEx, implemented as a Cytoscape plug-in, has been applied to the analysis of electrical power networks, Bayesian networks, and the Enron e-mail repository.

In this paper we focus on data from a NASA electrical power network, representative of those found in aerospace vehicles, to demonstrate how NetEx supports the analytical task of diagnosis. Specifically, results from a user study with 25 subjects suggest NetEx enables more accurate isolation of complex faults compared to an especially designed electrical power analysis software tool.

8654-3, Session 2

Visual analytics of cyber physical data streams using spatio-temporal radial pixel visualization

Ming C. Hao, M. Marwah, Hewlett-Packard Labs. (United States); Sebastian Mittelstaedt, Halldor Janetzko, Daniel A. Keim, Univ. Konstanz (Germany); Umeshwar Dayal, Cullen Bash, Carlos Felix, Chandrakant Patel, Meichun Hsu, Yuen Chen, Hewlett-Packard Labs. (United States); Michael Hund, Univ. Konstanz (Germany)

Cyber physical systems (CPS), such as smart buildings and data centers, are richly instrumented systems composed of tightly coupled computational and physical elements that generate large amounts of data. To explore CPS data and obtain actionable insights, we present a new approach called Radial Pixel Visualization (RPV); which uses multiple concentric rings to show the data in a compact circular layout of pixel cells, each ring containing the values for a specific variable over time and each pixel cell representing an individual data value at a specific time. RPV provides an effective visual representation of locality and periodicity of the high volume, multivariate data streams. RPVs may have an additional analysis ring for highlighting the results of correlation analysis or peak point detection. Our real-world applications demonstrate the effectiveness of this approach. The application examples show how RPV can help CPS administrators to identify periodic thermal hot spots, find root-causes of the cooling problems, understand building energy consumption, and optimize IT-services workloads.

8654-4, Session 2

Exploring large scale time-series data using nested timelines

Zaixian Xie, Oracle America Inc. (United States); Matthew Ward, Elke Rundensteiner, Worcester Polytechnic Institute (United States)

When data analysts study time-series data, an important task is to discover how data patterns change over time.

If the dataset is very large, this task becomes challenging. Researchers have developed many visualization techniques to help address this problem. However, little work has been done regarding the changes of multivariate patterns, such as linear trends and clusters, on time-series data. In this paper, we describe a set of history views to fill this gap. This technique works under two modes: merge and non-merge. For the merge mode, merge algorithms were applied to selected time windows to generate a change-based hierarchy. Contiguous time windows having similar patterns are merged first. Users can choose different levels of merging with the tradeoff between more details in the data and less visual clutter in the visualizations. In the non-merge mode, the framework can use natural hierarchical time units or one defined by domain experts to represent timelines. This can help users navigate across long time periods. Grid-based views were designed to provide a compact overview for the history data.

In addition, MDS pattern starfields and distance maps were developed to enable users to quickly investigate the degree of pattern similarity among different time periods. The usability evaluation demonstrated that most participants could understand the concepts of the history views correctly and finished assigned tasks with a high accuracy and relatively fast response time.

Conference 8654

8654-5, Session 3

Visibility-difference entropy for automatic transfer function generation

Philipp C. Schlegel, Renato B. Pajarola, Univ. of Zürich (Switzerland)

Direct volume rendering allows for interactive exploration of volumetric data and has become an important tool in many visualization domains. But the insight and information that can be obtained are dependent on the transfer function defining the transparency of voxels. Constructing good transfer functions is one of the most time consuming and cumbersome tasks in volume visualization. We present a novel general purpose method for automatically generating an initial set of best transfer function candidates. The generated transfer functions reveal the major structural features within the volume and allow for an efficient initial visual analysis, serving as a basis for further interactive exploration in particular of originally unknown data. The basic idea is to introduce a metric as a measure of the goodness of a transfer function which indicates the information that can be gained from rendered images by interactive visualization. In contrast to prior methods, our approach does not require a user feedback-loop, operates exclusively in image space and takes the characteristics of interactive data exploration into account. We show how our new transfer function generation method can uncover the major structures of an unknown dataset within only a few minutes.

8654-6, Session 3

Coherent view-dependent streamline selection for importance-driven flow visualization

Jun Ma, Chaoli Wang, Ching-Kuang Shene, Michigan Technological Univ. (United States)

Streamline visualization can be formulated as the problem of streamline placement or streamline selection. In this paper, we present an importance-driven approach to view-dependent streamline selection that guarantees coher- ent streamline update when the view changes gradually. Given a large number of randomly or uniformly seeded and traced streamlines and sample viewpoints, our approach evaluates, for each streamline, the view-dependent importance by considering the amount of information shared by the 3D streamline and its 2D projection as well as how stereoscopic the streamline's shape is reflected under each viewpoint. We achieve coherent view-dependent streamline selection following a two-pass solution that considers i) the relationships between local viewpoints and the global streamline set selected in a view-independent manner and ii) the continuity between adjacent viewpoints. We demonstrate the effectiveness of our approach with several synthesized and simulated flow fields and compare our view-dependent streamline selection algorithm with a na ?ıve algorithm that selects streamlines solely based on the information at the current viewpoint.

8654-7, Session 3

Single-pass GPU-raycasting for structured adaptive mesh refinement data

Ralf Kaehler, Tom Abel, SLAC National Accelerator Lab. (United States)

Structured Adaptive Mesh Refinement (SAMR) is a popular numerical technique to study processes with high spatial and temporal dynamic range. It minimizes memory requirements by adapting the lattice on which the underlying differential equations are solved to most efficiently represent the solution. Particularly in astrophysics and cosmology such simulations now can capture spatial scales ten orders of magnitude apart and more.

The irregular locations and extensions of the refined regions in the SAMR scheme and the fact that different resolution levels (partially) overlap, poses a challenge for graphics hardware-accelerated direct volume rendering methods. kD-trees have proven to be well-suited to (re-)organize SAMR data into blocks of equally sized cells, but previous GPU-supported approaches for SAMR data using this structure required a separate rendering pass for each of its nodes, limiting the applicability of advanced lighting schemes that access more than one block of cells to operate on. In this paper we present an efficient single-pass GPU-raycasting algorithm for SAMR data that is tailored to current graphics hardware architectures. It employs an kD-tree, efficiently encoded by a set of 3D-textures, to adaptively sample complete rays entirely on the GPU without any CPU interaction. We further discuss two different data storage strategies to access the grid data on the GPU and prove the usefulness of the presented approaches by applying them to several real-world SAMR datasets.

8654-8, Session 4

Multi-user smartphone-based interaction with large high-resolution displays

Lynn Nguyen, Jurgen Schulze, Univ. of California, San Diego (United States)

This abstract is for the Visualization and Data Analysis conference, Number EI107.
This conference expects full paper submissions, which won't fit into this box. The PDF with the full paper has been submitted.

8654-9, Session 4

Stereo frame decomposition for error-constrained remote visualization

Steven L. Martin, Han-Wei Shen, The Ohio State Univ. (United States)

As growth in dataset sizes continues to exceed growth in available bandwidth, new solutions are needed to facilitate efficient visual analysis workflows. Remote visualization can enable the colocation of visual analysis compute resources with simulation compute resources, reducing the impact of bandwidth constraints. While there are many off-the-shelf solutions available for general remoting needs, there is substantial room for improvement in the interactivity they offer, and none focus on supporting stereo remote visualization with programmable error bounds.

We propose a novel system enabling efficient compression of stereo video streams using standard codecs that can be integrated with existing remoting solutions, while at the same time offering error constraints that provide users with fidelity guarantees. By taking advantage of interocular coherence, the flexibility permitted by error constraints, and knowledge of scene depth and camera information, our system offers improved remote visualization frame rates.

8654-10, Session K1

Why high performance visual data analytics is both relevant and difficult (Keynote Presentation)

Edward W. Bethel, Prabhat Prabhat, Surendra Byna, Oliver Ruebel, Kesheng Wu, Michael F Wehner, Lawrence Berkeley National Lab. (United States)

No Abstract Available

8654-11, Session 5

Three-dimensional volume analysis of vasculature in engineered tissues

Mohammed Yousefhussien, Rochester Institute of Technology (United States); Kelley Garvin, Diane Dalecki, Univ. of Rochester (United States); Eli Saber, María Helguera, Rochester Institute of Technology (United States)

Three-dimensional textural and volumetric image analysis holds great potential in understanding the image data produced by multi-photon microscopy. In this paper, an algorithm that quantitatively analyzes the texture and the morphology of vasculature in engineered tissues is proposed. The investigated 3D artificial tissues consist of Human Umbilical Vein Endothelial Cells (HUVEC) embedded in collagen exposed to two regimes of ultrasound standing wave fields under different pressure conditions. Textural features were evaluated using the normalized Gray-Scale Co-occurrence Matrix (GLCM) combined with Gray-Level Run Length Matrix (GLRLM) analysis. To minimize error resulting from any possible volume rotation and to provide a comprehensive textural analysis, an averaged version of nine GLCM and GLRLM orientations is used. To evaluate volumetric features, an automatic threshold using the gray level mean value is utilized. Results show that our analysis is able to differentiate among the exposed samples, due to morphological changes induced by the standing wave fields. Furthermore, we demonstrate that providing more textural parameters than what is currently being reported in the literature, enhances the quantitative understanding of the heterogeneity of artificial tissues.

8654-12, Session 5

3D surface reconstruction and visualization of the Drosophila wing imaginal disc at cellular resolution

Linge Bai, Drexel Univ. (United States); Thomas Widmann, Centro de Genómica e Investigación Oncológica (Spain); Frank Jülicher, Max-Planck-Institut für Physik komplexer Systeme (Germany); Christian Dahmann, Dresden Univ. of Technology (Germany); David E. Breen, Drexel Univ. (United States)

Quantifying and visualizing the shape of developing biological tissues provide information about the morphogenetic processes in multicellular organisms. The size and shape of biological tissues depend on the number, size, shape, and arrangement of the constituting cells. To better understand the mechanisms that guide tissues into their final shape, it is important to investigate the cellular arrangement within tissues. Here we present a data processing pipeline to generate 3D volumetric surface models of epithelial tissues, as well as geometric descriptions of the tissues' apical cell cross-sections. The data processing pipeline includes image acquisition, editing, processing and analysis, 2D cell mesh generation, 3D contour-based surface reconstruction, cell mesh projection, followed by geometric calculations and color-based visualization of morphological parameters. In their first utilization we have applied these procedures to construct a 3D volumetric surface model at cellular resolution of the wing imaginal disc of Drosophila melanogaster. The ultimate goal of the reported effort is to produce tools for the creation of detailed 3D geometric models of the individual cells in epithelial tissues. To date, 3D volumetric surface models of the whole wing imaginal disc have been created, and the apicolateral cell boundaries have been identified, allowing for the calculation and visualization of cell parameters, e.g. apical cross-sectional area of cells. The calculation and visualization of morphological parameters show position-dependent patterns of cell shape in the wing imaginal disc. Our procedures should offer a general data processing pipeline for the construction of 3D volumetric surface models of a wide variety of epithelial tissues.

8654-13, Session 6

Visual exploration and analysis of human-robot interaction rules

Hui Zhang, Indiana Univ.-Purdue Univ. Indianapolis (United States); Michael J. Boyles, Indiana Univ. (United States)

We present a novel interaction paradigm for the visual exploration, manipulation and analysis of human-robot interaction (HRI) rules; our development is implemented using a visual programming interface and exploits key techniques drawn from both information visualization and visual data mining to facilitate the interaction design and knowledge discovery process. HRI is often concerned with manipulations of multi-modal signals, events, and commands that form various kinds of interaction rules. Depicting, manipulating and sharing such design-level information is a compelling challenge. Furthermore, the closed loop between HRI programming and knowledge discovery from empirical data is a relatively long cycle. This, in turn, makes design-level verification nearly impossible to perform in an earlier phase. In our work, we exploit a drag-and-drop user interface and visual languages to support depicting responsive behaviors from social participants when they interact with their partners. For our principal test case of gaze-contingent HRI interfaces, this permits us to program and debug the robots' responsive behaviors through a graphical data-flow chart editor. We exploit additional program manipulation interfaces to provide still further improvement to our programming experience: by simulating the interaction dynamics between a human and a robot behavior model, we allow the researchers to generate, trace and study the perception-action dynamics with a social interaction simulation to verify and refine their designs. Finally, we extend our visual manipulation environment with a visual data-mining tool that allows the user to investigate interesting phenomena such as joint attention and sequential behavioral patterns from multiple multi-modal data streams. We have created instances of HRI interfaces to evaluate and refine our development paradigm. As far as we are aware, this paper reports the first program manipulation paradigm that integrates visual programming interfaces, information visualization, and visual data mining methods to facilitate designing, comprehending, and evaluating HRI interfaces.

8654-14, Session 6

Emotion scents: a method of representing user emotions on GUI widgets

Daniel Cernea, Technische Univ. Kaiserslautern (Germany) and Linnaeus Univ. (Sweden); Christopher Weber, Achim Ebert, Technische Univ. Kaiserslautern (Germany); Andreas Kerren, Linnaeus Univ. (Sweden)

The world of desktop interfaces has been dominated for years by the concept of windows and standardized user interface (UI) components. Still, while supporting the interaction and information exchange between the users and the computer system, graphical user interface (GUI) widgets are rather one-sided, neglecting to capture the subjective facets of the user experience. In this paper we propose a set of design guidelines for visualizing user emotions on standard GUI widgets (e.g. buttons, check boxes, etc.) in order to enrich the interface with a new dimension of subjective information by adding support for emotion awareness as well as post-task analysis and decision making. We highlight the use of an EEG headset for recording the various emotional states of the user while he is interacting with the widgets of the interface. We propose a visualization approach, called emotion scents, that allows users to view emotional reactions corresponding to different GUI widgets without influencing the layout or changing the positioning of these widgets. A use case and an evaluation enforces our findings, suggesting that these emotion-enhanced UI components can enrich desktop interfaces by enabling emotion awareness and subjectivity analysis.

8654-15, Session 7

Visual analysis of situationally aware building evacuations

Jack Guest, Todd Eaglin, Kalpathi R. Subramanian, William Ribarsky, Univ. of North Carolina at Charlotte (United States)

No Abstract Available

8654-16, Session 7

Improving projection-based data analysis by feature space transformations

Matthias Schaefer, Leishi Zhang, Tobias Schreck, Andrada Tatu, Univ. Konstanz (Germany); John A. Lee, Michel Verleysen, Univ. Catholique de Louvain (Belgium); Daniel A. Keim, Univ. Konstanz (Germany)

Generating effective visual embedding of high-dimensional data is difficult - the analyst expects to see the structure of the data in the visualization, as well as patterns and relations. Given the high dimensionality, noise and imperfect embedding techniques, it is hard to come up with a satisfactory embedding that preserves the data structure well, whilst highlighting patterns and avoiding visual clutters at the same time. In this paper, we introduce a generic framework for improving the quality of an existing embedding in terms of both structural preservation and class separation by feature space transformations. A compound quality measure based on structural preservation and visual clutter avoidance is proposed to access the quality of embeddings. We evaluate the effectiveness of our approach by applying it to several widely used embedding techniques using a set of benchmark data sets and the result looks promising.

8654-17, Session 7

Does interactive animation control improve exploratory data analysis of animated trend visualization?

Felwa A. Abukhodair, Simon Fraser Univ. (Canada) and King Abdullah Univ. of Science and Technology (Saudi Arabia); Bernhard E. Riecke, Halil I. Erhan, Chris D. Shaw, Simon Fraser Univ. (Canada)

Effectively analyzing trends of temporal data becomes a critical task especially when the amount of data is large. Motion techniques (animation) for scatterplots make it possible to represent lots of data in a single view and make it easy to identify trends and highlight changes. These techniques have recently become very popular and to an extent successful in describing data in presentations. However, compared to static methods of visualization, scatterplot animations may be hard to perceive when the motions are complex. This paper studies the effectiveness of interactive scatterplot animation as a visualization technique for data analysis of large data. We compared interactive animations with non-interactive (passive) animations where participants had no control over the animation. Both conditions were evaluated for specific as well as general comprehension of the data. While interactive animation was more effective for specific information analysis, it led to many misunderstandings in the overall comprehension due to the fragmentation of the animation. In general, participants felt that interactivity gave them more confidence and found it more enjoyable and exciting for data exploration. In conclusion, interacting with animations of trend visualizations proved to be an effective technique for exploratory data analysis and was significantly more accurate than non-interactive animation alone. With these findings we aim at supporting and guiding the design of interactivity to effectively enhance data exploration in animated visualizations.

8654-18, Session K2

Social media analysis and platform *(Keynote Presentation)*

Meichun Hsu, Hewlett-Packard Labs. (United States)

No Abstract Available

8654-19, Session 10

iMap: a stable layout for navigating large image collections with embedded search

Chaoli Wang, John P. Reese, Huan Zhang, Jun Tao, Robert J. Nemiroff, Michigan Technological Univ. (United States)

Effective techniques for organizing and visualizing large image collections are in growing demand as visual search gets increasingly popular. Targeting an online astronomy archive with thousands of images, we present our solution for image search and clustering based on the evaluation image similarity using both visual and textual information. To lay out images, we introduce iMap, a treemap-based representation for visualizing and navigating image search and clustering results. iMap not only makes effective use of available display area to arrange images but also maintains stable update when images are inserted or removed during the query. We also develop an embedded visualization that integrates image tags for in-place search refinement. We show the effectiveness of our approach by demonstrating experimental results and conducting a comparative user study.

8654-21, Session 10

uVis Studio: an integrated development environment for visualization

Kostas Pantazos, Mohammad A. Kuhail, Soren Lauesen, Shangjin Xu, The IT Univ. of Copenhagen (Denmark)

A toolkit facilitates the visualization development process. The process can be further enhanced by integrating the toolkits in development environments. This paper describes how the uVis toolkit, a formula-based visualization toolkit, has been extended with a development environment, called uVis Studio. Instead of programming, developers apply a Drag-Drop-Set-View-Interact approach to enhance cognitive abilities. Developer bind controls to data, and the Studio gives immediate visual feedback in the design panel. This is a novel feature, called What-You-Bind-Is-What-You-Get. The Studio also provides Modes that allow developers to interact and view the visualization from a developer's and an end-user's perspectives without switching screens, and Data-Auto-Completion; a feature of the Property Grid that helps developers with suggestions to write the correct formulas that refer to controls and database information.

We conducted a usability study with six developers to evaluate if the Studio and its features enhance cognition and facilitate the visualization development. The results show that developers appreciated the Drag-Drop-Set-View-Interact approach, the What-You-Bind-Is-What-You-Get feature, the Data-Auto-Completion and the Modes. Several usability problems were identified, and some suggestions for improvement include: new panels, better presentation of Modes, and better error messages.

8654-22, Session 11

Interactive visual comparison of multimedia data through type-specific views

Russ E. Burtner IV, Shawn Bohn, Debbie A. Payne, Pacific Northwest National Lab. (United States)

When analyzing data sources such as news, blogs, and social media sites to discover actionable intelligence, a wealth of information is available. The data collected from these sources comprises all types of multimedia data: text, image, and video. To date, visual analytic tools generally provide analysis of one mode of data, or as a single mode while attaching additional modes as metadata. An example is a text analysis tool that offers visualizations based on processing the text and allows images to be displayed as auxiliary information attached to the text. While such a tool allows a user to dive deeply into the detail of a specific type of data, it does not provide a global perspective.

Conference 8654

Traditional video and image analytics include digital image processing, pattern recognition, edge finding, boundary and curve detection, region growing, shape identification, feature extraction within individual images or frames of video, content-based image retrieval, and content-based video clip retrieval. Traditional text analytics include information retrieval, lexical analysis, pattern recognition, tagging/annotation, entity extraction, and link and association analysis. These analysis techniques are used in individual visual analytic tools without bringing them together to provide additional insight into multimedia collections. The need for multimedia analysis is particularly acute when the analysis of a single mode of information fails to reveal the larger picture, for example; image analysis does not yet provide semantic understanding.

Canopy, is a visual analytic tool developed to aid information analysts in understanding collections of multimedia. To support the computational needs of multimedia processing, Canopy has been developed as an enterprise application. Canopy has a client application for visual display of multiple interactive visualizations and multiple server components that extract data, process text, process image and video data, and combine results. Canopy can help with analytic problems such as, "I have a large collection of data; help me investigate this collection to determine the most relevant files without my having to watch every movie, view every image, and read all the text"; "I have a picture of interest; help me find other pictures of the same subject or videos taken in the same location"; and "I have a name; help me determine if I have any images or videos of this person." Canopy bootstraps the analysis process by providing visual clues to potential data relationships and highlights connections, providing the user with understanding of all the data and additional structural context. This, in turn, facilitates discovery of previously unknown content as well as unexpected or non-obvious relationships.

Canopy extracts the base types from a compound document. For example, a PowerPoint file will be extracted into the text from the slides, metadata fields such as author, and the images on any slide. Canopy applies the appropriate analysis to each extracted portion of the file based on its mode: text, image, or video. Once a project has processed, the user is presented with multiple interconnected views that tie the extracted types through the relationships found in the structure of the documents. These views allow the user to explore the data and relationships within and across modalities.

Canopy is deployed in multiple locations, and information analysts are discovering previously hidden linkages among the text, image, and video media. Canopy is a novel application that takes multimedia analysis to a new level by applying cross media understanding through visualization and allowing analysts to see relationships within their data regardless of source or media type.

8654-23, Session 11

Evaluating multivariate visualizations on time-varying data

Mark A. Livingston, U.S. Naval Research Lab. (United States); Jonathan W Decker, US Naval Research Lab (United States); Zhuming Ai, U.S. Naval Research Lab. (United States)

Multivariate visualization techniques have been applied to a wide variety of visual analysis tasks and a broad range of data types and sources. Their utility has been evaluated in a modest range of simple analysis tasks. In this work, we extend our previous task to a case of time-varying data. We implemented five visualizations of our synthetic test data: three previously evaluated techniques (Data-driven Spots, Oriented Slivers, and Attribute Blocks), one hybrid of the first two that we call Oriented Data-driven Spots, and an implementation of Attribute Blocks that integrates the temporal slices. We conducted a user study of these five techniques. Our previous finding (with static data) was that users performed best when the density of the target (as measured by the visualization's symbols: spots, slivers, or blocks ``on") was either highest or had the highest ratio to non-target features. We present the results of our study and our conclusions for how the time-varying presentations affected these findings.

8654-24, Session 11

Multi-focus and multi-window techniques for interactive network exploration

Priya Krishnan Sundararajan, Ole J. Mengshoel, Ted Selker, Carnegie Mellon Univ. (United States)

Networks analysts often need to compare nodes in different parts of a network and keep in mind the network structure, node labels and internal details of nodes. Even a moderately-sized network on a screen shows coarse structure; unfortunately it may make detailed structure and node labels unreadable. Zooming in can be used to study details and read node labels but in doing this the network analyst may lose track of already studied details elsewhere in the network. We present multi-focus and multi-window techniques to support interactive exploration of networks. Our work supports the user in partitioning and selectively zooming in the network; data associated with each node can be further inspected using aligned or floating windows. Based on a user's selection of focus nodes, the network is enlarged near the focus nodes. The technique allows the user to simultaneously zoom in on upto 10-20 nodes and their neighborhood while retaining the larger network context. We demonstrate our technique and tool by showing how they support interactive debugging of a Bayesian network model of an electrical power system. In addition, we show that it can visually simplify comparisons across on different types of networks.

8654-25, Session IPI

Effective color combinations in isosurface visualization

Sussan Einakian, Timothy S. Newman, The Univ. of Alabama in Huntsville (United States)

Selection of suitable color combinations is a challenge for visualizations that utilize color as a cue, such as simultaneous visualization of multiple isosurfaces, each in its own color. Use of various color mapping schemes, including heat maps, color spectrum maps, etc., have been explored in information visualization [1], but less attention has focused on employing various artistic and psychological theories of color. Here, we consider applicability of color theories of harmonious, disharmonious, and opponent colors for multiple isosurface visualization. We consider this via user study of preferences for use of each such color theory in this visualization application domain. Little prior work has focused on color choice for isosurface visualization. The work here does build on some prior efforts to apply theories of color (dis)harmony and opponency in information visualization.

In one prior works [1], we have considered some aspects of using color theories (harmony, disharmony, and opponent color) in information visualization. In this work here we consider use of these color theories to drive chose of the color for display of multiple isosurfaces from volumetric datasets. We consider this choice via a user study that considers the practical determination of applicability of these color theories for this isosurface visualization task. Our focus in the study is how easily different regions of the datasets are distinguishable. In this user study the distinctness and preferences of colors were discovered based on user responses. The three datasets considered here are from the Volume Library website [3, 4], including the Engine, DTI (Diffusion Tensor Imaging), and H2O datasets.

We used OpenDx, the visualization open source package on Linux environment for rendering nested isosurfaces. A collection of isovalues and opacity were initially considered for each dataset to see if it was possible to create clear and distinct surfaces. In the end, we found use of a common opacity for all surfaces allowed presentation of a clear result for nested isosurfaces in these datasets that could be used for ready comparison of the theories and color choices. Forty participants were considered in this user study. Participants were asked to make repeated pair-wise comparisons of visualization rendering to report their preferences for each comparison. Full details of the steps of the study will be described in the full paper.

The statistical significance of distinctiveness of each disharmonious, opponent, and harmonious color combination was determined by application of a statistical sign test. Based on such testing, we found

there is a statistically significant difference between disharmony and opponent colors for the Engine and DTI isosurface renderings. In addition, our tests found a statistically significant difference between disharmonious and opponent colors for all three images. A third test was to consider the preferences for disharmonious versus harmonious colors, harmonious versus opponent colors, and disharmonious versus opponent colors for each image. A statistically significant difference was found between the disharmonious and harmonious color combinations for the Engine and H2O datasets. Also, a statistically significant difference was found between the disharmonious and the opponent color combinations for all images. For the Engine and DTI images there is also a statistically significant difference between the harmonious and opponent colors. There was not a statistically significant difference between these two theories in application to the H2O image.

In summary, we found that users appear to find disharmonious color combination and harmonious color combinations to have statistically significant differences (at the 99% confidence level) in perceived distinctiveness. Moreover, harmonious colors were found to be distinctive to the opponent color combinations (also at the 99% confidence level).

We also have considered results by gender and educational level of participants and found no statistically significant difference to exist. Based on the outcome of the user study it appears that the structure and shape of an instance is one of the main issues affecting the perception of suitability of harmonic versus disharmonic versus opponent color combinations. The Engine has a hard structure and there is no interference between different surfaces. The DTI dataset has tissues of brain and there is overlapping between two isosurfaces. The H2O dataset is based on three isosurfaces that have more overlap in the image therefore giving rise to more complex color and opacity interactions when viewed from some directions, resulting in there being no significant difference in user rating of opponent and harmonious combinations.

In conclusion, our study results suggest that the shape of structures in a volumetric dataset effects the perception of which combination is most suitable for nested isosurfaces. However, it appears that hard structures with no interference between isosurfaces were found disharmonious color combination statistically significant difference in distinctiveness. On the other hand, for soft structures, like body tissues, with overlap between isosurfaces the harmonious color combinations were found to be more distinct than disharmonious color combinations.

References:

[1] Einakian, S. and Newman, Timothy S.; Experiments on Effective Color Combination in Map-Based Information Visualization. Proc. Visualization and Data Analysis Conf. '10 (SPIE Vol. 7530), San Jose Jan. 2010.

[2] The Volume Library. [Online] http://schorsch.efi.fh-nuernberg.de/data/volume/. Accessed (March 2012).

[3] The Volume Library. [Online] http://www9.informatik.uni-erlangen.de/External/vollib/. Accessed (March 2012).

8654-26, Session IPI

Web tools for rapid experimental visualization prototyping

Jonathan W. Decker, Mark A. Livingston, U.S. Naval Research Lab. (United States)

Quite often a researcher finds themselves looking at spreadsheets of high-dimensional data generated by experimental models and user studies. We can use analysis to challenge or confirm hypothesis, but unexpected results can easily be lost in the shuffle. For this reason, it would be useful to visualize the results so we can explore our data and make new discoveries.

Web browsers have become increasingly capable for creating complex, multi-view applications. Javascript is quickly becoming a de facto standard for scripting, online and offline. This work demonstrates the use of web technologies as a powerful tool for rapid visualization prototyping.

We have developed two prototypes: One for high-dimensional results of the abELICIT - multi-agent version of the ELICIT platform tasked with collaborating to identify the parameters of a pending attack. Another prototype displays responses to a user study on the effectiveness of multi-layer visualization techniques. We created coordinated multiple views prototypes in the Google Chrome Browser written in Javascript, CSS and HTML. We will discuss the benefits and shortcomings of this approach.

8654-27, Session IPI

Time-based user-movement pattern analysis from location-based social network data

Huey Ling Chuan, Isaraporn Kulkumjon, Surbhi Dangi, Carnegie Mellon Univ. (United States)

Virtual social interactions play an increasingly important role in the discovery of places with digital recommendations. Our hypothesis is that people define the character of a city by the type of places they frequent. With a brief description of our dataset, anomalies and observations about the data, this paper delves into three distinct approaches to visualize the dataset addressing our two goals of: 1. Arriving at a time-based region specific recommendation logic for different types of users classified by the places they frequent. 2. Analyzing the behaviors of users that check-in in groups of two or more people. The study revealed that distinct patterns exist for people that are residents of the city and for people who are short-term visitors to the city. The frequency of visits, however, is both dependent on the time of the day as well as the urban area itself (e.g. eateries, offices, tourist attractions). The observations can be extended for application in food and travel recommendation engines as well as for research in urban analytics, smart cities and town planning.

Keywords: digital recommendations, social interactions, check-in behavior, frequency of visits, travel recommendation engines, urban analytics, smart cities.

8654-28, Session IPI

Visualizing vascular structures in virtual environments

Thomas Wischgoll, Wright State Univ. (United States)

In order to learn more about the cause of coronary heart diseases and develop diagnostic tools, the extraction and visualization of vascular structures from volumetric scans for further analysis is an important step. By determining a geometric representation of the vasculature, the geometry can be inspected and additional quantitative data calculated and incorporated into the visualization of the vasculature. To provide a more user-friendly visualization tool, virtual environment paradigms can be utilized. This paper describes techniques for interactive rendering of large-scale vascular structures within virtual environments. This can be applied to almost any virtual environment configuration, such as CAVE-type displays. Specifically, the tools presented in this paper were tested on a Barco I-Space and a large 62x108 inch passive projection screen with a Kinect sensor for user tracking.

8654-29, Session IPI

A combined multidimensional scaling and hierarchical clustering view for the exploratory analysis of multidimensional data

Paul Craig, Néna Roa-Seïler, Univ. Technológica de la Mixteca (Mexico)

This paper describes the a novel information visualization technique that combines multidimensional scaling and hierarchical clustering to support the exploratory analysis of multidimensional data. The technique displays the results of multidimensional scaling using a

scatter plot where the closeness of any two items' representation's are approximate to their similarity according to a Euclidean distance metric. The results of hierarchical clustering are overlaid onto this view by drawing smoothed outlines around each nested cluster. The difference in similarity between successive cluster combinations is used to colour code clusters and make stronger natural clusters more prominent in the display. When a cluster or group of items is selected, multidimensional scaling and hierarchical clustering are re-applied to a filtered subset of the data, and animation is used to smooth the transition between successive filtered views. As a case study we demonstrate the technique being used to analyse survey data relating to the appropriateness of different phrases to different emotionally charged situations.

8654-30, Session IPI

Visualization of decision processes using a cognitive architecture

Mark A. Livingston, Arthi Murugesan, Derek Brock, Wende K. Frost, Dennis Perzanowski, U.S. Naval Research Lab. (United States)

Cognitive architectures are computational theories of reasoning the human mind engages in as it processes facts and experiences. A cognitive architecture uses declarative and procedural knowledge to represent mental constructs that are involved in decision making. Employing a model of behavioral and perceptual constraints derived from a set of one or more scenarios, the architecture reasons about the most likely consequence(s) of a sequence of events. Reasoning of any complexity and depth involving computational processes, however, is often opaque and challenging to comprehend. Arguably, for decision makers who may need to evaluate or question the results of autonomous reasoning, it would be useful to be able to inspect the steps involved in an interactive, graphical format. When a chain of evidence and constraint-based decision points can be visualized, it becomes easier to explore both how and why a scenario of interest will likely unfold in a particular way.

In this paper, we present initial work on a scheme for visualizing autonomous reasoning that produces graphical representations of models run in the Polyscheme cognitive architecture. First, we give a brief overview of the architecture and note the key types of data that are critical for visual representations of cognitively-based, computational reasoning mechanisms. We propose an algorithm to generate visualizations of model-based reasoning, and discuss properties of our technique that pose challenges for our representation goals. Finally, we present example visualizations and simple interactions with the underlying chain of reasoning. We conclude with a summary of feedback solicited from domain experts and practitioners in the field of cognitive modeling.

8654-31, Session IPI

Vortex core timelines and ribbon summarizations: flow summarization over time and simulation ensembles

Alexis Yee Lyn Chan, Joohwi Lee, Russell M. Taylor II, The Univ. of North Carolina at Chapel Hill (United States)

This paper presents two new vortex-summarization techniques designed to portray vortex motion over an entire simulation and over an ensemble of simulations in a single image. Linear "vortex core timelines" with cone glyphs summarize flow over all time steps of a single simulation, with color varying to indicate time. Simplified "ribbon summarizations" with hue nominally encoding ensemble membership and saturation encoding time enable direct visual comparison of the distribution of vortices in time and space for a set of simulations. Feedback from domain experts indicates that they found these techniques to be important and effective for summarizing and comparing time-varying flows across an ensemble of simulations.

8654-32, Session IPI

X3DBio2: A visual analysis tool for biomolecular structure comparison

Hong Yi, Sidharth Thakur, Renaissance Computing Institute (United States); Latsavongsakda Sethaphong, Yaroslava G. Yingling, North Carolina State Univ. (United States)

A major problem in structural biology is the recognition of differences and similarities between related three dimensional (3D) biomolecular structures. Investigating these structure relationships is important not only for understanding of functional properties of biologically significant molecules, but also for development of new and improved materials based on naturally-occurring molecules. We developed a new visual analysis tool, X3DBio2, for 3D biomolecular structure comparison and analysis. The tool is designed for elucidation of structural effects of mutations in proteins and nucleic acids and for assessment of time dependent trajectories from molecular dynamics simulations. X3DBio2 is a freely downloadable open source software and provides tightly integrated features to perform many standard analysis and visual exploration tasks. We expect this tool can be applied to solve a variety of biological problems and illustrate the use of the tool on the example study of the differences and similarities between two proteins of the glycosyltransferase family 2 that synthesize polysaccharides oligomers. The size and conformational distances and retained core structural similarity of SpsA to K4CP represent significant epochs in the evolution of inverting glycosyltransferases.

8654-33, Session IPI

Improvement of web-based data acquisition and management system for GOSAT validation lidar data analysis

Hiroshi Okumura, Shoichiro Takubo, Takeru Kawasaki, Indra N. Abdullah, Saga Univ. (Japan); Osamu Uchino, Isamu Morino, Tatsuya Yokota, National Institute for Environmental Studies (Japan); Tomohiro Nagai, Tetsu Sakai, Takashi Maki, Meteorological Research Institute (Japan); Kohei Arai, Saga Univ. (Japan)

1.BACKGROUND

Global warming has become a very serious issue for human beings. Scientists have suggested that, at the rate the Earth's temperature is rising, an extreme form of global climate change could occur in a few centuries. In 1997, the Kyoto Protocol was adopted at the Third Session of the Conference of the Parties to the United Nations Framework Convention on Climate Change (COP3), making it mandatory for developed nations to reduce carbon dioxide emissions by six to eight per cent of their total emissions in 1990, and to meet this goal sometime between 2008 and 2012. Furthermore, the Global Climate Observation System (GCOS) has been proposed by the World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) in order to strengthen observations of land, ocean, and space conducted by each country. So far, the number of ground-based carbon dioxide observation points has been limited, and they have been distributed unequally throughout the world. Greenhouse gas Observation SATellite (GOSAT) enables the precise monitoring of the density of carbon dioxide by combining global observation data sent from space with data obtained on land, and with simulation models. In addition, observation of methane, another Greenhouse gas, has been considered. For validation of GOSAT data products, we have continued ground-base observation with Fourier Transform Spectrometer (FTS), aerosol lidar and ozone-DIAL lidar at Saga University, JAPAN since March, 2011.

2. DETAILS

For lidar data analysis, not only acquired lidar data but also meteorological data are required.

We develop an web-base data acquisition and management system for effective

acquisition of meteorological data and management of acquired lidar data.

The system consists of data acquisition part and data management part.

Data acquisition part written in Perl language acquires AMeDAS ground-

level meteorological data, RawinSonde high altitude meteorological data, ground-level oxidant data, skyradiometer data, skyview camera images, meteorological satellite IR image data and GOSAT validation lidar data.

AMeDAS (Automated Meteorological Data Acquisition System) is a high-resolution surface observation network developed by Japan Meteorological Agency (JMA) used for gathering regional weather data and verifying forecast performance. Begun operation on

November 1, 1974, the system consists of about 1,300 stations with automatic observation equipment. These stations, of which more than 1,100 are unmanned, are located at an average interval of 17 km throughout Japan.

These AMeDAS meteorological data are acquired from Saga city station every hour.

RawinSonde are essential to obtain high altitude temperature and air pressure data.

These RawinSonde meteorological data are acquired from Fukuoka city station twice (00Z and 12Z) a day.

Ground-level oxidant data are required for analysis of ozone-DIAL lidar data.

These ground-level oxidant data are acquired from Saga city station every hour.

Skyradiometer is required to observe optical thickness.

Skyradiometer data are acquired every hour.

Skyview camera images and meteorological satellite IR image data are required to confirm clouds above lidar.

Skyview camera images are acquired every 5 minutes in daytime and meteorological satellite IR image data are acquired every hour.

GOSAT validation lidar data are manually uploaded via secure FTP.

Data management part written in PHP language demonstrates satellite-pass date and all acquired data and provides interactive graphical user interface.

3. IMPROVEMENT OF THE SYSTEM

For higher performance and higher data usability, ozone vertical profile estimated with global ozone transportation model, optical thickness chart and automatic calculation function of Rayleigh molecule distinction and backscatter coefficient from actual RawinSonde data and US standard atmosphere model are implemented for aerosol lidar and ozone-DIAL lidar data analysis.

8654-35, Session IPI

Optimizing threshold for extreme scale analysis

Robert J. H. Maynard, Kitware, Inc. (United States); Kenneth Moreland, Sandia National Labs. (United States); Utkarsh Atyachit, Berk Geveci, Kitware, Inc. (United States); Kwan-Liu Ma, Univ. of California, Davis (United States)

As the HPC community starts focusing its efforts towards exascale, it becomes clear that we are looking at machines with a billion way concurrency. Although parallel computing has been at the core of the performance gains achieved until now, scaling over 1,000 times the current concurrency can be challenging. As discussed in this paper, even the smallest memory access and synchronization overheads can cause major bottlenecks at this scale. As we develop new software and adapt existing algorithms for exascale, we need to be cognizant of such pitfalls. In this paper, we document our experience with optimizing a fairly common and parallelizable visualization algorithm, threshold of cells based on scalar values, for such highly concurrent architectures. Our experiments help us identify design patterns that can be generalized for other visualization algorithms as well. We discuss our implementation within the Dax toolkit, which is a framework for data

analysis and visualization at extreme scale. The Dax toolkit employs the patterns discussed here within the framework's scaffolding to make it easier for algorithm developers to write algorithms without having to worry about such scaling issues.

8654-36, Session IPI

Perceptualization of geometry using intelligent haptic and visual sensing

Jianguang Weng, Zhejiang Univ. (China); Hui Zhang, Indiana Univ.-Purdue Univ. Indianapolis (United States)

We present a set of paradigms for investigating geometric structures using haptic and visual sensing. Our principal test cases include smoothly embedded geometry shapes such as knotted curves embedded in 3D and knotted surfaces in 4D, that contain massive intersections when projected to one lower dimension. One can exploit a touch-responsive 3D interactive probe to haptically override this conflicting evidence in the rendered images, by forcing continuity in the haptic representation to emphasize the true topology. In our work, we exploited a predictive haptic guidance, a "computer-simulated hand" with supplementary force suggestion, to support intelligent exploration of geometry shapes that will smooth and maximize the probability of recognition. The cognitive load can be reduced further when enabling an attention-driven visual sensing during the haptic exploration. Finally, we proposed a family of haptic based "rolling" methods to facilitate the haptic manipulation, for example, just as we can explore every part of an ordinary knot by rolling it between our hands, we can use this "rolling method" to optimize the combined viewable and touchable aspects of the geometry by cleverly exploiting interactive continuity. All these methods combine to reveal the full richness of the haptic exploration of geometric structures, and to overcome the limitations of traditional 4D visualization.

8654-37, Session IPI

Review of chart recognition in document images

Yan Liu, Xiaoqing Lu, Yeyang Qin, Zhi Tang, Peking Univ. (China); Jianbo Xu, Peking Univ. Founder Group Corp. (China)

As an effective information transmitting way, chart is widely used to represent scientific statistics datum in books, research papers, newspapers etc. Though textual information is still the major source of data, there has been an increasing trend of introducing graphs, pictures, and figures into the information pool. Text recognition techniques for documents have been accomplished using optical character recognition (OCR) software. Chart recognition techniques as a necessary supplement of OCR for document images are still an unsolved problem due to the great subjectiveness and variety of charts styles. This paper reviews the development process of chart recognition techniques in the past decades and presents the focuses of current researches. The whole process of chart recognition is presented systematically, which mainly includes three parts: chart segmentation, chart classification, and chart Interpretation. In each part, the latest research work is introduced. In the last, the paper concludes with a summary and promising future research direction.

Conference 8655: Image Processing: Algorithms and Systems XI

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8655 Image Processing: Algorithms and Systems XI

8655-2, Session 1

Robust face recognition algorithm for identifition of disaster victims

Wouter J. R. Gevaert, Eindhoven Univ. of Technology (Netherlands) and Univ. College West Flanders (Belgium); Peter H. N. de With, Eindhoven Univ. of Technology (Netherlands)

We present a robust face recognition algorithm for the identification of occluded, injured and mutilated faces with a limited training set per person. In such cases, the conventional face recognition methods fall short due to specific aspects in the classification. The proposed algorithm involves recursive Principle Component Analysis for reconstruction of affected facial parts, followed by a feature extractor based on Gabor wavelets and uniform multi-scale Local Binary Patterns. As a classifier, a Radial Basis Neural Network is employed. In terms of robustness to facial abnormalities, tests show that the proposed algorithm outperforms conventional face recog- nition algorithms like, the Eigenfaces approach, Local Binary Patterns and the Gabor magnitude method. To mimic real-life conditions in which the algorithm would have to operate, specific databases have been constructed and merged with partial existing databases and jointly compiled. Experiments on these particular databases show that the proposed algorithm achieves recognition rates beyond 95%.

8655-3, Session 1

Target re-identification in low-quality camera networks

Federica Battisti, Marco Carli, Giovanna Farinella, Alessandro Neri, Univ. degli Studi di Roma Tre (Italy)

Person re-identification trough camera network deal with the problem of finding correct link between consecutive observations of the same target among different cameras choosing the most probable correspondence among a set of possible matches. This paper can be divided in two parts. First of all, the typology of illumination changes viewed from a target while crossing our network is learned. The results show that the changing involves the intensity. Afterwards, a new links pruning system called TCS (Target Color Structure) is proposed: the link between two observations is kept if there is coherence in the color structure of clothes worn by targets.

The color structure is evaluated using a novel color descriptor called CSD (Color Structure Descriptor) which is intensity scale and shift invariant. The CSD describes the difference in term of dominant colors between two regions of interest: the shirt and the pants of a target.

The proposed TCS control can be run over whatever person re-identification algorithm; in this paper we used hue histogram intersection to assign a similarity score to the survived links.

Results show that the improvements achieved applying TCS control are up to 4% for the top rank and up to 16% considering the first eleven more similar candidates.

8655-5, Session 2

Improved image copyright protection scheme exploiting visual cryptography in wavelet domain

Alessio Meneghetti, Giulia Boato, Francesco G. B. De Natale, Univ. degli Studi di Trento (Italy)

The last few years have seen a massive increment in the use of the internet as a channel for sharing and transmitting data, so that several copyright protection schemes have been proposed. The idea of embedding a watermark directly in the data is however unacceptable in various fields of application, due to the intrinsic degradation introduced by non reversible watermarking schemes. Hence some zero watermarking schemes have been developed. In this work we propose an optimization of a recent watermarking method based on visual cryptography, by improving results against most common types of attacks and achieving a higher perceptual quality of the extracted mark.

8655-6, Session 2

HDR image multi-bit watermarking using bilateral-filtering-based masking

Vassilis Solachidis, Emanuele Maiorana, Patrizio Campisi, Univ. degli Studi di Roma Tre (Italy)

The present paper proposes a multi-bit watermarking method for High Dynamic Range (HDR) images. The proposed approach is designed in order to guarantee the mark imperceptibility in both the HDR marked image and its Low Dynamic Range (LDR) counterpart, and to be robust against significant non-linear distortions such as those performed by tone mapping operators (TMOs). The paper offers a detailed review on the state-of-the-art on HDR watermarking, and then presents the proposed scheme where the watermark is embedded in the wavelet transform of the Just Noticeable Difference (JND)-scaled space of the original HDR image. A visual mask taking into account specific aspects of the Human Visual System (HVS) regarding the sensibility to the modifications in the details part of an image, as well as the dependence of the perceived contrast on orientation and spatial location, is also defined and employed to modulate the watermark intensity for each wavelet coefficient. The procedure to detect and extract a multi-bit message is then highlighted, and an extensive set of experimental results is provided to testify the effectiveness of the proposed scheme in embedding watermarks into HDR images without affecting the visual quality of the original image, while being robust against TMOs.

8655-7, Session 2

Body-part estimation from Lucas-Kanade tracked Harris points

Vladimir Pribula, Roxanne L. Canosa, Rochester Institute of Technology (United States)

Skeleton estimation from single-camera grayscale images is generally accomplished using model-based techniques. Multiple cameras are sometimes used; however, skeletal points extracted from a single subject using multiple images are usually too sparse to be helpful for localizing body parts. For this project, we use a single viewpoint without any model-based assumptions to identify a central source of motion, the body, and its associated extremities. Harris points are tracked using Lucas-Kanade refinement with a weighted kernel found from expectation maximization. The algorithm tracks key image points and trajectories and re-represents them as complex vectors describing the motion of a specific body part. Normalized correlation is calculated from these vectors to form a matrix of graph edge weights, which is subsequently partitioned using a graph-cut algorithm to identify dependent trajectories. The resulting Harris points are clustered into rigid component centroids using mean shift, and the extremity centroids are connected to their nearest body centroid to complete

the body-part estimation. We collected ground truth labels from seven participants for body parts that are compared to the clusters given by our algorithm.

8655-8, Session 2

Hue processing in tetrachromatic spaces

Alfredo Restrepo, Univ. de los Andes (Colombia)

In the tetrachromatic hypercube the luminance is given by the midrange μ of the tetrad of coordinates of a colour point while the chromatic saturation is given by the range of the tetrad.

The boundary of the tetrachromatic hypercube consists of 8 solid cubes. Two types of closed surface that are unions of the square faces of the cubes in this boundary are to be found: topological spheres and topological tori . By making the coordinates of the points dependant in special ways on such surfaces, tetrachromatic hues result. A tetrachromatic hue is 2-dimensional and, together with the luminance and the chromatic saturation, it uniquely specifies a tetrad in the hypercube. These hue surfaces can be exploited to visualize tetra chromatic images in such a way that certain aspects are made conspicuous. The hue may then be processed by automorphisms either of a hue sphere or of a hue torus. It is possible to identify 8 (intersecting) dodecahedra and also three (intersecting) PL tori, so it is possible to define a hue in multiple ways. As the hue surfaces are rotated or otherwise automorphed, the colours of a tetrachromatic image are changed in interesting ways. The automorphisms respect the continuity; the rotations respect the antipodicity or complementary colours as well. By processing the hue of a tetrachromatic image (for example, a 4-spectral satellite image) and then visualizing it by feeding the RGB channels of a normal visualizing system with 3 of the 4 bands of the tetrachromatic image, as is commonly done in satellite imaginery, di?erent aspects

8655-9, Session 3

Decomposition of satellite derived images for the distinction of cloud types features

Jules R. Dim, Hiroshi Murakami, Japan Aerospace Exploration Agency (Japan)

Linear filtering methods using convolution techniques are applied in computer vision, to detect spatial discontinuities in the intensity of luminance in photographs images. These techniques are based on the principal that, a pixel's neighborhood in an image, contains information about its intensity. The variation of this intensity provides information about the distribution and possible decomposition of the image in specific features based on the relative position of the pixel at the edge or not in the image. The use of these principals on polar orbit satellite derived thermal images can provide alternative methods to cloud types' classifications, often based on optical or thermodynamic properties of the clouds. A product from satellite infrared images, the Cloud top temperatures (CTT), is used to compute local histogram edge gradients. The segmentation of these edge gradients histograms and their interpretation based on a satellite remote sensing analyses allows the distinction of morphological cloud features, which combined to the altitude of occurrence of the cloud yield a separation of the image in different cloud types. The interpretation of the distribution of these cloud morphological features, and their frequency is evaluated with another cloud classification method (based on cloud optical properties) in order to understand the implications of the results obtained on a general discussion about climate variations.

8655-10, Session 3

Locally tuned inverse sine nonlinear technique for color image enhancement

Saibabu Arigela, Vijayan K. Asari, Univ. of Dayton (United States)

In this paper, a novel inverse sine nonlinear transformation based

image enhancement technique is proposed to improve the visual quality of images captured in extreme lighting conditions. This method is adaptive to the local information of a pixel and simple to implement. The proposed technique consists of four main stages namely histogram adjustment, dynamic range compression, contrast enhancement and nonlinear color restoration. Histogram adjustment on each spectral band is performed to belittle the effect of illumination. Dynamic range compression is accomplished by an inverse sine nonlinear function with a locally tunable image dependent parameter based on the local statistics of each pixel's neighborhood regions of the luminance image. A nonlinear color restoration process based on the chromatic information and luminance of the original image is employed. A statistical quantitative evaluation is performed with the state of the art techniques to analyze and compare the performance of the proposed technique. The proposed technique is also tested on face detection in complex lighting conditions. The results of this technique on images captured in hazy/foggy weather environment are also presented. The evaluation results confirm that the proposed method can be applied to surveillance, security applications in complex lighting environments.

8655-12, Session 3

Fusing electro-optic and infrared signals for high resolution visible images: part I

Xiaopeng Huang, Stevens Institute of Technology (United States); Ravi Netravali, Columbia Univ. (United States); Hong Man, Victor Lawrence, Stevens Institute of Technology (United States)

No Abstract Available

8655-13, Session 4

Embedding high dynamic range tone mapping in JPEG compression

Jian Liu, The Univ. of Akron (United States); Firas Hassan, Ohio Northern Univ. (United States); Joan E. Carletta, The Univ. of Akron (United States)

A method that integrates tone mapping for high dynamic range (HDR) gray-scale images with JPEG compression is proposed. The tone mapping operator (TMO) is block-based, and structured so that the same discrete cosine transform (DCT) that is used for the JPEG compression serves to complete a major part of the tone-mapping operation. Simulations have been done on high dynamic range images from the Debevec library. Experimental results show the technique successfully tone maps and compresses simultaneously; the number of bits per pixel is reduced from 32 to an average of 0.67 by the compression, with an average PSNR of 56.3 dB for the compressed tone-mapped images compared to images that have been only tone-mapped. The output of the proposed method is an image that requires only limited storage space, and can be decompressed with a standard JPEG decoder.

8655-14, Session 4

Formulation, analysis, and hardware implementation of chaotic dynamics based algorithm for compression and feature recognition in digital images

Chance M. Glenn, Alabama A&M Univ. (United States) and Rochester Institute of Technology (United States); Srikanth Mantha, Sajin George, Deepti Atluri, Antonio Mondragon, Rochester Institute of Technology (United States)

In this presentation we will discuss the utilization of a set of waveforms derived from chaotic dynamical systems for compression and feature recognition in digital images. We will also describe the design and testing of an embedded systems implementation of the algorithm. We will show that a limited set of combined chaotic oscillations are

sufficient to form a basis for the compression of thousands of digital images. We will demonstrate this in the analysis of images extracted from the solar heliospheric observatory (SOHO), showing that we are able to detect coronal mass ejections (CMEs) in quadrants of the image data during a severe solar event. We undertake hardware design in order to optimize the speed of the algorithm, taking advantage of its parallel nature. We compare the calculation speed of the algorithm in compiled C, enhanced Matlab, Simulink, and in hardware.

8655-15, Session 4

Quality constraint and rate-distortion optimization for predictive image coders

Khouloud Samrouth, François Pasteau, Olivier Deforges, Institut National des Sciences Appliquées de Rennes (France)

Next generations of image and video coding methods should of course be efficient in terms of compression, but also propose advanced functionalities. Among these functionalities such as scalability, lossy and lossless coding, data protection, Rate Distortion Optimization (RDO) and Rate Control (RC) are key issues. RDO aims at optimizing compression performances, while RC mechanism enables to exactly compress at a given rate. A less common functionality than RC, but certainly more helpful, is Quality Control (QC): the constraint is here given by the quality. In this paper, we introduce a joint solution for RDO and QC applied to a still image codec called Locally Adaptive Resolution (LAR), providing scalability both in resolution and SNR and based on a multi-resolution structure. The technique does not require any additional encoding pass. It relies on a modeling and estimation of the prediction errors obtained in an early work. First, quality constraint is applied and propagated through the whole resolution levels called pyramid. Then, the quantization parameters are deduced considering inter and intra pyramid level relationships. Results show that performances of the proposed method are very close to an exhaustive search solution.

8655-16, Session 5

Visual quality analysis for images degraded by different types of noise

Nikolay Ponomarenko, Vladimir V. Lukin, Oleg Ieremeyev, National Aerospace Univ. (Ukraine); Karen O. Egiazarian, Jaakko Astola, Tampere Univ. of Technology (Finland)

Images corrupted by different types of noise are considered. Visual quality for equal input MSE but three types of the noise, namely, Poisson, additive and multiplicative, is inspected (assessed in experiments by observers) and characterized by visual quality metrics. Color images from the database TID 2008 are exploited. It is shown that metrics do not adequately describe image visual quality for the considered situation. Possible reasons for this are analyzed. The main reasons are masking effects in textural regions and observer attention to special objects present in images.

8655-17, Session 5

Graph cut and image intensity-based splitting improves nuclei segmentation in high-content screening

Muhammad Farhan, Pekka Ruusuvuori, Tampere Univ. of Technology (Finland); Mario Emmenlauer, Pauli Rämö, Univ. Basel (Switzerland); Olli Yli-Harja, Tampere Univ. of Technology (Finland); Christoph Dehio, Univ. Basel (Switzerland)

Quantification of phenotypes in high-content screening experiments depends on the accuracy of single cell analysis. In such analysis workflows, cell nuclei segmentation is typically the first step and is followed by cell body segmentation, feature extraction, and complicated data analysis workflows. Therefore, it is of utmost importance that the first steps of high-content analysis are done correctly in order to guarantee correctness of the final analysis results. Existing nuclei segmentation methods, when applied individually, tend to fuse as well as to split nuclei along with giving suboptimal separation of touching nuclei. This results in loss of many interesting biological phenotypes. Inability to find all the phenotypes causes their misclassification which leads to inaccurate subsequent biological analysis. Here, we proposed a novel cell nuclei segmentation framework which exploits robustness of graph cut to obtain initial segmentation for image intensity-based clump splitting method to deliver the accurate overall segmentation. By using quantitative benchmarks and qualitative comparison of real-world images with complicated multinucleate cells, we show that our method outperforms the other state-of-the-art nuclei segmentation methods. Moreover, we provide a modular and easy-to-use implementation of the method for a widely used platform to make it available for routine use in high-content image analysis.

8655-18, Session 5

Near real-time skin deformation mapping

Steve T. Kacenjar, Lockheed Martin Corp. (United States); Suzie Chen, Rutgers, The State Univ. of New Jersey (United States); Madiha Jafri, Lockheed Martin Corp. (United States); Brian Wall, Rutgers University (United States); Richard Pedersen, Lockheed Martin Corp. (United States); Richard Bezozo M.D., MoleSafe USA (United States)

A novel in vivo approach is described that provides large area mapping of the mechanical properties of the skin in human patients. Such information is important in the understanding of skin health, cosmetic surgery[1], aging, and impacts of sun exposure.

Currently, several methods have been developed to estimate the local biomechanical properties of the skin, including the use of a physical biopsy of local areas of the skin (in vitro methods) [2, 3, and 4], and also the use of non-invasive methods (in vivo) [5, 6, and 7]. All such methods examine localized areas of the skin.

Our approach examines the local elastic properties via the generation of field displacement maps of the skin created using time-sequence imaging [9] with 2D digital imaging correlation (DIC) [10]. In this approach, large areas of the skin are reviewed rapidly, and skin displacement maps are generated showing the contour maps of skin deformation. These maps are then used to precisely register skin images for purposes of diagnostic comparison.

This paper reports on our mapping and registration approach, and demonstrates its ability to accurately measure the skin deformation through a described nulling interpolation process. The result of local translational DIC alignment is compared using this interpolation process. The effectiveness of the approach is reported in terms of residual RMS, image entropy measures, and differential segmented regional errors.

8655-20, Session 6

Object segmentation using graph cuts based edges features

Yuki Masumoto, Weiwei Du, Nobuyuki Nakamori, Kyoto Institute of Technology (Japan)

The paper presents a simple graph cuts algorithm based edges features to object segmentation problems. The user gives some scribbles to background and foreground of an image. Gaussian mixture models(GMMs) are built based on the scribbles. The pixel without scribble belongs to the background or the foreground depending on the relative probability of each pixel. The contribution of our paper is to add edges features to GMMs. The approach is applied with images from the Grab cuts segmentation database. The approach is suitable for images with noise and in the foreground and background with similar colors.

Conference 8655

8655-21, Session 6

A hybrid skull-stripping algorithm based on adaptive balloon snake models

Hung-Ting Liu, Tony W. H. Sheu, Herng-Hua Chang, National Taiwan Univ. (Taiwan)

Skull-stripping is one of the most important preprocessing steps in neuroimage analysis. We proposed a hybrid algorithm based on an adaptive balloon snake model to handle this challenging task. The proposed framework consists of two stages: first, the fuzzy possibilistic c-means (FPCM) is used for voxel clustering, which provides a labeled image for the snake contour initialization. At the second stage, the contour is initialized outside the brain surface based on the FPCM result and evolves under the guidance of the balloon snake model, which drives the contour with an adaptive inward normal force to capture the boundary of the brain. The similarity indices indicates that our method outperformed the BSE and BET methods in skull-stripping the MR image volumes in the IBSR data set. Experimental results show the effectiveness of this new scheme and potential applications in a wide variety of skull-stripping applications.

8655-47, Session 6

Approximations to camera sensor noise

Keigo Hirakawa, Xiaodan Jin, Keigo Hirakawa, Univ. of Dayton (United States)

Noise is a "necessary evil" in image acquisition that is present in all image sensor data. Poisson distribution is said to model the stochastic nature of the photon arrival process, while it is common to approximate readout/thermal noise by additive white Gaussian noise (AWGN). Other sources of signal-dependent noise such as Fano and quantization also contribute to the overall noise profile. Question remains, however, about how best to model the combined sensor noise.

Though additive Gaussian noise with signal-dependent noise variance and Poisson corruption are two widely used models to approximate the actual sensor noise distribution, the the justification given to these types of models are based on limited evidence. The goal of this paper is to provide a more comprehensive characterization of random noise.

We concluded by presenting concrete evidence that Poisson model is a better approximation to real camera model than AWGN. We suggest further modification to Poisson that may improve the noise model.

8655-22, Session 7

Fast noise variance estimation by principal component analysis

Stanislav Pyatykh, Lei Zheng, Jürgen Hesser, Heidelberg Univ. (Germany)

Noise variance estimation is required in many image denoising, compression, and segmentation applications. In this work, we propose a fast noise variance estimation algorithm based on principal component analysis of image blocks. First, we rearrange image blocks into vectors and compute the covariance matrix of these vectors. Then, we use Bartlett's test in order to select the covariance matrix eigenvalues, which correspond only to noise. This allows estimating the noise variance as the average of these eigenvalues. Since the maximum possible number of eigenvalues corresponding to noise is utilized, it is enough to process only a small number of image blocks, which allows reduction of the execution time. The blocks to process are selected from image regions with the smallest variance. During our experiments involving seven state of the art methods, the proposed approach was significantly faster than the methods with similar or higher accuracy. Meanwhile, the relative error of our estimator was always less than 15%. We also show that the proposed method can process images without homogeneous areas.

8655-23, Session 7

Spatial-temporal noise reduction method optimized for real-time implementation

Ilya V. Romanenko, Apical (United Kingdom); Eran Edirisinghe, Loughborough Univ. (United Kingdom); Daniel Larkin, Apical (United Kingdom)

Image de-noising in spatial-temporal domain has been a problem studied in-depth in the field of digital image processing. However algorithmic complexity often leads to high hardware resource and memory usage, or increased computational complexity, making their practical use impossible. In our research we have addressed these problems with an implementation of a practical spatial-temporal de-noising algorithm. Spatial-temporal filtering is performed in Bayer RAW data space, which allows us to benefit from predictable sensor noise characteristics and reduce memory bandwidth requirements. Proposed algorithm efficiently removes different types of noise in a wide range of signal to noise ratios. In our algorithm the local motion compensation was performed in Bayer RAW data space, while preserving the resolution and effectively improving signal to noise ratios of moving objects.

The main challenge for the use of spatial-temporal noise reduction algorithms in video applications is the compromise made between the quality of the motion prediction and the complexity of the algorithm and required memory bandwidth (and consequentially the cost of integration). In photo and video applications it is very important that moving objects should stay sharp, while the noise is efficiently removed in both static background and moving object areas.

Taking into account the achievable improvement in PSNR (on the level of the best known noise reduction techniques) and low algorithmic complexity, enabling its practical use in commercial video applications, the results of our research can be very valuable.

8655-24, Session 7

Evolution of image regularization with PDEs toward a new anisotropic smoothing based on half kernels

Baptiste Magnier, Philippe Montesinos, Ecole des Mines d'Alès (France)

This paper is dedicated to a new anisotropic diffusion approach for image regularization based on a gradient and two diffusion directions obtained from half Gaussian kernels.

This approach results in smoothing an image while preserving edges.

From an anisotropic edge detector, built of half Gaussian derivative kernels, we introduce a new smoothing method preserving structures which drives the diffusion function of the angle between the two edge directions and the gradient value.

Due to the two directions diffusion used in the control function, our diffusion scheme enables to preserve edges and corners, contrary to other anisotropic diffusion methods.

Moreover, parameters of the Gaussian kernel can be tuned to be sufficiently thin extracting precisely edges whereas its length allows detecting in contour orientations which leads to a coherent image regularization.

Finally, we present some experimental results and discuss about the choice of the different parameters.

8655-25, Session 7

Poisson shot noise parameter estimation from a single scanning electron microscopy image

Stephen Kockentiedt, Otto-von-Guericke-Univ. Magdeburg (Germany) and Federal Institute for Occupational Safety and Health

(Germany); Klaus D. Tönnies, Otto-von-Guericke-Univ. Magdeburg (Germany); Erhardt Gierke, Nico Dziurowitz, Carmen Thim, Sabine Plitzko, Federal Institute for Occupational Safety and Health (Germany)

Scanning electron microscopy (SEM) has an extremely low signal-to-noise ratio leading to a high level of shot noise which makes further processing difficult. Unlike often assumed, the noise stems from a Poisson process and is not Gaussian but depends on the signal level.

A method to estimate the noise parameters of individual images should be found.

Using statistical modeling of SEM noise, a robust optimal noise estimation algorithm is derived.

A non-local means noise reduction filter tuned with the estimated noise parameters on average achieves an 18% lower root-mean-square error than the untuned filter on simulated images.

The algorithm is stable and can adapt to varying noise levels.

8655-26, Session 8

Parallel algorithms for fast subpixel detection in hyperspectral imagery

Chung M. Wong, John Shepanski, Stephanie Sandor-Leahy, Northrop Grumman Aerospace Systems (United States)

We present parallel algorithms for fast subpixel detection of targets in hyperspectral imagery produced by our Hyperspectral Airborne Tactical Instrument (HATI-2500). The parallel detection algorithm selected for processing the hyperspectral data cubes is based on the adaptive coherence/cosine estimator (ACE). The ACE detector is a robust detector that is built upon the theory of generalized likelihood ratio testing (GLRT) in implementing the matched subspace detector to unknown parameters such as the noise covariance matrix. Subspace detectors involve projection transformations whose matrices can be efficiently manipulated through multithreaded massively parallel processors on modern graphics processing units (GPU). The GPU kernels developed in this work are based on the CUDA computing architecture. We constrain the detection problem to a model with known target spectral features and unstructured background. The processing includes the following steps: 1) quick scale and offset correction to reduce system noise in the collected data, 2) update the mean and noise covariance matrices using most recently collected data; and 3) compute the generalized likelihood ratio for each pixel for binary hypothesis testing. We present the performance improvements obtained using GPU and discuss possible architectures for implementation of such processing system for general on-board airborne hyperspectral imaging applications.

8655-27, Session 8

Vasculature segmentation using parallel multi-hypothesis template tracking on heterogeneous platforms

Dongping Zhang, Lee Howes, Advanced Micro Devices, Inc. (United States)

We present a parallel multi-hypothesis template tracking algorithm on heterogeneous platforms using a layered dispatch programming model. The contribution of this work is two-fold: a novel approach to segment the vascular lumen network from volumetric CTA images in real-time; a layered dispatch programming model to free the developers from hand-crafting mappings to particularly constrained execution domains on the high throughput architecture. This abstraction is demonstrated through the vasculature segmentation application and can also be applied in other real-world applications.

8655-28, Session 8

IMPAIR: massively parallel deconvolution on the GPU

Michael J. Sherry, Andrew Shearer, National Univ. of Ireland, Galway (Ireland)

IMPAIR: Massively parallel deconvolution on the GPU

The IMPAIR software is a high throughput image deconvolution tool for processing large out-of-core datasets of images, varying from large images with spatially varying PSFs to large numbers of images with spatially invariant PSFs. IMPAIR implements a parallelised version of the tried and tested Richardson Lucy deconvolution algorithm regularised via a custom wavelet thresholding library. It exploits the inherantly parallel nature of the convolution operation to achieve quality results on consumer grade hardware: through the NVIDIA Tesla GPU implementation, the multicore OpenMP implementation, and the cluster computing MPI implementation of the software. IMPAIR addresses the problem of paralleisation in both top-down and bottom-up approaches: by managing the input data at the image level, and by managing the execution at the instruction level. These combined techniques lead to a scalable solution with minimal resource consumption and maximal load balancing. IMPAIR is being developed as both a stand-alone tool for image processing, and as a library which can be embedded into non-parallelised code to transparently provide parallel high throughput deconvolution.

8655-29, Session 9

Fast texture and structure image reconstruction using the perceptual hash

Viacheslav V. Voronin, Vladimir I. Marchuk, Vladimir A. Frantc, Don State Technical Univ. (Russian Federation); Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

This paper focuses on the fast texture and structure reconstruction of images. The proposed method, applied to images, consists of several steps. The first deals with the extracted textural features of the input images based on the Law's energy. The pixels around damaged image regions are clustered using these features, that allow to define the correspondence between pixels from different patches. Second, cubic spline curve is applied to reconstruct structure and connect edge and contour in damaged area. The choice of the current pixel to be recovered is decided using the fast marching approach. The Telea method or modifications of the exemplar based method are used after this depending on the classification of the regions where to-be-restored pixel is located. In modification to quickly find patches we use perceptual hash. Such strategy allows to get some data structure containing the hashes of similar patches. This enables us to reduce the search procedure to the procedure for "calculations" of the patch. The proposed method is tested on various samples of images, with different geometrical features and compared with state-of-the-art image inpainting methods; the proposed technique is shown to produce better results in reconstruction of missing small and large objects on the test images.

8655-30, Session 9

Improved multichannel up-sampling method for reconstruction based super-resolution

Seunghoon Jee, Moon Gi Kang, Yonsei Univ. (Korea, Republic of)

In reconstruction based super-resolution, it is an important issue to up-sample and to merge the high-frequency information contained in low-resolution images efficiently and without artifact. The conventional up-sampling methods, which used for registering low-resolution data to high-resolution grid, have the difference between them and ideal up-sampling in observation modeling. In this paper, we analyze the difference and propose a new up-sampling and merging method which is able to incorporate low-resolution images without loss of the high frequency data and minimizes the artifact caused by data insufficiency.

By this method, regions that the registered high frequency data do not cover are naturally regularized without using complex regularizers. The experimental results show that the choice of up-sampling significantly affects the quality of resulting images and the proposed up-sampling method gives better results compared with conventional simple up-sampling method.

8655-31, Session 9

Kernel-based image upscaling method with shooting artifact reduction

Chul Hee Park, Joonyoung Chang, Moon Gi Kang, Yonsei Univ. (Korea, Republic of)

This paper describes the interpolation algorithm which contains shooting or ringing artifact suppression based on windowed sinc interpolator. In general, the windowed sinc interpolator can achieve better performance by using wider window. However, using wide window causes more ripples that produce unwanted defects such as ringing or shooting artifact. Therefore, shooting reduction technique is proposed in this paper for using wider windows to improve the performance without shooting artifact. The proposed algorithm can suppress shooting artifact by using median sinc interpolator and it can be also used as a post processor for many kernel-based interpolation methods. The resulted image shows that the proposed algorithm can maintain local details and suppress shooting artifact in the image well.

8655-32, Session 10

A stochastic approach for non-rigid image registration

Ivan A. Kolesov, Jehoon Lee, Patricio Vela, Georgia Institute of Technology (United States); Allen Tannenbaum, Boston Univ. (United States)

This note describes a non-rigid image registration approach that parametrizes the deformation field by

an additive composition of a rigid transformation and a set of Gaussian radial basis functions. The bases' centers, variances, and weights are determined with a global optimization approach that is introduced. This approach is composed of simulated annealing with a particle filter based generator function to perform the optimization.The registration process is implicitly regularized by limiting the number of bases making up the deformation. Further, a constraint on the deformation is enforced whose role is to ensure physically meaningful fields (i.e., invertible). Results on 2D and 3D data demonstrate the algorithm's robustness to large deformations.

8655-33, Session 10

Video inpainting using scene model and object tracking

Vladimir A. Frantc, Vyatcheslav V. Voronin, Vladimir I. Marchuk, Don State Technical Univ. (Russian Federation); Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

This paper describes a novel approach for video restoration. Our method is based on scene model and object tracking. It also uses camera egomotion estimation for compensate camera change of camera position and parameters. Proposed approach allow to remove objects or restore missing or tainted regions present in a video sequence by utilizing spatial and temporal information from neighboring scenes. The algorithm iteratively performs following operations: achieve frame; update the scene model; update positions of moving objects (this step use the condensation algorithm); replace parts of the frame occupied by the objects marked for remove with use of a background model. As a possible model of background we considered running average, codebook and some others. The main advantage of this approach is the ability to work in streaming mode. There is no need to know the

whole sequence to restore it. We demonstrate the performance of a new approach via several examples, showing the effectiveness of our algorithm and compared with state-of-the-art video inpainting methods.

8655-34, Session 10

Fast DCT-based algorithm for signal and image accurate scaling

Leonid Bilevich, Leonid P. Yaroslavsky, Tel Aviv Univ. (Israel)

A new DCT-based algorithm for signal and image scaling by arbitrary factor is presented. The algorithm is virtually free of boundary effects and implements the discrete sinc-interpolation, which preserves the spectral content of the signal, and therefore is free from interpolation errors. Being implemented through the fast FFT-type DCT algorithm, the scaling algorithm has computational complexity of $O(N*\log(sigma*N))$ operations per output sample, where N and sigma*N are number of signal input and output samples, correspondingly.

8655-1, Session IPI

A new set of wavelet- and fractals-based features for Gleason grading of prostate cancer histopathology images

Clara M. Mosquera Lopez, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

Prostate cancer detection and staging is an important step towards patient treatment selection. Advancements in digital pathology allow the application of new quantitative image analysis algorithms for computer-assisted diagnosis (CAD) on digitized histopathology images. In this paper, we introduce a new set of features to automatically grade pathological images using the well-known Gleason grading system. The goal of this study is to classify biopsy images belonging to Gleason patterns 3, 4, and 5 by using a combination of wavelet and fractal features. For image classification we use pairwise coupling Support Vector Machine (SVM) classifiers. The accuracy of the system, which is close to 97%, is estimated through three different cross-validation schemes. The proposed system offers the potential for automating classification of histological images and supporting prostate cancer diagnosis.

8655-19, Session IPI

Method and architecture for quantification of bone structure using microscopic image slices

Sunderam Krishnan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States); Dana Mecke, Sergio Montelongo, Wang Xiaodu, University of texas san antonio (United States)

This paper presents a new system that reconstructs, visualizes and classifies trabecular bone structure by using microscopic image slices. In this study, we evaluated the structure of a trabecular bone using 3D X-ray imaging after passing through the special image enhancement and de-nosing algorithms. We propose a new simple imaging technique tool for the quantification of structural changes within the micro architecture of human bones by enhancing the characteristics attributes of the bone architecture from μ-CT scans. Computer simulation illustrates that the presented imaging technique has the potential to become a powerful tool to investigate the structure of trabeculae during in vivo measurements.

8655-35, Session IPI

Active shape models and depth for image registration

Colin Bellmore, Roxanne L. Canosa, Rochester Institute of Technology (United States)

Most current registration techniques rely on matching patches or points directly from information obtained from multiple images. This technique becomes infeasible when image sources are independent of one another and common landmarks cannot be guaranteed. To overcome the difficulties inherent in correlating image patches obtained from multiple sensors and in the absence of real landmarks, we propose using a non-rigid deformation technique based on statistical shape information. This technique is applied to real time video and uses an active shape model (ASM) to assess curvature and learn discriminating points on a per-channel basis. The combination of independent channels for registration is useful for other techniques that require well-registered images but do not have access to reliable landmarks. The algorithm was tested on ASMs trained on low-resolution facial images generated from an inexpensive commercially available RGB+D sensor. The mean fitting time of our 14-point ASM model was significantly reduced as compared to other current face finding models. Model annotation, model building, and model fitting all functioned correctly together to register the color and depth images. The system successfully generated a stream of filtered color and depth images with facial landmark points fitted by the trained ASM.

8655-36, Session IPI

Principal component analysis for through wall image enhancement

Muhammad M. Riaz, Abdul Ghafoor, National Univ. of Sciences and Technology (Pakistan)

No Abstract Available

8655-37, Session IPI

Parallel GPGPU stereo matching with an energy-efficient cost function based on normalized cross correlation

Sarala Arunagiri, Jaime Jaloma, Patricia J. Teller, The Univ. of Texas at El Paso (United States)

Stereo matching is a heavily investigated topic in computer vision because of its wide range of applicability and its high computational complexity. In this paper we explore an energy-efficient technique that is suitable for parallel (GPGPU) stereo matching of remotely-sensed radar images. A major issue with images captured by remote sensing, e.g., by airborne radars, is that they are likely to have speckle noise, which is undesired information that contaminates and distorts the image. Although there are filters that alleviate the effects of speckle noise, they introduce a certain amount of image distortion. It has been demonstrated that cost functions based on Normalized Cross Correlation (NCC) can be used for accurate stereo matching in the presence of speckle noise, thus, we explored such a function for passive stereo matching. Accordingly, this paper presents a new NCC-based cost function that minimizes the number of floating-point operations utilized for stereo matching and compares its performance to that of the original NCC cost function. This evaluation is achieved via experiments in which these cost functions are employed by GPGPU stereo matching codes that use the simulated annealing algorithm. Performance comparison is conducted in terms of power and energy consumption, execution time, and output quality.

8655-38, Session IPI

Hyperspectral bands prediction based on inter-band spectral correlation structure

Ayman Ahmed, NARSS (Egypt); Mohamed E. El. Sharkawy, Egypt-Japan Univ. of Science and Technology (Egypt); Salwa H. Elramly, Ain Shams Univ. (Egypt)

Hyperspectral imaging has been widely studied in many applications; notably in climate changes, vegetation, and desert studies. However, such kind of imaging brings a huge amount of data, which requires transmission, processing, and storage resources for both airborne and spaceborne imaging. Compression of hyperspectral data cubes is an effective solution for these problems. Lossless compression of the hyperspectral data usually results in low compression ratio, which may not meet the available resources; on the other hand, lossy compression may give the desired ratio, but with a significant degradation effect on object identification performance of the hyperspectral data. Moreover, most hyperspectral data compression techniques exploits the similarities in spectral dimensions; which requires bands reordering or regrouping, to make use of the spectral redundancy. In this paper, we analyze the spectral cross correlation between bands for AVIRIS and Hyperion hyperspectral data; spectral cross correlation matrix is calculated, assessing the strength of the spectral matrix, we propose new technique to find highly correlated groups of bands in the hyperspectral data cube based on "inter band correlation square", and finally, we propose a new technique of band regrouping based on correlation values weights for different group of bands as network of correlation.

8655-39, Session IPI

A GPU based implementation of direct multi-bit search (DMS) screen algorithm

Barry M. Trager, IBM Thomas J. Watson Research Ctr. (United States); Kartheek Chandu, InfoPrint Solutions Co. (United States); Chai Wah Wu, IBM Thomas J. Watson Research Ctr. (United States); Mikel J. Stanich, InfoPrint Solutions Co. (United States)

In this paper, we study the feasibility for using programmable Graphics Processing Unit (GPU) technology for image halftoning, in particular implementing the computationally intense Direct Multi-bit Search (DMS) Screen algorithm. Multi-bit screening is an extension of binary screening, in which every pixel in continuous-tone image can be rendered to one among multiple absorptance levels. For example a 2 bit printer is capable of printing with four different drop sizes. In our previous work, we have extended the Direct Binary Search (DBS) to the multi-bit case using Direct Multi-bit Search (DMS) where at every pixel the algorithm chooses the best drop absorptance level to create a visually pleasing halftone pattern without any user defined guidance. This process is repeated throughout the entire range of gray levels while satisfying the stacking constraint to create a high quality multi-bit screen (dither mask). In this paper, we illustrate how employing Graphics Processing Units (GPU) can speed-up intensive DMS image processing operations. Particularly, we illustrate how different modules can be been parallelized. The main goal of many of the previous articles regarding DBS is to decrease the execution time of the algorithm. One of the most common approaches is to decrease the neighborhood size or filter size. The proposed parallel approach allows us to use a large neighborhood and filter size, to achieve the highest halftone quality, while having minimal impact on performance. In addition, we also demonstrate processing several non-overlapping neighborhoods in parallel, by utilizing the GPU's parallel architecture, to further improve the computational efficiency.

8655-40, Session IPI

Ore minerals textural characterization by hyperspectral imaging

Giuseppe Bonifazi, Univ. degli Studi di Roma La Sapienza (Italy); Nicoletta Picone, Univ degli Studi di Roma La Sapienza (Italy);

Silvia Serranti, Univ. degli Studi di Roma La Sapienza (Italy)

The utilisation of hyperspectral detection devices, for natural resources mapping/exploitation through remote sensing techniques, dates back to the early 1970s. From the first devices utilising a one-dimensional profile spectrometer HyperSpectral Imaging (HSI) devices have been developed. Thus, from specific-customised devices, originally developed by Governmental Agencies (e.g. NASA, specialised research labs, etc.), a lot of HSI based devices are today available at commercial level. Several companies, in fact, develop hyperspectral sensors. Parallel to this huge increase of hyperspectral devices development/manufacturing addressed to airborne application, a strong increase also occurred in developing HSI based devices for "ground" utilisation that is sensing units able to play inside a laboratory, a processing plant and/or in an open field. Thanks to this diffusion more and more applications have been developed and tested in this last years also in the materials sectors. Such an approach, when successful, is quite challenging being usually reliable, robust and characterised by lower costs if compared with those usually associated to commonly applied analytical off- and/or on-line analytical approaches. In this paper such an approach is presented with reference to ore minerals characterization. According to the different phases and stages of ore minerals and products characterization, and starting from the analyses of the detected hyperspectral firms, it is possible to derive useful information about mineral flow stream properties and their physical-chemical attributes. This last aspect can be utilized to define innovative process mineralogy strategies and to implement on-line procedures at processing level. The present study discusses the effects related to the adoption of different hardware configurations, the utilization of different logics to perform the analysis and the selection of different algorithms according to the different characterization, inspection and quality control actions to apply.

8655-41, Session IPI

Colour modification and colour combination in double-cone colour space

Alfredo Restrepo, Univ. de los Andes (Colombia)

It is useful to have a means to predict the colour that will result when two beams of light converge on a given surface, for simulation purposes for example. Given the RGB coordinates of the colours of the light beams, we propose an arithmetic model to get the colour coordinates of the combination. We work in double-cone space, which is readiy derived from RGB cubic space. Double-cone space is of the type hue-saturation-luminance and the model gives the hue, the saturation and the luminance of the combination, provided those of the rays that are being combined.

Double-cone space results from the spinning of the range-midrange triangle about its base (the midrange axis). The range ? of the RGB triple, being a measure of the distance to the achromatic segment in the cube, measures the chromatic saturation while the midrange ?, a measure of the distance to the black point in the cube, measures the luminance. In addition, we assume there is a relation between luminance and intensity of the beam of the saturation type, such as a hyperbolic tangent or a Naka-Rushton curve; this is used to get the luminance of the combined beam and to appropriately weight the hue and saturation of the combining beams.

Also in double-cone space, we give formulae for the modification of the luminance and the saturation.

8655-42, Session IPI

Automatic image and video denoising based on Ising theory

Eliahu Cohen, Tel Aviv Univ. (Israel); Maya Carmi, Ron Heiman, Ofer Hadar, Asaf Cohen, Ben-Gurion Univ. of the Negev (Israel)

Statistical models, such as the Ising model, have been proven to be very useful in describing solid state systems in physics. Although their results relay on probabilistic basis, a very good agreement was found between them and experimental results obtained from examining physical systems. Furthermore, These models can be applied to a variety of problems in engineering, chemistry, biology and more, without losing their effectiveness, simplicity and intuitiveness.

In a previous work [1] we have greatly improved Kandes' model [2] and used an Ising-like model in order to restore colored images and videos damaged from various kinds of noise.

In our recent work we wish to present important algorithmic developments. Most of them obtained as a combination of better physical models and well known image restoration techniques. The proposed model analyzes automatically the noisy image and chooses the appropriate model parameters accordingly, without the need of manual support.

Finally, an average improvement of 3 dBs in comparison to the old model's PSNR is achieved and an overall improvement of up to 7 dBs in comparison to a standard denoising performed by median filter.

8655-43, Session IPI

A study of non-diagonal models for image white balance

Ching-Chun Huang, De-Kai Huang, National Kaohsiung Univ. of Applied Sciences (Taiwan)

White balance is an algorithm proposed to mimic the color constancy mechanism of human perception. However, as shown by its name, current white balance algorithms only promise to correct the color shift of gray tones to correct positions; for other color values, white balance algorithms process them as gray tones and therefore produce undesired color biases. To improve the color prediction of white balance algorithms, in this paper, we propose a 3-parameter non-diagonal model, named as PCA-CLSE, for white balance. Unlike many previous researches which use the von Kries diagonal model for color prediction, we proposed applying a non-diagonal model for color correction which aimed to minimize the color biases while keeping the balance of white color. In our method, to reduce the color biases, we proposed a PCA-based training method to gain extra information for analysis and built a mapping model between illumination and non-diagonal transformation matrices. While a color-biased image is given, we could estimate the illumination and dynamically determine the illumination-dependent transformation matrix to correct the color-biased image. Our evaluation shows that the proposed PCA-CLSE model can efficiently reduce the color biases.

8655-44, Session IPI

A comparison between space-time video descriptors

Luca Costantini, Licia Capodiferro, Fondazione Ugo Bordoni (Italy); Alessandro Neri, Univ. degli Studi di Roma Tre (Italy)

Video representation techniques are used in database retrieval, indexing, or scene identification. In the context of video representation and indexing, two main approaches are adopted . The first one is based on the key frame detection: after a temporal segmentation of the video, the detected key frames are processed as still images. Once the features are extracted from each key frame, retrieval, indexing or recognition operations can be performed. In the second approach, the video is represented by space-time descriptors. In this case, the video is represented as a set of points in a high dimensional space, in which every points is the representation of a space-time patch. This second approach is especially employed in the human actions recognition algorithms. These algorithms are very important in many computer vision applications such as video surveillance or human computer interfaces.

The 3D Zernike polynomials are employed to compute the 3D Zernike moments, that can be used as space-time descriptors in spherical coordinates. In this paper the author present a space-time descriptor in cylindrical coordinates. Our aim is to investigate which descriptor, between the spherical and the cylindrical, is more appropriate to build an efficient space-time descriptor. Our idea is that using two different space-time descriptor both based on the Zernike polynomials, but the first one describes a sphere, and the other one describes a cylinder, allows us to investigate only the differences between spherical descriptors and cylindrical descriptors.

8655-45, Session IPI

Face recognition based on logarithmic local binary patterns

Debashree Mandal, Karen A. Panetta, Tufts Univ. (United States); Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

This paper presents a novel approach to the problem of face recognition that combines the classical Local Binary Pattern (LBP) feature descriptors with image processing in the logarithmic domain and the human visual system. Particularly, we have introduced parameterized logarithmic image processing (PLIP) operators based LBP feature extractor. We also use the human visual system based image decomposition, which is based on the Weber's law to extract features from the decomposed images and combine those with the features extracted from the original images thereby enriching the feature vector set and obtaining improved rates of recognition. Comparisons with other methods are also presented. Extensive experiments clearly show the superiority of the proposed scheme over LBP feature descriptors. Recognition rates as high as 99% can be achieved as compared to the recognition rate of 96.5% achieved by the classical LBP using the AT&T Laboratories face database.

8655-46, Session IPI

Hyperspectral images lossless compression using the 3D binary EZW algorithm

Kai-jen Cheng, Jeffrey Dill, Ohio Univ. (United States)

This paper proposes a transform

based image lossless compression for hyperspectral images which is inspired by Shapiro (1993)'s EZW algorithm. The proposed compression method uses a hybrid transform which includes an integer Karhunrn-Loeve transform (KLT) and discrete wavelet transform (DWT). The integer KLT is employed to eliminate the presence of correlations among the bands of the hyperspectral image. The integer 2D discrete wavelet transform (DWT), which is implemented by lifting scheme, is applied to eliminate the correlations in the spatial dimensions and produce wavelet coefficients. These coefficients are then coded by a binary EZW algorithm. The binary EZW can eliminate the subordinate pass of conventional EZW by coding residual values and producing binary sequences. The binary EZW algorithm combines the merits from the conventional EZW and SPIHT, and it is computationally simpler for lossless compression. The proposed method is applied to AVIRIS images and compared to other state of art image compression techniques. The results show that the proposed lossless image compression is more efficient and it also has higher compression ratio than other algorithms.

Conference 8656: Real-Time Image and Video Processing 2013

Wednesday - Thursday 6 –7 February 2013

Part of Proceedings of SPIE Vol. 8656 Real-Time Image and Video Processing 2013

8656-11, Session IPI

Achieving real-time capsule endoscopy (CE) video visualization through panoramic imaging

Steven Yi, Jean Xie, Peter Mui, Xyken, LLC (United States); Jonathan Leighton, Mayo Clinic (United States)

In this paper, we present a novel and real-time capsule endoscopy (CE) video visualization algorithm based on panoramic imaging. Typical CE videos run about 8 hours and are manually reviewed by physicians to locate diseases such as bleedings and polyps. To date, there is no commercially available tool capable of providing stabilized and processed CE video that is easy to analyze in real time. The burden on physicians' disease finding efforts is thus big. In fact, since the CE camera sensor has a limited forward looking view and low image frame rate (typical 2 frames per second), and captures very close range imaging on the GI tract surface, it is no surprise that traditional visualization method based on tracking and registration often fails to work. This paper presents a novel concept for real-time CE video stabilization and display. Instead of directly working on traditional forward looking FOV (field of view) images, we work on panoramic images to bypass many problems facing traditional imaging modalities. Methods on panoramic image generation based on optical lens principle leading to real-time data visualization will be presented. In addition, non-rigid panoramic image registration method will be discussed.

8656-19, Session IPI

Analysis and characterization of embedded vision systems for taxonomy formulation

Muhammad Imran, Mid Sweden Univ. (Sweden); Khaled Benkrid, The Univ. of Edinburgh (United Kingdom); Khursheed Khursheed, Naeem Ahmad, Mattias O'Nils, Najeem Lawal, Mid Sweden Univ. (Sweden)

In this paper, we presented a system taxonomy with the aim of providing an abstract reference model of vision systems. Central to our proposed system taxonomy is a limited number of vision functions suffice to cover a large number of real visions systems reported in the literature. We have tested 20 visions systems from the literature against our proposed taxonomy. The evaluation criterion for the proposed system taxonomy was based on two parameters: 1) functions presence, and 2) the ordering of functions in the taxonomy in relation to the actual vision system ordering. Indeed, our system taxonomy focuses on the functionality of vision systems rather than the exact ordering of functions which depends on the details of the application at hand. We hope that our proposed taxonomy will provide a realistic model of embedded vision systems which can be developed in the future by the community at large. This will facilitate fairer benchmarking and evaluation of embedded vision systems, as well as the development of generic and efficient solutions for different grades or classes of embedded vision systems, reducing development cost of future embedded vision systems.

8656-20, Session IPI

Design and implementation of a real-time image registration in an infrared search and track system

Fu-yuan Xu, Guohua Gu, Nanjing Univ. of Science and Technology (China); Tiekun Zhao, XI'AN SICONG CHUANGWEIOPTO-

ELECTRONIC CO.,LTD. (China); Qian Chen, Weixian Qian, Nanjing Univ. of Science and Technology (China)

In this paper, an Infrared image registration method which combine the phase correlation registration and the Corner registration. This registration method is applied to infrared search and track system and realize motion compensation of the infrared image. This algorithm has small amount of calculation and high accuracy of matching . The experiments show that this algorithm can provide higher confidence motion compensation parameters in real-time infrared search and track system and effectively reduce the false alarm rate of the ground Small Target Detection on the motion platform.Therefore, the proposed algorithm with a strong practical and widely used in image stabilization, panorama stitching, and image mosaic.

8656-21, Session IPI

Binary video codec for data reduction in wireless visual sensor networks

Khursheed Khursheed, Naeem Ahmad, Muhammad Imran, Mattias O'Nils, Mid Sweden Univ. (Sweden)

Efficiency in both computation and compression rate of the Visual Sensor Node (VSN) is required for the energy constrained outdoor applications of Wireless Visual Sensor Networks (WVSN). We investigated the compression efficiency of the information reduction techniques such as image coding, change coding and Region of Interest (ROI) coding for data reduction in WVSN. The compression efficiency of both change coding and ROI coding is better than that of image coding for applications involving few changes in a continuous set of frames. But, the compression efficiency of both the change coding and ROI coding becomes worse than that of image coding if the change frame contains too many objects. In this paper, we analysed the compression efficiency of the Binary Video Codec (BVC) for various kinds of changes such as different shapes, sizes, locations and number of objects in a set of frames. We observed that the curve representing the average compressed file size of the BVC is always lower than that of change coding and ROI coding. We concluded that the compression efficiency of BVC is always better than that of change coding and is always better than or equal to that of ROI coding and image coding.

8656-22, Session IPI

Determinant of homography-matrix-based multiple-object recognition

Nagachetan Bangalore, Madhu Kiran, Anil Suryaprakash, Visio Ingenii Ltd. (United Kingdom)

Finding a given object in an image or a sequence of frames is one of the fundamental computer vision challenges. Humans can recognize a multitude of objects with little effort despite scale, lighting and perspective changes. A robust computer vision based object recognition system is achievable only if a considerable tolerance to change in scale, rotation and light is achieved. Partial occlusion tolerance is also of paramount importance in order to achieve robust object recognition in real-time applications. In this paper, we propose an effective method for recognizing a given object from a class of trained objects in the presence of partial occlusions and considerable variance in scale, rotation and lighting conditions. The proposed method can also identify the absence of a given object from the class of trained objects. Unlike the conventional methods for object recognition based on the key feature matches between the training image and a test image, the proposed algorithm utilizes a statistical measure from the homography transform based resultant matrix to determine an object match. The magnitude of determinant of the

homography matrix obtained by the homography transform between the test image and the set of training images is used as a criterion to recognize the object contained in the test image. The magnitude of the determinant of homography matrix is found to be very near to zero (i.e. less than 0.005) and ranges between 0.05 and 1, for the out-of-class object and in-class objects respectively. Hence, an out-of-class object can also be identified by using low threshold criteria on the magnitude of the determinant obtained. The proposed method has been extensively tested on a huge database of objects containing about 100 similar and difficult objects to give positive results for both out-of-class and in-class object recognition scenarios. The overall system performance has been documented to be about 95% accurate for a varied range of testing scenarios.

8656-23, Session IPI

Investigating the structure preserving encryption of high efficiency video coding (HEVC)

Zafar Shahid, William Puech, Univ. Montpellier 2 (France)

HEVC is the emerging video coding standard of ITU-T and ISO/IEC. HEVC achieves similar visual quality to its precedent H.264/AVC High Profile, with around 30% bit-rate reduction for low delay mode, and with around 20% bit-rate reduction for random access mode on average, but with lower complexity than H.264/AVC Baseline Profile. HEVC performs better because of some additional tools. Since digital content can be easily copied and modified, it is pertinent to analyze this standard regarding its protection and authentication. Selective encryption (SE) is used to restrict access of video data to only authenticated users. This paper presents a novel method for the real-time protection of HEVC video standard. Structure preserving selective encryption is being performed in CABAC entropy coding module of HEVC, which is significantly different from CABAC entropy coding of H.264/AVC. In CABAC of HEVC, exponential Golomb coding is replaced by truncated Rice (TR) up to a specific value for binarization of transform coefficients. Selective encryption is performed using AES cipher in cipher feedback mode on a plaintext of binstrings in a context aware manner. The encrypted bitstream has exactly the same bit-rate and is format complaint. Experimental evaluation and security analysis of the proposed algorithm is performed on several benchmark video sequences containing different combinations of motion, texture and objects.

8656-24, Session IPI

A computationally efficient approach to 3D point cloud reconstruction

Chih-Hsiang Chang, Nasser Kehtarnavaz, The Univ. of Texas at Dallas (United States)

This paper addresses improving the computational efficiency of the 3D point cloud reconstruction pipeline using un-calibrated image sequences. In the existing pipelines, a global bundle adjustment is carried out which is quite time consuming. The computational complexity of this module as part of the 3D point cloud reconstruction pipeline is, where denotes number of images in an entire sequence and number of matched points. Also, the searching and sorting aspects of the global bundle adjustment is of the computational cost [3]. A local refinement process is introduced in this paper in order to reduce the computational complexity by bypassing the searching and sorting aspects of the bundle adjustment via limiting to only 3.

8656-25, Session IPI

TDC-based readout electronics for real-time acquisition of high resolution PET bio-images

Nahema Marino, Sergio Saponara, Univ. di Pisa (Italy) and INFN sez. Pisa (Italy); G. Ambrosi, INFN sez. Perugia (Italy); Federico Baronti, Univ. di Pisa (Italy); Maria Giuseppina Bisogni, Univ. di Pisa (Italy) and INFN sez. Pisa (Italy); Piergiorgio Cerello, INFN sez. Torino (Italy); F. Ciciriello, Francesco Corsi, Politecnico di Bari (Italy) and INFN sez. Bari (Italy); Luca Fanucci, Univ. di Pisa (Italy) and INFN sez. Pisa (Italy); Maria Ionica, Univ. degli Studi di Perugia (Italy); F. Licciulli, Cristoforo Marzocca, Politecnico di Bari (Italy) and INFN sez. Bari (Italy); M. Morrocchi, Dept. of Physics, University of Pisa (Italy) and INFN sez. Pisa (Italy); F. Pennazio, Università di Torino (Italy) and INFN sez. Torino (Italy); Roberto Roncella, Univ. di Pisa (Italy); C. Santoni, INFN sez. Perugia (Italy); R. Wheadon, INFN sez. Torino (Italy); Alberto Del Guerra, Univ. di Pisa (Italy) and INFN sez. Pisa (Italy)

Positron emission tomography (PET) is a clinical and research tool for real-time metabolic imaging. The demand for better image quality entails continuous research to improve PET instrumentation. In clinical applications, PET image quality benefits from the Time of Flight (TOF) feature. Indeed, by measuring the photons arrival time on the detectors with a resolution less than 100 ps, the annihilation point can be estimated with sub-centimeter resolution. This leads to better noise level, contrast and clarity of detail in the images either using analytical or iterative reconstruction algorithms. This work discusses a SiPM-based magnetic-field compatible TOF-PET module with Depth of Interaction (DOI) correction. The detector features a 3D architecture with two tiles of SiPMs coupled to a single LYSO scintillator on both its faces. The real-time front-end electronics is based on a current-mode ASIC where a low input impedance, fast current buffer allows achieving the required time resolution. A pipelined Time to Digital Converter measures and digitizes the arrival time and the energy of the events with a timestamp of 100 ps and 400 ps, respectively. An FPGA clusters the data and evaluates the DOI, with a simulated z resolution of the PET image of 1.4 mm FWHM.

8656-26, Session IPI

A visibility improvement technique for fog images suitable for real-time application

Yoshitaka Toyoda, Daisuke Suzuki, Koichi Yamashita, Takashi Ito, Narihiro Matoba, Tetsuya Kuno, Hiroaki Sugiura, Mitsubishi Electric Corp. (Japan)

Cameras used in outdoor scenes require high visibility performance under various environmental conditions. We present a visibility improvement technique which can improve the visibility of images captured in bad weather such as fog and haze, and also applicable to real-time processing in surveillance cameras and vehicle cameras. Our algorithm enhances contrast pixel by pixel according to the brightness and sharpness of neighboring pixels. In order to reduce computational costs, we preliminary specify the adaptive functions which determine contrast gain from brightness and sharpness of neighboring pixels. We optimize these functions using the sets of fog images and examine how well they can predict the fog-degraded area using both qualitative and quantitative assessment. We demonstrate that our method can prevent excessive correction to the area without fog to suppress noise amplification in sky or shadow region, while applying powerful correction to the fog-degraded area. In comparison with other real-time oriented methods, our method can reproduce clear-day visibility while preserving gradation in shadows and highlights and also preserving naturalness of the original image. Our algorithm with low computational costs can be compactly implemented on hardware and thus applicable to wide-range of video equipments for the purpose of visibility improvement in surveillance cameras, vehicle cameras, and displays.

8656-27, Session IPI

Fast non-blind deconvolution based on 2D point spread function database for real-time ultrasound imaging

Jooyoung Kang, Sung-Chan Park, Kyuhong Kim, Jung-Ho Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

In this paper, we introduce a new method for non-blind de-convolution with pre-measured and estimated 2D PSFs database that acquired from the actual transducer being used. Our algorithm is based on two-stage reconstruction scheme, in which 2D PSF selection first and image restoration second. In the PSF selection block, minimum variance value of applied each different PSF in all depth of image are compared to derive the information of depth and speed of sound. In the de-convolution block, once the appropriated PSF of each depth selected, it restores a beam-formed uncompressed radio-frequency data using several de-convolution technique like Weiner and fast de-convolution technique using hyper-Laplacian priors. Therefore, using the real PSF from actual transducer being used, our algorithm produces a better restoration of ultrasound image than de-convolution by simulated PSF, and has low complexity for real-time ultrasound imaging.

8656-1, Session 1

Real-time, robust target tracking in videos via graph-cuts *(Invited Paper)*

Barak Fishbain, Israel Institute of Technology (Israel); Dorit S. Hochbaum, Yan T. Yang, Univ. of California, Berkeley (United States)

Real-time, robust and fast, polynomial time - integer-programming algorithm for target tracking in video sequences, that delineates a target of interest in a video from its background, is presented here. The tracking task is cast as a graph-cut problem, incorporating intensity and motion data into the formulation. Previously reported tracking algorithms used continuous models that are not appropriate for digital videos. Other algorithms utilized stochastic, iterative models which are computationally intense and do not guarantee optimal solution or consistency over sequential runs on the same input data. Furthermore, previous methods, which considered motion in the tracking process, did so by regarding it as a set of constraints rather than pixels' features. This approach resulted in complex, non-polynomial (NP-hard) problems. Finally, in previous tracking algorithms, motion is estimated by optical flow techniques. These optical flow methods are computationally intense, which makes them oftentimes unsuitable for real-time applications. Our method is highly robust in that it allows exploiting fast but noisy and coarse motion data such as MPEG-4 motion estimation schemes. The evaluation of the method on standard and non-standard benchmark videos clearly shows that the method is more efficient than existing techniques, and that it delivers good quality results.

8656-2, Session 1

Tracking yarns in high resolution fabric images: a real-time approach for online fabric flaw detection

Dorian Schneider, RWTH Aachen (Germany)

An algorithmic framework for real-time localization of single yarns within industrial fabric images is presented. The information about precise yarn locations forms the foundation for a fabric aw detection system which is based on individual yarn measurements. Matching a camera frame rate of 15 fps, we define the term "real-time" by the capability of tracking all yarns within a 5 megapixel image in less than 35 ms, leaving a time slot of 31ms for further image processing and defect detection algorithms. The processing pipeline comprises adaptive histogram equalization, Wiener deconvolution, normalized template matching and a novel feature point sorting scheme. To meet real-time requirements, extensive use of the NVIDIA CUDA framework is made. Implementation details are given and source code for selected algorithms is provided. Evaluation results show that wefts and warps can be tracked reliably and independently of the fabric material or binding. Video and image footage is provided on the project website to expand the paper content.

8656-3, Session 1

Real-time bicycle detection at signalized intersections using thermal imaging

Robin Collaert, Traficon N.V. (Belgium)

More and more governments and authorities around the world are promoting the use of bicycles in cities, as this is healthy for the bicyclist and improves the quality of life in general. Safety and efficiency of bicyclists has become a major focus. To achieve this, there is a need for a smarter approach towards the control of signalized intersections. Various traditional detection technologies, such as video, microwave radar and electromagnetic loops, can be used to detect vehicles at signalized intersections, but none of these can consistently separate bikes from other traffic, day and night and in various weather conditions.

As bikes should get a higher priority and also require longer green time to safely cross the signalized intersection, traffic managers are looking for alternative detection systems that can make the distinction between bicycles and other vehicles near the stop bar. In this paper, we present the drawbacks of a video-based approach and the benefits of a thermal-video-based approach for vehicle presence detection with separation of bicycles. Also, we highlight the specific technical challenges in developing a system that combines thermal image capturing, image processing and output triggering to the traffic light controller in near real-time and in one housing.

8656-4, Session 1

How fast can one arbitrarily and precisely scale images?

Leonid Bilevich, Leonid P. Yaroslavsky, Tel Aviv Univ. (Israel)

Image scaling is a frequent operation in video processing for optical metrology. In the paper, results of comparative study of the accuracy and computational complexity of different algorithms for arbitrary scaling of digital images are presented and discussed. The following algorithms were compared: different types of spatial domain processing (bilinear, bicubic, cubic spline) algorithms and DFT-based ones and a new DCT-based algorithm, which implements perfect (interpolation error free) scaling through discrete sinc-interpolation The comparison results enable evaluation of the feasibility of real-time implementation of the algorithms for arbitrary image scaling.

8656-5, Session 1

Digital ruler: real-time object tracking and dimension measurement using stereo cameras *(Invited Paper)*

James Nash, Kalin Atanassov, Sergio R. Goma, Vikas Ramachandra, Qualcomm Inc. (United States); Hasib Siddiqui, Qualcomm Inc (United States)

Stereo metrology is a method for obtaining spatial measurements of an object's perimeter using the disparity between boundary points. True 3D scene information is necessary to extract length measurements of an object's projection onto the 2D image plane. In stereo metrology the measurement is highly sensitive to object distance, baseline distance, calibration errors, and relative movement of the left and right demarcation points between successive frames. Real-time operation at video rates is difficult because traditional feature extraction and self-calibration routines are computationally expensive. In this paper we present a real-time implementation of a digital ruler using cell-phone stereo calibrated cameras on a mobile platform.

Conference 8656

8656-6, Session 2

FPGA design of a real-time edge enhancing smoothing filter *(Invited Paper)*

Chang Choo, Nimit Pandya, Bhavika Patel, San José State Univ. (United States)

Noise removal filters have an undesirable side effect of blurring edges, which is not tolerable for many image processing applications. To overcome this problem, we designed an edge enhancing smoothing filter and implemented it on an FPGA to reduce noise while sharpening edges for real-time image processing. It consists of a combination of the bilateral filter for edge preserving smoothing and the Shock filter for edge enhancement to achieve the desired result. The bit-accurate Matlab model for the edge-enhancing smoothing filter was converted to the Verilog hardware design on Altera Stratix III and Cyclone III devices. The FPGA implementation utilizes, for example, 21% of logic cells on Stratix III EP3SE50F484C2 device and runs at 126MHz. The resulting images have much sharper edges in terms of its gradient and reduced noise.

8656-7, Session 2

Large object extraction for binary images on the GPU

Gregory C. Huchet, Samsung Information Systems America, Inc. (United States)

Object filtering by size is a basic task in computer vision. A common way to extract the largest connected components in a binary image is to run the connected-component labeling (CCL) algorithm and to compute the area of each component. Selecting the components with the largest areas is then straightforward. Several CCL algorithms for the GPU have already been implemented but few of the related papers mention the time required to compute the component area. This extra step can be critical for real-time applications such as real-time video segmentation. The aim of this paper is to present a fast algorithm for the extraction of visually large objects in a binary image. It is implemented using CUDA (Compute Unified Device Architecture), a parallel computing architecture developed by NVIDIA.

8656-8, Session 2

Real-time structured light intraoral 3D measurement pipeline

Radu Gheorghe, Andrei Tchouprakov, Roman Sokolov, D4D Technologies, L.P. (United States)

Computer aided design and manufacturing (CAD/CAM) is increasingly becoming a standard feature and service provided to patients in dentist offices and denture manufacturing laboratories. Although the quality of the tools and data has slowly improved in the last years, due to various surface measurement challenges, practical, accurate, in-vivo, real-time 3D high quality data acquisition and processing still needs improving. Advances in GPU computational power have allowed for achieving near real-time 3D intraoral in-vivo scanning of patient's teeth. We explore in this paper, from a real-time perspective, a hardware-software-GPU solution that addresses all the requirements mentioned before. Moreover we exemplify and quantify the hard and soft deadlines required by such a system and illustrate how they are supported in our implementation.

8656-9, Session 2

Three-dimensional fuzzy filter in color video sequence denoising implemented on DSP

Volodymyr Ponomaryov, Instituto Politécnico Nacional (Mexico); Hector Montenegro, Instituto Politécnico Nacional (Mexico) and ESIME-CULHUAN (Mexico); Ricardo Peralta-Fabi, Univ. Nacional Autónoma de México (Mexico)

Noise produces system deficiencies during acquisition, broadcast or storage of color image sequences. In this paper, we propose an efficient fuzzy 3D approach for impulsive noise suppression in color video sequences. In contrast to current state-of-the-art fuzzy filters, the proposed framework gathers red, green and blue channels sequence data, uses fuzzy logic to analyze the basic pixel gradient value and several related pixel gradient values in eight directions, and processes two neighboring frames concurrently. Several video color sequences Miss America, Salesman, Flowers, Stefan, etc. with different color, texture, fine features and different movements were used to evaluate 3D fuzzy algorithms. The frames of the color video sequences were contaminated artificially by random impulsive noise of different intensities (0% to 30%) in each color channel independently.

The results of numerous simulations demonstrate that the proposed 3D filtering framework performs well in objective criteria (PSNR, MAE, NCD, and SSIM) and a human subjective analysis of the frames in the color video sequences. In addition, the results of implementation and efficiency analysis of the proposed and better denoising 3D algorithms in real time mode on DSP TMS320DM642 of TI, using MATLAB, TC6 software, CCS, IDE with the Real-Time Workshop software, and Simulink are presented.

8656-12, Session 3

Design of a pseudo-log image transform IP in an HLS-based memory management framework

Shahzad Ahmad Butt, Politecnico di Torino (Italy); Stéphane Mancini, Frédéric Rousseau, TIMA Lab. (France); Luciano Lavagno, Politecnico di Torino (Italy) and Cadence Design Systems, Inc. (United States)

The pseudo-log image transform is essentially a logarithmic transformation that simulates the distribution of the eye's photoreceptors and finds application in many important areas of real time image processing such as motion detection and estimation in robots, foveated space variant cameras, time of impact calculation, etc. It belongs to a family of non-linear image processing kernels in which references made to memory are non-linear function of loop indices.

Non-linear kernels require some form of memory management in order to achieve the required throughput, minimize on-chip memory requirement and maximize possible data re-use.

In this paper we present the design of a pseudo-log image processing IP block, integrated with different interpolation filtering techniques, using a framework that can automatically generate a memory hierarchy around the IP. This memory hierarchy reduces on-chip memory requirements, optimizes throughput and increases data-reuse. The design of the IP is fully performed at the algorithmic level in C/C++. The algorithmic description is profiled with a framework that creates a customized memory hierarchy, also described at algorithmic level, which can be synthesized using high level synthesis tools. Finally, high level synthesis is used to perform hardware design space exploration and performance estimation. The generated memory hierarchy is able to feed the IP with the highest bandwidth even in presence of high external memory latencies.

8656-13, Session 3

Real-time color/shape-based traffic signs acquisition and recognition system

Sergio Saponara, Univ. di Pisa (Italy)

A real-time system is proposed to acquire from an automotive fish-eye CMOS camera the traffic signs, and provide their automatic recognition on the vehicle network. Differently from the state-of-the-art, in this work color-detection is addressed exploiting the HSV color space which is robust to lighting changes. Hence the first stage of the processing system implements fish-eye correction and RGB to

HSV transformation. After color-based detection a noise deletion step is implemented and then, for the classification, a template-based correlation method is adopted to identify potential traffic signs, of different shapes, from acquired images. Starting from a segmented-image a matching with templates of the searched signs is carried out using a distance transform. These templates are organized hierarchically to reduce the number of operations and hence easing real-time processing for several types of traffic signs. Finally, for the recognition of the specific traffic sign, a technique based on extraction of signs characteristics and thresholding is adopted. Implemented on DSP platform the system recognizes traffic signs in less than 150 ms at a distance of about 15 meters from 640x480-pixel acquired images. Tests carried out with hundreds of images show a detection&recognition rate of about 90%, with false recognitions less than 4%

8656-14, Session 3

DSPACE hardware architecture for on-board real-time image/video processing in European space missions *(Invited Paper)*

Sergio Saponara, Massimiliano Donati, Luca Fanucci, Univ. di Pisa (Italy); Maximilian Odendahl, Reiner Leupers, RWTH Aachen (Germany); Walter Errico, SITAEL S.p.A. (Italy)

The increasing demand of on-board real-time image/video processing represents one of the critical issues in forthcoming scientific and commercial European space missions. To accomplish planetary observation, surveillance, Synthetic Aperture Radar imaging and telecommunication faster and faster signal and image processing algorithms are required. The only existing space-qualified European Digital Signal Processor (DSP) free of International Traffic in Arms Regulations restrictions (ATMEL TSC21020) faces a poor performance of 60 MFLOPS peak.

The DSPACE space-qualified DSP architecture fills the gap between the computational requirements and the available device. Its core leverages a pipelined and massively parallel architecture based on the very long instruction word (VLIW) paradigm: 64 registers, 4 arithmetic logic units, 2 multipliers and 2 address generation units are arranged into 2 identical data-paths with cross-path capabilities.

Both the synthesizable VHDL and the software development tools are generated from the LISA high-level description and then refined at RTL level. A Xilinx-XC5VLX110 FPGA is chosen to realize an engineering prototype, providing an easy migration to the space-qualified Xilinx-XQR5VFX130 FPGA. Finally, first synthesis results on ATMEL 180 nm standard cell ASIC technology show an area of around 300 kgates and a peak performance of 1 GOPS and 750MFLOPS at 125MHz.

8656-15, Session 4

Priority-based methods for reducing the impact of packet loss on HEVC encoded video streams

(Invited Paper)

James M. Nightingale, Qi Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom)

No Abstract Available

8656-16, Session 4

Low complexity DCT engine for image and video compression

Maher Jridi, Yousri Ouerhani , Ayman Alfalou, ISEN Brest (France)

In this paper, we defined a new scalable and reconfigurable 2D-DCT architecture. The latter is able to transform spatial pixels to spectral pixels while taking into account the constraints of the considered compression standard. Due to our new matrix decomposition, we could define one common 2D-DCT architecture. The constant multipliers can be configured to handle the case of RealDCT and/or IntDCT (multiplication by ±2). The proposed algorithm not only provides a reduction of computational complexity, but also leads to scalable pipelined design in systolic arrays. Indeed, the 8*8 StdDCT can be computed by using 4*4 StdDCT which can be obtained by calculating 2*2 StdDCT. Besides, the proposed structure can be extended to deal with higher number of N (i.e. 16*16 and 32*32). The FPGA implementation results show the performance of the proposed architecture when compared with conventional designs. In particular, for N=4, it is found that the proposed design have nearly third the area-time complexity of the existing DCT structures. This gain is expected to be higher for a greater size of 2D-DCT.

8656-17, Session 4

A CABAC codec of H.264AVC with secure arithmetic coding

Nihel Neji, Maher Jridi, Ayman Alfalou, ISEN Brest (France); Nouri Masmoudi, Univ. of Sfax (Tunisia)

This paper presents an optimized 1080p H.264/AVC decoder system for HDTV displays. For high quality streaming, the decoder uses Binary Arithmetic Decoding algorithm with high complexity and a JVCE scheme which was not used before. The proposed optimized allows us to encrypt the information using a pseudorandom number generator (PRNG). Thus we achieved the tow operations (compression and encryption) simultaneously and in a depended manner which is a novelty in this kind of architecture. Moreover, we investigated the hardware implementation of CABAC decoder. The proposed architecture is based on parallelism to handle significant pixel rates (i.e. 1080p) videos with low cost and high performance for most frequent SEs. This was checked using 5 video frames. The experimental results we obtained using an FPGA (Xilinx's ISE) show that our design is relevant to decode main profile HD video stream.

8656-18, Session 4

A modified prediction scheme of the H.264 multiview video coding to improve the decoder performance

Ayman M. Hamdan, Hussein A. Aly, Mohamed M. Fouad, Military Technical College (Egypt); Richard M Dansereau, Carleton University (Canada)

In this paper, we modify the interview prediction scheme of the standard Multiview Video Coding (MVC) to reduce the bit-rate with comparable video quality measures. The proposed scheme is compared to the MVC standard and its predecessor, namely, H.264/MPEG-4 AVC using real data sets. Clear improvements are shown using the proposed scheme over competing approaches in terms of bit-rate with a slight increase in the average PSNR values (dB), as well as a slight increase in the computational complexity.

Conference 8657: Computational Imaging XI

Tuesday - Thursday 5 –7 February 2013

Part of Proceedings of SPIE Vol. 8657 Computational Imaging XI

8657-1, Session 1

A unifying retinex model based on non-local differential operators

Dominique Zosso, Giang Tran, Stanley J. Osher, Univ. of California, Los Angeles (United States)

In this paper, we present a unifying framework for retinex that is able to reproduce many of the existing retinex implementations within a single model. The fundamental assumption, as shared with many retinex models, is that the observed image is a multiplication between the illumination and the true underlying reflectance of the object. Starting from Morel's 2010 PDE model for retinex, where illumination is supposed to vary smoothly and where the reflectance is thus recovered from a hard-thresholded Laplacian of the observed image in a Poisson equation, we define our retinex model in similar but more general two steps.

First, look for a filtered gradient that is the solution of an optimization problem consisting of two terms: The first term is a sparsity prior of the reflectance, such as the TV or H1 norm, while the second term is a quadratic fidelity prior of the reflectance gradient with respect to the observed image gradients. In a second step, since this filtered gradient almost certainly is not a consistent image gradient, we then look for a reflectance whose actual gradient comes close.

Beyond unifying existing models, we are able to derive entirely novel retinex formulations by using more interesting non-local versions for the sparsity and fidelity prior. Hence we define within a single framework new retinex instances particularly suited for texture-preserving shadow removal, cartoon-texture decomposition, color and hyperspectral image enhancement.

8657-2, Session 1

Subspace methods for computational relighting

Ha Q. Nguyen, Siying Liu, Minh N. Do, Univ. of Illinois at Urbana-Champaign (United States)

We propose a vector space approach for relighting a Lambertian convex object with distant light source, whose crucial task is the decomposition of the reflectance function into albedos (or reflection coefficients) and lightings based on a set of images of the same object and its 3-D model. Making use of the fact that reflectance functions are well approximated by a low-dimensional linear subspace spanned by the first few spherical harmonics, this inverse problem can be formulated as a matrix factorization, in which the basis of the subspace is encoded in the spherical harmonic matrix S. A necessary and sufficient condition on S for unique factorization is derived with an introduction to a new notion of matrix rank called nonseparable full rank. An SVD-based algorithm for exact factorization in the noiseless case is introduced. In the presence of noise, the algorithm is slightly modified by incorporating the positivity of albedos into a convex optimization problem. Implementations of the proposed algorithms are done on a set of synthetic data.

8657-3, Session 1

Bayesian demosaicing using gaussian scale mixture priors with local adaptivity in the dual tree complex wavelet packet transform domain

Bart Goossens, Jan Aelterman, Hiep Luong, Aleksandra Pi?urica, Wilfried Philips, Univ. Gent (Belgium)

In digital cameras and mobile phones, there is an ongoing trend to increase the image resolution, decrease the sensor size and to use lower exposure times. Because smaller sensors inherently lead to more noise and a worse spatial resolution, digital post-processing techniques are required to resolve many of the artifacts. Color filter arrays (CFAs), which use alternating patterns of color filters applied at every pixel position, are very popular because of price and power consumption reasons. However, color filter arrays require the use of a post-processing technique such as demosaicing to recover full resolution RGB images. Recently, there has been some interest in techniques that jointly perform the demosaicing and denoising. This has the advantage that the demosaicing and denoising can be performed optimally (e.g. in the MSE sense) for the considered noise model, while avoiding artifacts introduced when using demosaicing and denoising sequentially.

In this paper, we will continue the research line of the wavelet-based demosaicing techniques. These approaches are computationally simple and very suited for combination with denoising. Therefore, we will derive Bayesian Minimum Squared Error (MMSE) joint demosaicing and denoising rules in the complex wavelet packet domain, taking local adaptivity into account. As an image model, we will use Gaussian Scale Mixtures, thereby taking advantage of the directionality of the complex wavelets. Our results show that this technique is well capable of reconstructing fine details in the image, while removing all of the noise, at a relatively low computational cost. In particular, the complete reconstruction (including color correction, white balancing etc) of a 10 megapixel RAW image takes 4.4 sec on a recent mid-range GPU.

8657-4, Session 1

Demosaicing for RGBZ sensor

Lilong Shi, Ilia Ovsiannikov, Samsung Semiconductor, Inc. (United States); Dong-Ki Min, Yohwan Noh, SAMSUNG Electronics Semiconductor (Korea, Republic of); Wanghyun Kim, Sunhwa Jung, Joonho Lee, Deokha Shin, Hyekyung Jung, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Gregory Waligorski, Michelle Wang, Wendy Wang, Samsung Semiconductor, Inc. (United States); Yoondong Park, Chilhee Chung, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

In this paper, we proposed a new technique for demosaicing a unique RGBZ color-depth imaging sensor, which captures color and depth images simultaneously, with a specially designed color-filter-array (CFA) where two out of six RGB color rows are replaced by "Z" pixels that capture depth information but little color information. Therefore, in an RGBZ image, the red, green and blue colors are more sparsely sampled than in a normal Bayer image. Due to the missing rows in the data image, commonly used demosaicing algorithms for the standard Bayer CFA cannot be applied directly. To this end, our method first fills-in the missing rows to reconstruct a full Bayer CFA, followed by a color-selective adaptive demosaicing algorithm that interpolates missing color components. In the first step, unlike common bilinear interpolation approaches that tend to blur edges when estimating the missing pixels, our edge-based directional interpolation approach, derived from deinterlacing techniques, emphasizes on reconstructing more straight and sharp edges with fewer artifacts and thereby preserves the vertical resolution in the reconstructed the image. To avoid using the newly estimated pixels for demosaicing, the bilateral-filter-based approach interpolates the missing color samples based on weighted average of adaptively selected known pixels from the local neighborhoods. Tests show that the proposed method reconstructs full color images while preserving edges details, avoiding artifacts, and removing noise with high efficiency.

8657-5, Session 1

Auto zoom crop from face detection and facial features

Raymond Ptucha, David Rhoda, Brian Mittelstaedt, Eastman Kodak Co. (United States)

We propose an automatic algorithm that recomposes an original capture such that it represents the composition that the photographer had wished he/she had taken in the first place. Additionally, this algorithm enables trimming to differing aspect ratios for email, smart phone, tablet, or print viewing.

Our facial based algorithm utilizes automatic zoom and crop composition rules that are based upon head size, number of faces, position of faces, a hierarchical grouping of faces, and input to output aspect ratio. We utilize symmetric cropping and rule of thirds for final composition. We further introduce methods of auto zoom crop that utilize facial pose, blink, eye gaze, expression, and subject interrelationships. For example, if a subject is looking to the left, the crop region is biased towards the left. If a detected face exhibits eye blink, its weight is decreased.

We demonstrate through controlled psychophysical studies that our newly introduced face based algorithm significantly outperforms simpler rule based zoom/crop algorithms commonly in use today, sophisticated state of the art main subject detectors, and even main subject detectors aided by skin and face detection.

8657-6, Session 2

Optimal filters for high-speed compressive detection in spectroscopy

Gregery T. Buzzard, Bradley J. Lucier, Purdue Univ. (United States)

Recent advances allow for the construction of filters with precisely defined frequency response for use in Raman chemical spectroscopy. In this paper we give a probabilistic interpretation of the output of such filters and use this to give an algorithm to design optimal filters to minimize the mean squared error in the estimated photon emission rates for multiple spectra. Experiments using these filters demonstrate that detecting as few as 10 Raman scattered photons (in as little time as 30 microseconds) can be sufficient to positively distinguish chemical species. This speed should allow ``chemical imaging" of samples.

8657-8, Session 2

Neutron Imaging with Coded Sources: Design Pitfalls and the Implementation of a Simultaneous Iterative Reconstruction Technique

Hector J. Santos-Villalobos, Philip R. Bingham, Oak Ridge National Lab. (United States); Jens Gregor, The Univ. of Tennessee (United States)

The limitations in neutron flux and resolution (L/D) of current neutron imaging systems can be addressed with a Coded Source Imaging system with magnification (xCSI). More precisely, the multiple sources in an xCSI system can exceed the flux of a single pinhole system for several orders of magnitude, while maintaining a higher L/D with the small sources. Moreover, designing for an xCSI system reduces noise from neutron scattering, because the object is placed away from the detector to achieve magnification. However, xCSI systems are adversely affected by correlated noise such as non-uniform illumination of the neutron source, incorrect sampling of the coded radiograph, misalignment of the coded masks, mask transparency, and the imperfection of the system Point Spread Function (PSF). We argue that a model-based reconstruction algorithm can overcome these problems and describe the implementation of a Simultaneous Iterative Reconstruction Technique algorithm for coded sources. Design pitfalls that preclude a satisfactory reconstruction are documented.

8657-10, Session 3

Physics-based regularization

Jeffrey P. Simmons, Air Force Research Lab. (United States)

No Abstract Available

8657-11, Session 3

Reconstruction of bright field STEM with Bragg diffraction correction

Charles A. Bouman, Purdue Univ. (United States)

No Abstract Available

8657-35, Session K1

Petapixel photography and the limits of camera information capacity *(Keynote Presentation)*

David J Brady, Daniel L Marks, Steven Feller, Duke Univ. (United States); Michael E Gehm, Dathon Golish, Esteban Vera, University of Arizona (United States); David Kittle, Duke Univ. (United States)

The monochromatic single frame pixel count of a camera is limited by diffraction to the space-bandwidth product, roughly the aperture area divided by the square of the wavelength. We have recently shown that it is possible to approach this limit using multiscale lenses for cameras with space bandwidth product between 1 and 100 gigapixels. When color, polarization, coherence and time are included in the image data cube, camera information capacity increases by over 6 orders of magnitude. This talk reviews progress in the construction of DARPA AWARE gigapixel cameras and describes compressive measurement strategies that may be used in combination with multiscale systems to push camera capacity to near physical limits.

8657-12, Session 4

Sparse imaging for fast electron microscopy

Hyrum S. Anderson, Jovana Helms, Brandon Rohrer, Jason Wheeler, Kurt W. Larson, Sandia National Labs. (United States)

Scanning electron microscopes (SEMs) are used in neuroscience and materials science to image centimeters of sample area at nanometer scales. Since imaging rates are in large part SNR-limited, large collections can lead to weeks of around-the-clock imaging time. To increase data collection speed, we propose and demonstrate on an operational SEM a fast method to sparsely sample and reconstruct smooth images. To accurately localize the electron probe position at fast scan rates, we model the dynamics of the scan coils, and use the model to rapidly and accurately visit a randomly selected subset of pixel locations. Images are reconstructed from the undersampled data by compressed sensing inversion using image smoothness as a prior. We report image fidelity as a function of acquisition speed by comparing traditional raster to sparse imaging modes. Our approach is equally applicable to other domains of nanometer microscopy in which the time to position a probe is a limiting factor (e.g., atomic force microscopy), or in which excessive electron doses might otherwise alter the sample being observed (e.g., scanning transmission electron microscopy).

8657-13, Session 4

Building and enforcing shape priors for segmentation of alloy micrographs

Landis M. Huffman, The MITRE Corp. (United States); Jeffrey P. Simmons, Air Force Research Lab. (United States); Marc De Graef,

Carnegie Mellon Univ. (United States); Ilya Pollak, Purdue Univ. (United States)

Computer simulation of metal alloys is an emerging trend in materials development. Simulated replicas of fabricated alloys are based on the segmentations of alloy micrographs. Therefore, accurate segmentation of visible precipitates is paramount to simulation accuracy. Since the shape and size of precipitates are key indicators of physical alloy properties, automated segmentation algorithms must account for abundant prior information of precipitate shape. We present a new method for constructing a prior enforcing rectangular shape which can be applied within a min-cut framework for maximum a-posteriori segmentation.

8657-14, Session 5

Real-time dynamic range and signal to noise enhancement in beam-scanning microscopy by integration of sensor characteristics, data acquisition hardware, and statistical methods

Garth J. Simpson, Purdue Univ. (United States)

Despite the ubiquitous use of multi-photon and confocal microscopy measurements in biology, the core techniques suffer from fundamental compromises between signal to noise (S/N) and linear dynamic range (LDR). In this study, direct synchronous digitization of voltage transients coupled with statistical analysis is shown to allow S/N approaching the theoretical maximum throughout an LDR spanning 6 decades, limited only by the dark counts of the detector on the low end and by the intrinsic nonlinearities of the photomultiplier tube (PMT) detector on the high end. Synchronous digitization of each voltage transient represents a fundamental departure from established methods in confocal/multi-photon imaging, which are currently based on either photon counting or signal averaging. High information-density data acquisition (up to 3.2 GB/s of raw data) enables the smooth transition between the two modalities on a pixel-by-pixel basis and the ultimate writing of much smaller files (few kB/s). Modeling of the PMT response allows extraction of key sensor parameters from the histogram of voltage peak-heights, which in turn can be used for automated optimization of the thresholds and on the fly correction of 1/f noise in the baseline. Applications in second harmonic generation (SHG) microscopy are described requiring quantitation over large dynamic ranges.

8657-15, Session 5

Multi-modal electron tomography

Lawrence F. Drummy, UES, Inc. (United States)

No Abstract Available

8657-16, Session 5

Segmentation of materials images using 3D electron interaction modeling

Mary L. Comer, Dae Woo Kim, Purdue Univ. (United States)

No Abstract Available

8657-17, Session 5

A forward modeling approach to electron back-scatter diffraction patterns

Marc De Graef, Carnegie Mellon Univ. (United States)

No Abstract Available

8657-18, Session 5

Interactive grain image segmentation using graph cut algorithms

Jarrell W Waggoner, Youjie Zhou, Univ. of South Carolina (United States); Jeffrey P. Simmons, Air Force Research Lab. (United States); Ayman Salem, Materials Resources International (United States); Marc De Graef, Carnegie Mellon Univ. (United States); Song Wang, Univ. of South Carolina (United States)

Segmenting materials images is a laborious and time-consuming process and automatic image segmentation algorithms usually contain imperfections and errors. Interactive segmentation is a growing topic in the areas of image processing and computer vision, which seeks to find a balance between fully automatic methods and fully-manual segmentation processes. By allowing minimal and simplistic interaction from the user in an otherwise automatic algorithm, interactive segmentation is able to simultaneously reduce the time taken to segment an image while achieving better segmentation results. Given the specialized structure of materials images and level of segmentation quality required, we show an interactive segmentation framework for materials images that has two key contributions: 1) a multi-labeling framework that can handle a large number of structures while still quickly and conveniently allowing manual interaction in real-time, and 2) a parameter estimation approach that prevents the user from having to manually specify parameters, increasing the simplicity of the interaction. We show a full formulation of each of these contributions and example results from their application.

8657-31, Session IPI

Sub-pixel depth camera registration using gradient circle pattern

Seungkyu Lee, Samsung Advanced Institute of Technology (Korea, Republic of)

Recently, consumer depth cameras have widely applied to robotics, computer vision and graphics applications. Furthermore sensor fusion with color camera has provided multimodal 3-dimensional data for many research problems. As a natural advance in this field, now researcher started to combine multiple depth sensors for many reasons; additional noise elimination, more complete 3D reconstruction and robot navigation.

In this work, we propose a new sub-pixel external calibration and registration method of multiple depth cameras. We propose a new chessboard pattern fit for low resolution depth images. Instead of using corner point for correspondence detection between multiple depth images, we use our chessboard pattern structure matching scheme for higher accuracy of correspondence detection in such challenging low resolution condition. Rather than direct matching between the cues extracted from multiple images, we build a parameterized model of our new chessboard pattern and perform the matching between the model and each image. By doing this we can finely tune the matching location for sub-pixel external calibration. Simulation and experimental results on real depth images prove the proposed sub-pixel external calibration and registration showing around 40% of accuracy gain. The proposed method will contribute further use of multiple depth images of other researchers providing merged depth point cloud from multiple depth sensors.

8657-32, Session IPI

Efficient synthetic refocusing method from multiple coded aperture images for 3D user interaction

Sungjoo Suh, Changkyu Choi, Dusik Park, Chang-Yeong Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

In this paper, we propose an efficient synthetic refocusing method from multiple coded aperture images for 3D user interaction. The proposed

method is applied to a flat panel display with a sensor panel which forms lens-less multi-view cameras. To capture the scene in front of the display, the modified uniformly redundant arrays (MURA) patterns are displayed on the LCD screen without the backlight. Through the imaging patterns on the LCD screen, MURA coded images are captured in the sensor panel. Instead of decoding all coded images to synthetically generate a refocused image, the proposed method only decodes one coded image corresponding to the refocusing image at a certain distance after circularly shifting and averaging all coded images. Further, based on the proposed refocusing method, the depth of an object in front of the display is estimated by finding the most focused image for each pixel through a stack of the refocused images at different depth levels. Experimental results show that the proposed method captures an object in front of the display, generates refocused images at different depth levels, and accurately determines the depth of an object including real human hands near the display.

8657-33, Session IPI

Multiscale based adaptive contrast enhancement

Muhammad I. Abir, Missouri Univ. of Science and Technology (United States) and Idaho National Lab. (United States); Fahima F. Islam, Missouri Univ. of Science and Technology (United States); Daniel M. Wachs, Idaho National Lab. (United States); Hyoung Koo Lee, Missouri Univ. of Science and Technology (United States)

A contrast enhancement algorithm is developed for enhancing the contrast of x-ray images. The algorithm is based on Laplacian pyramid image processing technique. The image is decomposed into three frequency sub-bands- low, medium, and high. Each sub-band contains different frequency information of the image. The detail structure of the image lies on the high frequency sub-band and the overall structure lies on the low frequency sub-band. Apparently it is difficult to extract detail structure from the high frequency sub-bands. Enhancement of the detail structures is necessary in order to find out the calcifications on the mammograms, cracks on any object such as fuel plate, etc. In our proposed method contrast enhancement is achieved from high and medium frequency sub-band images by decomposing the image based on multi-scale Laplacian pyramid and enhancing contrast by suitable image processing. Standard Deviation-based Modified Adaptive contrast enhancement (SDMACE) technique is applied to enhance the low-contrast information on the sub-bands without overshooting noise. An alpha-trimmed mean filter is used in SDMACE for sharpness enhancement. After modifying all sub-band images, the final image is derived from reconstruction of the sub-band images from lower resolution level to upper resolution level including the residual image. To demonstrate the effectiveness of the algorithm an x-ray of a fuel plate and two mammograms are analyzed. Subjective evaluation is performed to evaluate the effectiveness of the algorithm. The proposed algorithm is compared with the well-known contrast limited adaptive histogram equalization (CLAHE) algorithm. Experimental results prove that the proposed algorithm offers improved contrast of the x-ray images.

8657-19, Session 6

An enhanced grid-based Bayesian array for target tracking

Qian Sang, Zongli Lin, Scott T. Acton, Univ. of Virginia (United States)

A grid-based Bayesian approach (GBA) for robust visual tracking has recently been developed, which proposes a novel method of deterministic sample generation and sample weighting for position estimation. In particular, a target motion model is constructed, predicting target position in the next frame based on estimations in previous frames. Samples are generated by gridding within an ellipsoid centered at the prediction. For localization, radial edge detection is applied for each sample to determine if it is inside the target boundary. Sample weights are then assigned according to the number of the edge points detected around the sample and its distance from the predicted

position. The position estimation is computed as the weighted sum of the sample set.

In this paper, we enhance the capacity of the GBA tracker in accommodating the tracking of targets in video with erratic motion, by introducing adaptation in the motion model and iterative position estimation. The improved tracking performance over the original GBA tracker are demonstrated in tracking a single leukocyte in vivo and ground vehicle target observed from UAV videos, both undergoing abrupt changes in motion. The experimental results show that the enhanced GBA tracker outperforms the original by tracking more than 10% of the total number of frames, and increases the number of video sequences with all frames tracked by greater than 20%.

8657-20, Session 6

A fourth-order active contour shape prior for multiple-instance object detection in images

Ikhlef Bechar, INRIA Sophia Antipolis - Méditerranée (France); Ian H. Jermyn, Durham Univ. (United Kingdom); Josiane B. Zerubia, INRIA Sophia Antipolis - Méditerranée (France)

No Abstract Available

8657-21, Session 6

Efficient occlusion reasoning for articulated tracking in monocular views

Landis M. Huffman, The MITRE Corp. (United States); Ilya Pollak, Purdue Univ. (United States)

Pose estimation and tracking of articulated objects like humans is particularly difficult due to the complex occlusions among the articulated parts. Without the benefit of multiple views, resolution of occlusions becomes both increasingly valuable and challenging. We propose a method for articulated 3D pose estimation from monocular video which uses nonparametric belief propagation and employs a novel and efficient approach to occlusion reasoning. We present a human tracking application, and evaluate results using the the Human Eva II data set.

8657-22, Session 6

An efficient optimizer for simple point process models

Ahmed Gamal Eldin, INRIA Rhône-Alpes (France); Guillaume Charpiat, Xavier Descombes, Josiane B. Zerubia, INRIA Sophia Antipolis - Méditerranée (France)

No Abstract Available

8657-23, Session 7

Texture mapping 3D planar models of indoor environments with noisy camera poses

Peter Cheng, Michael Anderson, Stewart He, Avideh Zakhor, Univ. of California, Berkeley (United States)

Automated 3D modeling of building interiors is useful in applications such as virtual reality and environment mapping. Applying textures to these models results in useful photorealistic visualizations of the indoor environment. The quality of such textures is largely determined by the camera poses used to project source images onto surfaces while texturing. In cases where camera poses are highly imprecise, strong misalignments are visible in areas where successive images are projected. We propose two approaches to reduce discontinuities in texture mapping 3D models made of planar surfaces. The first one is robust to images taken from greatly varying angles, and can be used

Conference 8657

for all manner of images and planes. The second approach produces more seamless textures by taking advantage of scenarios where camera headings are largely perpendicular to their respective surface to be textured. This approach typically results in more artifact-free texture mapping. We propose a scheme whereby the first method is used to texture map floors and ceilings, and the second one is used on walls. We demonstrate the effectiveness of this scheme to a number of data sets captured by an ambulatory 3D modeling system equipped with laser scanners and cameras.

8657-24, Session 7

Optical touch sensing: practical bounds for design and performance

Alexander Blaessle, UBC Okanagan (Canada); Bebart Janbek, Simon Fraser Univ. (Canada); Lifeng Liu, Univ. of Pittsburgh (United States); Kanna Nakamura, Univ. of Maryland, College Park (United States); Kimberly Nolan, Drexel Univ. (United States); Victor Paraschiv, Univ. of Victoria (Canada)

Touch screens are commonly used in many consumer applications, ranging in size from smartphones and tablets to touch-walls and collaborative-surfaces. Optical touch sensing technology is especially well suited for large-size touch surfaces, since cameras and light sources are usually placed on the perimeter of the touch area. Thus, the cost of the solution is proportional to the perimeter rather than the area of the screen. The number of cameras and light sources plays an important role in determining the cost and performance of the system. In this work, we analyze and determine various bounds and tradeoffs relating the number of sensors used and the resulting resolution and discrimination-capability of the system.

We start with analyzing the system's ability to discriminate two fingers touching the screen. Specifically, we determine the minimum number of cameras required to guarantee the ability to discern two circular objects on the screen. The next problem we consider is that of identifying polygons touching the screen and describe a method for placing the cameras that ensures this ability. We conclude with a direct numerical tool to relate the number of cameras and light sources and the resulting resolution of the touch screen using some practical cases.

8657-25, Session 8

Light field image denoising using a linear 4D frequency-hyperfan all-in-focus filter

Donald G. Dansereau, Daniel L. Bongiorno, Oscar Pizarro, Stefan B. Williams, The Univ. of Sydney (Australia)

Imaging in low light is problematic as sensor noise can dominate imagery, and increasing illumination or aperture size is not always effective or practical. Computational photography offers a promising solution in the form of the light field camera, which by capturing redundant information offers an opportunity for elegant noise rejection. We show that the light field of a Lambertian scene has a 4D hyperfan-shaped frequency-domain region of support, and by designing and implementing a filter with appropriately shaped passband we accomplish denoising with a single all-in-focus linear filter. Drawing examples from the Stanford Light Field Archive with synthetic noise, and real-world imagery captured using the Lytro consumer-grade plenoptic camera, we demonstrate that the hyperfan outperforms competing methods including synthetic focus, fan-shaped antialiasing filters, and a range of modern nonlinear image and video denoising techniques. We show the filter preserves depth of field, making it a single-step all-in-focus denoising filter suitable for general-purpose light field rendering. We include results for different noise types and levels, over a variety of metrics, and in real-world scenarios. Finally, we show that the hyperfan's performance scales with aperture count.

8657-26, Session 8

Computational imaging approach for fanbeam x-ray scatter imaging

Joseph A O'Sullivan, David G. Politte, Washington Univ. in St. Louis (United States); Kenneth MacCabe, Kalyani Krishnamurthy, Duke Univ. (United States); Ikenna Odinaka, Washington Univ. in St. Louis (United States); Anuj Kapadia, David J. Brady, Duke Univ. (United States)

In x-ray scatter imaging, tomographic measurements of the forward scatter distribution are used to infer scatter densities within a volume. A mask placed between the object and the detector array provides information about scatter angles. An efficient computational implementation of the forward and backward model facilitate iterative algorithms based upon a Poisson log-likelihood. Results are presented on simulated and Monte Carlo data.

8657-27, Session 8

Robust registration of electron tomography projections without fiducial makers

Viet Dung Tran, Maxime Moreaud, IFP Energies Nouvelles (France); Éric M. Thiébaut, Ctr. de Recherche Astronomique de Lyon (France); Loïc Dénis, Jean Marie Becker, Lab. Hubert Curien, CNRS (France)

A major issue in electron tomography is the misalignment of the projections contributing to the reconstruction. The current alignment techniques usually employ fiducial markers such as gold particles for a correct alignment of the images. When the use of markers is not possible, the correlation between adjacent projections is used to align them. However, this method sometimes fails. In this paper, we propose a new method for the alignment of transmission electron microscopy (TEM) images series without the need of fiducial markers. The proposed approach is composed of two steps. The first step consists of an initial alignment process, where relies on the minimization of a cost function based on robust statistics measuring the similarity of a projection to its previous projections in the series. It reduce strong shifts resulting from the acquisition between successive projections. The second steps aligns the projections finely. Pre-registered projections are used to initialize an iterative refinement alignment process, which aligns each projection with its simulated version obtained from reconstructed volume. We have successfully tested our method with real projections of a zeolite support catalyst with TEM images. It recovers in an accurate manner the changes in translation, rotation and scaling parameters.

8657-28, Session 9

Low signal noise modeling for statistical CT reconstruction

Jean-Baptiste Thibault, GE Healthcare (United States); Ken D. Sauer, Univ. of Notre Dame (United States); Charles A. Bouman, Purdue Univ. (United States)

No Abstract Available

8657-29, Session 9

Joint reconstruction and segmentation of electron tomography data

Ahmet Tuysuzoglu, William C. Karl, David A. Castañón, M. Selim Ünlü, Boston Univ. (United States)

No Abstract Available

8657-30, Session 9

Analysis of image color and effective bandwidth as a tool for assessing air pollution at urban spatiotemporal scale

Yael Etzion-Cohen, David M. Broday, Barak Fishbain, Technion-Israel Institute of Technology (Israel)

Size and concentration of airborne particulate matter (PM) are important indicators for air pollution events and public health risks, since exposure to fine particles smaller than 2.5 ?m (PM2.5) has been associated with adverse health effects. It is therefore important to monitor the spatiotemporal variations of PM size resolved concentrations in the ambient air near the ground at an urban-scale resolution. The suggested method here utilizes quantitative measures of urban scene image characteristics for predicting PM concentrations. Our study explores different quantitative characteristics of the image contrast as potential prediction variables for PM concentrations, including Fishbain's et al. (2008) image effective bandwidth (IEB) and yellow intensity level, which is physically related to scattering of radiance in the visible range by fine particles smaller than 1 ?m. For validating the suggested method, we have assembled a large data set that consists of time series imaging of mixed scenery (industrial facilities, roads and residential neighborhoods) and measurements from air quality monitoring stations in the studied region that monitor PM concentrations and meteorological data. Quantitative and qualitative statistical evaluation of the suggested method shows that PM dynamics and concentrations can be inferred and studied from the acquired images.

Conference 8658: Document Recognition and Retrieval XX

Tuesday - Thursday 5 –7 February 2013

Part of Proceedings of SPIE Vol. 8658 Document Recognition and Retrieval XX

8658-1, Session 1

Semi-structured document image matching and recognition

Olivier Augereau, Nicholas Journet, Jean-Philippe Domenger, LaBRI (France)

This article presents a method to recognize and to localize semi-structured documents such as ID cards, tickets, invoices, etc. Standard object recognition methods based on interest points work well on natural images but fail on document images because of repetitive patterns like text. In this article, we propose an adaptation of object recognition for image documents. The advantages of our method is that it does not use character recognition or segmentation and it is robust to rotation, scale, illumination, blur, noise and local distortions. Furthermore, tests show that an average precision of 97.2% and recall of 94.6% is obtained for matching 7 different kinds of documents in a database of 2155 documents.

8658-2, Session 1

Rotation-robust math symbol recognition and retrieval using outer contours and image subsampling

Siyu Zhu, Lei Hu, Richard Zanibbi, Rochester Institute of Technology (United States)

This paper presents an unified recognition and retrieval system for isolated offline printed mathematical symbols for the first time. The system is based on nearest neighbor scheme and uses modified Turning Function and Grid Features to calculate the distance between two symbols based on Sum of Squared Difference (SSD). An unwrap process and an alignment process are applied to modify Turning Function to deal with the horizontal and vertical shifts caused by the changing of staring point and rotation. This modified Turning Function makes our system robust against rotation of the symbol image. The system obtains top-1 recognition rate of 96.90% and 47.27% Area Under Curve (AUC) of Precision/Recall plot on the InftyCDB-3 dataset. Experiment result shows that the system with modified Turning Function performs much better than the system with original Turning Function on the rotated InftyCDB-3 dataset.

8658-3, Session 1

NESP: Nonlinear enhancement and selection of plane for optimal segmentation and recognition of scene word images

Deepak Kumar, M. N. Anil Prasad, A. G. Ramakrishnan, Indian Institute of Science (India)

In this paper, we report a breakthrough result on the difficult task of segmentation and recognition of coloured text from the word image dataset of ICDAR robust reading competition challenge 2: reading text in scene images. We split the word image into individual colour, gray and luminous planes; We then enhance the contrast of each of these planes independently by a power-law transform. We compute the discrimination factor of each plane. The plane that has maximum discrimination factor is selected for segmentation. We use the trial version of Omnipage OCR on the binarized words for recognition. We compare our recognition results with those reported on ICDAR

2011 word dataset and also on ICDAR 2003 word dataset. The word recognition rate of NESP method is 72.77% and 66.22% for ICDAR 2011 and 2003 word datasets, respectively. We have created ground-truth for each image at the pixel level to benchmark these datasets. We have used our MAST toolkit for creation of ground-truth images. The recognition rate of benchmarked images is 86.73% and 83.87% for ICDAR 2011 and 2003 datasets, respectively.

8658-4, Session 2

Combining evidence using likelihood ratios in writer verification

Sargur N. Srihari, Univ. at Buffalo (United States); Dimitry Kovalenko, Yi Tang, Gregory Ball, Univ. at Buffalo, SUNY (United States)

Forensic identification is the task of determining whether or not observed evidence arose from a known source. It involves determining a likelihood ratio (LR) -- the ratio of the joint probability of the evidence and source under the identification hypothesis (that the evidence came from the source) and under the exclusion hypothesis (that the evidence did not arise from the source). In LR- based decision methods, particularly handwriting comparison, a variable number of input evidences is used. A decision based on many pieces of evidence can result in nearly the same LR as one based on few pieces of evidence. We consider methods for distinguishing between such situations. One of these is to provide confidence intervals together with the decisions and another is to combine the inputs using weights. We suggest a new method that generalizes Bayesian approach and uses an explicitly defined discount function. Empirical evaluation with three data sets including large-scale synthesized data and specific examples of handwriting comparison shows greater flexibility of the designed method.

8658-5, Session 2

Handwritten word preprocessing for database adaptation

Cristina A. Oprean, Laurence Likforman-Sulem, Telecom ParisTech (France); Chafic Mokbel, University of Balamand (Lebanon)

Handwriting recognition systems are typically trained using publicly available databases, where data have been collected in controlled conditions (image resolution, paper background, noise level, ..). Since this is not often the case in real-world scenarios, classification performance can be affected when novel data is presented to the word recognition system. To overcome this problem, we present in this paper a new approach called database adaptation. It consists of processing one set (training or test) in order to adapt it to the other set (test or training, respectively). We consider a specific case of processing, namely stroke thickness normalization and pixel intensity uniformization. The advantage of such approach is that we can re-use the existing recognition system trained on controlled data. We conduct several experiments with the Rimes 2011 word database and a real-world database. We adapt either the test set or the training set. Results show that training set adaptation achieves slightly better results than test set adaptation, at the cost of a second training stage on the adapted data. Accuracy of data set adaptation is increased by 2% to 3% in absolute value over no adaptation.

8658-6, Session 2

Optimal policy for labeling training samples

Lester Lipsky, Univ. of Connecticut (United States); Daniel Lopresti, Lehigh Univ. (United States); George Nagy, Rensselaer Polytechnic Institute (United States)

Confirming the labels of automatically classified patterns is generally faster than entering new labels or correcting incorrect labels. Most labels assigned by a classifier, even if trained only on relatively few pre-labeled patterns, are correct. Therefore the overall cost of human labeling can be decreased by interspersing labeling and classification. Given a parameterized model of the error rate as an inverse power law function of the size of the training set, the optimal splits can be computed rapidly. Projected savings in operator time are over 60% for a range of empirical error functions for hand-printed digit classification with ten different classifiers.

8658-7, Session 2

Evaluation of lexicon size variations on a verification and rejection system based on SVM, for accurate and robust recognition of handwritten words

Yann Ricquebourg, Bertrand Couasnon, IRISA / INSA de Rennes (France); Laurent Guichard, E2I SAS (France)

The transcription of handwritten words remains a still challenging and difficult task remains. When processing full pages, approaches are limited by the trade-off between automatic recognition errors and the tedious aspect of human user verification. In this article, we present our investigations to improve the capabilities of an automatic recognizer, so as to be able to reject unknown words (not to take wrong decisions) while correctly rejecting (i.e. to recognize as much as possible from the lexicon of known words).

This is the active research topic of developing a verification system that optimize the trade-off between performance and reliability. To minimize the recognition errors, a verification system is usually used to accept or reject the hypotheses produced by an existing recognition system. Thus, we re-use our novel verification architecture1 here: the recognition hypotheses are re-scored by a set of the support vector machines, and validated by a verification mechanism based on multiple rejection thresholds. In order to tune these (class-dependent) rejection thresholds, an algorithm based on dynamic programming has been proposed which focus on maximizing the recognition rate for a given error rate.

Experiments have been carried out on the RIMES database in three steps. The first two showed that this approach results in a performance superior or equal to other state-of-the-art rejection methods. We focus here on the third one showing that this verification system also greatly improves results of keywords extraction in a set of handwritten words, with a strong robustness to lexicon size variations (21 lexicons have been tested from 167 entries up to 5,600 entries) which is particularly relevant to our application context cooperating with humans, and only made possible thanks to the rejection ability of this proposed system. The proposed verification system, compared to a HMM with simple rejection, improves on average the recognition rate by 57% (resp. 33% and 21%) for a given error rate of 1% (resp. 5% and 10%).

Keywords: Handwritten word recognizer, Verification system, SVM re-scoring, Rejection method, Lexicon size variation, RIMES database

8658-8, Session 3

Comic image understanding based on polygon detection

Luyuan Li, Yongtao Wang, Zhi Tang, Dong Liu, Peking Univ. (China)

Comic image understanding aims to automatically decompose scanned comic page images into storyboards and then identify the reading order of them, which is the key technique to produce digital comic documents that are suitable for reading on mobile devices. We propose a novel comic image understanding method based on polygon detection in this paper. First, we segment a comic page images into storyboards by finding the polygonal enclosing box of each storyboard. Then, each storyboard can be represented by a polygon, and the reading order of them is determined by analyzing the relative geometric relationship between each pair of polygons. The proposed method is tested on 2000 comic images from ten printed comic series, and the experimental results demonstrate that it works well on different types of comic images.

8658-9, Session 3

Context modeling for text/non-text separation in free-form online handwritten documents

Adrien Delaye, Cheng-Lin Liu, Institute of Automation (China)

Free-form online handwritten documents contain a high diversity of content, organized without constraints imposed to the user. The lack of prior knowledge about content and layout makes the modeling of contextual information of crucial importance for interpretation of such documents. In this work, we present a comprehensive investigation of the sources of contextual information that can benefit the task of discerning textual from non-textual strokes in handwritten online documents. An in-depth analysis of interactions between strokes is conducted through the design of various pairwise clique systems that are combined within a Conditional Random Field formulation of the stroke labelling problem. Our results demonstrate the benefits of combining complementary sources of context for improving the text/non-text recognition performance.

8658-10, Session 3

Annotating image ROIs with text descriptions for multimodal biomedical document retrieval

Daekeun You, National Library of Medicine (United States); Matthew Simpson, National Library of Medicine/NIH (United States); Sameer K. Antani, Dina Demner-Fushman, George R. Thoma, National Library of Medicine (United States)

No Abstract Available

8658-11, Session 3

Graphic composite segmentation for PDF documents with complex layouts

Canhui Xu, Zhi Tang, Xin Tao, Cao Shi, Peking Univ. (China)

Converting the PDF books to re-flowable format has recently attracted various interests in the area of e-book reading. Robust graphic segmentation is highly desired for increasing the practicability of PDF converters. To cope with various layouts, a multi-layer concept is introduced to segment the graphic composites including photographic images, drawings with text insets or surrounded with text elements. Both image based analysis and inherent digital born document advantages are exploited in this multi-layer based layout analysis method. By combining low-level page elements clustering applied on PDF documents and connected component analysis on synthetically generated PNG image document, graphic composites can be segmented for PDF documents with complex layouts. The experimental results on graphic composite segmentation of PDF document pages have shown satisfactory performance.

8658-12, Session 4

A classification-free word-spotting system

Nikos Vasilopoulos, Ergina Kavallieratou, Univ. of the Aegean (Greece)

In this paper, a classification-free Word-Spotting system, appropriate for the retrieval of printed historical document images is proposed. The system skips many of the procedures of a common approach. It does not include segmentation, feature extraction or classification. Instead it treats the queries as compact shapes and uses image processing techniques in order to localize a query in the document images. Our system was tested on a historical document collection with many problems and a Google book, printed in 1675.

8658-13, Session 4

Combining geometric matching with SVM to improve symbol spotting

Nibal Nayef, Thomas M. Breuel, Technische Univ. Kaiserslautern (Germany)

Symbol spotting is important for automatic interpretation of technical line drawings. Current spotting methods are not reliable enough for such tasks due to low precision rates. In this paper, we combine a geometric matching- based spotting algorithm with an SVM classifier to improve the spotting's precision. In symbol spotting, a query symbol is to be located within a line drawing. Candidate matches can be found, however, the found matches may be true or false. To distinguish a false match, an SVM classifier is used. The classifier is trained on true and false matches of a query symbol. The matches are represented as vectors that indicate the qualities of how well the query features are matched, those qualities are obtained via geometric matching. Using the classification, the precision of the spotting improved from an average of 76.6% to an average of 97.2% on a dataset of technical line drawings.

8658-14, Session 4

Segmentation-free keyword spotting framework using dynamic background model

Gaurav Kumar, Safwan R. Wshah, Venu Govindaraju, Univ. at Buffalo (United States); Ramachandrula Sitaram, Hewlett-Packard Labs. India (India)

We propose a segmentation free word spotting framework using Dynamic Background Model. The proposed approach is an extension to our previous work where dynamic background model was introduced and integrated with a segmentation based recognizer for keyword spotting. The dynamic background model uses the local character matching scores and global word level hypotheses scores to separate keywords from non-keywords. We integrate and evaluate this model on Hidden Markov Model (HMM) based segmentation free recognizer which works at line level without any need for word segmentation. We outperform the state of the art line level word spotting system on IAM dataset.

8658-15, Session K1

History of the Tesseract OCR engine: what worked and what didn't *(Keynote Presentation)*

Ray W Smith, Google (United States)

The development history and current state of the Tesseract open source OCR Engine is laid out and compared to changes in the OCR field over the last 20 years or so. The use and abandonment of statistical classification approaches is examined in particular. There are significant similarities between the non-statistical approaches used in Tesseract and the methods used in Hidden Markov Model OCR approaches, and the paper draws attention to the recent use of the terms "ad-hoc" and "principled" that attempt to separate these strangely similar approaches. A recurring theme throughout the paper is the lessons learned from developing an industrial strength OCR system over a period of 20 years, not just covering feature and classifier design, but also the very important topics of testing and language coverage. The discussion of languages provides a minimal set that covers the significant problems that a truly international industrial strength OCR system must address. Results are provided on 9 different languages with a total test set size exceeding 350 million characters.

8658-16, Session 5

Data acquisition from cemetery headstones

Cameron S. Christiansen, William A. Barrett, Brigham Young Univ. (United States)

Data extraction from engraved text is discussed rarely, and nothing in the open literature discusses data extraction from cemetery headstones. Headstone images present unique challenges such as engraved or embossed characters (causing inner-character shadows), low contrast with the background, and significant noise due to inconsistent stone texture and weathering. Current systems for extracting text from outdoor environments (billboards, signs, etc.) make assumptions (i.e. clean and/or consistently-textured background and text) that fail when applied to the domain of engraved text. The ability to extract the data found on headstones is of great historical value. This paper describes a novel and efficient feature-based text zoning and segmentation method for the extraction of noisy text from a highly textured engraved medium. This paper also demonstrates the usefulness of constraining a problem to a specific domain. The transcriptions of images zoned and segmented through the proposed system have a precision as high as 87% compared to 2% precision without zoning and a 70% recall compared to 50%.

8658-17, Session 5

Automated recognition and extraction of tabular fields for the indexing of census records

Robert Clawson, Kevin Bauer, Glen Chidester, Milan Tyler-Pohontsch, Douglas J. Kennard, Jongha Ryu, William A. Barrett, Brigham Young Univ. (United States)

We describe a system for indexing of census records in tabular documents with the goal of recognizing the content of each cell, including both headers and handwritten entries. Each document is automatically rectified, registered and scaled to a known template following which lines and fields are detected and delimited as cells in a tabular form. Whole-word or whole-phrase recognition of noisy machine-printed text is performed using a glyph library, providing greatly increased efficiency and accuracy (approaching 100%), while avoiding the problems inherent with traditional OCR approaches. Constrained handwriting recognition results for a single author reach as high as 98% and 94.5% for the Gender field and Birthplace respectively. Multi-author accuracy (currently 82%) can be improved through an increased training set. Active integration of user feedback in the system will accelerate the indexing of records while providing a tightly coupled learning mechanism for system improvement.

8658-18, Session 5

Old document image segmentation using the autocorrelation function and multiresolution analysis

Maroua Mehri, Univ. de La Rochelle (France) and Univ. of Rouen

(France); Petra Gomez-Krämer, Univ. de La Rochelle (France); Pierre Héroux, Univ. of Rouen (France); Rémy Mullot, Univ. de La Rochelle (France)

Recent progress in the digitization of heterogeneous collections of historical documents has rekindled new challenges in information retrieval in digital libraries and document layout analysis. Therefore, in order to control the quality of historical document image digitization and to meet the need of a characterization of their content using intermediate level metadata, we propose an automatic pixel-based layout segmentation method of historical document images content based on features computed with the autocorrelation function and performed using multiresolution analysis. The algorithm proposed in this article has advantage to be performed without expressing any hypothesis: either about the document model (physical structure), or typographical parameters (logical structure). It is also parameter-free since it automatically adapts to the image content. We demonstrate that is possible to automatically find the homogeneous regions defined by similar indices of autocorrelation without a priori knowledge using adapted clustering approach. To assess our method, we apply our algorithm on heterogeneous corpus content. Moreover, we define a new evaluation metric, the homogeneity measure, which aims at evaluating the segmentation and characterization accuracy of our proposal. The overall results are quite satisfying. Such analysis would help providing a similarity measure between pages and constructing a computer-aided categorization tool of pages.

8658-19, Session 5

Lexicon-supported OCR of eighteenth century Dutch books: a case study

Jesse de Does, Katrien Depuydt, Leiden Univ. (Netherlands)

IMPACT (2008-2012) is a project funded by the European Commission. Its aim was to significantly improve access to historical text and to take away the barriers that stand in the way of the mass digitization of the European cultural heritage. For that purpose IMPACT aimed to improve the quality of OCR (Optical Character Recognition) for historical documents and to enhance their accessibility. There are many aspects involved in dealing with the problems addressed by IMPACT. Image processing, which tries to remedy typical problems like skewed, warped or otherwise noisy data; better segmentation procedures and adaptive OCR aim to overcome the irregularities of historical typography. The present contribution focuses on enhancement of OCR results by using the appropriate historical lexica. We report on a case study on OCR of eighteenth century books conducted in the IMPACT project. After introducing the IMPACT project and its approach to lexicon building and deployment, we zoom in to the application of IMPACT tools and data to the Dutch EDBO collection. The results are exemplified by detailed discussion of various practical options to improve text recognition beyond a baseline of running an uncustomized Finereader 10. In particular, we discuss recognition of long s.

8658-20, Session 6

Character feature integration of Chinese calligraphy and font

Cao Shi, Jianguo Xiao, Wenhua Jia, Peking Univ. (China); Canhui Xu, Peking Univ. (China) and Beijing Founder Apabi Technology Ltd. (China)

A framework is proposed in this paper to effectively generate a new hybrid character type by means of integrating local contour feature of Chinese calligraphy with structural feature of font in computer system. To explore traditional art manifestation of calligraphy, multi-directional spatial filter is applied for local contour feature extraction. Then the contour of character image is divided into sub-images. The sub-images in the identical position from various characters are estimated by Gaussian distribution. According to its probability distribution, the dilation operator and erosion operator are designed to adjust the boundary of font image. And then new Chinese character images are generated which possess both contour feature of artistical calligraphy and elaborate structural feature of font. Experimental results demonstrate the new characters are visually acceptable, and the proposed framework is an effective and efficient strategy to automatically generate the new hybrid character of calligraphy and font.

8658-21, Session 6

A segmentation-free approach to Arabic and Urdu OCR

Nazly Sabbour, The German Univ. in Cairo (Egypt); Faisal Shafait, DFKI GmbH (Germany)

We present a generic Optical Character Recognition system for Arabic script languages called Nabocr. Nabocr uses OCR approaches specific for Arabic script recognition. Performing recognition on Arabic script text is relatively

more difficult than Latin text due to the nature of Arabic script, which is cursive and context sensitive. Moreover, Arabic script has different writing styles that vary in complexity. Nabocr is initially trained to recognize both Urdu Nastaleeq and Arabic Naskh fonts. However, it can be trained by users to be used for other Arabic script languages. We have evaluated our system's performance for both Urdu and Arabic. In order to evaluate Urdu recognition, we have generated a dataset of Urdu text called UPTI (Urdu Printed Text Image Database), which measures different aspects of a recognition system.

The performance of our system for Urdu clean text is 91%. For Arabic clean text, the performance is 86%. Moreover, we have compared the performance of our system against Tesseract's newly released Arabic recognition, and the performance of both systems on clean images is almost the same.

8658-22, Session 6

Local projection-based character segmentation method for historical Chinese documents

Linjie Yang, Liangrui Peng, Tsinghua Univ. (China)

Digitization of historical Chinese documents includes two key technologies, character segmentation and character recognition. This paper focuses on developing Character Segmentation algorithm. As the original historical document images contain various noises, we propose an EM clustering algorithm to remove the noise-like annotations and seals. We use projective information to remove the page border lines. We also apply a linear fitting method in a local window to remove the rule lines between the character rows. This paper adopts the Otsu's method to binarize the image. This paper proposes a new character segmentation algorithm based on projection of Cost Image in a local window which aims at segmenting single character from the binarized image. First, we use morphological operation to extract the skeleton image; Then we derive the stroke information from the skeleton image and get the bounding box of every stroke. This algorithm combines the information of stroke bounding box and the skeleton image to form a Cost Image, and use it to segment the characters. We evaluate the proposed algorithm based on matching degree of character bounding boxes between the analysis result and the ground-truth, which proves the effectiveness of the algorithm.

8658-34, Session IPI

A super resolution framework for low resolution document image OCR

Di Ma, Gady Agam, Illinois Institute of Technology (United States)

No Abstract Available

8658-35, Session IPI

A robust pointer segmentation in biomedical images toward building a visual ontology for biomedical article retrieval

Daekeun You, National Library of Medicine (United States); Matthew Simpson, National Library of Medicine/NIH (United States); Sameer K. Antani, Dina Demner-Fushman, George R. Thoma, National Library of Medicine (United States)

No Abstract Available

8658-36, Session IPI

Combining multiple thresholding binarization values to improve OCR output

William B. Lund, Douglas J. Kennard, Brigham Young Univ. (United States); Eric K. Ringger, Brigham Young Univ. (United States) and Computer Science Dept. (United States)

For noisy, historical documents, a high optical character recognition (OCR) word error rate (WER) can render the OCR text unusable. Since image binarization is often the method used to identify foreground pixels, a body of research seeks to improve image-wide binarization directly. This paper presents new research and a novel approach by not relying on any one adaptive binarization but leverages information from multiple simple thresholding binarizations of the same image to improve text output. Using a new corpus of 19th century newspaper grayscale images for which the text transcription is known, we observe WERs of 13.8% and higher using current binarization techniques and a state-of-the-art OCR engine. Our novel approach combines the OCR outputs from multiple thresholded images by aligning the text output and producing a lattice of word alternatives from which a lattice word error rate (LWER) is calculated. Our results show a LWER of 7.6% when aligning two threshold images and a LWER of 6.8% when aligning five. From the word lattice we commit to one hypothesis by applying the methods of Lund et al. (2011) achieving an improvement over the original OCR output and a 8.41% WER result on this data set.

8658-37, Session IPI

Goal-oriented evaluation of binarization algorithms for historical document images

Tayo Obafemi-Ajayi, Univ. of Missouri-Columbia (United States); Gady Agam, Illinois Institute of Technology (United States)

Binarization is of significant importance in document analysis systems. It is an essential first step, prior to further stages such as Optical Character Recognition (OCR), document segmentation, or enhancement of readability of the document after some restoration stages. Hence, proper evaluation of binarization methods to verify their effectiveness is of great value to the document analysis community. In this work, we perform a detailed goal- oriented evaluation of image quality assessment of the 18 binarization methods that participated in the DIBCO 2011 competition using the 16 historical document test images used in the contest. We are interested in the image quality assessment of the outputs generated by the different binarization algorithms as well as the OCR performance, where possible. We compare our evaluation of the algorithms based on human perception of quality to the DIBCO evaluation metrics. The results obtained provide an insight into the effectiveness of these methods with respect to human perception of image quality as well as OCR performance.

8658-38, Session IPI

Document segmentation via oblique cuts

Jeremy P. Svendsen, Alexandra Branzan-Albu, Univ. of Victoria (Canada)

This paper presents a novel solution for the layout segmentation of graphical elements in Business Intelligence documents. We propose a generalization of the recursive X-Y cut algorithm, which allows for cutting along arbitrary oblique directions. An intermediate processing step consisting of line and solid region removal is also necessary due to presence of decorative elements. The output of the proposed segmentation is a hierarchical structure which allows for the identification of primitives in pie and bar charts. The algorithm was tested on a database composed of charts from business documents. Results are very promising.

8658-39, Session IPI

Preprocessing document images by resampling is error prone and unnecessary

George Nagy, Rensselaer Polytechnic Institute (United States)

Integrity tests are proposed for image processing algorithms that should yield essentially the same output under 90 degree rotations, edge-padding and monotonic gray-scale transformations of scanned documents. The tests are demonstrated on built-in functions of the Matlab Image Processing Toolbox. Only the routine that reports the area of the convex hull of foreground components fails the rotation test. Ensuring error-free preprocessing operations like size and skew normalization that are based on resampling an image requires more radical treatment. Even if faultlessly implemented, resampling is generally irreversible and may introduce artifacts. Fortunately, advances in storage and processor technology have all but eliminated any advantage of preprocessing or compressing document images by resampling them. Using floating point coordinate transformations instead of resampling images yields accurate run-length, moment, slope, and other geometric features.

8658-40, Session IPI

Multilingual artificial text detection and extraction from still images

Ahsen Raza, Ali Abidi, National Univ of Sciences and Technology (Pakistan); Imran A. Siddiqi, Bahria Univ. (Pakistan)

We present a novel method for multilingual artificial text extraction from still images. We propose a lexicon independent, block based technique that employs a combination of spatial transforms, texture, edge and gradient based operations to detect unconstrained textual regions from still images. Finally, some morphological and geometrical constraints are applied for fine localization of textual content. The proposed method was evaluated on two standard and three custom developed datasets comprising a wide variety of images with artificial text occurrences in five different languages namely English, Urdu, Arabic, Chinese and Hindi.

8658-41, Session IPI

A proposal system for historic Arabic manuscript transcription and retrieval

Abdelaziz Labben, Afef K. Echi, Unité de Recherche en Technologies de l'Information et de la Communication (Tunisia); Abdel Belaïd, LORIA (France)

In this paper, we propose a computer-assisted transcription system of old registers, handwritten in Arabic from the 19th century onwards, held in the National Archives of Tunisia (NAT). The proposed system assists the human supervisor to complete the transcription task as efficiently as possible. This assistance is given at all different recognition levels. Our system addresses different approaches for transcription of document images. It also implements an alignment method to find mappings between word images of a handwritten document and their respective words in its given transcription.

8658-42, Session IPI

Evaluation of document binarization using eigen value decomposition

Deepak Kumar, M. N. Anil Prasad, A. G. Ramakrishnan, Indian Institute of Science (India)

A necessary step for the recognition of scanned documents is binarization, which is essentially the segmentation of the document. In order to binarize a scanned document, we can find several algorithms in the literature. What is the best binarization result for a given document image? To answer this question, a user needs to check different binarization algorithms for suitability, since different algorithms may work better for different type of documents. Manually choosing the best from a set of binarized documents is time consuming. Can we come up with a metric which evaluates these binarized documents? To automate the selection of the best segmented document, either we need to use ground-truth of the document or propose an evaluation metric. If ground-truth is available, then precision and recall can be used to choose the best binarized document. What is the case, when ground-truth is not available? Hence, we propose a metric to evaluate binarized document images using eigen value decomposition. We have evaluated this measure on DIBCO and H-DIBCO datasets. The proposed method chooses the best binarized document that is close to the ground-truth of the document.

8658-43, Session IPI

Efficient symbol retrieval by building a symbol index from a collection of line drawings

Nibal Nayef, Thomas M. Breuel, Technische Univ. Kaiserslautern (Germany)

Symbol retrieval is important for content-based search in digital libraries and for automatic interpretation of line drawings. In this work, we present a complete symbol retrieval system. The proposed system has an off-line content-analysis stage, where the contents of a database of line drawings are represented as a symbol index , which is a compact indexable representation of the database. Such representation allows efficient on-line query retrieval. Within the retrieval system, three methods are presented. First, a feature grouping method for identifying local regions of interest (ROIs) in the drawings. The found ROIs represent symbols' parts. Second, a clustering method based on geometric matching, is used to cluster the similar parts from all the drawings together. A symbol index is then constructed from the clusters' representatives. Finally, the ROIs of a query symbol are matched to the clusters' representatives. The matching symbols' parts are retrieved from the clusters, and spatial verification is performed on the matching parts. By using the symbol index we are able to achieve a query look-up time that is independent of the database size, and dependent on the size of the symbol index. The retrieval system achieves higher recall and precision than the state-of-the-art methods.

8658-23, Session 8

Structural analysis of online handwritten mathematical symbols based on support vector machines

Foteini Simistira, Vassilis M. Papavassiliou, Vassilis Katsouros, Institute for Language and Speech Processing (Greece); George Carayannis, National Technical Univ. of Athens (Greece)

Mathematical expression recognition is still a very challenging task for the research community mainly because of the two-dimensional (2d) structure of mathematical expressions (MEs). In this paper, we present a novel approach for the structural analysis between two on-line handwritten mathematical symbols of a ME, based on spatial features of the symbols. We introduce six features to represent the spatial affinity of the symbols and compare two multi-class classification methods that employ support vector machines (SVMs): one based on the "one-against-all" technique and one based on the "one-against-all", in identifying the relation between a pair of symbols (i.e. subscript, numerator, etc). A dataset containing 1906 spatial relations derived from the Competition on Recognition of Online Handwritten Mathematical Expressions (CROHME) 2012 training dataset is constructed to evaluate the classifiers and compare them with the classifier of the ILSP-1 system participated in the contest. The experimental results give an overall error rate of 2.61% for the "one-against-one" SVM approach, 6.57% for the "one-against-all" SVM technique and 12.31% error rate for the ILSP-1 classifier.

8658-24, Session 8

Using online handwriting and audio streams for mathematical expressions recognition: a bimodal approach

Sofiane Medjkoune, Univ. de Nantes (France) and Le Mans Univ, LIUM Lab. (France); Harold Mouchère, Univ. de Nantes (France); Simon Petitrenaud, Univ. du Maine (France); Christian Viard-Gaudin, Univ. de Nantes (France)

This paper deals with the problem of mathematical expressions recognition. The main goal is to set up a multimodal system dedicated to this task. In the proposed system, the hypotheses formulated in both specialized systems are combined to increase as possible the reliability of the final hypotheses which would be considered during the structural analysis to formulate the mathematical expression solution. Various fusion methods are explored. Performances evaluated on the HAMEX dataset show a significant improvement compared to a single modality (handwriting) based system.

8658-25, Session K2

What does the future hold for search user interfaces? *(Keynote Presentation)*

Marti Hearst, Univ. of California, Berkeley (United States)

No Abstract Available

8658-26, Session 9

Using clustering and a modified classification algorithm for automatic text summarization

Abdelkrime Aries, Houda Oufaida, Ecole Nationale Supérieue d'Informatique (Algeria); Omar Nouali, Ctr. de recherche sur l'Information Scientifique et Technique (Algeria)

In this paper we describe a modified classification method destined for extractive summarization purpose. The classification in this method doesn't need a learning corpus; it uses the input text to do that. First, we cluster the document sentences to exploit the diversity of topics, then we use a learning algorithm (here we used Naive Bayes) on each cluster considering it as a class. After obtaining the classification model, we calculate the score of a sentence in each class, using a scoring model derived from classification algorithm. These scores are used, then, to reorder the sentences and extract the first ones as the output summary.

We conducted some experiments using a corpus of scientific papers, and we have compared our results to another summarization system called UNIS. Also, we experiment the impact of clustering threshold tuning, on the resulted summary, as well as the impact of adding more features to the classifier. We found that this method is interesting, and gives good performance, and the addition of new features (which is simple using this method) can improve summary's accuracy.

8658-27, Session 9

Evaluating supervised topic models in the presence of OCR errors

Daniel D. Walker IV, Eric K. Ringger, Kevin D. Seppi, Brigham Young Univ. (United States)

Supervised topic models are promising tools for text analytics that simultaneously model topical patterns in document collections and relationships between those topics and document metadata, such as timestamps. We to examine empirically the effect of OCR noise on the ability of supervised topic models to produce high quality output through a series of experiments in which we evaluate three supervised topic models and a naive baseline on synthetic OCR data with various levels of degradation and on real OCR data from two different decades. The evaluation includes experiments with and without feature selection. Our results suggest that supervised topic models are no better, or at least not much better in terms of their robustness to OCR errors, than unsupervised topic models and that feature selection has the mixed result of improving topic quality while harming metadata prediction quality.

8658-28, Session 9

Rule-based versus training-based extraction of index terms from business documents: how to combine the results

Daniel Schuster, Marcel Hanke, Klemens Muthmann, Daniel Esser, Technische Univ. Dresden (Germany)

Current systems for automatic extraction of index terms from business documents either take a rule-based or training-based approach. As both approaches have their advantages and disadvantages it seems natural to combine both methods to get the best of both worlds. We present a combination method with the steps selection, normalization, and combination based on comparable scores produced during extraction. Furthermore, novel evaluation metrics are developed to support the assessment of each step in an existing extraction system. Our methods were evaluated on an example extraction system with three individual extractors and a corpus of 12,000 scanned business documents.

8658-29, Session 9

Post processing with first- and second-order hidden Markov models

Kazem Taghva, Univ. of Nevada, Las Vegas (United States); Srijana Poudel, Univ. of Nevada (United States)

In this paper, we present the implementation and evaluation of first order and second order Hidden Markov Models to identify and correct OCR errors in the post processing of books. Our experiments show that the first order model approximately corrects 10% of the errors with 100% precision, while the second order model corrects a higher percentage of errors with much lower precision.

8658-30, Session 9

Combining discriminative SVM models for the improved recognition of investigator names in medical articles

Xiaoli Zhang, Jie Zou, Daniel X. Le, George R. Thoma, National Library of Medicine (United States)

Investigators are people who are listed as members of corporate organizations but not entered as authors in an article. Beginning with journals published in 2008, investigator names are required to be included in a new bibliographic field in MEDLINE citations. Automatic extraction of investigator names is necessary due to the increase in collaborative biomedical research and the large number of such names. We implemented two discriminative SVM models, i.e., SVM and structural SVM, to identify named entities such as the first and last names of investigators from online medical journal articles. Both approaches achieve good performance at the word and name chunk levels. We further conducted error analysis and found that SVM and structural SVM can offer complementary information about the patterns to be classified. Hence, we combined the two independently trained classifiers where SVM is chosen as a base learner with its output further enhanced by the predictions from structural SVM. The overall performance especially the recall rate of investigator name retrieval improves as compared to standalone SVM model.

8658-31, Session 10

Adaptive detection of missed text areas in OCR outputs: application to the automatic assessment of OCR quality in mass digitization projects

Ahmed Ben Salah, Bibliothèque nationale de France (France) and LITIS-Univ. de Rouen (France); Nicolas Ragot, Univ. de Tours (France); Thierry Paquet, Univ. de Rouen (France)

The French National Library (BnF) has started many mass digitization projects in order to give access to its collection of documents. The indexation of digital documents on Gallica (the digital library of the BnF) is done through their textual content obtained thanks to service providers that use Optical Character Recognition softwares (OCR).

OCR softwares have become increasingly complex systems composed of several subsystems dedicated to the analysis and the recognition of the elements in a page. However, the reliability of these systems is always an issue at stake. Indeed, in some cases, we can find errors in OCR outputs that occur because of an accumulation of several errors at different levels in the OCR process. One of the frequent errors in OCR outputs is the missed text components. The presence of such errors may lead to severe defects in digital libraries.

In this paper, we intend to examine the detection of missed text components to control the OCR results in the collections of the French National Library. Our verification approach uses local information within the pages based on Radon transform descriptors and Local Binary Patterns descriptors (LBP) coupled with OCR results to control their consistency. The experimental results show that our method detects 84.15% of the missed textual components, by comparing the OCR ALTO files outputs produced by the service providers to the input image of the document.

8658-32, Session 10

Evaluating structural pattern recognition for handwritten math via primitive label graphs

Richard Zanibbi, Rochester Institute of Technology (United States); Harold Mouchère, Christian Viard-Gaudin, Univ. de Nantes (France)

Currently evaluation of pattern recognition systems that produce graphs (i.e. parsers) is often performed using metrics comparing objects and object relationships in ground truth and those detected by an algorithm, e.g. recognition rates, recall and precision. We present a structure representation using labels for all input primitives (e.g. handwritten strokes) and all primitive pairs based on a bipartite graph over primitives. We define a number of new evaluation metrics using Hamming distances over structure strings. We demonstrate these metrics through comparing and contrasting a stroke-level evaluation with a symbol-level evaluation for the CROHME 2012 handwritten math recognition competition systems. Benefits of structure strings include the ability to represent classification, segmentation and parsing errors within proper distance metrics, and to easily identify and completely characterize sources of error. A library providing implementation of primitive and object-level (e.g. recall and precision) metrics will be

made publicly available. Structure strings may also be useful for information retrieval and machine learning applications.

8658-33, Session 10

WFST-based ground truth alignment for difficult historical documents with text modification and layout variations

Mayce Al Azawi, Technische Univ. Kaiserslautern (Germany); Marcus Liwicki, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany); Thomas M. Breuel, Technische Univ. Kaiserslautern (Germany)

This work proposes several approaches that can be used for generating correspondences between real scanned books and their transcriptions which might have different modifications and layout variations under consideration of OCR errors. Our approaches for the alignment between the manuscript and the transcription are based on weighted finite state transducers (WFST). In particular, we propose adapted WFSTs to represent the transcrip- tion to be aligned with the OCR lattices. The character-level alignment has edit rules to allow edit operations (insertion, deletion, substitution). Those edit operations support the transcription model to be adapted regard- ing to the OCR segmentation and recognition errors and also the ability of aligning with different text editions. We implemented an alignment model which has a hyphenation model so it can adapt the non hyphenated tran- scription. Our models can also work with Fraktur ligatures which is typical for historical Fraktur documents. We evaluated the our approach on Fraktur documents from "Wanderungen durch die Mark Brandenburg" volumes (1862-1889) and observed the performance of those models under OCR errors. We compare the performance of our model for three different scenarios: having no information about the correspondence at the word (i), line (ii), sentence (iii) or page (iv) level.

Conference 8659: Sensors, Cameras, and Systems for Industrial/ Scientific Applications XIV

Wednesday - Thursday 6 –7 February 2013

Part of Proceedings of SPIE Vol. 8659 Sensors, Cameras, and Systems for Industrial and Scientific Applications XIV

8659-1, Session 1

Fundamental performance differences of CMOS and CCD imagers: part V

James R. Janesick, Sarnoff Corp. (United States)

This paper is a continuation of several previous sequential papers focused on fundamental performance differences of scientific CMOS and CCD imagers. Discussions in this paper include: 1). data for custom ultra low noise scientific CMOS 5T/6T PPD pixel imagers fabricated on various new Sandbox fab runs, 2). fundamental ionization and bulk CMOS pixel radiation damage characteristics at dose levels up to 10 Mrd along with CCD data for damage comparisons (protons, electrons, Co-60), 3). progress report on ultra large stitched CMOS imagers developed for the Mk x Nk CMOS imager family, 4). fabrication of a 4k x 4k CMOS imager for NRL's SoloHi flight instrument on the Solar Orbiter Mission, 5). high performance buried channel CCDs based on CMOS processing, 6). RTN and radiation hardness comparison of p and n pixels, 7). various miscellaneous fundamental pixel topics and 11). upcoming development work expected.

8659-2, Session 1

Kirana: a solid-state megapixel uCMOS image sensor for ultrahigh speed imaging

Jamie Crooks, Ben Marsh, Renato Turchetta, STFC - Rutherford Appleton Lab. (United Kingdom); Keith Taylor, Wai Chan, Specialised Imaging Ltd. (United Kingdom); Assaf Lahav, Amos Fenigstein, TowerJazz Semiconductor Ltd (Israel)

This paper describes a solid-state sensor for ultra high-speed (UHS) imaging. The 'Kirana' sensor was designed and manufactured in a 180 nm CMOS technology to achieve full-frame 0.7 Megapixel video capture at speeds greater than 1MHz. The 30um pixels contain a pinned photodiode, a set of 180 low-leakage storage cells, a floating-diffusion, and a source follower output structure. Both the individual cells and the way they are arranged in the pixel are novel. The pixel architecture allows correlated double sampling for low noise operation. In the fast mode, the storage cells are operated as a circular buffer, where 180 consecutive frames are stored until receipt of a trigger; up to 5 video-bursts per second can be read out. In the 'slow' mode, the storage cells act like a pipeline; the sensor can be read out like a conventional sensor at a continuous frame rate of 1,000 fps. The sensor architecture is fully scalable in resolution since memory cells are located inside each pixel. The pixel architecture is scalable in memory depth (number of frames) as a trade-off with pixel size, dependent on application. The present implementation of 0.7Mpixels has an array focal plane which is optimised for standard 35mm optics, whilst offering a competitive 180-frame recording depth. The sensor described has been manufactured and is currently being characterised. Operation of the sensor in the fast mode at 2 million frames per second has been achieved. Details on the camera/sensor operation will be presented together with first experimental results.

8659-3, Session 1

Back-side-illuminated image sensor with burst capturing speed of 5.2 Tpixel per second

Toshiki Arai, Jun Yonai, Tetsuya Hayashida, Hiroshi Ohtake, Japan Broadcasting Corp. (Japan); Harry van Kuijk, Teledyne DALSA (Netherlands); Takeharu G. Etoh, Ritsumeikan Univ. (Japan)

We developed a back-side illuminated image sensor with burst capturing speed of 5.2 Tpixel per second. The image sensor has a special charge-coupled device (CCD) with in-situ CCD memories. The total pixel counts are 411 pixels vertically and 760 pixels horizontally. The area of each pixel was 46.8 ?m?46.8 ?m. The opposite angle length of the image area was 40.4 mm. The size of each memory element was 3.0 ?m?3.6 ?m. The number of recordable images was 139 frames. The results of a driving experiment indicated that the sensitivity of the sensor was 13 times higher than that of front-side illuminated image sensor. The sensitivity of the evaluation camera system using the monochrome sensor was 2000 lux F90. The resolution of the sensor was 410 TV lines. A saturation signal level of 100% was maintained over 10 Mfps. The maximum frame rate was 16.7 Mfps. The product of the pixel count and maximum frame rate is often used as a figure of merit in high-speed imaging devices. In this case, 312,360 pixels times the maximum frame rate of 16.7 Mfps yields 5.2 Tpixel per second. The burst capturing speed of this image sensor was thus 5.2 Tpixel per second.

8659-4, Session 2

A custom CMOS imager for multi-beam laser scanning microscopy and an improvement of scanning speed

Min-Woong Seo, Keiichiro Kagawa, Keita Yasutomi, Shoji Kawahito, Shizuoka Univ. (Japan)

Multi-beam confocal microscopy with a 256?256-pixel custom CMOS imager performing focal-plane pinhole effect, in which any rotating disk is not required, is demonstrated. A specimen is illuminated by 32?32 diffraction-limited light spots whose wavelength and pitch are 532 nm and 8.4 ?m, respectively. The spot array is generated by a microlens array, which is scanned by two-dimensional piezo actuator according to the scanning of the image sensor. The frame rate of the prototype is 0.26 Hz, which is limited by the actuator. The confocal effect has been confirmed by comparing the depth of focus in the confocal imaging mode with that of the normal imaging mode. The depth of focus in the confocal mode by the full width at half maximum (FWHM) is reduced by 61.8% compared with that of the normal mode. Through this result, the confocal effect of the proposed imager is confirmed. The focal-plane pinhole effect in the confocal microscopy with the proposed CMOS imager has been demonstrated at low frame rate. An improvement of the scanning speed and a CMOS imager with photo-sensitivity modulation pixels suitable for high-speed scanning are also discussed.

8659-5, Session 2

An ultrafast ultracompact sensor for diffuse correlation spectroscopy

Kinia Barjean, Eric Tinet, Dominique Ettori, Univ. Paris 13 (France); Antoine Dupret, Commissariat à l'Énergie Atomique (France); Marius Vasiliu, Univ. Paris-Sud 11 (France); Jean-Michel Tualle, Univ. Paris 13 (France)

Diffuse Correlation Spectroscopy (DCS) is based on time-resolved correlations of the light that have diffused through a biological media.

Measurements must be made on a small coherence area of the size of a speckle grain. Summing independent measurement enables to increase the SNR as the square root of the number of detectors. We present a bidimensionnal pixel CMOS detector array specially designed for this task, with parallel in-pixel demodulation and time-resolved correlation computation. Optical signal can be processed at a rate higher than 10,000 samples per second with demodulation frequencies in the MHz range.

8659-6, Session 2

A 3D image sensor with adaptable charge subtraction scheme for background light suppression

Jungsoon Shin, Byong Min Kang, Keechang Lee, James D. K. Kim, Samsung Electronics (Korea, Republic of)

We present a 3D ToF (Time-of-Flight) image sensor with adaptive background light storage and subtraction scheme for background light suppression. The proposed sensor can alternately capture high resolution color image and high quality depth map in each frame. In depth-mode, the sensor requires enough integration time for accurate depth acquisition, but saturation will occur in high background light illumination. We propose to divide the integration time into N sub-integration times adaptively. During the first sub-integration time, our sensor captures an image only from background light without IR modulated light and stores it in frame memory. The stored image is subtracted from subsequent captured image frames with IR modulated light and same sub-integration duration. The subtraction results are cumulated N times obtaining a final result image without background illumination at full integration time. Experimental results with our own ToF sensor show high background suppression performance. We also propose in-pixel storage and column-level subtraction circuit for chip-level implementation of the proposed method. We believe the proposed scheme will enable 3D sensors to be used in out-door environment.

8659-7, Session 2

High performance 7.4-micron interline transfer CCD platform for applied imaging markets

Douglas A. Carpenter, James A. DiBella Sr., Robert Kaser, Truesense Imaging, Inc. (United States); Brent J. Kecskemety, Truesense Imaging, Inc. (Andorra); Stephen L. Kosman, John P. McCarten, Christopher Parks, Truesense Imaging, Inc. (United States)

Technology developed for a 5.5 μm pixel interline transfer CCD family has been incorporated into a new family of high-performance 7.4 μm pixel CCDs, providing significant improvements in several key performance parameters compared to both the 5.5 μm family as well as the previous generation of 7.4 μm pixel products. Smear in the new platform has been reduced to -115 dB, and frame rate has been doubled relative to the previous generation of 7.4 μm pixel products. Dynamic range in normal operation has been improved to 70 dB, and the platform supports a new extended dynamic range mode which provides 82 dB when binning 2 x 2. The new family leverages the package and pin-out configurations used in the 5.5 μm pixel family, allowing easy integration into existing camera designs.

8659-8, Session 3

A 33M-pixel wide color gamut image capturing system using four CMOS image sensors at 120 Hz

Takuji Soeno, NHK Science & Technology Research Labs. (Japan); Kohei Omura, NHK Engineering Services, Inc. (Japan); Takayuki Yamashita, Ryohei Funatsu, Masayuki Sugawara, Makoto Okui, NHK Science & Technology Research Labs. (Japan)

We have been developing an extremely high resolution video system with 33M pixels. Experimental results regarding human perception show that the frame frequency for this system should be at least 120 Hz to achieve a significant improvement in motion portrayal. We have therefore investigated a 33M-pixel image-capturing system at 120 Hz. The system consists of four CMOS image sensors that can not only read out 33M pixel at 60 Hz but also read out data at 120 Hz from only odd or even lines. Two image sensors are used for the green channel (G1 and G2), and one is used for each of the red and blue channel. G1 sensor outputs the odd lines, while G2 sensor outputs the even lines. 33M-pixel green images are produced by combining G1 and G2. The red and blue sensors scan the odd lines and the even lines, respectively. Subsequently, the lines whose data have not been read are interpolated in the vertical direction. We have build prototype, and evaluated its resolution and image lag. The performance of the image-capturing system has been found to be sufficient to capture 33M-pixel images at the frequency of 120 Hz.

8659-9, Session 3

A 3 mpixel ROIC with 10um pixel pitch and 120 Hz frame rate digital output

Elad Ilan, Niv Shiloah, Shimon Elkind, Roman Dobromislin, Willie Freiman, Alex Zviagintsev, Itzik Nevo, Oren Cohen, Fanny Khinich, Amnon Adin, Ron Talmor, Yaakov Milstain, SCD Semiconductor Devices (Israel)

A 1920x1536 matrix readout IC (ROIC) for 10x10 um2 P-on-N InSb photodiode array is reported. The ROIC features several conversion gain options implemented at the pixel level. A 2-by-2 pixel binning feature is implemented at the pixel level as well, improving SNR and enabling higher frame rates by a factor of four. A dedicated column ADC is designed for low noise and low power consumption, while reaching 95 kSps sampling rate. Since 3840 column ADCs are integrated on chip, the total conversion rate is over 360Mpxl/sec. The ROIC achieves 120 Hz frame rate at the full format, with power consumption of less than 400mW. A high speed serial video interface is developed to output the required data bandwidth at a reasonable pin count.

8659-11, Session 3

Dynamic capability of sensors with nonlinear pixels utilized by security cameras

Anders Johannesson, Henrik Eliasson, Sebastian Fors, Axis Communications AB (Sweden)

We have previously proposed a framework containing a typical security camera use case and have discussed how well this is handled by linear image sensors with various characteristics. The findings were visualized graphically, using a simple camera simulator generating images under well-defined conditions. In order to successfully render low-contrast objects together with large intra-scene variations in illuminance, the sensor requirements must include a high dynamic range combined with a comparably high signal-to-noise ratio. In this paper we reuse the framework and extend the discussion by including also sensors with non-linear pixel responses.

The obvious benefit of a non-linear pixel is that it generally can cope with a higher scene dynamic range and that in most cases the exposure control can be relaxed. Known drawbacks are, for example, that the noise level can be fairly high. More specifically, the spatial noise levels are high due to variable pixel-to-pixel characteristics and lack of on-chip corrections, like correlated double sampling.

In this paper we ignore the spatial noise, since some of the related issues have been addressed recently. Instead we focus on the temporal noise and dynamic resolution issues involved in non-linear imaging on a system level. Since the requirements are defined by our selected use case, and since we have defined a visual framework for analysis, it is straightforward to compare our findings with the results for linear image sensors. As in the previous paper, the image simulations are based on sensor data obtained from our own measurements.

Conference 8659

8659-27, Session IPI

Optical high-altitude balloon tracking and imaging for optical communications

Jeremy Straub, Ronald A. Fevig, Univ. of North Dakota (United States)

An optical communications system has been proposed which utilizes an LED on each face of a CubeSat for downlink communications from deep space. Testing of this system via high altitude balloon (HAB) will be conducted to validate key assumptions and system functionality. The receiving system incorporates two components. A high-resolution, low-spatial-coverage telescope and CCD receives the data, which is decoded by a computer and compared to the pre-programmed messages that the HAB transmitter sends during the testing. A lower-resolution, wide-field-of-view sensor is used to track the point-light-source and issue commands to the telescope's orientation system to keep the transmitter centered in the frame and to determine what area of the CCD should be read to maximize data throughput.

A predictive algorithm, based on paths of movement, is used to attempt to keep ahead of the movement faced by the HAB. This algorithm utilizes knowledge of the exact pixels occupied by the payload on the high-resolution CCD and computed frame-to-frame movement vectors to predict the position of the HAB payload in the next frame, and reposition the telescope accordingly. Each subsequent prediction is based on the actual movement recorded and the newly computed movement vector. Correction for miss-predictions is also incorporated.

8659-28, Session IPI

Characterization of a solid state air corona charging device

Michael Young, Baomin Xu, Steve Buhler, Palo Alto Research Center, Inc. (United States); Karl Littau, Stanford Univ. (United States)

Two new solid state devices which produced an atmospheric air corona discharge for generating and depositing a layer of static charge for Xerographic imaging have been fabricated and characterized. One type was a parallel plate capacitive structure and the other was an interdigitated capacitive structure. It was determined that the interdigitated capacitive structure performed better than the parallel plate capacitive structure in terms of power consumption, charging current level stability and device reliability. Several metal electrode material alternatives were investigated and gold electrodes performed the best. The air corona's light emission peaks were measured to be in the 350 nm to 500 nm range. Ozone gas generation to ~ 13 ppm was detected for an active surface area of 5 cm^2. Charge deposition on to an imaging drum with a significant charging current density of 1.6E-4 A/cm^2 has been successfully demonstrated.

8659-29, Session IPI

Creation of North-East Indian face database for human face identification

Kankan Saha, Priya Saha, Mrinal K. Bhowmik, Tripura Univ. (India); Debotosh Bhattacharjee, Mita Nasipuri, Jadavpur Univ. (India)

This paper describes the design and development of the North-East Indian (NEI) Face Database. It also gives a detailed description of the image capturing equipments' (i.e. lights, cameras etc.) setup and their different angles and distances from the subject. The database images have been captured with four illumination types, eight expressions (including neutral), and faces with glasses, in a controlled environment. The NEI Face Database contains face images of 292 individuals with 95 images per individual. All the participants in this database are students, faculty members and other staffs of various educational institutes. Participants include both Indian tribes and non-tribes of Mongolian origin collected from the three north-eastern states of India, aging from 18 to 63. Some baseline algorithms like Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have also been tested using the Support Vector Machines (SVM) classifier on our own database, which may be used as the control algorithm performance score by other researchers. A comparative study of the various 2D face image databases based on the properties of the images, conditions used for capturing the faces, and based on the performance score is also provided.

8659-30, Session IPI

Optical characterization parameters by study and comparison of subwavelength patterns for color filtering and multispectral purpose

Jacques Mantanga, Yvon Lacroute, Pierre Gouton, Eric Bourillot, Le2i (France)

A comparative study of realization processes of color filters for solid state imager has been developed. The unprecedented ability of nanometallic structures to concentrate light into deep subwavelength volumes has propelled their use in a vast array of nanophotonics technologies and research endeavors. Plasmonic light concentrators can elegantly interface diffraction-limited dielectric optical components with nanophotonic structures. So large interest was focused on Surface Plasmon Resonant (SPR) based sensors and applications. Recently, Resonant Waveguide Metallic Grating (RWMG) and Nanoholes Filtering Array (NFA) have been developed with high transmission factor. Different kind of 2D periodic structures is observed and quantified in this paper, with various metals to check the consistency of optical parameters in order to achieve a series of on-chip paper, suitable spectral filter with multispectral purpose. Simulation, and experimentally tests using the Rigorous Coupled Wave Analysis (RCWA) gives significant results in 2D nanostructure shapes and thin films, spectrum of each structure base RCWA parameters. We compared amorphous and crystallized structures and an improvement occurred by crystalline state. Most of SPR based sensors consist in a stack of superposed layer. These results, thereof, will potentially be used in the realization of the upper layer of accurate resonant spectral filter.

8659-13, Session 4

Empirical formula for rates of hot pixel defects based on pixel size, sensor area, and ISO

Glenn H. Chapman, Rohit Thomas, Simon Fraser Univ. (Canada); Zahava Koren, Israel Koren, Univ. of Massachusetts, Amherst (United States)

Image sensors measurements show a continuous development of in-field permanent hot pixel defects over time. Experimental tests have been accumulated on cameras ranging from large area (>300 sq mm) DSLR's, medium sized (~40 sq mm) point and shoot, and small (20 sq mm) cell phone cameras. These results show that the defect rate depends on the technology (APS or CCD) and on design parameters like imager area, pixel size (from 1.5-7 um), and gain (from ISO100 to 1600). Comparing different sensor sizes with similar pixel sizes showed that defect rates scale linearly with sensor area, suggesting the metric of defects/year/sq mm. A search was made to model that defect rate using the parameters of both pixel size and ISO. The best empirical curve fit was for CCD imagers with a power law with defect rates proportional to the pixel size to the power of -2.24 times, the ISO raised to 0.68, times a constant. For APS (CMOS) sensors the power law had pixel sizes to -3.6 times, the ISO raised to 0.56. Extending our empirical formula to include ISO allows us to predict the expected defect development rate for a wide set of sensor parameters.

8659-14, Session 4

A statistical evaluation of low frequency noise of in-pixel source follower-equivalent transistors with various channel types and body bias

Rihito Kuroda, Akihiro Yonezawa, Akinobu Teramoto, Tsung-Ling Li, Yasuhisa Tochigi, Shigetoshi Sugawa, Tohoku Univ. (Japan)

For a further performance improvement of low-light-level imaging CMOS image sensors, a reduction of low frequency noise of in-pixel source followers is indispensable. For this purpose, we evaluated statistical low frequency noise characteristics of in-pixel source follower-equivalent transistors with various channel types and body bias conditions. The evaluated transistor types are buried channel (BC) and surface channel (SC) transistors with or without an isolated well. T The gate width/length of the evaluated transistors was 0.32/0.32?m. The gate oxide thickness was 7.6 nm.

The BC transistors without isolated well exhibits a noise distribution having a much lower noise level in the tail parts compared to the SC transistors as previously reported. Different impacts of the isolated well were observed for BC and SC transistors. For SC transistors, both threshold voltage variation and noise distribution did not change much among transistors with or without isolated well. For the BC transistors with isolated well, threshold voltage variation was grater that that of ones without isolated well. In addition, noise distribution became a larger level for BC transistors with isolated well than without it.

8659-15, Session 4

New analog readout architecture for low noise CMOS image sensors using column-parallel forward noise-canceling circuitry

Tsung-Ling Li, Yasuyuki Goda, Shunichi Wakashima, Rihito Kuroda, Shigetoshi Sugawa, Tohoku Univ. (Japan)

In this paper, a new type of analog pixel readout architecture for low-noise CMOS image sensors has been presented. Comparing with conventional readout architecture, a proposed forward noise-canceller is applied in our readout architecture to provide a sharper noise filtering. The main column circuits consist of x16 low-noise amplifiers with correlated-double-sampling operation, proposed forward noise-cancellers, and analog memories for storage of readout reset and signal voltage. Through the high-gain column amplifier together with the column forward noise-canceller, this readout architecture significantly improves random noise of in-pixel source follower and column amplifier as well as temporal line noise from supplies and pulse lines. A 400(H) x 250(V) CMOS image sensor using the new readout architecture has been implemented in 0.18um 1P3M CMOS technology with pinned-photodiodes. Both the pixel pitch and the column circuit pitch are 4.5um. The area of the proposed forward noise-canceller is only 4.5um x 136um which is area-efficient. The pixel conversion gain is 67uV/e-. With the proposed architecture, the random noise of the in-pixel amplifier and column amplifier is reduced by 32% and 40%, respectively. The power supply rejection ratio improves nearly 90%. The measured random noise is reduced by about 35% compared to the conventional architecture.

8659-16, Session 4

A novel pixel design with hybrid type isolation scheme for low dark current in CMOS image sensor

Sung Ho Choi, Sungkyunkwan Univ. (Korea, Republic of) and Samsung Electronics Co., Ltd. (Korea, Republic of); Yi Tae Kim, Min Seok Oh, Samsung Electronics Co., Ltd. (Korea, Republic of); Young Hwan Park, Samsung Advanced Institute of Technology (Korea, Republic of); Jeong Jin Cho, Young Heub Jang, Hyung Jun Han, Jong Won Choi, Ho Woo Park, Sang Il Jung, Hoon Sang Oh, Jung Chak Ahn, Hiroshige Goto, Chi Young Choi, Samsung Electronics Co., Ltd. (Korea, Republic of); Yonghan Roh, Sungkyunkwan Univ. (Korea, Republic of)

New isolation scheme for CMOS image sensor pixel is proposed and the improved dark current performance is reported. It is well known STI(Shallow Trench Isolation), which is commonly used for electrical isolation of semiconductor devices, is one of the major sources of dark current in imager pixel due to the existence of interfacial defects at STI/Si interface.

To suppress the dark current generated at the STI/Si surface, the implantation of born atoms near the STI/Si surface is generally used to provide the sufficient amount of holes and to annihilate dark electrons.

In this work, we did experiment to remove the STI itself from the peripheral of photodiode in order to remove fundamental source of dark current, while remaining STI around in-pixel transistors. Experimental results are reviewed in this paper including isolation ability of in-pixel transistors.

8659-17, Session 5

Continuous fabrication technology for improving resolution in RGB-stacked organic image sensor

Toshikatsu Sakai, Hokuto Seo, Satoshi Aihara, Misao Kubota, NHK Science & Technology Research Labs. (Japan); Mamoru Furuta, Kochi Univ. of Technology (Japan)

To develop compact, high-resolution color cameras, we have been researching a novel image sensor with three stacked organic photoconductive films sensitive to only one of the primary color components, each of which has a signal readout circuit. We have previously fabricated three separate elements by forming red, green and blue organic layers with a ZnO thin-film transistor (TFT) array as a readout circuit on three glass substrates and stacked these elements to produce an image sensor (128?96 pixels). In the prototype sensor, however, when the optical image is focused on the middle of the three layers, the images formed on other layers are blurred because 0.7-mm-thick glass substrates exist between organic layers. This reduces the resolution in output images.

We developed a continuous fabrication technology that entails placing three organic layers close to each other, separated by 1-2-?m-thick interlayer insulators, on a single glass substrate. To prevent thermal damage to the organic layers, we fabricated a TFT and an interlayer insulator at a temperature lower than 150°C. We also explained a test production of a continuously-stacked organic image sensor using these low-temperature fabrication processes.

8659-18, Session 5

Biological tissue identification using a multispectral imaging system

Céline Delporte, Univ. Pierre et Marie Curie (France); Sylvie Sautrot, Univ. Denis Diderot (France) and Univ. Pierre et Marie Curie (France); Mohamed Ben Chouikha, Univ. Pierre et Marie Curie (France); Françoise Viénot, Muséum national d'Histoire naturelle (France); Georges Alquié, Univ. Pierre et Marie Curie (France)

This paper presents a multispectral imaging system (MIS) enabling biological tissue identifying and differentiation. The system would be installed in surgical room and should provide the surgeon an intraoperative diagnosis aid. In a previous work we introduced calibration of the multispectral imaging system and we validate its measurementof tissue spectral luminance, B(L). Even though the system was calibrated, we found wide variability of? B(L) values in the case of biological tissues. The dispersion was studied and evaluated for several tissues. As these media are complex and non homogenous, B(L)?variability reaches 20 %. This may be a big limitation for tissue identifying and differentiation. This wok focuses on the study of the

MIS ability to identify biological tissues and to differentiate them. Several identification methods are proposed and evaluated in order to assess their relevance. Each method uses specific criteria applied to the multispectral images. The MIS produces a B(L)?spectral cube of biological tissues in the spectral range of 500nm to 980nm. The tissue images were obtained according to the 0°/45° CIE standard geometry. An image processing algorithm extract at each pixel of the image the corresponding B(L)???A ?database of B(L) for various tissues can be established. We carried out measurement of B(L) spectral cube, for four tissue types present in the same scene : beef muscle, pork muscle, turkey muscle and beef. The experiment was repeated with the same types of tissue from different animals. The identification methods were applied to the multispectral images of the scenes. The method relevance is evaluated by calculating identification rate and over-detection rate for each tissue at each scene. The first method we studied is based on the correlation of spectral luminance factors B(L). Detection rates ranging from 64% to 89%, depending on tissue type, were obtained. This correlation criterion maintains low rates of over-detection, ie 0.5% to 5.5%. A second method was defined by calculating the derivative of each of the spectral luminance factors. The singular characteristics of tissues were thus highlighted. The correlation factor is calculated on the B(L) derivatives. Detection rates have improved and vary from 79% to 89% . A very low over-detection rate ranging from 0.1% to 0.2% is obtained. To improve detection rates, a third identification method is proposed and implemented. Some relevant wavelengths have been selected to display

8659-19, Session 5

A CMOS image sensor using floating capacitor load readout operation

Shunichi Wakashima, Yasuyuki Goda, Tsung-Ling Li, Rihito Kuroda, Shigetoshi Sugawa, Tohoku Univ. (Japan)

In this paper, a CMOS image sensor using floating capacitor load readout operation has been discussed. The floating capacitor load readout operation is used during pixel signals readout. And this operation has two features; one, in-pixel source followers are driven by load capacitor without current sources, two, parasitic capacitor of vertical signal line is used as storage capacitor. This operation produces three advantages. Smaller chip size, lower output noise and lower power consumption than conversional CMOS image sensors are achieved. The prototype CMOS image sensor was fabricated using a 0.18um 1P3M CMOS technology with pinned photodiodes. The die size is 2.5mm(H)x2.5mm(V), the pixel size is 4.5um(H)x4.5um(V), and the number of pixels is 400(H)x300(V). The conversion gain is 67uV/e-. This image sensor consists of only pixel array, scanning circuits, column source followers of which height is as low as that of the pixel and two output buffers. The size of peripheral circuit is reduced by 81% of conventional CMOS image sensors. The dark random noise of pixels except reset noise (after CDS) is 160uVrms. The noise of conventional image sensor is 340uVrms; therefore, 53% reduction is achieved. The power consumption for pixel output readout reduced 97% of conventional CMOS image sensor.

8659-20, Session 5

A UV Si-photodiode with almost 100% internal Q.E. and high transmittance on-chip multilayer dielectric stack

Yasumasa Koda, Rihito Kuroda, Taiki Nakazawa, Yukihisa Nakao, Shigetoshi Sugawa, Tohoku Univ. (Japan)

Ultraviolet light (UV-light) sensing, especially in about 200-300nm wavelength, is required. Sensitivity (external quantum efficiency) is expressed by internal quantum efficiency times transmittance. Recently, a technology to fabricate photodiode with almost 100% internal quantum efficiency and high stability to UV-light has been reported [1] [2]. We introduce a high transmittance on-chip multilayer dielectric stack above this type of photodiode and demonstrate the improved sensitivity and high stability to UV-light.

We focused on SiO2 and Si3N4 which have been used as interlayer dielectric in LSI. We employed a microwave excited high density plasma CVD equipment to form high quality Si3N4. [3] By decreasing gas flow rates of SiH4 and H2, we obtained a low extinction coefficient of 0.013 at 215nm wavelength.

By using the developed Si3N4 film, we fabricated photodiode with on-chip multilayer dielectric stack; Si/SiO2(8nm)/Si3N4(13nm)/SiO2(44nm)/Si3N4(25nm) with band pass filter-type high transmittance to UV-light. We obtained the peak of 86.9% external quantum efficiency at 238nm wavelength and sensitivity-degradation in UV-light region did not appear.

In conclusion, by optimizing the structure and thickness of the stack using SiO2 and low extinction coefficient Si3N4 with the high UV-light sensitivity photodiode technology, high external quantum efficiency and high stability to UV-light ware successfully obtained.

[1] R.Kuroda et al., "High Ultraviolet Light Sensitive and Highly Reliable Photodiode with Atomically Flat Si Surface" International Image Sensor Workshop, pp.38-41, 2011

[2] T.Nakazawa et al., "Photodiode dopant structure with atomically flat Si surface for high sensitivity to UV-light" The International Society for Optical Engineering, 2012

[3] Y.Nakao et al., "High Quality and Low Thermal Budget Silicon Nitride Depsition Using PECVD for Gate Spacer, Silicide Block and Contact Etch Stopper" Solid State Devices and Materials, 2011

8659-21, Session 6

High sensitivity analysis of speckle patterns: a technological challenge for biomedical optics *(Invited Paper)*

Jean-Michel Tualle, Kinia Barjean, Eric Tinet, Dominique Ettori, Univ. Paris 13 (France)

Diffuse light in tissue can be a very interesting tool for medical diagnosis, especially if one considers the fluctuations of the speckle pattern. Of course, speckle analysis suffers from the low spatial coherence of speckle patterns, and multi-pixel detection is required in order to increase the signal to noise ratio. There is therefore a need of a setup with high sensitivity, capable of outputting a signal from noise through averaging on a high number of pixels, as the signal can be lower than the photon level for one image and one frame. Furthermore, such a processing has to be done at a very high acquisition rate. "Smart-pixels" arrays can represent a major breakthrough in this field.

8659-22, Session 6

Gesture recognition on smart cameras

Aziz Dziri, Mehdi Darouich, Stéphane Chevobbe, CEA (France)

No Abstract Available

8659-23, Session 6

3DS-colorimeter based on a mobile phone camera for industrial applications

Jari Miettinen, Birgitta Martinkauppi, VTT Electronics (Finland); Pekka Suopajaervi, VTT Technical Research Center of Finland (Finland)

We have developed a novel mobile phone based measurement accessory device which is called as 3DScolorimeter. This is a new, low-cost consumer- or industrial-level product. It is suitable almost any mobile phone with a camera. When the 3DScolorimeter is attached on the mobile phone, then this system is capable to do colorimetric measurement and evaluation. As the 3D part of the name indicates, the system can do also 3D surface evaluation. This paper starts with an introduction of the device operation in which we demonstrate it suitability especially for industrial use. The device is also suitable

for a consumer because the price of the device is quite low and its use is very easy. The goal of this paper is, however, to study the repeatability and reliability of the 3D and colorimetric measurements for industrial use. We also evaluate the accuracy of colorimetric results with a Macbeth chart. These two factors, repeatability and accuracy are important for any device intended for reliable measurements. It is also possible to save, compare and send data and thus document the measurements.

8659-24, Session 6

A single lens with no moving parts for rapid high-resolution 3D image capture

Daniel C. Gray, Hongquiang Chen, Joseph Czechowski, Kang Zhang, Jilin Tu, Frederick Wheeler, Masako Yamada, Juan Pablo Cilia, Russell DeMuth, Esmaeil Heidari, Gil Abramovich, Kevin G. Harding, GE Global Research (United States)

Presently available methods to capture 3D data (stereo cameras, mechanical focus shifting, and structured illumination), are limited in speed, complexity, and transverse resolution. We present a novel lens system with high-resolution, wide field-of-view, and rapid multi-focus capture for 3D imaging. The imainge system was designed and manufactured with over 1000 pixel-per-inch resolution across a 300mm field-of-view at a 700mm working distance. The lens incorporates a single diffraction limited imaging objective coupled with an electro-optics unit to provide rapid focal shifts across a 30mm volume. Using a 12 megapixel, high speed CMOS camera, high-resolution, volumetric images are captured in less than 0.5 seconds. A novel depth-from-defocus algorithm was developed to generate 3D point clouds from the focal stack. The performance of the lens system is characterized using standard optical metrology methods and compared against the theoretical design. Three dimensional test targets are imaged to demonstrate the utility and performance of the assembled system. This material was prepared with the support of the U.S. Army under Award No. W911NF-11-C-0256. However, any opinions, findings, conclusions or other recommendations expressed herein are those of the author(s) and do not necessarily reflect the views of the U.S. Army.

8659-25, Session 6

Measurement and description method for image stabilization performance of digital cameras

Norihiro Aoki, Nikon Corp. (Japan); Hiroya Kusaka, Panasonic Corp. (Japan); Hiroyuki Otsuka, Canon Inc. (Japan)

Image stabilization functionality is widely acknowledged as an automated function of digital cameras. However, because unified methods of measurement of stabilization performance had not been developed, the Camera & Imaging Products Association (CIPA) standardized the measurement and description methods for image stabilization performance.

For standardization, CIPA formulated two vibration waveforms that almost faithfully simulate the characteristics of camera shake that occur when users actually shoot with the digital camera in hand. Using these vibration waveforms, vibration is applied to the digital camera which is mounted on a vibratory apparatus. The specified chart is photographed in this state, and from the motion blur amount of the image, image stabilization performance is measured. Image stabilization performance is calculated in "stops" converted into shutter speeds.

By using vibration waveforms, measurement man-hours are greatly reduced, and measurement is also made possible for digital cameras in which the image stabilization function cannot be deactivated. Moreover, by introducing a method that removes the effects of bokeh caused by the optical performance of the lens from motion blur, a high accuracy of measurement is realized. The image stabilization performance attained when users shoot is confirmed to be favorably measured in verification experiments of commercial digital cameras.

8659-26, Session 6

Interconnected network of cameras

Mahdad Hosseini Kamal, Hossein Afshari, Yusuf Leblebici, Alex Schmid, Pierre Vandegheynst, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The real-time development of multi-camera systems is a great challenge. Synchronization and large data rates of the cameras adds to the complexity of these systems as well. The complexity of such system also increases as the number of their incorporating cameras increases. The customary approach to implementation of such system is a central type, where the entire raw stream from the camera are first stored then processed for their target application. An alternative approach is to embed smart cameras to these systems instead of ordinary cameras with limited or no processing capability. Smart cameras with intra and inter camera processing capability and programmability at the software and hardware level will offer the right platform for distributed and parallel processing for multi-camera systems real-time application development. Inter camera processing requires the interconnection of smart cameras in a network arrangement. A novel hardware-emulating platform is introduced for demonstrating the concept of the interconnected network of cameras. A methodology is demonstrated for the interconnection network of camera construction and analysis. A sample application is developed and demonstrated.

Conference 8660: Digital Photography IX

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8660 Digital Photography IX

8660-1, Session 1

Multichannel camera calibration

Wei Li, Julie Klein, RWTH Aachen (Germany)

For the latest computer vision applications, it becomes more and more popular to take advantage of multichannel cameras (RGB cameras, etc.) to obtain not only gray values but also color information of pixels. The currently most common approach for multichannel camera calibration is the straightforward application of methods developed for calibration of single channel cameras. These conventional calibration methods may give quite poor performances including color fringes and displacement of features, especially for high-resolution multichannel cameras. In this paper, to suppress the undesired effects, a novel multichannel camera calibration approach, considering each single channel individually and involving different transversal chromatic aberration models, is introduced and evaluated. In comparison to the standard approach, the proposed approach provides more accurate calibration results in most cases and should lead subsequently to more reliable estimation results for computer vision issues. Moreover, besides the existing transversal chromatic aberration (TCA) model, further TCA model and correction methods are introduced which are superior to the existing ones. Since the proposed approach is based on the most popular calibration routine, only minimal modifications have to be made to the existing approaches to obtain the improved calibration quality.

8660-2, Session 1

Direct spatio-spectral datacube reconstruction from raw data using a spatially adaptive spatio-spectral basis

Yusuke Monno, Masayuki Tanaka, Masatoshi Okutomi, Tokyo Institute of Technology (Japan)

Spectral reflectance is an inherent property of objects that is useful for many computer vision tasks. The spectral reflectance of a scene can be described as a spatio-spectral (SS) datacube, in which each value represents the reflectance at a spatial location and a wavelength. In this paper, we propose a novel method that reconstructs the SS datacube directly from raw data, in which spectral measurements are spatially sampled. We first describe the SS datacube as a linear combination of SS basis vectors. Then, based on the raw data, we estimate the linear coefficients of the SS basis vectors. In contrast to the previous methods in which spatially invariant SS basis vectors are used for reconstructing the SS datacube, our method adaptively generates the SS basis vectors for each spatial location. Experimental results demonstrate that our proposed method can accurately reconstruct the SS datacube compared to the methods using the spatially invariant SS basis vectors.

8660-3, Session 1

Joint focus stacking and high dynamic range imaging

Qinchun Qian, Bahadir Gunturk, Louisiana State Univ. (United States); Aziz Umit Batur, Texas Instruments Inc. (United States)

Focus stacking and high dynamic range (HDR) imaging are two well-known computational photography paradigms; the former aims to create an extended depth of field image from multiple images with limited focus depth, whereas the latter aims to create a high dynamic range image from multiple images with low dynamic range. In this paper, we discuss the idea of combining focus stacking and HDR creation so that the resulting image has both higher dynamic range and extended depth of field. We developed an Android smartphone app that allows selecting multiple regions of interest to determine the optimal focus and exposure settings. Images captured with different focus regions and exposure times are then merged to form an extended depth of field and HDR image. We have considered both local sharpness and exposure information during the merge process. We have also developed an algorithm to estimate both photometric mapping function and dense optical flow. Experimental results obtained on an Android smartphone show that the proposed method is reliable and efficient to be implemented as a mobile imaging application.

8660-4, Session 1

FUJIFILM X10 white orbs and DeOrbIt

Henry G. Dietz, Univ. of Kentucky (United States)

The FUJIFILM X10 is a high-end enthusiast compact digital camera using an unusual sensor design. Unfortunately, upon its Fall 2011 release, the camera quickly became infamous for the uniquely disturbing "white orbs" that often appeared in areas where the sensor was saturated. FUJIFILM's first attempt at a fix was firmware released on Februrary 25, 2012 -- it had little effect. In April 2012, a sensor replacement reduced the problem to an acceptable level.

This paper explores the "white orb" phenomenon in detail. After FUJIFILM's attempt at a firmware fix failed, the author decided to create a post-processing tool that automatically could repair existing images. DeOrbIt was released as a free tool on March 7, 2012. To better understand the problem and how to fix it, the WWW form version of the tool logs images, processing parameters, and evaluations by users. The current paper describes the technical problem, the novel computational photography methods used by DeOrbit to repair affected images, and the public perceptions revealed by this experiment.

8660-5, Session 1

Automated calibration and optical testing of the AWARE-2 gigapixel multiscale camera

David S. Kittle, Daniel L. Marks, David J. Brady, Duke Univ. (United States)

Gigapixel-class cameras present new challenges in calibration, mechanical testing, and optical performance evaluation. The AWARE-2 gigapixel camera has nearly one-hundred micro-cameras spanning a 120 degree wide by 40 degree tall field of view, with one pixel spanning an 8 arcsec field angle. The 98 individual micro-camera images are captured to within one-tenth of a second for all the cameras in the array then stitched into a composite panoramic image. This makes the AWARE-2 system ideal for capturing snapshot, high-resolution imagery of dynamic events.

Stitching the images requires pixel-level accurate mapping over the entire FOV, therefore a testbed has been developed to automatically calibrate and test each micro-camera in the array. MTF measurements, pointing vectors of the micro-cameras for compositing, focus calibration, distortion mapping, and illumination profile measurements are detailed for the 1-gigapixel AWARE-2 prototype camera system. This data then builds a look up table for near real-time compositing and viewing, significantly improving the system model and overall reconstruction quality.

8660-6, Session 2

Across-resolution adaptive dictionary learning for single-image super-resolution

Masayuki Tanaka, Ayumu Sakurai, Masatoshi Okutomi, Tokyo Institute of Technology (Japan)

This paper proposes a novel adaptive dictionary learning approach for a single-image super-resolution based on a sparse representation. The adaptive dictionary learning approach of the sparse representation is very powerful, for image restoration such as image denoising. The existing adaptive dictionary learning requires training image patches which have the same resolution as the output image. Because of this requirement, the adaptive dictionary learning for the single-image super-resolution is not trivial, since the resolution of the input low-resolution image which can be used for the adaptive dictionary learning is essentially different from that of the output high-resolution image. It is known that natural images have high across-resolution patch redundancy which means that we can find similar patches within different resolution images. Our experimental comparisons demonstrate that the proposed across-resolution adaptive dictionary learning approach outperforms state-of-the-art single-image super-resolutions.

8660-7, Session 2

Aesthetic scoring of digital portraits for consumer applications

Sebastiano Battiato, Marco Moltisanti, Fabrizio Ravì, Univ. degli Studi di Catania (Italy); Arcangelo R Bruna, Filippo Naccari, STMicroelectronics (Italy)

Computational Aesthetics applied on digital photography is becoming an interesting issue in different frameworks (e.g., photo album summarization, imaging acquisition devices). Although it is widely believed and can often be experimentally demonstrated that aesthetics is mainly subjective, we aim to find some formal or mathematical explanations of aesthetics in photographs. We propose a scoring function to give an aesthetic evaluation of digital portraits and group pictures, taking into account faces aspect ratio, their perceptual goodness in terms of lighting of the skin and their position. Also well known composition rules (e.g., rule of thirds) are considered especially for single portrait. Both subjective and quantitatively experiments have confirmed the effectiveness of the proposed methodology.

8660-8, Session 2

Correcting saturated pixels in images based on human visual characteristics

Jun Fu, Hui Peng, Xi'an Jiaotong Univ. (China); Xi Chen, Xi'an Jiaotong University (China); Xuanqin Mou, Xi'an Jiaotong Univ. (China)

This paper presents a correcting method for saturated images which is operated in the YCbCr color space. The algorithm is based on two human visual characteristics, one is the visual sensitivities to color differences and the other is the Hunt effect. During the process of correcting colors, MacAdam ellipse model mapped to the YCbCr color space is used to search the nearest color. So that corrected pixels' chromatic aberration is less to that of their neighborhood. And during the process of the quantification of the YCbCr components for digital implementation, the regions with high luminance are set to have less saturation based on the Hunt effect. Experimental results show that the proposed method is effective in correcting face saturated and antagonistic color images, and the color and the luminance of unsaturated pixels in original image are maintained as much as possible, especially for the optimization of the region with less luminance and more colorfulness.

8660-9, Session 3

Spectrally tunable pixel sensors

Giacomo Langfelder, Cesare Buffa, Antonio F. Longoni, Federico Zaraga, Politecnico di Milano (Italy)

After the demonstration of the basic working principle, of potential applications and of a suitable active pixel topology, they are here reported the development and experimental results of fully operating matrices of spectrally tunable pixels based on the Transverse Field Detector (TFD).

Unlike several digital imaging sensors based on color filter arrays or layered junctions, the TFD has the peculiar feature of having electrically tunable spectral sensitivities. In this way the sensor color space is not fixed a priori but can be real-time adjusted, e.g. for a better adaptation to the scene content or for multispectral capture.

These advantages come at the cost of an increased complexity both for the photosensitive elements and for the readout electronics. The challenges in the realization of a matrix of TFD pixels will be analyzed in this work. First experimental results on an 8x8 (x 3 colors) and on a 64x64 (x 3 colors) matrix will be presented and analyzed in terms of colorimetric and noise performance, and compared to simulation predictions.

8660-10, Session 3

Design, simulation, and evaluation of imaging oximeters

Steve Dai, Ye Tian, Joyce E. Farrell, Stanford Univ. (United States)

Computer simulations have played an important role in the design and evaluation of imaging sensors with applications in remote sensing [1] and consumer photography [2-7]. In this paper, we provide an example of computer simulations used to guide the design of imaging sensors for a biomedical application: We consider how sensor design, illumination, measurement geometry, and skin type influence the ability to detect blood oxygen saturation from non-invasive measurements of skin reflectance. The methodology we describe in this paper can be used to design, simulate and evaluate the design of other biomedical imaging systems.

8660-11, Session 3

A collection of hyperspectral images for imaging systems research

Torbjorn Skauli, Norwegian Defense Research Establishment (Norway); Joyce E. Farrell, Stanford Univ. (United States)

A set of hyperspectral image data are made available, intended for use in modelling of imaging systems. The set contains images of faces, landscapes, buildings and paintings. The data cover wavelengths from 0.4 to 2.5 micrometers, spanning in the visible, NIR and SWIR electromagnetic spectral ranges. The images have been recorded with two HySpex line-scan imaging spectrometers covering the spectral ranges 0.4 to 1 micrometers and 1 to 2.5 micrometers. The respective across-track pixel count of the two sensors is 1600 and 320. The hyperspectral data set includes measured illuminants and software for converting the radiance data to estimated reflectance. Full sensor calibration data are available, enabling estimation of signal-dependent noise from the physical sensor model. The images are being made available from the Stanford Center for Image Systems Engineering (SCIEN).

8660-12, Session 3

Evaluation of multispectral plenoptic camera

Lingfei Meng, Ting Sun, Rich Kosoglow, Kathrin Berkner, Ricoh Innovations, Inc. (United States)

Plenoptic cameras enable capture of a 4D lightfield, allowing digital refocusing and depth estimation from data captured with a compact portable camera. Whereas most of the work on plenoptic camera design has been based a simplistic geometric-optics-based characterization of the optical path only, little work has been done of optimizing end-to-end system performance for a specific application. Such design optimization requires design tools that need to include careful parameterization of main lens elements, as well as microlens array and sensor characteristics. In this paper we are interested in evaluating the performance of a multispectral plenoptic camera, i.e. a camera with spectral filters inserted into the aperture plane of the main lens. Such a camera enables single-snapshot spectral data acquisition. We first describe in detail an end-to-end imaging system model for a spectrally coded plenoptic camera. Different performance metrics are defined to evaluate the spectral reconstruction quality. We then present a prototype that is developed based on a modified DSLR camera containing a lenslet array on the sensor and a filter array in the main lens. Finally we evaluate the spectral reconstruction performance of a multispectral plenoptic camera.

8660-13, Session 4

Median filtering in multispectral filter array demosaicking

Xingbo Wang, Univ. de Bourgogne (France) and Gjøvik Univ. College (Norway); Jean-Baptiste Thomas, Univ. de Bourgogne (France); Jon Y. Hardeberg, Gjøvik Univ. College (Norway); Pierre Gouton, Univ. de Bourgogne (France)

No Abstract Available

8660-14, Session 4

A novel blind image quality assessment metric and its feature selection strategy

Ying Chu, Xi'an Jiaotong Univ. (China) and Shenzhen Univ. (China); Xuanqin Mou, Wei Hong, Xi'an Jiaotong Univ. (China); Zhen Ji, Shenzhen Univ. (China)

We recently proposed a natural scene statistics based image quality assessment (IQA) metric named STAIND, which extracts nearly independent components from natural image, i.e., the divisive normalization transform (DNT) coefficients, and evaluates perceptual quality of distortion image by measuring the degree of dependency between neighboring DNT coefficients. To improve the performance of STAIND, its feature selection strategy is thoroughly analyzed in this paper.

The basic neighbor relationships in STAIND include scale, orientation and space. By analyzing the joint histograms of different neighborhoods and comparing the IQA modal performances of diverse feature combination schemes in the publicly available databases such as LIVE, CSIQ and TID2008, we draw the following conclusions: 1) Spatial neighbor relationship contributes most to the modal design, scale neighborship takes second place, and orientation neighbors might introduce negative effects; 2) In space domain, second order spatial neighbors are beneficial supplements to first order spatial neighbors; 3) Combining scale and space neighborships and introducing spatial parent neighbors could be helpful for blind IQA metric design.

8660-15, Session 4

Edge patterns extracted from natural images and their statistics for reduced-reference image quality assessment

Wenting Shao, Xuanqin Mou, Xi'an Jiaotong Univ. (China)

Natural images are of plenty of edges which results in the scaling law. This fact hints that the spatial distribution of edge points of natural images are of some specific patterns. Based on this thought, we proposed to learn a number of groups of the named edge patterns to represent the distribution of zero-crossings (ZC) maps of both the natural images and their distortion counterparts and then successfully proposed a RR (Reduced Reference) IQA (Image Quality Assessment) metric. In this paper, we focus on the issues that whether there exist some specific edge patterns being good at represent the ZC maps of natural images, what the rule is by which the edge patterns are selected, and if the above is true, how we could use those edge patterns to perform an IQA model. In this investigation, we raise a process for the edge patterns selection by proposing the smoothness constraint and the error tolerance in the statistic of edge patterns representation, and collect RR features. Experiments show that the selected edge patterns can well represent the ZC maps of natural images, as well as the distortion images with error tolerance. The IQA model evaluation shows that the proposed metric works fairly good compared to its competitor.

8660-16, Session 5

Mobile-phone-based clinical microscopy (Invited Paper)

Daniel A. Fletcher, Univ. of California, Berkeley (United States)

Microscopy is a critical tool for disease research, screening, and diagnosis. Presently, use of microscopy for health care is often limited to well-equipped medical laboratory settings staffed by qualified personnel. In the developing world and other underserved regions, the lack of equipment and expertise required for diagnostic microscopy contributes to poor health, spread of treatable diseases, and emergence of drug-resistant disease strains. While medical resources are scarce in many developing countries and rural communities, the widespread availability of wireless communication and camera-enabled mobile phones has the potential to fundamentally change the way medical diagnoses are performed. A compact and portable microscopy system based on a mobile phone and capable of image capture, image processing, and communication with medical experts could dramatically increase access to basic health care by delivering services closer to where patients live and work. This talk will describe recent progress developing and implementing such a device, which we call CellScope, to improve diagnosis of infectious diseases.

8660-17, Session 5

Computational microscopy, sensing, and diagnostics on a cellphone (Invited Paper)

Aydogan Ozcan, Univ. of California, Los Angeles (United States)

Today there are close to 6 billion cell-phone users in the world, and the majority of these cellphones are being used in the developing parts of the world. This massive volume of wireless phone communication brings an enormous cost-reduction to cellphones despite their sophisticated hardware and software capabilities. Utilizing this advanced state of the art of the cell phone technology toward point-of-care diagnostics, sensing and microscopic imaging applications can offer numerous opportunities to improve health care especially in the developing world where medical facilities and infrastructure are extremely limited or even do not exist. For this end, here we review our recent progress on the use of cellphones as computational microscopes, bio-chemical sensors as well as medical diagnostic tools.

8660-18, Session 5

Optics in pathology: from superresolution microscopy to point-of-care devices (Invited Paper)

Sebastian Wachsmann-Hogiu, NSF Ctr. for Biophotonics Science and Technology (United States)

As society struggles to continue to make strides in basic biological research and to continue to meet global demand for medical care at

decreasing cost, optical technologies offer compelling solutions by providing detailed chemical and structural information about biological samples, often without the need for costly chemical consumables or labor and expertise-intensive biopsy procedures. In this talk I will present our group's work to develop and apply novel optical technologies for pathology.

One example is the development of two attachments to a commercial cell phone that transform the phone's integrated lens and image sensor into a 350x microscope and visible-light spectrometer. The microscope is capable of transmission and polarized microscopy modes and is shown to have 1.5 micron resolution and a usable field-of-view of ~150x150 microns with no image processing, and approximately 350x350 microns when post-processing is applied. The spectrometer has a 300 nm bandwidth with a limiting spectral resolution of close to 5nm. We show applications of the devices to medically relevant problems. In the case of the microscope, we image both stained and unstained blood-smears showing the ability to acquire images of similar quality to commercial microscope platforms, thus allowing diagnosis of clinical pathologies. With the spectrometer we demonstrate acquisition of a white-light transmission spectrum through diffuse tissue as well as the acquisition of a fluorescence spectrum.

8660-19, Session 6

SPCA: a no-reference image quality assessment based on the statistic property of the PCA on nature images

Yun Zhang, Chao Wang, Xuanqin Mou, Xi'an Jiaotong Univ. (China)

Despite the acceptable performance of current full-reference image quality assessment (IQA) algorithms, the need for a reference signal limits their application, and calls for reliable no-reference algorithms. Most no-reference IQA approaches are distortion specific, aiming to measure image blur, JPEG blocking or JPEG2000 ringing artifacts respectively. In this paper, we proposed a no-reference IQA algorithm based on the property of principal component analysis on nature image, which does not assume any specific type of distortion of the image. The method gets statistics of discrete cosine transform coefficients from the distort image's principal components. Those features are trained by?-support vector regression method and finally test on LIVE database. The experimental results show a high correlation with human perception of quality (averagely over 90% by scores of SROCC), which is fairly competitive with the existing no-reference IQA metrics.

8660-20, Session 6

Local binary pattern statistics feature for reduced reference image quality assessment

Min Zhang, Gifu Univ. School of Medicine (Japan); Xuanqin Mou, Xi'an Jiaotong Univ. (China); Hiroshi Fujita, Gifu Univ. School of Medicine (Japan); Lei Zhang, The Hong Kong Polytechnic Univ. (China); Xiangrong Zhou, Gifu Univ. School of Medicine (Japan); Wufeng Xue, Xi'an Jiaotong Univ. (China)

Measurement of visual quality is of fundamental importance for numerous image and video processing applications. This paper presented a novel and concise reduced reference (RR) image quality assessment prototype. Statistics of local binary pattern (LBP) is introduced as a similarity measure to form a novel RR image quality assessment (IQA) method for the first time. First, the test image is decomposed with a multi-scale transform. Second, LBP encoding maps are extracted for each subband images. Third, the histograms are extracted from the LBP encoding map to form the RR features. In this way, image structure primitive information for RR features extraction can be reduced greatly. Hence, new RR IQA method is formed with only at most 56 RR features. The experimental results on two large scale IQA databases show that the statistic of LBPs is fairly robust and reliable to RR IQA task. The proposed method shows strong correlations with subjective quality evaluations.

8660-21, Session 7

Evaluation of a high dynamic range video camera with non-regular sensor

Michael Schöberl, Friedrich-Alexander Univ. Erlangen-Nürnberg (Germany) and Fraunhofer-Institut für Integrierte Schaltungen (Germany); Joachim Keinert, Matthias Ziegler, Fraunhofer-Institut für Integrierte Schaltungen (Germany); Jürgen Seiler, Friedrich-Alexander Univ. Erlangen-Nürnberg (Germany); Marco Niehaus, Gerald Schuller, Technische Univ. Ilmenau (Germany); André Kaup, Friedrich-Alexander Univ. Erlangen-Nürnberg (Germany); Siegfried Fößel, Fraunhofer-Institut für Integrierte Schaltungen (Germany)

Although there is steady progress in sensor technology, imaging with a high dynamic range (HDR) is still difficult for motion imaging with high image quality. This paper presents our new approach for video acquisition with high dynamic range. The principle is based on optical attenuation of some of the pixels of an existing image sensor. This well known method traditionally trades spatial resolution for an increase in dynamic range. In contrast to existing work, we use a non-regular pattern of optical ND filters for attenuation. This allows for an image reconstruction that recovers high resolution images. The reconstruction is based on the assumption that natural images can be represented nearly sparse in transform domains, which allows for recovery of scenes with high detail. The proposed combination of non-regular sampling and image reconstruction leads to a system with an increase in dynamic range without sacrificing spatial resolution. In this paper, a further evaluation is presented on the achievable image quality. In our prototype we found that crosstalk is present and significant. The discussion thus shows the limits of the proposed imaging system.

8660-22, Session IPI

Spectral characterization of COTS RGB cameras using a linear variable edge filter

Daniel L. Bongiorno, Mitch Bryson, Donald G. Dansereau, Stefan B. Williams, The Univ. of Sydney (Australia)

The spectral response of the color channels of a commercial-off-the-shelf (COTS) Red/Green/Blue (RGB) camera is often unknown. Knowledge of this response can be useful for a wide variety of applications such as simulating animal vision, color correction and color space transformations. COTS cameras are widely used due to their low cost and ease of implementation. We investigate a method using a Linear Variable Edge Filter (LVEF) and a low-cost spectrometer to characterize an RGB camera; this method has the advantage over previous methods in the simplicity and high speed of the spectral characterization. Results are presented for three cameras: a consumer-level digital SLR; the other two are point-and-shoot consumer grade cameras with one of them being an underwater camera.

8660-23, Session IPI

Focus finding using scale invariant patterns

David P. Morgan-Mar, Matthew R. Arnison, Canon Information Systems Research Australia Pty. Ltd. (Australia)

We present a test-chart based focus finding system for use in image quality measurement with digital SLR cameras. We establish an absolute measurement of focus position independent of any mechanical instability in the focus drive motor by combining optical focus breathing, high accuracy scale invariant alignment patterns, and blur-invariant phase correlation. This combination allows accurate measurement of focus position under conditions of chart scaling, test rig misalignment, and defocus. The resulting absolute focus scale allows an efficient closed-loop search for the best focus position, followed by a return to the found position for subsequent image quality measurement. This method rapidly drives the lens to the best focus position to within one focus motor pulse, which is significantly more accurate and reliable than camera auto-focus systems, and an order of magnitude faster than an exhaustive focus sweep. We evaluated

our method with 8771 focus finding operations over a range of focal lengths and found that it allows the peak focus MTF to be measured consistently to within 1.4% RMS.

8660-24, Session IPI

Frequency division multiplexed imaging

Bahadir Gunturk, Martin Feldman, Louisiana State Univ. (United States)

In this paper, we describe frequency division multiplexed imaging (FDMI), where multiple images are captured simultaneously in a single shot as a multiplexed image and later extracted from the multiplexed image through frequency domain filtering. This is achieved by spatially modulating the images so that they are placed at different locations in the Fourier domain. The technique assumes that the images are band-limited and placed at non-overlapping frequency regions through a spatial light modulation process. The FDMI technique could be used to in a variety of applications, including stereoscopic image capture with a single sensor and sub-exposure motion analysis. We present optical setups and provide experimental results with real and simulated data to illustrate the idea.

8660-25, Session IPI

Bayesian estimation of device spectral sensitivities and its application for improvement of color accuracy using color balancing filter

Yusuke Murayama, Jay Arre O. Toque, Pengchang Zhang, Ari Ide-Ektessabi, Kyoto Univ. (Japan)

We proposed a Bayesian method to recover the system spectral sensitivities of a color imaging device from an acquired color chart image, and a scheme to predict a color balancing filter that needs to be attached to the device to improve color accuracy based on the recovered sensitivities. The recovery method is derived by introducing the non-negativity, the smoothness and the zero boundary of the sensitivity curves as prior information, and is ran without manual tuning by adopting the marginalized likelihood criterion. The sensitivities are modified by available filters computationally and their color accuracy are estimated. An experiment was carried out to test the performance of the proposed method for predicting the color accuracy improvement using a scanner and two kinds of commercial color charts, IT8 target as a training chart, and ColorChecker Classic as a test chart. The average color difference was improved from 4.27 to 2.57 in the test chart by attaching the predicted optimal filter. The proposed method to recoverfor recovering the system sensitivities will be able to be applied to choose the camera and illuminant, or to optimize calibration method.

8660-26, Session IPI

Achieving equal image quality at lower bit rates using evolved image reconstruction transforms

Brendan J. Babb, Frank W. Moore, Univ. of Alaska Anchorage (United States)

Several recent NASA missions have used the state-of-the-art wavelet-based ICER Progressive Image Compressor for lossy image compression. In this paper, we describe a methodology for using evolutionary computation to optimize wavelet and scaling numbers describing reconstruction-only multiresolution analysis (MRA) transforms that are capable of accepting as input test images compressed by ICER software at a reduced bit rate (e.g., 0.99 bits per pixel [bpp]), and producing as output images whose average quality, in terms of mean squared error (MSE), equals that of images produced by ICER's reconstruction transform when applied to the same test images compressed at a higher bit rate (e.g., 1.00 bpp). This improvement can be attained without modification to ICER's compression, quantization, encoding, or dequantization algorithms, and with very small modifications to existing ICER reconstruction filter code. As a result, future NASA missions will be able to transmit greater amounts of information (i.e., a greater number of images) over channels with equal bandwidth, thus achieving a no-cost improvement in the science value of future NASA missions.

8660-27, Session IPI

Computer graphics solutions for pottery colors specification

Filippo D. Stanco, Anna M. Gueli, Univ. degli Studi di Catania (Italy)

A main issue in the archaeological research is to deal with colors of soils and artefacts, especially pottery. To avoid risks of a too subjective recognition Munsell system is commonly used in archaeology. In this paper a semi-automatic method of color detection on selected regions of digital images of ancient pottery is presented. The starting point of the experiment was to take digital pictures of specimens together with the Gretag-Macbeth Color Checker Chart. The digital image is processed with color balancing techniques aimed to restore the original value of Macbeth patches, in order to eliminate distortions coming from acquisition process. After the color correction, several regions of interest are selected for the identification of surface color, the algorithm converts RGB values in Munsell data. The reliability of our tool is also verified comparing these chromatic values with the color specification of pottery sherds performed with a spectrocolometer using the CIELAB space to evaluate the differences.

8660-28, Session IPI

Investigating the lateral resolution in a plenoptic capturing system using the SPC model

Mitra Damghanian, Roger Olsson, Mårten Sjöström, Mid Sweden Univ. (Sweden); Hector Navarro, Manuel Martinez-Corral, Univ. de València (Spain)

Complex multidimensional capturing setups such as plenoptic cameras (PC) introduce a trade-off between various system properties. Established capturing properties, like image resolution, need, therefore, to be described thoroughly for these systems. Hence models and metrics that assist exploring and formulating this trade-off are highly beneficial for studying as well as designing of complex capturing systems. This work demonstrates the capability of our previously proposed sampling pattern cube (SPC) model to extract the depth dependent lateral resolution. The SPC carries both ray information as well as focal properties of the capturing system it models. The proposed operator extracts the lateral resolution from the SPC model throughout an arbitrary number of depth planes giving a depth-resolution profile. We have validated the resolution operator by comparing the results with those from other models and from empirical data. The lateral resolution predicted by the SPC model agrees with the results from experimental setups and strengthens the conclusion that the SPC fills the gap between ray-based models and real system measurements, by including the focal information of the system as a model parameter. The SPC is proven a simple yet efficient model for extracting depth-based lateral resolution as a high-level property of complex plenoptic capturing system.

Conference 8661: Image Processing: Machine Vision Applications VI

Tuesday - Wednesday 5 –6 February 2013

Part of Proceedings of SPIE Vol. 8661 Image Processing: Machine Vision Applications VI

8661-1, Session 1

A polynomial phase-shift algorithm for high precision three-dimensional profilometry

Fuqin Deng, The Univ. of Hong Kong (Hong Kong, China); Chang Liu, Wui-Fung Sze, Jiangwen Deng, Shunming K. Fung, ASM Pacific Technology Ltd. (China); Edmund Y. Lam, The Univ. of Hong Kong (China)

The perspective effect is common in real optical systems for machine vision applications. In conventional phase-shifting algorithms, the frequencies of the projected sinusoidal patterns are assumed to be uniform at different height levels, therefore, the error caused by the perspective effect from projection system becomes pronounced in phase-measuring profilometry, especially for some high precision metrology applications such as measuring the surfaces of the tiny semiconductors at micrometer level in the semiconductor industry. In this research, we investigate the perspective effect on phase-measuring profilometry when reconstructing the surfaces of moving objects. To generalize the phase-measuring profilometry model and solve the phase reconstruction problem effectively, we use a polynomial to approximate the phase distribution under a perspective projection system and name it polynomial phase-measuring profilometry (P-PMP) model. Under this model, we characterize how the frequency of the projected pattern changes according to height variations and how the phase of the projected pattern distributes in the measuring space. After calibrating the phase distribution, we propose a polynomial phase-shifting algorithm (P-PSA) to correct the error due to imperfect projection system during phase reconstruction. Simulation experiments show that the proposed method can improve the reconstruction quality both visually and numerically.

8661-2, Session 1

High-temperature dual-band thermal imaging by means of high-speed CMOS camera system

Wolfgang Hauer, Gerald Zauner, FH OÖ Forschungs & Entwicklungs GmbH (Austria)

In this paper we present a novel approach to determine surface radiation temperatures (up to 4000 °K) at a very high time resolution (~10 µs) by means of a CMOS high speed camera. As governed by Plank's law of radiation, the amount of light emission in the visible and near infrared range of the electromagnetic spectrum at temperatures above 1500 °K is sufficient to be detected with a CMOS sensor even at very short integration times (high speed imaging).

With the aid of a 3D lens and two distinct spectral narrow band pass filters, we apply the so called dual-band method (or two-color pyrometry), where the intensity ratio of thermal light emission in two different spectral bands is used to deduce the surface temperature of an object. Due to slightly different camera viewing angels of each lens, a stereo calibration process is applied in order to correctly register corresponding image pixels. Additionally, we apply image denoising algorithms, based on a noise characterization of the camera, to further improve the measurement quality.

With the presented system we successfully imaged and measured the contact surface temperature decay of low-voltage circuit breakers - a very demanding metrological task, as temperatures of 4000 °K can occur at time scales of less than 100 µs.

8661-3, Session 1

A state observer for using a slow camera as a sensor for fast control applications

Reinhard Gahleitner, Univ. of Applied Sciences Upper Austria (Austria); Martin Schagerl, University of Applied Sciences Upper Austria (Austria)

The contribution deals with a problem that often arises in vision based control, if a camera is used as a sensor for fast control applications, or more precise, when the sample rate of the control loop is higher than the frame rate of the camera. In control applications for mechanical axes, e.g. in robotics or automated production, a camera and some image processing can be used as a sensor to detect positions or angles. The sample time in this applications typically is below 10 ms or even 1 ms and this demands the use of a camera with a frame rate above 100 fps or 1000 fps, respectively.

The presented solution is a special state observer that can work with a slower and therefore cheaper camera to estimate the state variables at the higher sample rate of the control loop. To simplify the image processing for the determination of positions or angles and make it more robust some LED markers are applied to the plant. Simulation and experimental results show that the concept can be used even if the plant is unstable like the inverted pendulum.

8661-4, Session 2

Multiple-level patch-based object tracking using the MLBP-based integral histogram

Jirui Yuan, Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

This paper presents a novel multi-level patch-based approach for object tracking using Modified Local Binary Pattern (MLBP) histograms. The initial template is divided into overlapping rectangular patches, and each of these patches is tracked independently by finding the most similar match within a search region. Every patch votes on the possible locations of the object in the current frame, by comparing its MLBP histogram with the correspondence in the target frame. To reduce the individual tracking error of a given patch due to partial occlusions, the idea of multi-level patch partitioning is further developed. And the similarity between template and target object is compared patch-by-patch, level-by-level. The comparison starts from the highest level and progressively feeds to the lowest level through a median operation.

The proposed algorithm provides additional robustness and effectiveness in several ways. First, the spatial relationship among patches is improved by this overlapping partitioning manner. Second, by introducing MLBP operator, the tracking accuracy is significantly improved. Third, the median operation utilized in the multi-level vote-combining process provides additional robustness with respect to outliers resulting from occluded patches and pose changes. The proposed method is evaluated using both face and pedestrian sequences, and comparison is made w.r.t. several state-of-art tracking algorithms, including frag-track and mean-shift methods. Experimental results show that the proposed method significantly outperforms in case of occlusions and pose changes. Besides, the tracking in case of scale changes additionally proves the effectiveness and efficiency of the proposed method.

8661-5, Session 2

Periodicity estimation of nearly regular textures based on discrepancy norm

Gernot Stübl, Peter Haslinger, Volkmar Wieser, Software Competence Ctr. Hagenberg (Austria); Josef Scharinger, Johannes Kepler Univ. Linz (Austria); Bernhard Moser, Software Competence Ctr. Hagenberg (Austria)

Sliding window based processing of images is a crucial step in various image processing applications for example in template matching based methods.

The choice of an optimal window size is not always straightforward. Particularly, in the context of nearly regular textured images this question turns out to deserve special consideration. Typically such textures play an important role in quality inspection of textile fabrics.

This paper proposes a novel approach to determine the texture periodicity, the texture element size and further characteristics like the area of the basin of attraction in the case of computing the similarity of a test image patch with a reference.

The presented method utilizes the properties of a novel metric, the so-called discrepancy norm. In contrast to Minkowski norms this norm is based on the evaluation of partial sums by which the discrepancy norm becomes ordering dependent yielding a highly asymmetric unit ball. This metric distinguishes by monotonicity as well as a Lipschitz continuity property that allow robust computation at the presence of noise and variations in appearance.

The general form of the proposed approach relies on the generation of discrepancy norm induced similarity maps at random positions in the image. Through applying standard image processing operations like Watershed and blob analysis on the similarity maps a robust estimation of the characteristic periodicity can be computed. As a byproduct of this image analysis one gets a segmentation which specifies convergence properties for template matching.

Even though the discrepancy norm uses only add/max operations and can be implemented with O(n) with n as number of pixels, computing a full similarity map is still computationally expensive. Therefore a faster version of the general approach tailored to orthogonal aligned textures is presented.

Due to the Lipschitz and the monotonicity property the discrepancy norm distinguishes itself from other metrics by well-formed and stable convergence regions. Both the periodicity and the convergence regions are closely related and have an immediate impact on the performance of a subsequent template matching and evaluation step.

In an experimental setup the estimation performance is tested on samples of standardized image databases and is compared with state-of-the-art methods. Results show that the proposed method is applicable to a wide range of nearly regular textures and shows robustness to noise disturbed images.

Summing up it can be said that through the use of a novel metric the presented approach needs only basic image processing techniques to estimate the characteristic periodicity for near regular textures and furthermore gives important configuration information for subsequent optimization or template matching steps. This is demonstrated on real world examples and an implementation concept for industrial usage is given.

8661-6, Session 2

Gradient feature matching for in-plane rotation invariant face sketch recognition

Ann Theja Alex, Vijayan K. Asari, Alex Mathew, Univ. of Dayton (United States)

In this work we propose a new method for face sketch recognition that does not require any synthesis step and allows recognition across modalities. The proposed method uses the edge features of a face sketch and face photo image to create a feature string called 'edge-string'. The method uses edge detection algorithms to extract the edge features. The thresholds used for edge detection are determined automatically using the Otsu's algorithm. The next step is to represent edge features in polar coordinate representation. To represent features as polar coordinates, we need to determine the centroid and a reference line. The reference line is defined as the line passing through eye centers. The method uses Circular Hough transform to effectively localize the iris location and the eye center is determined as the center of the iris. The center of this reference line is defined as the centroid. The edge-strings of the face photo and face sketch are compared using the Smith-Waterman algorithm for local string alignments. The face photo that gave the highest similarity score is the photo that matches the test face sketch input. Preliminary results on the CUHK database give promising results and prove the effectiveness of the proposed approach.

8661-7, Session 2

An iris segmentation algorithm based on edge orientation for off-angle iris recognition

Mahmut Karakaya, Del Barstow, Hector J. Santos-Villalobos, Christopher Boehnen, Oak Ridge National Lab. (United States)

Iris recognition is known as one of the most accurate and reliable biometric measure. However, the accuracy of iris recognition system relies on a well-controlled setup and is affected by several factors such as off-angle, occluded, and dilated images. Therefore, non-ideal iris recognition is a new emerging research field and got attentions from many research groups. In this paper, we present a segmentation algorithm for off-angle iris images by using edge detection, edge elimination, edge classification, and ellipse fitting techniques. In our approach, we first detect the all possible edges in the iris image by using the canny edge detector which may contain edges from not only iris and pupil but also eyelash, eyelids, iris texture etc. Therefore, we eliminate the edges that cannot be part of the iris or pupil by using the edge orientation. Third, we classify the remaining edge points into two sets as pupil edges and iris edges. Finally, we randomly generate subsets of in each set, fit ellipses for subset of each iris and pupil edge points and select ellipses with similar parameters and average resultant ellipses. Based on the results from real experiments, the proposed method shows effectiveness in segmentation for off-angle iris images.

8661-8, Session 3

Dense sampling of shape interiors for improved representation

Vittal Premachandran, Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore)

In this paper, we identify some of the limitations of current-day shape matching techniques. We provide examples of how contour-based shape matching techniques cannot provide a good match for certain visually similar shapes. To overcome this limitation, we propose a perceptually motivated variant of the well-known shape context descriptor. We identify that the interior properties of the shape play an important role in object recognition and develop a descriptor that captures these interior properties. Our descriptor captures the properties of the shape in its entirety. We show that our method can easily be augmented with any other shape matching algorithm. We also show from our experiments that the use of our descriptor can significantly improve the retrieval rates.

8661-9, Session 3

Efficient defect detection with sign information of Walsh Hadamard transform

Qiang Zhang, Arizona State Univ. (United States); Peter J. L. van Beek, Chang Yuan, Cindy Xu, Hae-jong Seo, Sharp Labs. of America, Inc. (United States); Baoxin Li, Arizona State Univ. (United States)

Conference 8661

We propose a method for defect detection based on taking the sign information of Walsh Hadamard Transform (WHT) coefficients. The core part of the proposed algorithm only involves three steps, applying forward WHT, computing sign of transform coefficients and taking an inverse WHT on the sign information, which can all be implemented very efficiently. Our implementation takes only 7 milliseconds for a 512x512 image on a PC platform. As a result, the proposed method is more efficient than PHase Only Transform (PHOT) and other methods in literature. In addition, the proposed approach is capable of detecting defects of varying shapes, by combining 2-Dimensional WHT and 1-Dimensional WHT; and can detect defects in images with strong objects boundaries by utilizing a reference image. The proposed algorithm is robust over different image patterns and varying illumination conditions. Good results of the proposed methods are demonstrated on defect detection applications.

8661-10, Session 3

Improving the performance of interest point detectors with contrast stretching functions

Prabakar Gunashekhar, Bahadir Gunturk, Louisiana State Univ. (United States)

Interest point detection that is robust against variations in illumination, blur, compression, and geometric transformations is an important research problem in computer vision. In a recent paper, we proposed a method, called illumination robust feature extraction transform (IRFET), to improve the illumination robustness of Harris corner detector. In this paper, we demonstrate that IRFET improves not only the illumination robustness but also the robustness against blur, compression, and geometric transformations. The IRFET method is not limited to Harris corner detector; and the performance boost-up is demonstrated with multi-scale Harris and multi-scale Hessian interest point detectors in standard datasets.

8661-11, Session 3

Object detection using feature-based template matching

Simone Bianco, Marco Buzzelli, Raimondo Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

Pattern matching, also known as template matching, is a computationally intensive problem aimed to localize the instances of a given template within a query image.

In this work we present a fast exhaustive technique for template matching, able to use histogram-based similarity measures. In particular we will focus on Color Histograms (CH), Histograms of Oriented Gradients (HOG), and Bag of visual Words histograms (BOW). To obtain a BOW description of an image, we extract SIFT local image descriptors, and the resulting descriptors are vector quantized using a K-entry codebook of visual word prototypes.

In order to introduce spatial information among the sub-parts of the template, we have also integrated the histogram-based representation with the spatial pyramid representation, and other spatial representation within our template matching approach.

The proposed solution is compared with state of the art approaches in terms of both accuracy and execution time on different standard datasets. A detailed analysis of the results will be also provided to put in evidence the strengths of the proposed solutions with respect to the objects characteristics and scene complexity and variability.

8661-12, Session 4

Touch sensing analysis using multi-modal acquisition system

Jeffrey S. King, Zachi I. Baharav, Corning Incorporated (United States)

Touch sensing is quickly becoming a ubiquitous feature in consumer products, and in many practical implementations the user's finger physically interacts with a glass layer that protects the display and device. This layer has a significant impact on the touch sensing mechanism and performance, be it projected capacitive sensing or other methods. In this work, we present a system for the measurement and evaluation of various factors in the finger-glass interaction. This system enables synchronous capture of different modalities of the interaction, and post–processing allows correlation among them. We compare with available models of the finger-glass interactions, and describe both qualitative and quantitative valuable results. For example, high-resolution visible light images of the interface allow study of the effect of finger wetting and its related time constant on the performance of a touch screen. We describe the system and the design considerations involved, analyze the results, and discuss insights gained and implications to practical system design.

8661-13, Session 4

OrphanpPoint relocation and 3D super-resolution

Seungkyu Lee, Samsung Advanced Institute of Technology (Korea, Republic of)

Recently, consumer depth cameras have widely applied to robotics, computer vision and graphics applications. Furthermore sensor fusion with color camera has provided multimodal 3-dimensional data for many research problems. 3D point cloud data processing becomes more important to correctly deal with the depth image, however many conventional 2D image processing method have applied on depth images. Even though it is simple and effective in some limited condition to use such 2D image processing method, most of 3D applications such as 3D reconstruction, interaction shows many artifacts after 2D image processing application.

In this paper, we proposed two main 3D image processing methods for low resolution and low accuracy ToF depth images. First we relocate the orphan points back to either foreground or background rather than eliminating them as most of prior work do. Based on the proximity score calculated from calibrated color cameras and local depth point distance, we decide the 3D relocation direction. Secondly, we decide sub-pixel level foreground and background boundaries before we perform our superresolution. Extensive on various challenging 3D reconstruction experimental results at each step show that the proposed method shows significant improvements in dealing with 3D point cloud data.

8661-14, Session 4

Structural deformation measurement via efficient tensor polynomial calibrated electro-active glass targets

Christoph Gugg, Matthew J. Harker, Paul L. O'Leary, Montan Univ. Leoben (Austria)

This paper describes the physical setup and mathematical modelling of a device for the measurement of structural deformations over large scales, e.g., a mining shaft. Image processing techniques are used to determine the deformation by measuring the position of a target relative to a reference laser beam. It is necessary to compensate for optical distortion if high accuracy is to be achieved in a compact hardware design where lenses with short focal lengths are used. This paper presents a new coordinate mapping procedure based on a tensor product of discrete orthogonal polynomials. Both lens distortion and the projection are compensated by a single linear transformation. We have developed a new hierarchical approach based on a quad-tree subdivision to increase the measurement's performance. Cross-validation tests verify the validity, demonstrating that the proposed method accurately models both the optical distortion as well as the projection. The achievable accuracy is e < +/-0.01 [mm] in a field of view of 150 [mm] x 150 [mm] at a distance of the laser source of 120 [m].

8661-15, Session 4

Machine vision system for the control of tunnel boring machines

Michael Habacher, Matthew J. Harker, Paul L. O'Leary, Montan Univ. Leoben (Austria)

This paper presents a machine vision system of the control of Dual-Shield Tunnel Boring Machines (TBM). A robust system has been implemented both from a mechanical as also from a computing point of view. This is necessary to guarantee reliable functionality in the extremely harsh environment involved in tunneling. A redundant configuration with four retro-reflector targets has been chosen. A full measurement can still be preformed with the loss of one target, and a reduced measurement (a Euclidean transformation) can be achieved even with the loss of two targets. A contouring algorithm detects the edges of the targets, each contour is then segmented into N overlapping sub-segments, to which N circles are fit. The circles fits are classified with the aim of identifying points on the contour which belong to the expected circle. This delivers a robust circle identification algorithm. A new SVD based implementation of a covariance weighted circle fitting algorithm is presented, the new method is efficient and exhibits lower fitting bias than previous methods when fitting circular arcs. During the service phase the exact position of the machine is measured before and after moving the gripper shield. In this manner two full sets of data are available for the positions of the targets and the seen-positions of the targets. This enables a complete 3D camera calibration. The system was tested under laboratory conditions using an automatic positioning table. A resolution of 0.3 mm was achieved for extensions in the range of 2500 [mm] to 4200 [mm].

8661-16, Session 5

Eliminating illumination effects by discrete cosine transform (DCT) coefficients' attenuation and accentuation

Shan Du, Mohamed Shehata, Wael Badawy, Choudhury A Rahman, Intelliview Technologies Inc. (Canada)

Illumination variation is one of the most significant factors affecting the performance of face recognition and has received much attention in recent years. In this paper, a face image pre-processing approach was proposed that deals with the illumination problem to make face recognition robust to illumination variations. This method eliminates the illumination effects and extracts illumination insensitive features at the same time by attenuating low-frequency DCT coefficients and accentuating high-frequency DCT coefficients in logarithm domain. This is because illumination variations mainly concentrate on the low-frequency band and under poor illuminations, the high-frequency features become more important in recognition. To evaluate the effectiveness of the proposed method, face images with large illumination variations are used. Yale B, Extended Yale B and CMU-PIE databases are selected for evaluation. This method does not require modeling and model fitting steps. It can be directly applied to single face image, without any prior information of 3D shape or light sources. By experiments, we can see that the proposed method outperforms other existing methods.

8661-17, Session 5

Non-rigid ultrasound image registration using generalized relaxation labeling process

Jong-Ha Lee, Yeong Kyeong Seong, Moon Ho Park, Kyoung-Gu Woo, Samsung Advanced Institute of Technology (Korea, Republic of); Jeonghun Ku, Hee-Jun Park, Keimyung University (Korea, Republic of)

This research proposes a novel non-rigid registration method for ultrasound images. The most predominant anatomical features in ultrasound images. The most predominant anatomical features in medical images are tissue boundaries, which appear as edges. In ultrasound images, however, other features can be identified as well due to the specular reflections that appear as bright lines superimposed on the ideal edge location. In this work, an image's local phase information (via the frequency domain) is used to find the ideal edge location. The generalized relaxation labeling process is then formulated to align the feature points extracted from the ideal edge location. In this work, the original relaxation labeling method was generalized by taking "n" compatibility coefficient values to improve non-rigid registration performance. This contextual information combined with a relaxation labeling process is used to search for a correspondence. Then the transformation is calculated by the thin plate spline (TPS) model. These two processes are iterated until the optimal correspondence and transformation are found. We have tested our proposed method and the state-of-the-art algorithms with synthetic data and bladder ultrasound images of in vivo human subjects. Experiments show that the proposed method improves registration performance significantly, as compared to other state-of-the-art non-rigid registration algorithms.

8661-18, Session 5

Mammogram CAD, hybrid registration, and iconic analysis

Arnaud Boucher, Florence Cloppet, Nicole Vincent, Univ. Paris Descartes (France)

This paper aims to develop a computer-aided diagnosis (CAD) comprising two steps: a methodology to register and analyze pairs of temporal mammograms. The concept of "file", including all the past information on a patient, enables joint analysis of different acquisitions taken at different times, and the detection of significant modifications. The registration developed method is designed to superimpose at best the different breast anatomical structures. The registration is adapted in order to make disappear deformation undergone by the acquisition process. In order to reach this goal, a referent image is computed from control points based on anatomical features that are extracted automatically. Then the second image of the couple is realigned on the referent image, using a coarse-to-fine approach based on two 1D linear transforms, that allows the definition of a complex deformation field, which simulates mammographic like deformations. The joint analysis detects the evolution between two images representing the same scene. In order to achieve this, it is important to take into account the registration error limits to adapt the observation scale. The approach used in this paper is based on an image sparse representation. Decomposed in regular patterns, the images are analyzed under a new angle. The CAD is evaluated using recall and precision of differences in mammograms.

8661-19, Session 6

Neutron imaging for geothermal energy systems

Philip R. Bingham, Larry Anovitz, Yarom Polsky, Oak Ridge National Lab. (United States); Luc Dessieux, The Univ. of Tennessee (United States); Hassina Bilheux, Oak Ridge National Lab. (United States)

Geothermal energy systems transfer heat from the earth through water or other fluids and convert the heat to energy through a generator system. These systems are successfully being used for energy production in regions with natural hydrothermal circulation. In an effort to increase potential energy production for hydrothermal, an effort is underway to develop enhanced geothermal energy systems (EGS) that would allow energy production from hot dry rock areas. Development of EGS capabilities requires characterization of fluid flow in fractured rock at the high pressures and temperatures experienced in hot dry rock regions. Due to the high cross section of Hydrogen for neutrons, neutron imaging provides high contrast between water and rock. The authors have performed initial experiments to quantify fluid flow through rock cores and have developed an environmental chamber for neutron imaging of cores with fluid flow at pressures experience

in nature. This paper will present the imaging system, environmental chamber, neutron imaging data, and quantiative results calculated from neutron radiographs.

8661-20, Session 6

Wave front distortion based fluid flow imaging

Emishaw D. Iffa, Wolfgang Heidrich, The Univ. of British Columbia (Canada)

In this paper, a transparent flow surface reconstruction based on wave front distortion is investigated. A camera lens is used to focus the image formed by the micro-lens array to the camera imaging plane. The irradiance of the captured image is transformed to frequency spectrum and then the x and y spatial components are separated. A rigid spatial translation followed by low pass filtering yields a single frequency component of the image intensity. Index of refraction is estimated from the inverse Fourier transform of the spatial frequency spectrum of the irradiance. The proposed method is evaluated with synthetic data of a randomly generated index of refraction value and used to visualize a fuel injection volumetric data.

8661-21, Session 6

Autonomous ship classification using synthetic and real color images

Deniz Kumlu, B. Keith Jenkins, The Univ. of Southern California (United States)

Autonomous ship recognition is an active area for military and commercial applications. This work classifies color images of ships attained using cameras mounted on ships and in harbors. Our data-sets contain 9 different types of ship with 18 different perspectives for our training set, development set and testing set. The training data-set contains modeled synthetic images; development and testing data-sets contain real images. The database of real images was gathered from the internet, and 3D models for synthetic images were imported from Google 3D Warehouse. A key goal in this work is to use synthetic images to increase overall classification accuracy. We present a novel approach for autonomous segmentation and feature extraction for this problem. Support vector machine is used for multiclass classification. This work reports three experimental results for multiclass ship classification problem. First experiment trains on a synthetic image data-set and test on a real image data-set, and obtained accuracy is 87.8%. Second experiment trains on a real image data-set and tests on a separate real image data-set, and obtained accuracy is 87.8%. Last experiment trains on real + synthetic image data-sets (combined data-set) and tests on a separate real image data-set, and obtained accuracy is 93.3%.

8661-22, Session 6

Fast and flexible 3D object recognition solutions for machine vision applications

Ira Effenberger, Jens U. Kühnle, Fraunhofer-Institut für Produktionstechnik und Automatisierung (Germany); Alexander Verl, Fraunhofer-IPA (Germany)

In automation and handling engineering, the feeding of work pieces between the different stages of the production process chain is of special interest. Often the parts are stored unordered in bins or lattice boxes and hence have to be separated and ordered for feeding purposes. This can be done with mechanics specifically adapted to the geometric properties of the parts. A space-saving alternative is using a robot to grip the work pieces out of a bin or from a belt. For this kind of applications reliable and precise object recognition and localization systems are needed. In this paper we present fast and flexible 3D object recognition algorithms for handling applications. Based on 3D best-fit algorithms for geometric primitives and pre-

processing methods for segmentation of the 3D measure data a fast object localization has been developped which can be easily adapted to modified or even new parts. The object recognition has been tested on different real industrial parts and the results will be shown and summarized.

8661-23, Session 7

Low complexity smile detection technique for mobile devices

Valeria Tomaselli, Mirko Guarnera, Claudio Marchisio, Simone Moro, STMicroelectronics (Italy)

A low complexity smile detection technique, able to detect smiles in a variety of light conditions, face positions and image resolutions is proposed.

The approach firstly runs face detection and, consequently, mouth detection. The mouth detection is a simple application of geometrical proportions between face and mouth, which have been statistically estimated. The mouths are then described by a combination of two descriptors. The first one is the Pyramid Histogram of Orientation Gradients (PHOG), calculated on a limited number of orientation bins. The edge orientations are computed at each pixel of the mouth region and not on edges only to be invariant to illumination changes. The second descriptor is the Pyramid Local Binary Pattern (PLBP). PLBP is the concatenation of Local Binary Pattern histograms at different pyramid levels, and it is a powerful mean of texture description.

For classification we use a minimum distance classifier from the centroids of the smiling and non-smiling classes. In particular, Mahalanobis distance has been chosen, because it takes into account the correlations between features. This technique produces the 94% of area under the ROC (AUROC) curve. The proposed approach obtains better results than other more complex techniques, in terms of hit rate and AUROC.

8661-24, Session 7

Density-induced oversampling for highly imbalanced datasets

Daniel Fecker, Volker Märgner, Tim Fingscheidt, Technische Univ. Braunschweig (Germany)

The problem of highly imbalanced datasets with only sparse data of the minority class in the context of two-class classification is investigated. The examined application in this paper is optical on-line process monitoring of laser brazing with only rare sporadic occurring defects.

A novel synthetic data oversampling technique is proposed which utilizes estimations of the probability density distribution in the feature space. First, a Gaussian mixture model (GMM) from the data of the well-sampled majority class is generated and with its help a new GMM is approximated with the sparse minority class data by Bayesian adaptation. Random synthetic data is generated from the adapted GMM and an additional assignment rule assigns this data to either the minority class or else discards it. The obtained synthetic data is employed in combination with the available original data to train a support vector machine classifier.

Experiments with different amounts of minority class data samples and comparisons to other methods show that this approach performs very well for highly imbalanced datasets.

8661-25, Session 7

Coherent image layout using an adaptive visual vocabulary

Scott E. Dillard, Google (United States); Michael J. Henry, Shawn Joel Bohn, Luke J. Gosink, Pacific Northwest National Lab. (United States)

When querying a huge image database containing millions of images, the result of the query may still contain many thousands of images that need to be presented to the user. We consider the problem of arranging such a large set of images into a visually coherent layout, one that places similar images next to each other. Image similarity is determined using a bag-of-features model, and the layout is constructed from a hierarchical clustering of the image set by mapping an in-order traversal of the hierarchy tree into a space-filling curve. This layout method provides strong locality guarantees so we are able to quantitatively evaluate performance using standard image retrieval benchmarks. Performance of the bag-of-features method is best when the vocabulary is learned on the image set being clustered. Because learning a large, discriminative vocabulary is a computationally demanding task, we present a novel method for efficiently adapting a generic visual vocabulary to a particular dataset. We evaluate our clustering and vocabulary adaptation methods on a variety of image datasets and show that adapting a generic vocabulary to a particular set of images improves performance on both hierarchical clustering and image retrieval tasks.

8661-26, Session 7

Shape recognition for capacitive touch display

Ivana Guarneri, Alessandro Capra, STMicroelectronics (Italy); Giovanni M. Farinella, Sebastiano Battiato, Univ. degli Studi di Catania (Italy)

With the growing diffusion of touch screen based consumer devices the development of algorithms able to discriminate among the different shapes obtained by touching the device display becomes very important. For instance the detection and recognition of the fingers represents fundamental information in many touch based user applications. These algorithms are also extremely useful to recognize accidental touches in order to avoid involuntary touch functionalities.

Taking into account the aforementioned motivations, in this paper we present a technique able to classify the five most common classes of shapes acquired with a capacitive touch display: finger, ear, cheek, half ear half cheek and hand hold. The main steps of the proposed algorithm are summarized in the following. Firstly a noise removal filter is applied to each acquired capacitance image to eliminate noisy spikes. Secondly, the features used for shape classification are extracted. Finally, through a based decisional tree (C4.5) the final classification is performed.

The technique has been properly tested on a dataset of 1500 capacitive maps for each of the five classes. The dataset has been populated by involving both female and male individuals, presenting different anatomies and the recognition accuracy achieved by the proposed technique is 91.12%.

8661-27, Session IPI

An elliptic phase-shift algorithm for high speed three-dimensional profilometry

Fuqin Deng, Univ. of Hong Kong (Hong Kong, China); Zhao Li, Istituto Italiano di Tecnologia (Italy); Jia Chen, Harbin Institute of Technology (China); Jiangwen Deng, Shunming K. Fung, ASM Pacific Technology Ltd. (China); Edmund Y. Lam, The Univ. of Hong Kong (Hong Kong, China)

A high throughput is often required in many machine vision systems especially on the assembly line in the semiconductor industry. To develop a non-contact three-dimensional dense surface reconstruction system for real-time surface inspection and metrology applications, in this work, we project sinusoidal patterns onto the inspected objects and propose a high speed phase-shifting algorithm for dense surface profilometry. First, we use an illumination-reflectivity-focus (IRF) model to characterize the factors on forming the intensities in phase-measuring profilometry. Second, by visualizing and analyzing the characteristic intensity locus projected onto the intensity space, we build a two-dimensional phase map to store the phase information for

each point in the intensity space. Third, we develop an efficient elliptic phase-shifting algorithm (E-PSA) for high-speed surface profilometry. In this novel method, instead of calculating the time-consuming inverse trigonometric functions, we only need to normalize the measured intensity data and index the built two-dimensional phase map beforehand during real-time phase reconstruction. Finally, experimental results show that it is more than two times faster than conventional phase-shifting algorithm.

8661-29, Session IPI

An incompressible fluid flow model with mutual information for MR image registration

Leo Tsai, Herng-Hua Chang, National Taiwan Univ. (Taiwan)

Image registration is one of the fundamental and essential tasks within image processing. It is a process of determining the correspondence between structures in two images, which are called the template and reference images. The challenge of registration is to find an optimal geometric transformation between corresponding image data. This paper develops a new MR image registration algorithm that uses a closed incompressible viscous fluid model associated with mutual information. In our approach, we treat the image pixels as the fluid elements of a viscous fluid flow governed by the nonlinear Navier-Stokes partial differential equation (PDE). We replace the pressure term with the body force mainly used to guide the transformation with a weighting coefficient, which is expressed by the mutual information between the template and reference images. To solve this modified Navier-Stokes PDE, we adopted the fast numerical techniques proposed by Seibold [MIT, 2008]. The registration process of updating the body force, the velocity and deformation fields is repeated until the mutual information reaches a prescribed threshold. We applied our approach to BrainWeb and real MR images. As consistent with the theory of the proposed fluid model, we found that our method accurately transformed the template images into the reference images based on the intensity flow. Experimental results indicate that our method is of potential in a wide variety of medical image registration applications.

8661-31, Session IPI

Improved skin detection method by iteratively eliminating pseudo-skin colors through combined skin filter

Ohyeol Kwon, Kyungpook National Univ. (Korea, Republic of); Kyung-Ah Kim, Kyungpook National University (Korea, Republic of); Sung-Il Chien, Kyungpook National Univ. (Korea, Republic of)

In order to detect more reliably the skin color from pseudo-skin colors which are similar to the skin color, we propose an improved skin detection method, which can detect by iteratively eliminating pseudo-skin colors through a combination of the skin color detection filters. First, we classify images into the normal and the abnormal skin images depending on whether the pseudo-skin criterion exceeds a threshold value or not. In case of normal skin images, we determine a final skin color region by merging two regions detected by the two skin filters. As for the abnormal skin images, we detect the skin color by using two combined filters. Then, the detected skin color pixels from each combined filter are translated into Cheddad's feature respectively and we can create a histogram of each detected skin region. The skin color is detected by eliminating the pseudo-skin colors through Otsu's histogram binarization. Finally, we determine a final skin color region by merging two skin regions detected by applying the histogram binarization. Experimental results show that our proposed method can detect the skin color from the normal and the abnormal skin images more accurately than that of the conventional method.

8661-32, Session IPI

A modified hierarchical graph cut based video segmentation approach for high frame rate video

Xuezhang Hua , Nanjing University of Post and Telecommunications (China); Sumit Chakravarty, New York Institute of Technology (United States); Qi She, Boyu Wang, Nanjing Univ. of Posts and Telecommunications (China)

Video object segmentation entails selecting and extracting objects of interest from a video sequence. Video Segmentation of Objects (VSO) is a critical task which has many applications, such as video edit, video decomposition and object recognition. The core of VSO system consists of two major problems of computer vision, namely object segmentation and object tracking. These two difficulties need to be solved in tandem in an efficient manner to handle variations in shape deformation, appearance alteration and background clutter. Along with segmentation efficiency computational expense is also a critical parameter for algorithm development. Most existing methods utilize advanced tracking algorithms such as mean shift and particle filter, applied together with object segmentation schemes like Level sets or graph methods. As video is a spatiotemporal data, it gives an extensive opportunity to focus on the regions of high spatiotemporal variation. We propose a new algorithm to concentrate on the high variations of the video data and use modified hierarchical processing to capture the spatiotemporal variation.

8661-33, Session IPI

Power and performance tradeoffs of GPGPU vs. multicore: a case study for stereo matching

Sarala Arunagiri, Jaime Jaloma, Ricardo Portillo, Patricia J. Teller, The Univ. of Texas at El Paso (United States)

GPGPUs and multicore processors have become commonplace with their wide usage in traditional high-performance computing systems as well as mobile-computing devices. A significant speedup can be achieved for a variety of general-purpose applications by employing the parallelism made available by these technologies. Unfortunately, this speedup is often accompanied by high power and/or energy consumption. Thus, energy conservation is increasingly becoming a major concern in designing these computing devices. The main driver for energy efficiency in computer systems is the cost of power and cooling, as well as the environmental impact of ubiquitous PCs and massive data centers. On the other hand, for mobile devices energy conservation is driven by the need to extend battery life and power capping is mandated by the restrictive power budget of mobile platforms such as unmanned aerial vehicles (UAVs). Our focus is to understand the performance tradeoffs involved in executing Army applications on mobile computing platforms. In this paper we explore the tradeoffs involved in GPGPU vs. multicore computing in terms of execution time, output quality, and power and energy consumption for stereo matching, which is a heavily-investigated compute-intensive candidate Army application for mobile computing platforms.

8661-34, Session IPI

An efficient algorithm for food quality control based on multispectral signatures

Juan Carlos Valdiviezo-Navarro, Carina Toxqui-Quitl, José Francisco Solís-Villareal, Alfonso Padilla-Vivanco, Univ. Politécnica de Tulancingo (Mexico)

Multispectral imaging has motivated new applications related to food quality monitoring due to its capability of analysis based on spectral signatures. This manuscript introduces a fast and efficient algorithm that is used in combination with a multispectral system for the unsupervised determination of food quality. Our spectral classification algorithm is based on the min-W and the max-M lattice associative memories to determine the centers of corresponding clusters. Given two classes of food, previously identified as high and low quality products, we first register a multispectral image from them and perform a dimensionality reduction by taking into account those spectral bands that involve their most significant absorption and reflection spectra. From the reduced set of images, the min-W and the max-M lattice associative memories are computed and a subset of their column vectors are used as the most representative spectral signatures for each class, which will be used for later classification. The classification process will be accomplished by means of the spectral angle computed between each representative spectra and any other pixel in the image, and whose results allow us to classify regions in the image with similar spectral responses. The technique has been applied for the unsupervised classification of low and high quality coffee grains. The achieved results state that our procedure is efficient for food quality control and inspection.

8661-35, Session IPI

Bottle inspector based on machine vision

Carina Toxqui-Quitl, Juan A. Cardenas-Franco, Alfonso Padilla-Vivanco, Juan Carlos Valdiviezo-Navarro, Univ. Politécnica de Tulancingo (Mexico)

A PET bottle inspector for quality control based on machine vision is presented. The optical setup for imaging requires five cameras, an illumination system and a software interface. To achieve a complete inspection of the bottle, it is carried out in three modules. With the above, it inspects the finish, wall and base of the PET bottle for the detection of defects such as tears, cracks, or stains. In order to capture the image of the finish, the system utilizes direct light illumination with polarizers to avoid reflexions and two cameras from different angles for a whole scanning. A transmission illumination is used for imaging the base and wall of the bottle. The inspection of the wall and finish of the bottle requires a specific algorithm to search cracks and tears. While the inspection of the base finish detects dirt particles and evaluate the size of the base. Therefore, different algorithms are required. The algorithm for the finish and wall inspection detects checks and cracks. The method is based on the discrete wavelet transform (DWT). A multiresolution analysis allows detect edges and fine details. And the small defects connected with edges usually need higher resolution to display. Also, the DWT filtering using a low pass filter and a high pass filter, this allows remove background information. On the other hand, the algorithm for the base inspection that detects dirt particles or sticky objects is based on a pixel or region variation. The size of the bottle base is calculated using geometric moments.

From this, three proposed defect maps are using to train a neuronal network for classification of bottles. This method classifies PET bottles into two classes, with and without fault.

Experimental results show that the bottle inspector achieves a correct classification rate of 78 %, 90% and 98% when the system inspects finish, wall and base respectively. The proposed method increases considerably the speed of detection.

8661-36, Session IPI

Static sign language recognition using neural networks

José Francisco Solís-Villareal, Oscar Morales-Álvarez, Carina Toxqui-Quitl, Juan Carlos Valdiviezo-Navarro, Univ. Politécnica de Tulancingo (Mexico)

In sign language recognition process, segmentation is one of the most challenging issues, many works based on computer vision system, use special clothes, tinted gloves or uniform background for leading to this matter. For this work an infrared sensor was used for discard the background just as in previous report [Solís, et al., 2012], where it was use binary images to extract one-dimensional descriptors, however, to recognize the entire alphabet is necessary to extract information of the visible spectrum, so a two cameras configuration was chosen for this work, one for segmenting and the other for sign data extraction. A

database was created in the Laboratory of Optics and Vision Systems of the University, 23 static signs were picked-up from American Sign Language, other signs were discarded because they are expressed with movement, different versions of same person signs were stored. The purpose of this work is to achieve static American Sign Language Recognition with a computer vision based system, without the use of tinted gloves, special clothing or a particular background color; the main disadvantage is to avoid passing the hand in front of the face. The image preprocessing of the database begins with scale and translation transforms to get the hand intensity data in visible spectrum segmented by an infrared sensor, this allows the system to compute filters in spatial and frequency domains, at this point, wavelet transform was selected to improve accurate data length reduction in order to process less information in an artificial neural network, the multilayer perceptron with sigmoidal neurons achieve a good classification performance with k-fold cross-validation.

8661-37, Session IPI

Defect inspection technology for a gloss-coated surface using patterned illumination

Tsuyoshi Nagato, Takashi Fuse, Tetsuo Koezuka, Fujitsu Labs., Ltd. (Japan)

In this paper, we discuss the development of an inspection system for a gloss-coated surface using patterned illumination. Since an irregular defect on a UV-curable coating surface is transparent, it is difficult to observe it in typical illumination. Thus, by considering the nature of the specular reflection on the gloss-coated surface, we developed the inspection system which is characterized by patterned illumination and phase-shifting method. This system takes multiple specular reflection images of the stripe pattern generated by patterned illumination, and calculates the phase and modulation of brightness per pixel according to phase-shifting method. The phase and modulation images can be regarded as the distribution of specular reflection angle and intensity, respectively. Our system can obtain the distribution of specular reflection angle and intensity. Hence, it can observe shapes of the transparent irregular defect and spots such as fingerprints on the gloss-coating surface. This system can allow the inspection of the gloss-coated surface which was difficult in a conventional method. Additionally, it can be also applied to a mirror surface and a transparent object.

8661-38, Session IPI

HOG and CS-LBP based detector for surveillance of birds

Mihreteab A. Kidane, Thang Q. Nguyen, Nagaoka Univ. of Technology (Japan)

The gathering of large population of detrimental birds such as crows causes noise and environmental pollution to the vicinity areas. Hence, a robust image processing technique capable of detecting and localizing birds in images plays an important role in many computer vision applications. In this paper, a feature set detector based on Histogram of Oriented Gradients (HOG) and Center-Symmetric Local Binary Pattern (CS-LBP) for crow birds is proposed. HOG is robust in capturing local shape information or edge directions and CS-LBP descriptor captures gradient based information and texture information, thus, the combination of these two features is capable of capturing better bird features. By combining HOG and CS-LBP as a feature set, detection of crows under different lighting conditions could be carried out. A data set of crow birds with a wide range of poses and backgrounds was prepared and learned using Support Vector Machine (SVM). Experimental results on various test images under different lighting conditions show that HOG and CS-LBP based detector achieved a detection result of 91% at 0.02 FPPI and on average a detection result of 87% at 0.01 to 0.04 FPPI. The detection performance achieved using our HOG and CS-LBP based detector outperforms some of the state of the art algorithms.

8661-39, Session IPI

A semi-automatic annotation tool for cooking video

Gianluigi Ciocca, Raimondo Schettini, Simone Bianco, Paolo Napoletano, Univ. degli Studi di Milano-Bicocca (Italy); Roberto Margherita, Gianluca Marini, Giorgio Gianforme, Giuseppe Pantaleo, AlmavivA S.p.A. (Italy)

In order to create a cooking assistant application to guide the users in the preparation of the dishes relevant to their profile diets and food preferences, it is necessary to accurately annotate the video recipes, identifying and tracking the foods and the actions of the cook. These videos present particular annotation challenges such as frequent occlusions, food appearance changes, and high variability of fine-grained actions.

Manually annotate the videos is a time-consuming, tedious and error-prone task. Fully automatic tools that integrate computer vision algorithms to extract and identify the elements of interest are not error free, and false positive and false negative detections need to be corrected in a post-processing stage.

We present an interactive, semi-automatic tool for the annotation of cooking videos that integrates computer vision techniques under the supervision of the user. The annotation accuracy is increased with respect to completely automatic tools and the human effort is reduced with respect to completely manual ones.

The performance and usability of the proposed tool are compared with other annotation tools on the basis of the time and effort required to annotate the same video sequences. The accuracy of the computer vision modules are also reported and discussed.

8661-40, Session IPI

Intensity and color descriptors for texture classification

Claudio Cusano, Univ. degli Studi di Milano Bicocca (Italy); Paolo Napoletano, Raimondo Schettini, Univ. degli Studi di Milano-Bicocca (Italy)

In this paper we present a novel texture descriptor for color texture analysis specially designed to deal with changes in illumination. The descriptor, that we called Intensity Color Contrast Descriptor (ICCD), is based on a combination of the LBP approach with a measure of color contrast defined as the angle between two color vectors in an orthonormal color space. The ICCD descriptor is invariant to several transformations of the color space. Its robustness with respect to global changes in lighting conditions has been experimentally demonstrated by comparing it on standard data sets against several other color texture descriptors in the state of the art.

Conference 8662: Intelligent Robots and Computer Vision XXX: Algorithms and Techniques

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8662 Intelligent Robots and Computer Vision XXX: Algorithms and Techniques

8662-1, Session 1

Control issues and recent solutions for voltage controlled piezoelectric elements utilizing artificial neural networks *(Invited Paper)*

Marko Kauppinen, Juha Röning, Univ. of Oulu (Finland)

Performing actuation in nanomanipulation at the necessary accuracy is largely possible thanks to the many new piezoelectric actuation systems. However, there are still room for accuracy and speed improvements with the aid of better piezoelectric element models and better sensors.

In this work, we will cover some of the control issues related especially to piezoelectric actuation in nanomanipulation tasks. We will also take a look at some of the recent improvements made possible by methods utilizing artificial neural networks for improving the generalization capability and the accuracy of piezoelectric hysteresis models used in inverse modelling and control of the solid-state voltage controlled piezoelectric actuators.

We will also briefly discuss the problem areas that the piezoelectric control method research should be especially focused on and the biggest problem areas for many of the existing methods. In addition, some of the common issues related to testing and result representations are discussed.

8662-2, Session 1

The 20th annual intelligent ground vehicle competition: building a generation of robotists *(Invited Paper)*

Bernard L. Theisen, Andrew D. Kosinski, 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 (AUVSI). 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 20 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 80 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.

8662-3, Session 2

Visual homing with a pan-tilt based stereo camera

Paramesh Nirmal, Damian Lyons, Fordham Univ. (United States)

Visual homing is a navigation method based on comparing a stored image of the goal location and the current image (current view). It is theorized that insects, such as ants and bees, employ visual homing methods to return to their nest [1]. Visual homing has been applied to autonomous robot platforms using two main approaches: holistic and feature-based. Both methods aim at determining distance and direction to the goal location. Navigational algorithms using Scale Invariant Feature Transforms (SIFT) have gained great popularity in the recent years due to the robustness of the feature operator. Churchill and Vardy [2] have developed a visual homing method using scale change information (Homing in Scale Space, HiSS) from SIFT.

HiSS uses scale change information from SIFT features to determine distance between the robot and the goal location. Since the scale component is discrete with a small number of elements [3], the result is a rough measurement with limited accuracy. We have developed a method that uses stereo data, resulting in better homing performance. Our approach utilizes a pan-tilt based stereo camera, which is used to build wide-field images. We use the wide-field images combined with stereo-data obtained from the stereo camera to extend the keypoint vector described in [3] to include a new parameter, depth (z). Using this info, our algorithm determines the distance and orientation from the robot to the goal location.

We compare our method with HiSS in a set of indoor trials using a Pioneer 3-AT robot equipped with a BumbleBee2 stereo camera. We evaluate the performance of both methods using a set of performance measures described in this paper.

8662-4, Session 2

Motion control for intelligent ground vehicles based on the selection of paths using fuzzy inference

Shiwei Wang, Taskin Padir, Worcester Polytechnic Institute (United States)

This paper describes a motion planning technique for intelligent ground vehicles using a path selection algorithm based on fuzzy inference. The approach extends on the tentacle motion planning algorithm. Tentacles which are speed-dependent sets of virtual drivable paths represent a set of pre-calculated trajectories in the coordinate frame of the vehicle located at its center of gravity. In our approach, we utilize a fuzzy inference based selection to decide on the path to be followed by the autonomous vehicle. We formulate the mapping between the inputs and outputs using fuzzy logic without a mathematical model. The paper illustrates results from both a simulation carried out using MATLAB's Fuzzy Logic Toolbox as well as field tests performed on a Clearpath Husky A100 ground vehicle which is programmed using the Robot Operating System (ROS). The simulation takes into account the road model including obstacles and lanes, the calculation of speed-dependent sets of tentacles and implements the fuzzy inference based selection mechanism and generates a planned path in an obstacle course for the vehicle to follow. The practical implementation of the fuzzy inference rule on Husky A100, a 6-wheeled skid-steered ground vehicle operated in an off-road GPS-denied environment covered with trees, relies on a 2D mapping of the robot's environment using

a scanning laser range finder. The paper provide a discussion of the measurement methods and procedures used in this research.

8662-5, Session 2

Panoramic stereo sphere vision

Weijia Feng, Tianjin Univ. (China); Baofeng Zhang, Tianjin University of Technology (China); Juha Röning, Univ. of Oulu (Finland); Xiaoning Zong, YI Tian, Tianjin University of Technology (China)

Conventional stereo vision systems have a small field of view (FOV) which limits their usefulness for certain applications. While panorama vision is able to "see" in all directions of the observation space, scene depth information is missed because of the mapping from 3D reference coordinates to 2D panoramic image. In this paper, we present an innovative vision system which builds by a special combined fish-eye lenses module, and is capable of producing 3D coordinate information from the whole global observation space and acquiring no blind area 360??360? panoramic image simultaneously just using single vision equipment with one time static shooting. It is called Panoramic Stereo Sphere Vision (PSSV). We proposed the geometric model, mathematic model and parameters calibration method. Specifically, video surveillance, robotic autonomous navigation, virtual reality, driving assistance, multiple maneuvering target tracking, automatic mapping of environments and attitude estimation are some of the applications which will benefit from PSSV.

8662-6, Session 2

Stabilization and control of quad-rotor helicopter using a smartphone device

Alok Desai, Dah Jye Lee, Jason Moore, Yung-Ping Chang, Brigham Young Univ. (United States)

In recent years, autonomous micro unmanned aerial vehicles (micro-UAVs), or more specifically hovering micro-UAVs have proved to be suitable for many promising applications such as exploring of unknown environments and search and rescue operations. They have become an important tool in the field of indoor as well as outdoor UAV research. The early versions of UAVs had no on-board control capabilities, and were difficult for manual control from a ground station. Many UAVs now are equipped with on-board control systems that reduce the amount of control required from the ground-station operator. However, the limitations on payload and power consumption and control without human interference remain the biggest challenges.

This paper proposes to use a smartphone as the sole computational device to stabilize and control a quad–rotor. The goal is to use the readily available sensors in a smartphone such as GPS, accelerometer, rate-gyros, and camera to assist in some vision-related tasks such as flight stabilization, estimation of height above ground, target tracking, obstacle detection, and surveillance. We use a quad-rotor platform that has been built in the Robotic Vision Lab at Brigham Young University for our development and experiments. An Android smartphone is connected through the USB port to an external hardware that has a microprocessor and circuitries to generate pulse width modulation signals to control the brushless servomotors on the quad-rotor. The high-resolution camera on the smartphone is used to detect and track features to maintain a desired altitude level. The vision algorithms implemented include template matching, Harris feature detector, RANSAC similarity-constrained homography, color segmentation. Other sensors are used to control yaw, pitch, and roll of the quad-rotor. This smartphone-based system is able to stabilize and control and is ideal for micro-UAVs that have size, weight, and power limitations.

8662-7, Session 2

Loop closure detection using local Zernike moment patterns

Evangelos Sariyanidi, Istanbul Teknik Üniv. (Turkey); Onur Sencan, Istanbul Teknik Üniversitesi (Turkey); Hakan Temelta, Istanbul Teknik Üniv. (Turkey)

Loop closure, as very important problem in mobile robotics, can be defined as the correct identification of a previously visited location. This paper presents a new loop closure detection technique, which relies on visual landmark detection. The landmarks are detected by combining very small visual features, called Local Zernike Moment (LZM) patterns. The LZM transformation, which is used to extract the LZM patterns, has proven to be a useful representation for certain challenging recognition tasks like face recognition. In this paper, the LZM transformation is used to extract certain local features called LZM patterns, which are extracted by learning the appearance probability of the LZM features on an offline dataset. Once the appearance probability of the patterns is learnt, the next step is to find a rare combination of the extracted LZM patterns and use it as a visual landmark. In order to find a unique feature combination, a landmark extraction algorithm that has been introduced in a previous study is utilized. The loop closure detection framework that is being proposed is computationally efficient, to the point that it can operate near real-time. The overall approach has been evaluated on two datasets, and the achieved results are quite promising.

8662-8, Session 3

Optimizing feature selection strategy for adaptive object identification in noisy environment

Sagar Pandya, The Univ. of Southern California (United States); Thomas T. Lu, Tien-Hsin Chao, Jet Propulsion Lab. (United States)

We present the development of a multi-stage automatic target recognition (MS-ATR) system for computer vision in robotics. This paper discusses our work in optimizing the feature selection, and feature extraction strategies of the MS-ATR system. We discuss a modified saliency algorithm that accepts the guidance of target features in locating regions of interest (ROI). By introducing target related input parameters, saliency becomes more focused and task oriented. It is used as an initial stage for the fast ROI detection method. We further present a novel method for optimizing feature selection by combining the principal component analysis (PCA), the Independent Component Analysis (ICA), the saliency and geometric algorithms to extract the feature vectors. Genetic algorithm and several optimizing methods are used to select a minimum set of features that best distinguishes the target from the background and false targets. The feature vectors are passed to the last stage for target identification and verification. The results of the MS-ATR system have shown increased detection rate and reduced false alarm rate in noisy sonar and video imagery.

8662-9, Session 3

GPU-based real-time trinocular stereo vision

Yuanbin Yao, RJ Linton, Taskin Padir, Worcester Polytechnic Institute (United States)

Stereo vision has been extensively used in autonomous ground vehicle navigation. Traditionally, most stereo vision applications are binocular which uses information from a horizontal 2-camera array to perform stereo matching and compute the depth image. Trinocular stereo-vision with a 3-camera array has been proved to provide higher accuracy in stereo matching which could benefit applications like distance finding, object recognition and detection. However, the presence of the third image increases the computational burden and hence not practical in many time critical applications like autonomous robot navigation. Due to the nature of its highly parallelized SIMD (Single Instruction Multiple Data) architecture, GPGPU (General Purpose GPU) computing can effectively be used to parallelize the large data processing and greatly accelerate the computation of algorithms used in trinocular stereo vision. So, the implementation of trinocular stereo vision on a GPGPU is an effective method for the development of stereo vision application.

This work focuses on designing and implementing a real-time trinocular stereo vision algorithm with GPU. The implementation makes use of

Conference 8662

Open Source Computer Vision Library (OpenCV) in C++ and NVidia CUDA GPGPU solution. A set of image processing algorithms were developed within this framework and a winner-take-all method is applied to perform fusion of disparities in different directions. The results are compared in accuracy and computation speed to verify the improvement.

8662-12, Session 3

Remotely controlling of mobile robots using gesture captured by the Kinect and recognized by machine learning method

Roy Chaoming Hsu, Jhih-Wei Jian, Chih-Chuan Lin, National Chiayi Univ. (Taiwan); Chien-Hung Lai, National Chiayi University (Taiwan); Cheng-Ting Liu, National Chiayi Univ. (Taiwan)

It is a big breakthrough to the game industry and in remotely playing of the game by using Kinect sensor of body sensation research, while remotely controlling of mobile robots is the most popular subject in the field of robotics. The main purpose of this paper is to use Kinect and its body sensation technology to design simple, convenient, and for humanity operation remote robot control system. In this study, we use Kinect sensor to capture the human body skeleton with depth information, to design gesture training and identification technology, and finally to command a mobile robot for certain actions. Experimental results show that the designed remotely controlling of mobile robots using gesture captured by the Kinect sensor can achieve, on an average, more than 96% of accurate identification of 7 types of gestures and can control a simulated e-puck robot for find the way out of a simulated maze.

8662-13, Session 4

Relating vanishing points to catadioptric camera calibration

Wenting Duan, Univ. of Lincoln (United Kingdom); Hui Zhang, United International College (China); Nigel Allinson, Univ. of Lincoln (United Kingdom)

This paper presents the analysis and derivation of the geometric relation between vanishing points and camera parameters of central catadioptric systems. These vanishing points correspond to the three mutually orthogonal directions of 3D real world coordinate system (i.e. X, Y and Z axes). Compared to vanishing points (VPs) in the perspective projection, the advantages of VPs under central catadioptric projection are that there are normally two vanishing points for each set of parallel lines, since lines are projected to conics in the catadioptric image plane. Also, their vanishing points are usually located inside the image frame. We show that knowledge of the VPs corresponding to XYZ axes from a single image can lead to simple derivation of both intrinsic and extrinsic parameters of the central catadioptric system. This derived novel theory is demonstrated and tested on both synthetic and real data with respect to noise sensitivity.

8662-14, Session 4

Natural image understanding using algorithm selection and high-level feedback

Martin Lukac, Michitaka Kameyama, Tohoku Univ. (Japan); Kosuke Hiura, Tohoku University (Japan)

Natural Image processing and understanding encompasses hundreds or even thousands of different algorithms. Each algorithm has a certain peak performance for a particular set of input features and configurations of the objects/regions of the input image (environment). To obtain the best possible result of processing, we propose an algorithm selection approach that permits to always use the most appropriate algorithm for the given input image. This is obtained by at first selecting an algorithm based on low level features such as color intensity, histograms, spectral coefficients. The resulting high level image description is then analyzed for logical inconsistencies (contradictions) that are then used to refine the selection of the processing elements. The selection stops when the high level inconsistencies are all resolved or no more different algorithms can be selected.

8662-15, Session 4

Improving shape context using geodesic information and reflection invariance

Vittal Premachandran, Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore)

Object recognition is usually performed by extracting features and then matching the features to test images. Color, texture and shape are important features of an object. Shape is a rich information that has not been well exploited. In this paper, a new way for extracting the shape information from a given object, is proposed. To describe the object's shape in a meaningful way, the object's contours are first extracted and sampled into a set of uniformly spaced points. Shape contexts are generated at each of these uniformly spaced points. A shape context is a 2D histogram of distances and angles. The proposed variant of the shape context makes use of the geodesic distance between two sampled points. The use of geodesic distance better captures the shape information than the traditional Euclidean distance. Experimental results on a standard database shows that the use of geodesic information helps in obtaining better object retrieval rates.

8662-16, Session 4

A proposed super-fast scheme for instant-detect-instant-kill of a ground-to-air missile

Chialun J. Hu, Southern Illinois Univ. at Carbondale (United States)

We can apply a slightly modified local polar edge detection (LPED) method the author used in the last 3 years to a binary IR picture captured by an IR camera operated at room temperature to get the central of mass point (CMP) of the high temperature, enemy surface to air missile (or a SAM) against the low temperature environment INSTANTLY when the SAM is fired. Then it will automatically track the CMP of the fast flying SAM by predicting ahead the location of the CMP in the next IR picture frame. When this is automatically confirmed in the next picture frame, the operator screen will flash "TARGET LOCKED" sign. The operator can then press a red button to fire either a laser gun or an "induction-ignited air-to-air (or ATA) missile" at the predicted rendezvous point of the enemy SAM and kill the SAM instantly. For laser gun operation, the rendezvous point is the predicted CMP of the next SAM position. For the induction-ignited ATA missile, the rendezvous point is the center of a influence sphere such that when the ATA flies into the influence sphere of the SAM, not necessarily hitting the SAM, the ATA will explode to shoot the enemy SAM down.

To enhance further the accuracy, we should use two helicopters to cross-fire cooperatively with helicopter A targeting at the SAM aiming at helicopter B and helicopter B targeting at the SAM aiming at helicopter A. This cross-firing strategy is used to improve the target cross section by aiming at the target from "side view" instead of from "head-on view".

8662-17, Session 4

Video surveillance with alerts

Robert J. Wahlstedt, Whitworth Univ. (United States)

VSWA is a program that integrates computer vision to gather non-verbal clues of suspicious behavior. This paper discusses social aspects that characterize that are inconsistent that pose a threat so an alert can go to an authority. These situations can range from a bank holdup to drug trafficking. This program uses body posture confidence,

walking pace, scanning for weapons in images, and facial expression, as well as scanning for threatening gestures. This program does not alert based on one variable alone but rather it has an index of factors that suggest suspicion or guilt. This paper discusses the algorithms and methodology that is in the program. This program has the goal of making transit, libraries, schools, and banks safer and less invasive.

8662-18, Session 4

Finger tracking for hand-held device interface using profile-matching stereo vision

Yung-Ping Chang, Dah Jye Lee, Jason Moore, Alok Desai, Beau J. Tippetts, Brigham Young Univ. (United States)

Hand-held communication or computing devices have become an important part of human life. A large percentage of populations use them regularly. Although convenient for the user to dial a phone number of select music to play, they cause safety problems, especially when using these devices while driving. As the intended design, the user has to use fingers to touch the screen to operate them. These touch-screen operations distract drivers and put them in danger or make them dangerous to other drivers or pedestrians. A non-contact control interface will allow the user to control the device without touching the screen and to focus more on driving. The challenges of building such interface are how to localize and track fingers in real time.

We recently developed a new stereovision algorithm called Intensity Profile Shape-Matching for 3-D human gesture analysis and obstacle detection. This algorithm is able to extract 3-D information from a pair of intensity profiles from stereo images in real time. It matches intensity profile shape row by row, and produces a disparity map for analysis. By choosing a certain interval of disparity, object in a certain distance range can be segmented. In other words, we detect object not by its color, intensity, or texture but by distance. The advantage of this profile shape-matching algorithm is that detection of correspondences relies on the shape of profile not on intensity values which are subjected to lighting variations. Based on the resulting 3-D information, the movement of fingers in space from a specific distance can be determined. Finger location and movement can then be analyzed for non-contact control.

This algorithm, like others, faces the tradeoff between accuracy and processing speed. Higher accuracy requires more resources and time. For the proposed finger localization and tracking application, we are able to find a balance between accuracy and speed and find that it is suitable for the proposed application.

8662-19, Session 5

Training industrial robots with gesture recognition techniques

Jennifer R. Piane, DePaul Univ. (United States) and Yaskawa America, Inc. (United States); Daniela Raicu, Jacob Furst, DePaul Univ. (United States)

This paper describes that system that applies gesture recognition techniques and machine learning methods to build a system that generates a robot motion program from a pair of stereo videos. The videos used to evaluate the system are 'highly cluttered' and in some cases blurry to support a 'real' usage environment. This paper applies a color-thresholding model, naïve Bayes and SVM to detect the human hand. It performs Stereo Matching on the region where the hand was detected to find relative 3D coordinates. The system finally analyzes the list of coordinates to determine a path for the robot to move. With large amounts of clutter and some blur, the system was 77% effective in detecting the hand. The analysis of relative 3D coordinates was able to compensate for errors and produce effective robot motion.

8662-20, Session 5

A restrained-torque-based motion instructor: forearm flexion/extension-driving exoskeleton

Takuya Nishimura, Yoshihiko Nomura, Ryota Sakamoto, Mie Univ. (Japan)

The authors introduced a cost-effective idea: it would be enough to instruct the timing of start-stop and to accelerate or decelerate joint angle rotation because such instructions, in essentials, do not need to exert power and energy but just need to transmit motion information. This concept is very important point that differentiates our study from the ordinary "highly-powered strong arm" type of powered suits.

This paper presents a result of a fundamental experiment for constructing a guideline of an external torque-representing mechanical specification: taking an example with the start-up of a forearm bending motion, quantities of stimulus threshold for having the forearm move passively by the externally exerted torque were examined through psychophysical experiments using a prototype of a mechanical interface: a torque-based exoskeleton instructor for inducing forearm flexion motion.

8662-21, Session 5

3D recovery of human gaze in natural environments

Lucas Paletta, Katrin Santner, Gerald Fritz, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria); Heinz Mayer, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Germany)

In marketing and usability engineering, the tracking of human eye movements, i.e. eye-tracking, has been the central technology for capturing visual attention. Within the last couple of years miniaturized mobile eye-tracking systems have become available and been successfully applied in different areas of research with the major advantage to evaluate attention in the field where the task of interest is performed. A major disadvantage of existing eye-tracking technology is the need to manually analyze the huge amount of collected video data. In our work we seek to attain alignment of the human gaze data within a 3D reconstruction of the environment by recovering the full 6 Degrees of Freedom (DOF) pose and the sensor should provide three-dimensional information of the environment. This means that the human gaze is fully reconstructed and mapped towards the 3D model of the environment and once the environment has been annotated before, the semantics of the gaze trajectory can be reconstructed thereafter as well. For this purpose we apply the Simultaneous Localization and Mapping (SLAM) framework. On a standardized test dataset, we outperform EKF based solutions in terms of localization accuracy by at least a factor of two.

8662-22, Session 6

CANINE: a robotic mine dog

Brian Stancil, Jeff Hyams, Jordan Shelley, Kartik Babu, Neya Systems LLC (United States); Hernan Badino, Aayush Bansal, Daniel F. Huber, Carnegie Mellon Univ. (United States); Parag Batavia, Neya Systems LLC (United States)

The purpose of this effort is to develop the sensing and processing capabilities to enable a small mobile robot (CANINE) to act as a mine dog. Mine dogs were used to look for dangerous objects such as mines, unexploded ordnance, and trip wires. For this effort, the concept of operations is to show CANINE an object of interest, have it learn the salient features of that object, and then search for and recover the object after it has been thrown in a field containing distractor objects and other obstacles.

8662-23, Session 6

Development of dog-like retrieving capability in a ground robot

Douglas C. MacKenzie, Mobile Intelligence Corp. (United States); Rahul Ashok, James M. Rehg, Georgia Institute of Technology (United States); Gary Witus, Wayne State Univ. (United States)

The Mobile Intelligence Team's approach to addressing the CANINE outdoor ground robot competition will be presented. The competition required developing a robot that provided retrieving capabilities similar to a dog, while operating fully autonomously in unstructured environments. The vision team consisted of Mobile Intelligence, the Georgia Institute of Technology, and Wayne State University. Important computer vision aspects of the project were the ability to quickly learn the distinguishing characteristics of novel objects, searching images for the object as the robot drove a search pattern, identifying people near the robot for safe operations, correctly identifying the object among distractors, and localizing the object for retrieval.

8662-24, Session 6

Multidisciplinary unmanned technology teammate (MUTT)

Nenad Uzunovic, Anne Schneider, Alberto Lacaze, Karl Murphy, Robotic Research LLC (United States); Mark Del Giorno, Del Services (United States)

The unique approach combined with the latest technical developments in autonomous mobility and operator interaction guided the Robotic Research, LLC-led team to a first place win in TARDEC's CANINE competition. The goal of the competition was to develop a robotic system that would display useful capabilities similar to those found in an Army dog – such as fetching objects thrown by an operator. Using only gestures or voice commands, the robots would learn a new object at the start of each phase, find the object after it was thrown into a field, and return the object to the operator. Each of the six phases became more difficult, including clutter of the same color or shape as the object, moving and stationary obstacles, and finding the operator who moved to a new location. The Robotic Research Team integrated techniques in computer vision, speech recognition, object manipulation, and autonomous navigation. A multi-filter computer vision solution reliably detected the objects while rejecting objects of similar color or shape, even while the robot was in motion. A speech-based interface with short commands provided close to natural communication of complicated commands from the operator to the robot. An innovative gripper design allowed for efficient object pickup. A robust autonomous mobility and navigation solution for ground robotic platforms provided fast and reliable obstacle avoidance and course navigation. The research approach focused on winning the competition while remaining cognizant and relevant to real world applications.

8662-25, Session 6

R-MASTIF: robotic mobile autonomous system for threat interrogation and object fetch

Aveek K. Das, SRI International Sarnoff (United States); Dinesh Thakur, James F Keller, University of Pennsylvania (United States); Sujit Kuthirummal, Zsolt Kira, SRI International Sarnoff (United States); Mihail Pivtoraiko, University of Pennsylvania (United States)

Autonomous robotic "fetch" operation, where a robot is shown a novel object and then asked to locate it in the field, successfully retrieve it and bring it back to the human operator, is a challenging problem that is of interest to the military. The CANINE competition presented a forum for several research teams to tackle this challenge using state of the art in robotics technology. The SRI-UPenn team fielded a modified Segway RMP 200 robot with multiple cameras and lidars. We implemented a unique computer vision based approach for colored object training and detection to robustly located previously unseen objects out to 15 meters on moderately flat terrain. We integrated SRI's state of the art Visual Odometry for GPS-denied localization on our robot platform. We also designed a unique scooping mechanism which allowed retrieval of up to basketball sized objects with a reciprocating four-bar linkage mechanism. Further the software framework was developed using ROS (Robot Operating System) which is open source and well adopted by the robotics community. We present a description of the system, our key technical contributions and experimental results.

8662-26, Session 6

LABRADOR: a learning autonomous behavior-based robot for adaptive detection and object retrieval

Brian M. Yamauchi, Mark Moseley, iRobot Corp. (United States); Jonathan Brookshire, MIT (United States)

As part of the TARDEC-funded CANINE (Cooperative Autonomous Navigation in a Networked Environment) Program, iRobot developed LABRADOR (Learning Autonomous Behavior-based Robot for Adaptive Detection and Object Retrieval). LABRADOR was based on the rugged, man-portable, iRobot PackBot unmanned ground vehicle (UGV) equipped with an explosives ordnance disposal (EOD) manipulator arm and a custom gripper. For LABRADOR, we developed a vision-based object learning and recognition system that combined a TLD (track-learn-detect) filter based on object shape features with a color-histogram-based object detector. Our vision system was able to learn in real-time to recognize objects presented to the robot. We also implemented a waypoint navigation system based on fused GPS, IMU (inertial measurement unit), and odometry data. We used this navigation capability to implement autonomous behaviors capable of searching a specified area using a variety of robust coverage strategies – including outward spiral, random bounce, random waypoint, and perimeter following behaviors. While the full system was not integrated in time to compete in the CANINE competition event, we developed useful perception, navigation, and behavior capabilities that may be applied to future autonomous robot systems.

8662-27, Session IPI

Method and application of active visual tracking based on illumination invariants

Jie Su, Harbin Univ. of Science and Technology (China); Gui-sheng Yin, Harbin Engineering Univ. (China); Lei Wang, Yong-qian Sun, Zheng Liu, Harbin Univ. of Science and Technology (China)

To improve identification rate and track rate for quickly moving target, expand tracking scope and lower sensitivity to illumination varying, an active visual tracking system based on illumination invariants is proposed. Camera motion pre-control method based on particle filter pre-location is used to improve activity and accuracy of track for quickly moving target by forecasting target position and control camera joints of Tilt and Pan. Pre-location method using particle sample filter according to illumination invariants of target is used to reduce the affect of varying illumination during tracking moving target and to improve algorithm robust. The correctness and accuracy were tested though experiments. Activity, real time and accuracy are improved while sensitivity to illumination variation is reduced.

8662-28, Session IPI

Supervised linear dimensionality reduction with robust margins for object recognition

Fadi Dornaika, Univ. del País Vasco (Spain); Ammar Assoum, Lebanese Univ. (Lebanon)

Linear Dimensionality Reduction (LDR) techniques have been increasingly important in computer vision and pattern recognition since they permit a relatively simple mapping of data onto a lower

dimensional subspace, leading to efficient classification strategies. Recently, a linear method called Average Neighbors Margin Maximization (ANMM) was proposed and shown to have powerful discrimination properties. Although good results were obtained with this method, it suffers from two shortcomings: i) it requires the setting of two parameters-the neighbors sizes for homogeneous and heterogeneous samples, and ii) it can be very sensitive to the label outliers since a margin average is used.

In this paper, we propose a robust linear embedding method whose margins are computed over median distances. Experimental results performed on three face databases (Extebded Yale, PIE, and PF01) show that the proposed approach can give better generalization performance than the ANMM method. Our proposed method is much less sensitive to outliers than the ANMM method and the Nearest miss/Nearest hit method. The proposed approach could also be applied to other category of objects characterized by large variations in their appearance.

8662-29, Session IPI

Using a multi-port architecture of neural-net associative memory based on the equivalency paradigm for parallel cluster image analysis and self-learning

Vladimir G. Krasilenko, Vinnitsa Social Economy Institute (Ukraine); Alexander A. Lazarev, Vinnitsa National Technical Univ. (Ukraine); Sveta K. Grabovlyak, Diana V. Nikitovich, Vinnitsa Social Economy Institute (Ukraine)

First, we consider equivalency models, including matrix-matrix and matrix-tensor and with the dual adaptive-weighted correlation, multi-port neural-net auto-associative and hetero-associative memory (MP NN AAM and HAP), which are equivalency paradigm and the theoretical basis of our work. We make a brief overview of the possible implementations of the MP NN AAM and of their architectures proposed and investigated earlier by us. The main base unit of such architectures is a matrix-matrix or matrix-tensor equivalentor.

We show that the MP NN AAM based on the equivalency paradigm and optoelectronic architectures with space-time integration and parallel-serial 2D images processing have advantages such as increased memory capacity (more than an order of the number of neurons!), high performance in different modes (10^{10} -10^{12} connections per second!) and the ability to process, store and associatively recognize highly correlated images.

Next, we show that with minor modifications, such MP NN AAM can be successfully used for high-performance parallel clustering process of images. We show simulation results of using these modified for clustering and learning models and algorithms for cluster analysis of specific images and divide them into categories of the array.

Show example of a cluster division of 32 images (40 x 32 pixels) letters and graphics for 12 clusters with simultaneous formation of the output-weighted space allocated images for each cluster.

We will discuss algorithms for learning and self-learning in such structures and their comparative evaluations based on MathCad simulations are made.

It will be shown that, unlike the traditional Kohonen self-organizing maps, time of learning in the proposed structures of multi-port neuronet classifier/clusterizer (MP NN C) on the basis of equivalency paradigm, due to their multi-port decreases by orders and cannot exceed, in some cases, just a few epochs.

Estimates show that in the test clustering, selected for example, 32 1280-element images into 12 groups, the formation of neural connections of the matrix dimension of 128 * 120 elements occurs to tens of iterative steps (some epochs), and for a set of learning patterns consisting of 32 such images, and at time of processing of 1-10 microseconds, the total learning time does not exceed a few milliseconds.

We will offer criteria for evaluate the quality of patterns clustering with such MP NN AAM.

Conference 8663: Video Surveillance and Transportation Imaging Applications

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8663 Video Surveillance and Transportation Imaging Applications

8663-1, Session 1

Group localisation and unsupervised detection and classification of basic crowd behaviour events for surveillance applications

Nadejda S. Roubtsova, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

Recently, in response to the need for automatic monitoring of densely populated high-risk public places, interest to crowd behaviour analysis has been growing. Instead of tackling complex surveillance scenarios directly through machine learning, we present an optical-flow-based unsupervised detection and classification algorithm of three elementary crowd behaviour events: splits, merges and lateral slide-by motions. Our system is motion-based, rather than object-detection-based, and inherently localises moving groups at frame intervals. This group-based, rather than individual-based, approach allows for scalability and a degree of robustness against artifacts as infallible human detection is no longer essential. The system's performance is characterised by high detection and classification recall scores.

The presented framework can be built upon and its applications extend beyond the originally intended scope. Firstly, basic behaviour events, localised both spatially and temporally, together with the current group position information by themselves indirectly help sketch the monitored environment through the crowd's reaction to it (e.g. an obstacle causing a crowd split). Secondly, the presented framework, alone with minor extensions and/or when integrated with static content analysis algorithms (e.g. traffic sign detection), can serve for analysis of complex surveillance scenarios such as mass evacuation, unregulated road crossing or unauthorised overcrowding.

8663-2, Session 1

Gaussian mixtures for anomaly detection in crowded scenes

Habib Ullah, Univ. degli Studi di Trento (Italy); Lorenza Tenuti, University of Trento (Italy); Nicola Conci, Univ. degli Studi di Trento (Italy)

No Abstract Available

8663-3, Session 1

Space-time correlation filters for human action detection

Joseph A. Fernandez, B. V. K. Vijaya Kumar, Carnegie Mellon Univ. (United States)

To automate video surveillance systems, algorithms must be developed to automatically detect and classify basic human actions. Many traditional approaches focus on the classification of actions, which usually assumes prior detection, tracking, and segmentation of the human figure from the original video. On the other hand, action detection is a more desirable paradigm, as it is capable of simultaneous localization and classification of the action. This means that no prior segmentation or tracking is required, and multiple action instances may be detected in the same video.

Correlation filters have been traditionally applied for object detection in images. In this paper, we report the results of our investigation using correlation filters for human action detection in videos. Correlation filters have previously been explored for action classification, but this is the first time they are evaluated for the more difficult task of action detection. In addition, we investigate several practical implementation issues, including parameter selection, reducing computational time, and exploring the effects of preprocessing and temporal occlusion (i.e., loss of video frames) on performance.

8663-4, Session 1

Recognition of two-person interaction in multi-view surveillance video via proxemics cues and spatio-temporal interest points

Bo Zhang, Paolo Rota, Nicola Conci, Univ. degli Studi di Trento (Italy)

No Abstract Available

8663-5, Session 1

Weighted symbolic analysis of human behavior for event detection

Andrea Rosani, Giulia Boato, Francesco G. B. De Natale, Univ. degli Studi di Trento (Italy)

Video understanding has become a high interest research topic in recent years, with various applications, such as video browsing, content-based video indexing and visual surveillance. The automation of this process is still a challenging task. Symbolic approaches introduce a great simplification in the data, transforming them into chains of meaningful patterns. This allows to overcome a common drawback resulting from the clutter produced by low-level processing operations, embedding significant contextual information into the data, as well as using simple syntactic approaches to perform the matching between actual sequences and models. In this paper we propose a novel framework that employs a symbolic representation of complex activities through sequences of atomic actions based on a weighted Context-Free Grammar (SVM-CFG). The video understanding is performed at different abstraction layers, retrieving anomalous activities and limiting the influence of noise.

8663-6, Session 2

A resource allocation framework for adaptive selection of point matching strategies

Quentin De Neyer, Li Sun, Christophe De Vleeschouwer, Univ. Catholique de Louvain (Belgium)

Our work investigates the PTZ autotracking problem, which builds on the close-view videos captured by a moving camera. This case is indeed fundamentally different from conventional tracking scenarios, generally dealing with wide field of views, in which the object of interest can be tracked by discriminating it from its surrounding background. In clear words, the conventional tracking algorithms, which rely on the discrimination between the target and its background, are not relevant in our close-view context. Hence, we propose to adopt a point-based model of the target, so as to track the object based on the matching of points between consecutive frames of the video.

Conference 8663

The paper introduces an original cost-benefit framework to adapt the number of points and their individual matching metrics to maximize the global frame registration quality under a complexity constraint. Handling the complexity is of particular importance in the autotracking scenario since complexity directly affects the delay affecting the PTZ control loop. Our point-based representation model appears to be both effective and efficient in registering pairs of frame.

8663-7, Session 2

Human behavior analysis using trajectory data in a PTZ surveillance system

Sanjeev Kumar, Balasubramanian Raman, Indian Institute of Technology Roorkee (India)

This paper presents a modern video surveillance system composed by a PTZ camera for human motion analysis. The trajectory is obtained from the video stream of human motion and classified as normal or anomalous. Background subtraction approach has been used to obtain the trajectory of human motion. The trajectory is represented by a non uniform rational B-spline (NURBS) curve in terms of a polynomial. Each coordinate of NURBS polynomial is represented by a function of time. In this context, the position of human can be obtained just by the time information with an accurate localization. The control points and weights of the NURBS curves are used to generate the feature vector for classification. A binary support vector machine is used as the classifier for the trajectory classification. Experimental results are presented from the video streams of a lobby where a number of people are moving in different directions. Besides showing the analysis of the obtained results, we compare our results with some existing trajectory clustering approaches.

8663-8, Session 2

Collaborative real-time scheduling of multiple PTZ cameras for multiple objects tracking in video surveillance

Yu-Che Liu, Chung-Lin Huang, National Tsing Hua Univ. (Taiwan)

The video surveillance systems with many CCTV cameras perform poorly on tasks that require higher resolution facial images for biometric identification. One promising solution is using the Pan-Tilt-Zoom (PTZ) cameras to cover a potentially much larger area and obtain a higher resolution image through zoom capability. So, some surveillance systems combine traditional passive CCTV cameras with the active PTZ cameras in master/slave configurations.

Video surveillance system with fixed CCTV cameras cannot provide higher resolution images of the target such as close-up facial images. Once the object approaches some blind spots, it cannot be observed. Therefore, many surveillance systems combine the passive CCTV cameras with the PTZ cameras in master/slave configuration. The active PTZ cameras should track and observe the designated target object, and then switch their target objects (or hand-off) once the observation goal changes.

The camera network can resolve the conflicts of tasking multiple cameras to observe several pedestrian simultaneously. The main challenge arises when the number of subjects and activities exceeds the number of PTZ cameras so that the scheduling and controlling of the PTZ cameras become nontrivial. The multi-PTZ cameras system decides the best allocation of the camera to observe different human subjects in the mean while satisfies the observation criterion that requires seamless close-up video of the objects of interest. Here, we formula PTZ cameras assignment and handoff as a planning problem whose solution achieves optimal camera utilization for the predefine observation goal.

Many literatures in video surveillance systems focus on multi-people multi-camera (MPMC) object tracking in a large heterogeneous camera network. Using fixed CCTV cameras performs poorly on tasks that require higher resolution facial images for biometric identification. The MPMC tracking system needs to consider the following conditions: (1) the number of pedestrian and activities appears in the scene may

exceeds the number of available PTZ cameras, (2) the dual camera pairs required to cover all the area to avoid the occurrence of blind spots, (3) The computation complexity of object tracking and location increases.

This paper formulates PTZ camera assignment and hand-off as a planning problem. Our goal is to optimize the camera utilization to record seamless, high quality video of roaming individual by using only multiple PTZ cameras.

8663-9, Session 2

Tracking small targets in wide area motion imagery data

Alex Mathew, Vijayan K. Asari, Univ. of Dayton (United States)

In this paper, we propose a method to track very small targets such as pedestrians, in Wide Area Motion Imagery (WAMI) data. Our method is evaluated on AFRL Columbus Large Image Format (CLIF) WAMI data. The approach involves four steps – image registration, feature extraction, location estimation and template update. Speeded-Up Robust Feautres (SURF) is used for image registration. For feature extraction, we use a novel method in which the image is represented as a matrix of histograms. A Kalman filter, in combination with the proposed feature extraction technique, is used for tracking. A sliding window is used to search for possible location of the target within the search region given by the filter. Instead of treating the problem as an object detection problem, we treat this as an image retrieval problem and use Earth Mover's Distance (EMD) to compute a 'dissimilarity value'. The target location is taken as the one that produces the least value. The problem of template drift is solved by localizing the target with a blob detection algorithm. We show the robustness of the algorithm by giving a comparison with other methods such as SURF, SIFT, HoG and mean-shift tracking.

8663-10, Session 2

Feature descriptors for object matching in real-time tracking applications

Gernot Loibner, Oliver Sidla, SLR Engineering OG (Austria)

This work presents a real-time tracking framework which is designed to achieve tracking with at least 8-10 frames per second on a low power smart camera. Applications of the proposed tracking framework is pedestrian and vehicle tracking in indoor and outdoor environments. Our proposed tracking framework is kept simple in order to avoid unnecessary computation, but it has proven to be effective in low to medium crowded situations. This work concentrates on appearance based similarity measures used which are used to verify matches from frame to frame for i) the decision whether a detection actually matches with the extrapolated position from the previous frame, ii) not only between frames but also between different cameras, iii) to decide whether objects might be ‚good object to track' which significantly differ from the average.

The object descriptors which have been tested are all fast enough so that they cane used in our embedded system parallel to LK optical flow and HOG detector modules. Specifically we have tested the following feature descriptors:

- BRIEF
- ORB
- BRISK
- FREAK
- HOG
- DCT
- color histograms

Based on a set of several hundred manually extracted object samples (pedestrians only) we test every feature descriptor on its ability to

i) match identical objects from different frames

ii) to describe each object and separate it from other objects in feature-space

Conference 8663

Each test will report also on the relative computation cost on the embedded hardware under consideration. The tracking-framework will has been tested on real-world videos with annotated ground truth trajectories in order to evaluate its performance.

8663-11, Session 3

Identification of overloaded ships based on video surveillance in inland river

Lei Xie, Wuhan Univ. of Technology (China)

To reduce water traffic accident, identification of overloaded ships has been of major importance for enforcing inland river management, especially in some developing countries. Driven by rising oil prices or other economic benefits, the ship overload phenomenon continued to occur in China. Therefore, overloaded ship detection has been of a key factor of shipping safety. This paper presented a robust method for detecting overloaded ship and the proposed algorithm included three stages: ship detection, ship tracking and overloaded ship identification. Ship detection was a key step and the concept of ship tracking is built upon the ship-segmentation method, in which the algorithm about background estimation, background updating, background subtraction and ship detection has been described. According to the segmented ship shape, a predict method based on Kalman filter has been proposed to track each ship. The described identification system included a video camera and a high definition camera, which led to a necessary coordinate transformation in system model of Kalman filter. The data of ship length and ship speed could be used to identify overloaded ship. The proposed method has been tested on a number of monocular ship image sequences and the experimental results showed that the algorithm was robust and real-time.

8663-12, Session 3

Ship detection in port surveillance based on context and motion saliency analysis

Xinfeng Bao, Svitlana Zinger, Technische Univ. Eindhoven (Netherlands); Rob Wijnhoven, ViNotion B.V. (Netherlands); Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

Automatic ship detection is crucial for port surveillance in ship traffic monitoring and management. It is a challenging and complex task due to the dynamic water background and large variations of ship appearances. In this paper, we present a novel ship detection approach based on the analysis of contextual information and motion saliency in the video frames. For each frame, the water region is first labeled using a combination of graph-based segmentation with region-level recognition. The detected water region is then used as contextual information to locate the candidate ships, which are connected components (blobs) labeled as non-water. We assume that moving ships are the blobs having salient motion, where the saliency measure is defined by comparing the motion of candidate ships and the surrounding background. The main advantages of our approach are that it requires no prior knowledge of ship appearances and can delineate the contours of each detected ship. The experimental results show that our approach can detect moving ships with large variations in characteristic (shape, size, speed), with high recall and precision rates.

8663-13, Session 3

A CubeSat-size surveillance satellite concept

Jeremy Straub, Debabrata Ghosh, Naima Kaabouch, Ronald A. Fevig, The Univ. of North Dakota (United States)

An architecture for a high-resolution remote sensing satellite designed to be launched into low Earth orbit is presented. At an altitude of 300 kilometers, the proposed satellite is capable of producing images with a spatial resolution of below one meter through the combination of limited onboard optics and computer image enhancement.

The satellite incorporates passive stabilization and limited attitude control capabilities. It is powered by solar panels on five faces and operates in an extremely limited mode when not over its designated target area, allowing its batteries to be charged over the two-thirds of its orbit where it is not in eclipse and discharged to provide the required power when over the target and/or ground station. The satellite incorporates significant onboard computational capabilities for image processing. An onboard mosaicking system is used to reduce data transfer requirements (via not transmitting overlapping areas) and provide a ready-for-user wide-area image.

Surveillance applications (e.g., military, intelligence, defense assistance to civil authorities, emergency management, and Earth science) can benefit dramatically from a low-cost, rapidly-deployable, visible light remote sensing satellite. The CubeSat form factor is ideal for these applications. Rapid launch mechanisms are poised to allow in-the-field CubeSat deployment, in the imminent future.

8663-14, Session 3

An edge directed super resolution technique for multimedia applications

Osborn de Lima, Sreenath Rao Vantaram, Sankaranarayanan Piramanayagam, Eli Saber, Rochester Institute of Technology (United States); Kurt R. Bengtson, Hewlett-Packard Co. (United States)

Surveillance applications in today's world, requires images with superior visual quality and clear distinguishable features in order to make informed decisions with respect to security and law enforcement. The demand for high resolution images in different applications along with a stipulation of reduced cost has paved way to super resolution enhancement techniques. Super resolution can be viewed as a software solution that impacts hardware limitations. The proposed algorithm is a multiple pass iterative algorithm aimed at producing better defined images with sharper edges. The basic premise behind this algorithm is interpolating along the edge direction and thus reducing blur that comes from traditional interpolation techniques which operate across the edge in some instances. The process utilizes multiple images shifted by a sub pixel amount and yields a single higher resolution image. A polynomial warping registration approach determines the relative shifts between the frames which in turn leads the algorithm to produce a high resolution image with improved quality. The algorithm developed was applied to grayscale and color images and showed favorable results on a wide variety of datasets obtained from surveillance and printing applications.

8663-15, Session 4

A smart camera based traffic enforcement system: experiences from the field

Oliver Sidla, Gernot Loibner, SLR Engineering OG (Austria)

Railway crossings, even when secured by automated barriers, pose a threat to drivers day and night. Especially drivers which try to cross railway tracks in the last moments before a barrier closes have a relatively high probability to cause severe accidents. In order to monitor unsecured railway crossings and to detect and fine red-light jumpers, we have developed a vision based system which is completely integrated into a smart camera. Triggered by an electrical signal from the railway, the camera begins to observe the crossing. As soon as a vehicle moves over the stopping line and the vehicle is well over this line, image sequences are recorded and stored on the system. System architecture This work concentrates on the system architecture and setup in terms of HW and SW modules. The smart cameras need to provide readable license plate images of those cars which violate the traffic rules at the installation site. To this end we had to take several measures to be able achieve this goal due to the difficult camera geometry in the field:

- optical, geometric setup. IR illumination and proper filters are used to minimize the effect of environmental light.

- a SW controlled imaging control on the camera to offset effects of strong direct sunlight and very high contrast scenarios during the dark hours.

Algorithm for moving vehicle detection

The detection of moving vehicles is based on sparse optical flow We will report on results of the LK tracker with FAST features and the AGHAST corner point detector as comparison.

The moving vehicle detection works as follows:

For every video frame:

i) LK optical flow update

ii) Remove tracked points which do not move

iii) Remove tracked points which do not move along direction of the street lane

iv) Cluster points according to similar velocity, direction, and distances

v) Valid clusters are added to a ring buffer

vi) If valid clusters exist for a duration of more than 300 ms, we flag the detection of a vehicle

Experimental results and detector evaluation

The detection capability and false alarm rate of the systems is evaluated using several data sets:

- video sequences form the site for evaluation of the absolute detection rate

- actual detections from the installation site in order to evaluate the false alarm rate

- evaluation of onsite video sequences to asses the detection capability during night-time

8663-16, Session 4

Algorithm design for automated transportation photo enforcement camera image and video quality diagnostic check modules

Ajay Raghavan, Bhaskar Saha, Palo Alto Research Center, Inc. (United States)

Photo enforcement devices for traffic rules such as red lights, stops, and speed are increasingly being deployed in cities around the world to ensure smooth traffic flow and public safety. These are typically unattended fielded systems, and so it is important to periodically check them for potential image/video quality problems that might interfere with their intended functionality. There is interest in automating such checks to reduce the operational overhead and human error involved in manually checking large camera device fleets. Examples of problems affecting such camera devices include exposure issues, focus drifts, obstructions, misalignment, download errors, and motion blur. Furthermore, in some cases, in addition to the sub-algorithms for individual problems, one also has to carefully design the overall algorithm and logic to check for and accurately classifying these individual problems. Examples include camera misalignment that can cause some scene elements to go out of focus for wide-area scenes or download errors that can be misinterpreted as an obstruction. Therefore, the sequence in which the sub-algorithms are utilized is also important. This paper presents an overview of these problems along with image and video quality solutions to detect and classify such faults.

8663-17, Session 4

Full-body occlusion handling in vision-based traffic density counting

Evangelos Palinginis, Keitaro Kamiya, Man Woo Park, Ioannis Brilakis, Jorge A. Laval Jr., Randall Guensler, Georgia Institute of Technology (United States)

Vision-based traffic surveillance systems are amidst most reliable, inexpensive and highly applicable methodologies in surveying traffic conditions. However, the implementation of these strategies is limited under occluding and illuminating conditions. Given that the aforementioned conditions lead to either over-counted or under-counted traffic data, the proposed motion-based framework aims at overcoming these limitations by employing a methodology not associated with the trajectory of cars' movement. Within the methodology, the background subtraction is followed by the Histogram of Oriented Gradients (HOG) which is trained by linear Support Vector Machine (SVM); and Haar-like features trained by Adaboost. Both are implemented so as to detect each vehicle close the ground truth. Accordingly, their location is estimated the automatically detected data are converted into the real word coordinate system. The proposed method will be tested onto various 15-minutes long video from different location along the Georgia highway corridor at the time of a day where severe traffic jam is observed. Preliminary results indicate the reliability and effectiveness of the proposed methodology in providing traffic density counts. Future work may rely on the extension of the proposed methodology in order to transform the data of the removed vehicle into the data of the new vehicle automatically.

8663-20, Session 5

Vehicle identification in two non-overlapping views

Tzung-Yu Hsieh, National Tsing Hua Univ. (Taiwan); Chung-Lin Huang, Asia Univ. (Taiwan)

Vehicle identification in two non-overlapped views is a non-trivial problem because the appearance of the same vehicle in two cameras may not be similar. Direct matching can be difficult due to dramatic illumination/appearance and aspect changes. Instead of directly finding the probability of match between two observations of the same vehicle appearing in two non-overlapping views, we apply the so-called non-metric distance embeddings of vehicle observations within one camera and their corresponding embeddings in another camera to characterize the similarities and difference between the vehicles across cameras. The embedding are with respect to examples pre-selected for each camera, and the output is the probability of the same vehicle. The embedding process is based on concept that the observations of the vehicles appearing in the same camera will be similarly affected by the same illumination and aspect changes. Therefore, robust alignment and matching within the same camera can accurately capture the similarity and difference amongst the vehicles. The vehicle in any camera can be represented as a vector of distance to the respective examples within that camera. Then we can two such vector representations to verify whether the two observations represent the same vehicle or not.

To find the similarity distance, we find the edge map of the target vehicle and compare with example vehicle to find the edge distance. Besides the edges, we also compare the color similarity of the target vehicle and examples. After embedding, we train the Support Vector Machine (SVM) by using a set of similar vector pairs and non-similar vector pairs. Finally, in the testing process, after embedding, we have a vector pair which can be verify by applying the SVM to identify whether it is a similar vector pair or not.

Our system consists of (1) Foreground vehicle segmentation, (2) Feature (edges and color) extraction, (3) embedding vector generation based on the example pairs, and (4) SVM classification based on the embedding vector pair. In the experiments, we demonstrate the acceptable recognition rate for the vehicle pairs appearing in two scenes with totally different views and illuminations. Our method demonstrates a better performance than the conventional thresholding method.

8663-21, Session 5

Vehicle classification for road tunnel surveillance

Andrés Frias-Velázquez, Peter Van Hese, Aleksandra Pi?urica, Wilfried Philips, Univ. Gent (Belgium)

Vehicle classification for tunnel surveillance aims to not only retrieve vehicle class statistics, but also prevent accidents by recognizing

vehicles carrying dangerous goods. In this paper, we describe a method to classify vehicle images that experience different geometrical variations and challenging photometrical conditions such as those found in road tunnels. Unlike previous approaches, we propose a classification method that does not rely on length and height estimation of the vehicles. Alternatively, we propose a novel descriptor based on trace transform signatures to extract salient and non-correlated information of the vehicle images. Also, we propose a metric that measures the shape complexity of the vehicles based on corner point detection. As a result, these features describe the vehicle's appearance and shape complexity independently of the scale, pose, and illumination conditions. Experiments with vehicles captured from three different cameras confirm the saliency and robustness of the features proposed, achieving an overall accuracy of 97.5% for the classification of four different vehicle classes. For vehicles transporting dangerous goods, our classification scheme achieves an average recall of 97.6% at a precision of 98.6% for the combination of lorries and tankers, which is a very good result considering the scene conditions.

8663-36, Session 5

Using visual analytics model for pattern matching in surveillance data

Mohammad S Habibi, Tennessee State Univ. (United States)

In a persistent surveillance system huge amount of data is collected continuously and significant details are labeled for future references. In this paper, we explain a method to summarize video data as a result of identifying events based on such tagged information, leading to a concise description of behavior within a section of extended recordings. An efficient retrieval of various events thus becomes the foundation for determining a pattern in surveillance system observations, both in its extended and fragmented versions. The patterns consisting of spatiotemporal semantic contents are extracted and classified by application of video data mining on generated ontology, and can be matched based on analysts interest and rules set forth for decision making. The proposed extraction and classification method used in this paper uses query by example for retrieving similar events containing relevant features, and is carried out by data aggregation. Since structured data forms majority of surveillance information our Visual Analytics model employs KD-Tree approach to group patterns in variant space and time, thus making it convenient to identify and match any abnormal burst of pattern detected in a surveillance video. Several experimental video were presented to viewers to analyze independently and were compared with the results obtained in this paper to demonstrate the efficiency and effectiveness of the proposed technique.

8663-37, Session 5

Situation exploration in a persistent surveillance system with multidimensional data

Mohammad S Habibi, Tennessee State Univ. (United States)

There is an emerging need for fusing hard and soft sensor data in an efficient surveillance system to provide accurate estimation of situation awareness. These mostly abstract, multi-dimensional and multi-sensor data pose a great challenge to the user in performing analysis of multi-threaded events efficiently and cohesively. To address this concern an interactive Visual Analytics (VA) application is developed for rapid assessment and evaluation of different hypotheses based on context-sensitive ontology spawn from taxonomies describing human/human and human/vehicle/object interactions. A methodology is described here for generating relevant ontologies in a Persistent Surveillance System (PSS) and demonstrates how they can be utilized in the context of PSS to track and identify group activities pertaining to potential threats. The proposed VA system allows for visual analysis of raw data as well as metadata that have spatiotemporal representation and content-based implications. Additionally in this paper, a technique for rapid search of tagged information contingent to ranking and confidence is explained for analysis of multi-dimensional data. Lastly

the issue of uncertainty associated with processing and interpretation of heterogeneous data is also addressed.

8663-22, Session 6

Vehicle presence analysis for law enforcement applications and parking lot management

Yuriy Lipetski, Oliver Sidla, SLR Engineering OG (Austria)

The efficient and robust detection of the presence of vehicles in restricted parking areas is important for applications in law enforcement as well as for the enforcement of parking rules on private property. We present our work towards this goal aimed at the application of vehicle detection in urban environments. The method is to be suited for smart cameras which shave to operate autonomously over extended periods of time. Our system is developed as part of a bigger research effort which combines onsite vehicle presence detection and an associated web management system which is intended to monitor, steer and reroute delivery vehicles. Our detection approach is based on appearance based modeling with area based feature computation and a subsequent classification algorithm. The work presented here will show evaluation results onto sites in the city of Vienna from the operation of two smart cameras over an extended period of time.

8663-23, Session 6

Video-based parking occupancy detection

Michael Deruytter, Kevin Anckaert, Traficon N.V. (Belgium)

ITS technologies are increasingly used for near-road applications, such as parking area management. At least two use cases show that there is a need for a smarter approach towards parking management. First of all, because of a shortage of truck parks along highways truck drivers are forced to park on the hard shoulder, which increases the risk of accidents. Secondly, cities are coping with large amounts of vehicle traffic, generated by motorists that are driving around in search for a parking sport. In both cases, parking guidance and management systems can alleviate the problem. In the case of trucks, such systems allow truck drivers to dynamically plan their route and prevent them from having to park on the hard shoulder. In the case of city passenger traffic, motorists can significantly reduce their travel distance in search for a parking spot by relying on information generated by a parking guidance system. A critical factor of parking systems is occupancy detection. Various technologies, such as video detection, microwave radar detection and electromagnetic loops, can be used to detect vehicle presence. In this paper, we present the benefits of a video-based approach for truck and car park occupancy detection.

8663-24, Session 6

Motorcycle detection and counting using stereo camera, IR camera, and microphone array

Bo Ling, Migma Systems, Inc. (United States); David R. P. Gibson, Federal Highway Administration (United States); Dan Middleton, Texas A&M Transportation Institute (United States)

The detection, classification, and characterization are the key to enhancing motorcycle safety, motorcycle operations and motorcycle travel estimation. Average motorcycle fatalities per Vehicle Mile Traveled (VMT) are currently estimated at 30 times those of auto fatalities. Although it has been an active research area for many years, motorcycle detection and classification still remain a challenging task. Working with FHWA, we have developed a hybrid motorcycle detection and counting system using a suite of sensors including stereo camera, thermal IR camera and unidirectional microphone array. The IR thermal camera can capture the unique thermal signatures associated with motorcycle's exhaust pipes that often show bright elongated blobs in IR images. The stereo camera in the system is used to detect the

motorcyclist who can be easily windowed out in the stereo disparity map. If motorcyclist is detected through his or her 3D body recognition, motorcycle is detected. Microphones are used to detect motorcycles that often produce low frequency acoustic signals. All three microphones in the microphone array are placed in strategic locations on the sensor platform to minimize the interferences of background noises from sources such as rain and wind. Test results show that this system has an excellent performance.

8663-25, Session 6

Vehicle-driven video compression/decompression for fast and efficient search in large video databases

Orhan Bulan, Xerox Research Ctr. (United States); Edgar Bernal, Robert Loce, Xerox Research Center (United States); Wencheng Wu, Xerox Research Ctr. (United States) and Xerox Research Ctr. (United States)

Video cameras have been commonly deployed at highways, toll booths, traffic lights, stop signs, or local roads by authorized entities for monitoring traffic or law enforcement. The videos captured by these cameras are typically compressed and stored in large databases. Performing searches for a specific vehicle (e.g., vehicle involved in a crime) within these large databases of compressed videos is often required. In this paper, we propose a video compression and decompression algorithm for fast and efficient vehicle searching in large video databases. The proposed algorithm selects the reference frames (i.e., I-frames) based on a vehicle position in the scene while compressing a video sequence. A specific vehicle in the compressed video is then searched only across the reference frames, which do not require to decompress the whole video sequence. Our experimental results across the videos captured in a local road shows that the proposed algorithm significantly reduces the space for searching a vehicle in a compressed video captured especially in light traffic.

8663-26, Session 7

Vehicle occupancy detection camera position optimization using design of experiments and standard image references

Peter Paul, Martin Hoover, Mojgan Rabbani, Xerox Corp. (United States)

One challenge in vehicle occupancy detection is achiving camera line of sight into a vehicle cab in order to detect occupants. Seats, pillars, roof tops, vehicle body frames, and other occupants are some of the occluding items. In addition, differing vehicle shapes and sizes, as well as variations in occupants also lead to difficulties. An approach to camera setup for vehicle occupancy detection which is robust to these effects is proposed here. The method uses rigrous statistical techniques from the field of design of experiments (DOE). In addition, the technique of using a standard image reference (SIR) is used to determine a quantified output to be optimized. The approach is performed on a vehicle rear seat occupancy detection task. Experimental results show significant improvement over the baseline.

8663-27, Session 7

HOV/HOT occupancy detection system

Wayne Daley, Jack W. Wood, John M. Stewart, Colin Usher, Omar Arif, John Turgeson, Erin Hanson, Georgia Tech Research Institute (United States)

One technique to better utilize existing roadway infrastructure is the use of HOV and HOT lanes. Technology to monitor the use of these lanes would assist managers and planners in efficient roadway operation. There are no available occupancy detection systems that perform at acceptable levels of accuracy in permanent field installations. The main goal of this research effort is to assess the

possibility of determining passenger use with imaging technology. This is especially challenging because of recent changes in the glass types used by car manufacturers to reduce the solar heat load on the vehicles. We describe in this research a system to use multi-plane imaging with appropriate wavelength selection for sensing passengers in the front and rear seats of vehicles travelling in HOV/HOT lanes. The process of determining the geometric relationships needed, the choice of illumination wavelengths, and the appropriate sensors are described, taking into account driver safety considerations. The paper will also cover the design and implementation of the software for performing the window detection and people counting utilizing both image processing and machine learning techniques. The integration of the final system prototype will be described along with the performance of the system operating at a representative location.

8663-28, Session 7

Joint histogram between color and local extrema patterns for object tracking

Subrahmanyam Murala, Q. M. J. Wu, Univ. of Windsor (Canada); Balasubramanian Raman, R. P. Maheshwari, Indian Institute of Technology Roorkee (India)

In this paper, a new algorithm meant for object tracking application is proposed using local extrema patterns (LEP) and color features. The standard local binary pattern (LBP) encodes the relationship between reference pixel and its surrounding neighbors by comparing gray level values. The proposed method differs from the existing LBP in a manner that it extracts the edge information based on local extrema between center pixel and its neighbors in an image. Further, the joint histogram between RGB color channels and LEP patterns has been build which is used as a feature vector in object tracking. The performance of the proposed method is compared with Ning et al. on three benchmark video sequences. The results after being investigated proposed method show a significant improvement in object tracking application as compared to Ning et al.

8663-29, Session 8

Adaptive real-time road detection using VRay and A-MSRG in complex environments

SunHee Weon, SungIl Joo, HyungIl Choi, Soongsil Univ. (Korea, Republic of)

This paper proposes an adaptive detection method for detecting road regions that have ambiguous boundaries within natural images. The proposed method achieves reliable partitioning of the road region within a natural environment where noise is present through the following two stages. In the first stage, we separate out candidate regions of the road by detecting the road's boundary through the Radial region split method using VRay(Vanishing point-constrained ray). In the second stage, we apply so called Adaptive-Multiple Seed Region Growing(A-MSRG) approach into the separated candidate region in order to identify the road region in real time. The A-MSRG is an enhanced version of the Seed Region Growing(SRG). For performance evaluation, this study assessed efficiency based on the results of region detection achieved through the proposed combination of the Radial region split method and A-MSRG. We also conducted comparisons against the existing SRG and MSRG methods to confirm the validity of the proposed method.

The main task of the first stage is to detect the road boundary and the candidate region in real time within a complex natural environment. This task is accomplished through the Radial region split method, which includes three steps. They are the step of adaptive edge detection, the step of vanishing point detection using Hough Transform and the step of detecting the road boundary using VRay.

The second stage is the stage of road region detection using the A-MSRG method. In this stage, the road candidate region formed by the extracted optimal boundaries is adaptively partitioned into blocks using the viewpoint's altitude to extract the multiple seed blocks. Then we apply the local weighted value based on the measurement of the block's non-uniformity and we use the A-MSRG method for merging

and expanding the block region to perform the process of adaptively navigating the road region. To ensure its strong performance in response to feature distortions that are caused by changes in perspective depending on the point of view and in response to changes in the form of the sidewalk.

8663-30, Session 8

Intensity estimation method of LED array for visible light communication

Takanori Ito, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan); Arai Shintaro, Kagawa National College of Technology (Japan); Takaya Yamazato, Hiraku Okada, Toshiaki Fujii, Nagoya Univ. (Japan)

This paper focuses on a road-to-vehicle visible light communication (VLC) system using LED traffic light as the transmitter and camera as the receiver. The traffic light is composed of a hundred of LEDs on two dimensional plain. In this system, data is sent as two dimensional brightness patterns by controlling each LED of the traffic light individually, and they are received as images by the camera. Here, there are problems that neighboring LEDs on the received image are merged due to less number of pixels in case that the receiver is distant from the transmitter, and/or due to blurring by defocus of the camera. Because of that, bit error rate (BER) increases due to recognition error of intensity of LEDs

To solve the problem, we propose a method that estimates the intensity of LEDs by solving the inverse problem of communication channel characteristic from the transmitter to the receiver. The proposed method is evaluated by BER characteristics which are obtained by computer simulation and experiments. In the result, the proposed method can estimate with better accuracy than the conventional methods, especially in case that the received image is blurred a lot, and the number of pixels is small.

8663-31, Session 8

An improved background segmentation method for ghost removals

Waqas Hassan, Philip Birch, Rupert Young, Chris Chatwin, Univ. of Sussex (United Kingdom)

Video surveillance has become common for the maintenance of security in a wide variety of applications. However, the increasingly large amounts of data produced from multiple video camera feeds is making it increasingly difficult for human operators to monitor the imagery for activities likely to give rise to threats. This has led to the development of different automated surveillance systems that can detect, track and analyze video sequences both online and offline and report potential security risks. Segmentation of objects is an important part of such systems and numerous background segmentation techniques have been used in literature. One common challenge faced by these techniques is adaption in different lighting environments. A new improved background segmentation technique has been presented in this where the main focus is to accurately segment potentially important objects by reducing the overall false detection rate. Historic edge maps and tracking results are analyzed for this purpose. The idea is to obtain an up to date edge map of the segmented region highlighted as foreground areas and compare them with the stored results. The edge maps are obtained using a novel adaptive edge orientation based technique where orientation of the edge is used. Experimental results have shown that the discussed technique gives more than a 95% detection success rate when tested in different indoor and outdoor environments.

8663-32, Session 9

Retail video analytics: an overview and survey
(Invited Paper)

Quanfu Fan, Prasad Gabbur, Norman Haas, Sharathchandra Pankanti, Hoang Trinh, IBM Thomas J. Watson Research Ctr. (United States)

Today retail video analytics has gone beyond the tradi- tional domain of security and loss prevention by providing retailers insightful business intelligence such as store traffic statistics and queue data. Such information allows for en- hanced customer experience, optimized store performance, reduced operational costs, and ultimately higher profitabil- ity. This paper gives an overview of various camera-based applications in retail as well as the state-of-the-art computer vision techniques behind them. It also presents some of the promising technical directions for exploration in retail video analytics.

8663-33, Session 9

Video-CRM: understanding customer behaviors in stores

Ismail Haritaoglu, Anvato, Inc. (United States); Myron D. Flickner, David Beymer, IBM Almaden Research Ctr. (United States)

This paper describes two real-time computer vision systems created 10 years ago that detect and track people in stores to obtain insights of customer behavior while shopping. The first system uses a single color camera to identify shopping groups in the checkout line. Shopping groups are identified by analyzing the inter-body distances coupled with the cashier's activities to detect checkout transactions start and end times. The second system uses multiple overhead narrow-baseline stereo cameras to detect and track people, their body posture and parts to understand customer interactions with products such as "customer picking a product from a shelf". In pilot studies both systems demonstrated real-time performance and sufficient accuracy to enable more detailed understanding of customer behavior and extract actionable real-time retail analytics.

8663-35, Session 9

Human object articulation for CCTV video forensics

Iffat Zafar, Muhammad Fraz, Eran Edirisinghe, Loughborough Univ. (United Kingdom)

Video forensics applications often require the articulation of the key object of interest, namely a human object. For example it may be required to identify person wearing a red coloured top at the highest level of articulation. However it may also be useful to perform a further detailed articulation such as person wearing a short sleeved, white and blue top that includes text (or a logo) and a black trouser.

Articulating human objects in video forensics when performed in CCTV footage is challenging due to the presence of noise, limitations of resolution, changing levels and sources of scene illumination, variations of direction of view etc. We propose the use of image pre-processing pipeline that includes a multitude of enhancement operations that deals with the above challenges prior to performing object articulation. The objects of interests are identified by moving object detection followed by the binary classification of such objects into humans and non-humans. The identified human objects are further articulated by dividing the human figure into head, upper body and a torso area. The head area is checked for detecting the presence of skin tone and facial features to identify whether the person is walking towards the camera from the camera or alternatively identify the colour of hair. The top part of body is clustered into dominant colours that are used to determine the presence of skin tone to identify individuals wearing short sleeved shirts and also to determine multiple colours and presence of text in the top being worn. A similar analysis is carried out for the torso area enabling detailed articulation.

We provide detailed experimental results and an in-depth analysis providing subjective and objective metrics to prove level of performance achieved by the proposed image processing application on real CCTV footage.

Conference 8664: Imaging and Printing in a Web 2.0 World IV

Monday - Tuesday 4 –5 February 2013

Part of Proceedings of SPIE Vol. 8664 Imaging and Printing in a Web 2.0 World IV

8664-1, Session 1

Videos in tangible products

Reiner Fageth, Ralf Wieting, CeWe Color AG & Co. OHG (Germany)

High quality videos can be taken with nearly every camera, digital point and shoot cameras, DSLRs as well as smartphones. High quality means video with HD resolutions as well as high quality with respect to story-telling. These videos can be viewed on nearly every screen and easily shared via social networks. All of these use cases do not allow to monetize the large number of videos (snippets) taken. These videos are not linked to other personal data such as pictures, maps, text descriptions or any kind of tangible products.

Particularly these videos cannot be combined with the current state-of-the-art story telling, printed photo books. In the proposed paper an approach of implementing user-defined pictures out of the video implemented in the photo book to reach an even more compelling story as well as implementing the whole video via QR codes printed in the photo book is described. The QR technology is often used to link directly to a website, but not to personalized content. The implementation is done via a desktop based software to generate CEWE PHOTO BOOKS. Additionally the proposed technology also enables printed products to be easily personalized with personal messages, e.g. greeting cards or business cards.

8664-2, Session 1

HP Smart Marketing Suite: the future of dynamic content publishing

Galo Gimenez, Eduardo Argollo, Hewlett-Packard Co. (Spain); Jitendra Kumar, G. K. Tejaswi, Hewlett-Packard Labs. India (India)

In order to produce and deliver enterprise documents, the enterprise publishing space handles a wide range and a great load of structured, semi-structured and non-structured content. Because of their single touch-point focus, silo process orientation or data centralized organization, existent Content Management Systems fall behind on being a complete publishing solution, especially when focusing on medium and large enterprises. In this paper we introduce HP Smart Marketing Suite, a content centric platform that does content organization and publishing for large enterprises the right way. HP-SMS first client is HP itself where it provides real-time access to more than 7 million assets (including digital media and structured production information) stored in multiple distributed repositories. HP SMS services 30 thousand internal and 240 thousands external users. It is estimated that HP will save over $100 million during 3 years with the deployment of HP SMS.

8664-3, Session 1

TopicSelect: a platform for automated publications of Web content

Ildus Ahmadullin, Hewlett-Packard Labs. (United States); Yu Zhang, Xiao-Wei Wu, Hewlett-Packard China Co., Ltd. (China); Jerry Liu, Hewlett-Packard Labs. (United States)

No Abstract Available

8664-4, Session 1

The CloudBoard Research Platform: an interactive whiteboard for corporate users

John Barrus, Edward L. Schwartz, Ricoh Innovations, Inc. (United States)

No Abstract Available

8664-5, Session 2

Automatic page composition with nested sub-layouts

Andrew A. Hunter, Hewlett-Packard Labs. (United Kingdom)

This paper extends our previous work on the layout of cropped images. The new work demonstrates a solution to the more general challenge of laying out content with no inherent size or shape. The solution is particularly effective for deeply nested content such as magazine pages where layouts often contain other layouts in complex content hierarchies.

As we demonstrated at last year's conference, some images, especially consumer photographs, are best treated as mutable elements with aspect ratios and potential layouts that are mutually dependent. When elements being laid out are mutable sub-layouts of subsidiary content, the issues are similar. At each level of a document hierarchy, the space allocated for the elements depends both on the layout requirements for the other elements at the same level and also on the internal layout options of the elements at the next level down the hierarchy.

We will describe a solution in which elements provide data to the next level up the hierarchy that enables an initial allocation of space to balance the space distribution followed by a secondary allocation to optimize the layout aesthetics.

8664-6, Session 2

Graph-based layout analysis for PDF documents

Canhui Xu, Zhi Tang, Xin Tao, Yun Li, Cao Shi, Peking Univ. (China)

To increase the flexibility and enrich the reading experience of e-book on small portable screens, in this paper, a graph based method is proposed to segment the PDF pages which are difficult to cope with when using traditional XY cut and whitespace methods. Unlike the image based document analysis, digital born document has its inherent advantages like representing characters and images in explicit form, which can be straightforwardly exploited for layout analysis. To integrate the application of image-based document layout analysis methods and the inherent meta-data structure information provided by digital-born documents like PDF, all the elements in each page are processed, including text, image, graphic elements or decorations. Graph-based method developed can capture certain perceptually important non-local image characteristics for segmentation purposes. Instead of using pixel points or CC state, each page element or primitive corresponding to a vertex are constructed in the graph. All the text elements can be connected by establishing a neighborhood system. Delaunay tessellation is applied in this regard. Implementation of this method by Delaunay tessellation and feature-based classification is presented. Graph based method serving as a bottom-up method is used to group the page elements according to edge weights like the proximity and compactness. The font type and font size are used to in a top-down manner to obtain the desired block

segmentation. Hierarchically, text line detection is then accomplished by statistically filtering the edges with orientation. The experimental results on complex layout pages of PDF books have shown satisfactory performance.

8664-7, Session 2

Learning probabilistic document template models via interaction

Ildus Ahmadullin, Niranjan Damera-Venkata, Hewlett-Packard Labs. (United States)

No Abstract Available

8664-8, Session 2

Interaction paradigms for photo book layout on mobile devices

Philipp Sandhaus, CeWe Color AG & Co. OHG (Germany)

Getting from a collection of photos to the final printed photo book means a lot of effort for the end user: Photos have to be selected, be placed on the individual pages and have to be laid out in a nice and pleasing way. With the advent of mobile devices this process is also getting mobile. Here we are faced with a couple of challenges, e.g. limited screen size and less fine-grained manipulation capabilities of screen objects because of touch-based interaction. On the other hand, touch-based interaction is generally considered as much more intuitive as the objects on the screen are directly manipulated in contrast to the mouse as an input device. The challenge is to compensate these weaknesses.

With our system the user is able to very efficiently, roughly place text and photos on the single photo book pages and an automatic layout system automatically adjusts these rough layouts with the help of an automatic layout engine based on a set of carefully selected layout rules. The goal of this system is to provide the user with an easy way to very fast design a photo book on a mobile device in an intuitive way.

8664-9, Session 3

Non-Manhattan layout extraction algorithm

Aziza Satkhozhina, Purdue Univ. (United States); Ildus Ahmadullin, Hewlett-Packard Labs. (United States); Jan P. Allebach, Purdue Univ. (United States)

No Abstract Available

8664-10, Session 3

Automatic design of colors for magazine covers

Ali Jahanian, Jan P. Allebach, Purdue Univ. (United States); Jerry Liu, Qian Lin, Daniel Tretter, Eamonn O'Brien-Strain, Seungyon C. Lee, Nic Lyons, Hewlett-Packard Labs. (United States)

We propose a recommendation system for the automatic design of colors for the automatic process of magazine cover design. In the design of a magazine cover, making a set of decisions regarding the color distribution of the cover image and the colors of other graphical and textural elements is considered to be the color design concept. This concept addresses a number of subjective challenges: what set of colors is aesthetically pleasing yet also contributes to the purpose of the design, the legibility of textual elements, and the consistency of style with the class of magazine. Our solutions to automatic color design comprise the quantification of these challenges by deploying a number of well-known color theories. Furthermore, we use these color theories as scales to analyze the color design of some well-known magazine covers to delineate rules for automatic color design. As a
result, our system allows the user to upload a number of images, and it tells him/her which image is better, color-wise, for a design. This recommendation is done by a semantic extraction of a 3-color palette from the image. Based on this color palette, the system also makes a number of decisions for the color of the text.

8664-11, Session 3

Content-aware automatic cropping for consumer photos

Henry Tang, Daniel Tretter, Qian Lin, Hewlett-Packard Labs. (United States)

Consumer photos are typically authored once, but need to be retargeted for reuse in various situations. These include printing a photo on different size paper, changing the size and aspect ratio of an embedded photo to accommodate the dynamic content layout of web pages or documents, adapting a large photo for browsing on small displays such as mobile phone screens, and improving the aesthetic quality of a photo that was badly composed at the capture time. In this paper, we propose a novel, effective, and comprehensive content-aware automatic cropping (hereafter referred to as ``autocrop") method for consumer photos to achieve the above purposes. Our autocrop method combines the state-of-the-art context-aware saliency detection algorithm, which aims to infer the likely intent of the photographer, and the ``branch-and-bound" efficient subwindow search optimization technique, which seeks to locate the globally optimal cropping rectangle in a fast manner. Unlike most current autocrop methods, which can only crop a photo into an arbitrary rectangle, our autocrop method can automatically crop a photo into either a rectangle of arbitrary dimensions or a rectangle of the desired aspect ratio specified by the user. The aggressiveness of the cropping operation may be either automatically determined by the method or manually indicated by the user with ease. In addition, our autocrop method is extended to support the cropping of a photo into non-rectangular shapes such as polygons of any number of sides. It may also be potentially extended to return multiple cropping suggestions, which will enable the creation of new photos to enrich the original photo collections. Our experimental results show that the proposed autocrop method in this paper can generate high-quality crops for consumer photos of various types.

8664-12, Session 3

Automatic photobook: focusing on image selection and image layout based on content and composition

Shao-Fu Xue, Purdue Univ. (United States); Henry Tang, Daniel Tretter, Qian Lin, Hewlett-Packard Labs. (United States); Jan P. Allebach, Purdue Univ. (United States)

Automatic photo book generation should consider the diversity, coverage, aesthetics of the photos being selected and placement of the selected photos on the photobook. It is important that the algorithm identifies the photos that are important to the user and the algorithm generates the layouts which are aesthetically pleasing.

In this project, we proposed a complete workflow for generating auto photo books. Based on the proposed similarity measure, which takes account of the time of exposure, face, and theme information, the algorithm automatically finds representative photos that cover the concepts of the original input collection of photos. Each of the selected base photo goes into one page of the album. Then a few similar photos are selected with respect to each of the base photo as neighboring photos.

For all the potential images to put on the same page, dominant semantic lines and salient regions are identified in the photos. These visual clues are used to identify the major photo to put on the dominant position of the page, as well as to determine whether to put other potential photos onto the page. The decision is made from the visual path, visual distance and other aesthetics measures.

8664-13, Session 4

Visualization and exploration for recommender systems in enterprise organization

Zachi Karni, Hewlett-Packard Labs. Israel Ltd. (Israel); Lior Shapira, Microsoft Research, (United States)

We present a method for visualization, navigation and exploration of information assets in enterprise organizations. There has been a proliferation of information assets such as, emails, documents, presentations and more in enterprises. Tools are required which assist users in locating relevant data for their work, in the context of their organization. Several methods such as search engines and recommender systems have been developed to handle big-data and efficiently locate relevant assets. In these the results are usually presented with no context and the user has no ability to influence, refine or to further explore. In our visualization method a user can see the assets located or recommended for him, but in addition also gains an understanding why. Focusing on a user, we render a planar visualization listing a set of recommended assets. The assets are organized such that similar assets reside nearby on the screen, centered around real-time generated categories. We use a combination of iconography, text and tag clouds, with maximal use of screen real estate, and keep items from overlapping to produce our results. Our system enables visual exploration of the results and lets the user refine the quires according to his needs.

8664-14, Session 4

Augmented distinctive features with color and scale invariance

Yan Liu, Xiaoqing Lu, Yeyang Qin, Zhi Tang, Peking Univ. (China); Jianbo Xu, Peking Univ. Founder Group Corp. (China)

For objects with the same texture but different colors, it is difficult to discriminate them with the traditional scale invariant feature transform descriptor (SIFT), because it works for grayscale images only. Thus it is important to keep a high probability to make sure that the used key points are couples of correct pairs. In addition, mean distributed key points are much more expected than over dense and clustered key points for image match and other applications. In this paper, we analyze these two problems. First, we propose a scale-invariant method to extract a more mean distributed key points relying on illumination intensity invariance but object reflectance sensitivity variance variable. Second, we modify the key point's canonical direction accumulated error by dispersing each pixel's gradient direction on a relative direction around the current key point. At last, we build the descriptors on a Gaussian pyramid and match the key points with our enhanced two-way matching regulations. Experiments are performed on the Amsterdam Library of Object Images dataset and some synthetic images manually. The results show that the extracted key points have better distribution character and larger number than SIFT. The feature descriptors can well discriminate images with different color but with the same content and texture.

8664-15, Session 4

Video text tracking and recognition

Baokang Wang, Changsong Liu, Xiaoqing Ding, Tsinghua Univ. (China)

No Abstract Available

8664-16, Session 4

Image vectorization using blue-noise sampling

Jiaojiao Zhao, Jie Feng, Bingfeng Zhou, Peking Univ. (China)

Abstract:

Current image vectorization techniques mainly deal with images with simple and plain colors. For full-color photographs, many difficulties still exist in object segmentation, feature line extraction, and color distribution reconstruction, etc.

In this paper, we propose a high-efficiency image vectorization method based on blue-noise sampling. The method first generates a set of sampling points on the image plane by an improved adaptive error-diffusion sampling algorithm. The point set has ideal blue-noise property and can well preserve the features in the image. Then after a triangulation on this point set, the information such as colors and gradient in the image are recorded in the resulting triangle mesh to form a vector image. After certain image editing, e.g. scaling or transforming, the whole image can be reconstructed by color interpolating inside each triangle.

Experiments show that the method has high performing efficiency and abilities in feature-preserving and image reconstruction. This image vectorization method will bring benefits to many applications, e.g. image compressing, editing, transmitting and resolution enhancement, especially in the environments with limited computational resources such as mobile devices.

8664-17, Session 4

Non-iterative normalized feature extraction in large viewpoint variances based on PCA of gradient

Jian Zhang, Song Cao, Di Wen, Tsinghua Univ. (China)

Effective local feature extraction is fundamental for content-based image analysis and retrieval in multimedia application. However, it is difficult to achieve distinguishable local features in large viewpoint variances. In this paper, we propose a novel non-iterative approach of normalized feature extraction in large viewpoint variances, which adapts local regions to rotation, scale variance and rigid distortion from affine transformation. Our approach is based on two key ideas: 1) Localization and scale selection can be directly achieved with the centroid and covariance matrix of pixels distribution in a local region. 2) Principal Component Analysis (PCA) on gradient gives information on texture, thus it can be used to get a resampled region which is isotropic in terms of variance of gradient. Experiments demonstrate that our normalized approach has significant improvement on feature extraction in large viewpoint variances.

8664-18, Session 5

Geometric enumerated chrominance watermark embed for spot colors

Alastair M. Reed, Robert Lyons, John Stach, Digimarc Corp. (United States)

The majority of packaging is printed using spot colors mainly for reasons of cost and color consistency, and to achieve a wide color gamut on the package. Most watermarking techniques are designed to embed a watermark in either cyan, magenta, yellow or black for printed images or red, green and blue for digital images that are being displayed.

Our method is designed to address the problem of watermarking spot color images. An image containing 2 or more spot colors is embedded with a watermark in 2 of the spot colors, with the maximum signal strength within a user selectable visibility constraint.

The user can insert the maximum watermark signal, while meeting the required visibility constraint. The method has been applied to the case of two spot colors and images have been produced which are more than twice as robust to Gaussian noise as a single color image which is embedded using a luminance only watermark to the same visibility.

8664-19, Session 5

Dual resolution 2D color barcode

Zhigang Fan, Xerox Corp. (United States); Yonghui Zhao, Apple Inc. (United States); Shenge Wang, Qalcomm Inc. (United States); Hengzhou Ding, Xerox Corp (United States)

Quick Response (QR) codes and other 2-dimensional (2-D) barcodes have become a widely accepted method for automatically identifying and tracking objects. The ability of mobile devices to read the encoded information has made these codes increasingly popular in a growing list of applications, such as healthcare, electronics and product information on packages.

Currently, the typical QR code is around 2cm x 2cm in size. However, many small packages require that 2-D codes be 0.75cm x 0.75cm. Unfortunately, detection of these small codes is challenging for many mobile devices. Blurring, mainly caused by optical limitations of the camera systems and hand motion, magnifies its impact on small images contain minute details.

In this paper, a barcode system is presented which includes an encoder that configured to generate a 2-D color barcode with a dual resolution structure. It contains a high resolution layer that is coded in luminance and is in consistency with the conventional 2-D barcode, and a low resolution layer providing additional error checking information, that is coded in chrominance and is robust to blurring. The proposed barcode is compatible to its underlying conventional black and white 2-D barcode as it can be read by their decoders. Its advantage is additional reliability when a color decoder is used. In particular, it enhances the decoding accuracy for devices such as mobile devices for barcodes printed in small sizes.

8664-20, Session 5

HP EOS cross platform mobile solution

Junhua Fu, Hua Zhang, Hewlett-Packard China Co., Ltd. (China); Jerry Liu, Hewlett-Packard Labs. (United States)

No Abstract Available

8664-21, Session K1

Tangible imaging systems *(Keynote Presentation)*

James A. Ferwerda, Rochester Institute of Technology (United States)

No Abstract Available

8664-22, Session 7

Event detection from the microblog

Jingxuan Li, Florida International Univ. (United States); Wei Peng, Tong Sun, Xerox Corp. (United States)

Event detection has been attracting researchers' attention for a long time. To detect events from the microblog is becoming a hot topic recently because it reflects what happened in the world in a timely fashion, thus can help people monitor real time events. For example, a mobile service provider may want to monitor users' feedback through the microblog. However, given the collection of microblog documents, e.g., Tweets from Twitter, performing event detection is non-trivial for the natural of microblog documents – short. Meanwhile, online microblog documents usually means text streams available on the web for every minute. How to quickly detect current events from microblog documents is a challenge. In this work, we propose and implement a "New" and "Anomalous" event detection framework which can be efficiently applied over microblog streams.

8664-23, Session 7

Person-based video summarization and retrieval by tracking and clustering temporal face sequences

Tong Zhang, Hewlett-Packard Labs. (United States); Di Wen, Tsinghua Univ. (China); Xiaoqing Ding, Tsinghua University (China)

People are often the most important subjects in videos. It is highly desired to automatically summarize the occurrence of different people in a large collection of video and quickly find the video clips containing a particular person among them. In this paper, we present a person-based video summarization and retrieval system, VideoWho, which extracts temporal face sequences in videos and groups them into clusters, with each cluster containing video clips of the same person. This is accomplished based on advanced face detection and tracking algorithms, together with a semi-supervised face clustering approach. The system achieved good clustering accuracy when tested on a hybrid video set including home video, TV plays and movies. On top of this technology, a number of applications can be built, such as automatic summarization of major characters in videos, person-related video search on the Internet and personalized UI systems etc. Especially, we are making this as a cloud service for developers and consumers to manage video content.

8664-24, Session 7

Building an internet-scale web clipping system

Eamonn O'Brien-Strain, Hewlett-Packard Labs. (United States)

No Abstract Available

8664-25, Session K2

That thing in your pocket is really a computer!: the future of mobile imaging *(Keynote Presentation)*

Edward J. Delp III, Purdue Univ. (United States)

No Abstract Available

Conference 8665: Media Watermarking, Security, and Forensics 2013

Tuesday - Thursday 5 –7 February 2013

Part of Proceedings of SPIE Vol. 8665 Media Watermarking, Security, and Forensics 2013

8665-1, Session K1

Converting watermark research into a technology startup *(Keynote Presentation)*

Jaap Haitsma, Civolution (United States)

In this talk I will present my personal experience of how the watermark and later on fingerprint research results of just a small team of scientists in Philips Research (of which I was one) in the late 90's was converted into a company called Civolution that is currently employing over a hundred people that work on providing watermarking and fingerprinting solutions to customers worldwide. Furthermore it will touch upon the different applications and deployments of watermarking and fingerprinting technology and provide some insights in the challenges ahead. I hope this talk can give inspiration and some advice to other scientists to convert their research results into a technology startup. Having your own algorithms solve problems or give enhanced experiences to real customers is a very gratifying experience.

8665-2, Session 1

Insertion deletion robust audio watermarking: a set theoretic, dynamic programming approach

Andrew Nadeau, Gaurav Sharma, Univ. of Rochester (United States)

To date, desynchronization attacks are a key problem for audio watermarks, which have limited their success in applications such as content authentication and digital rights management. This paper proposes an audio watermarking scheme robust to both desynchronization from random insertions and deletions, and non-desynchronizing perturbations from MP3 compression. The proposed scheme embeds multiple spread spectrum (SS) watermarks in the poly-phase quadrature modulated filterbank (PQMF) subbands of an audio signal using a set theoretic framework for informed embedding. The set theoretic framework allows constraints on MP3 robustness, perceptual fidelity, and detectability to be enforced concurrently for the multiple watermarks. Blind correlation detection of short watermark blocks followed by dynamic time warping (DTW) efficiently searches for a optimal sequence of block detections in spite of insertions and deletions. Unlike prior art, DTW is integrated into SS detection and does not require the original audio signal for alignment.

Novel contributions of the proposed audio watermarking scheme include:

1) robustness to non-desynchronizing perturbations using a set theoretic framework, 2) insertion and deletion robustness using DTW as part of blind detection, and 3) reduction of false positives typical of high dimensional searches by cross validation of multiple watermarks.

8665-3, Session 1

Impeding forgers at photo inception

Matthias Kirchner, International Computer Science Institute (United States); Peter Winkler, Hany Farid, Dartmouth College (United States)

We describe a new concept for making photo tampering more difficult and time consuming, and therefore more prone to error. We record the camera preview and camera motion in the moments just prior to image capture. This information is packaged along with the full resolution image. Any subsequent manipulation of the final image will have to be propagated to be consistent with this data – a decidedly more difficult undertaking.

8665-4, Session 1

Watermark embedding in optimal color direction

Robert Lyons, Alastair M. Reed, John Stach, Digimarc Corp. (United States)

To watermark spot color packaging images one changes available color inks to create a watermark signal. By changing the ink densities at a point one can embed the watermark signal in different color directions. We describe the optimal color direction that embeds the maximum signal while keeping the visibility below some acceptable value. The optimal color direction depends on the definition of the watermark signal, the starting color for the image region and the ink density constraints.

After a description of the general problem we shall describe two ink embedding and try to find the optimal direction that will maximize robustness at a given visibility. If we define the watermark signal as $(R + G + B) / 3$, then the optimal two color embedder will insert color with low or zero luminance changes. We compare the optimal color embedder to a single color embedder.

8665-5, Session 1

Video game watermarking

Waldemar Berchtold, Marcel Schäfer, Fraunhofer-Institut für Sichere Informations-Technologie (Germany); Huajian Liu, Fraunhofer Institute SIT (Germany); Fábio Touceira Takahashi, University of Porto (Portugal); André Schmitz, waza! services UG (Germany); Sascha Zmudzinski, Martin Steinebach, Fraunhofer-Institut für Sichere Informations-Technologie (Germany); Jonas Wieneke, Technical University of Darmstadt (Germany)

The publishers of video games suffer from illegal piracy and information leakage caused by end-consumers, "release groups"or insiders shortly after or even before the offcial release of a new video game. Mechanisms to prevent or at least postpone this illegal redistribution are DRM or copy protection mechanisms. However, these mechanisms are very unpopular, because they restrict the customers in playing the game and demand a high administration effort from the developers and/or distributors. Even worse, most copy protection mechanisms have proven to be insecure as "patches" for circumvention usually are available quickly and easy to get. To satisfy the challenges of security and usability, this work introduces the idea of using digital watermarking to protect all available and suitable media types and software binaries contained in a video game. A three-layered watermarking deployment approach along the production chain is proposed to detect leakage in the release phase as well as during the development process of a boxed video game. The proposed approach features both copyright watermarking and collusion secure fingerprints embedded as transaction watermark messages in components of video games. We discuss the corresponding new challenges and opportunities. In addition, a prototype watermarking algorithm is presented to demonstrate the adaption necessity of classical image watermarking when applied to video games to satisfies the requirements for transparency, security as well as performance. The watermark capacity is significantly increased while inter-media and inter- file embedding is enabled and the associated synchronization challenge is solved by robust hashes.

Conference 8665

8665-6, Session 2

Banknote authentication with mobile devices

Volker Lohweg, Jan Leif Hoffmann, Helene Dörksen, Roland Hildebrand, Eugen Gillich, Ostwestfalen-Lippe Univ. of Applied Sciences (Germany); Jürg Hofmann, Johannes Schaede, KBA-NotaSys S.A. (Switzerland)

Maintaining confidence in security documents, especially banknotes, is and remains a major concern for the central banks in order to maintain the stability of the economy around the world. In this paper we describe an image processing and pattern recognition approach which is based on the Sound-of-Intaglio conceptfor the usage in smart devices such as smartphones. Today, in many world regions smartphones are in use. These devices become more and more computing units, equipped with resource-limited, but effective CPUs, cameras with illumination, and flexible operating systems. Hence, it appears to be obvious, to apply those smartphones for banknote authentication, especially for visually impaired persons. Our approach shows that those devices are capable of processing the data under the constraints of image quality and processing power. Even, pointing out strictly that a mobile device as such is not an industrial product for harsh environments, it is possible to use mobile devices for banknote authentication. The concept is based on a new strategy for constructing adaptive wavelets for the analysis of different print patterns on a banknote. Furthermore, a banknote specific feature vector is generated which describes an authentic banknote effectively. A multi-stage Linear-discriminant-analysis classifier generates stable and reliable output.

8665-7, Session 2

Visibility enhancement and validation of segmented latent fingerprints in crime scene forensics

Andrey Makrushin, Otto-von-Guericke-Univ. Magdeburg (Germany); Tobias Kiertscher, Fachhochschule Brandenburg (Germany); Mario Hildebrandt, Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany); Claus Vielhauer, Fachhochschule Brandenburg (Germany)

Forensic investigators are permanently looking for novel technologies for fast and effective recovering of latent fingerprints at a crime scene. Traditionally, this work is done manually and therefore considered very time consuming. Highly skilled experts apply chemical reagents to improve visibility of traces and use digital cameras or adhesive tape to lift prints. Through an automation of the surface examination, larger areas can be investigated faster. This work amplifies the experimental study on capabilities of a chromatic white-light sensor (CWL) regarding the contact-less lifting of latent fingerprints from differently challenging substrates. The crucial advantage of a CWL sensor compared to taking digital photographs is the simultaneous acquisition of luminance and topography of the surface, extending the standard twodimensional image processing to the analysis of three-dimensional data. The paper focuses on the automatic validation of localized fingerprint regions. In contrast to statistical features from luminance data, previously used for localization, we propose the streakiness of a pattern as the basic feature indicating the fingerprint presence. Regions are analyzed for streakiness using both luminance and topography data. As a result, the human experts significantly save time by dealing with a limited number of approved fingerprints. The experiments show that the validation performance in terms of equal error rate does not exceed 6% even on very challenging substrates regarding high-quality fingerprints.

8665-8, Session 2

Printed fingerprints at crime scenes: a faster detection of malicious traces using scans of confocal microscopes

Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

Fingerprint traces are an important part of forensic investigations to identify potential perpetrators. With the advent of the possibility of printing traces for quality assurance purposes it is also possible to place malicious traces on crime scenes. In forensics examiners are aware of multiple identical traces, e.g. produced by stamping fingerprints. The technique of printing fingerprints using artificial sweat allows for creating different versions of the same fingerprint, similar to the residue from a finger which is almost never 100 percent identical to an other latent fingerprint. Hence, Kiltz et al. (2011) introduce a first framework for the detection of such malicious traces in subjective evaluations based on dot patterns of amino acid. Hildebrandt et al. (2012) introduce a first automated approach for the detection of printed fingerprints using high resolution scans from a Chromatic White Light sensor. However, the reported recognition accuracy is insufficient for forensic investigations.

In this paper we propose an improved feature extraction for scans using a confocal microscope to reduce the overall analysis time and to increase the recognition accuracy. Our first preliminary evaluation on 55 printed and 55 real fingerprints on an overhead foil indicates a significant reduction of the necessary analysis time. The recognition accuracy in this first test set is increased to 100 percent.

8665-9, Session K2

YouTube content ID: background, challenges, and the roadmap ahead *(Keynote Presentation)*

Thabet Alfishawi, Google (United States); David Erb, Google Zürich (Switzerland)

YouTube launched Content ID in 2007 as a tool to enable content owners to identify content they own automatically and applying their desired policy on it: monetize, block, takedown or track. Over the years, Content ID evolved to become an integral part of YouTube's ecosystem. Content ID is among the most comprehensive audiovisual databases in the world as it scans over 100 years of user-uploaded video every day against more than 100 years of reference files.

In this talk we will provide background on Content ID, origins, evolution and inner workings. Then we'll take a deeper dive to focus on the current challenges facing Content ID breaking them down in three categories: technical, business and behavioral. Finally we'll close with an overview of the roadmap ahead for Content ID.

8665-10, Session 3

Sensor fingerprint digests for fast camera identification from geometrically distorted images

Miroslav Goljan, Jessica Fridrich, Binghamton Univ. (United States)

In camera identification using sensor fingerprint, it is absolutely essential that the fingerprint and the noise residual from a given test image be synchronized. If the signals are desynchronized due to a geometrical transformation, fingerprint detection becomes significantly more complicated. Besides constructing the detector in an invariant transform domain (which limits the type of the geometrical transformation) a more general approach is to apply a generalized likelihood ratio detector, which involves a search for the unknown nuisance parameters of the transform. This search, however, is in general very expensive as it requires numerous resamplings of the entire image (or fingerprint). In this abstract, we propose a measure that significantly reduces the search complexity by reducing the need to resample the entire image to a much smaller subset of the signal called the fingerprint digest. The technique can be applied to an arbitrary geometrical distortion, including non-linear lens-distortion correction. The full version of this paper will include extensive experiments and a mathematical framework for selecting the digest that minimizes the detection error for selected common geometrical transforms.

8665-11, Session 3

Case studies and further improvements on source camera identification

Kenji Kurosawa, Kenro Kuroki, Ken'ichi Tsuchiya, Naoaki Igarashi, Norimitsu Akiba, National Research Institute of Police Science (Japan)

The authors have conducted various forensic examinations of images in the police laboratory. In this paper, actual case examples and further improvements on source camera identification are shown. The specific topics are following (a) to (d).

(a) In order to improve performance of source camera identification, the hybrid identification method using dark current and PRNU noise is proposed. The preliminary experimental result showed that identification performance was improved with the proposed method.

(b) The source camera identification has been performed in the actual five criminal cases, such as the murder case, and so on. The procedure was a sequential examination of hot pixel coordinates validation followed by the similarity evaluation of sensor noise pattern. The authors could clearly prove that the questioned criminal scenes had been recorded by the questioned cameras in four cases of the five.

(c) The experiment results with twenty CCD modules of the same model showed that individual camera identification for dark images was possible by using non-uniform nature of CCD charge transfer.

(d) It was also shown that the source camera identification method can be applied for detecting forgery on the time axis of video sequences, such as insertion of a scene obtained by the other cameras.

8665-12, Session 3

Forensic analysis of interdependencies between vignetting and radial lens distortion

André Fischer, Thomas Gloe, Technische Univ. Dresden (Germany)

This abstract starts a discussion on the forensic analysis of interdependencies between lens aberrations on the example of vignetting and radial lens distortion. Starting with a brief discussion on the background, we assume the optical centre of vignetting and radial lens distortion similar and expect further dependencies in the appear- ance of both. Estimating aberration parameters in combination can help to increase the accuracy and makes investigations more reliable. After a short discussion of our test setup, we give a rough list of experimental results. We propose a novel method to make the characteristic of vignetting inherent in an image visible using a vector plot. This allows forensic investigators to inspect and compare patterns of vignetting visually. Our preliminary results document the influence of camera rotation and indicate previously unknown challenges in the forensic analysis. However, estimating camera rotation might help to authenticate metadata tag 'orientation'. Unexpectedly, the estimated optical centre for vignetting and radial lens distortion do not coincide, which might be a problem of the employed models or an imperfect illumination of the checkerboard images. Space limitations prohibit a complete documentation of our study and we will provide detailed results in the final paper in case of acceptance.

8665-13, Session 3

A sneak peek into the camcorder path

Cherif Ben Zid, Séverine Baudry, Technicolor S.A. (France); Bertrand Chupeau, Technicolor R&D France (France); Gwenaël Doërr, Technicolor S.A. (France)

The camcorder path refers to the communications channel where a video file is rendered on a display device and re-captured by a camera recorder. Several signal processing primitives are expected to be robust against such distortion. For instance, forensic video watermarks should be recovered even if the pirate camcords the TV/cinema screen onto which the movie is displayed. Video-based second-screen applications implicitly expect the underlying watermarking or

fingerprinting technology to survive camcording. Still, robustness to this communications channel is hardly evaluated nowadays because benchmarking campaigns simply prove to be too cumbersome in practice.

Camcording displayed content modifies the video essence in a number of ways: geometric distortions, temporal transforms, non-uniform and varying luminance transformations, saturation, color alteration, etc. It is therefore challenging to design a simulator that would mimic the camcorder path and that could thus be used in a systematic manner for benchmarking while avoiding logistics hurdles. As a first step in this direction, we focus in this study solely on luminance transforms. In particular, we investigate three different phenomena, namely: (i) the steady state luminance response, (ii) the spatial non uniformity, and (iii) the transient luminance response.

8665-14, Session 4

Ballistic examinations based on 3D data: a comparative study of probabilistic Hough Transform and geometrical shape determination for circle-detection and segmentation of cartridge bottoms

Robert Fischer, Fachhochschule Brandenburg (Germany); Claus Vielhauer, Otto-von-Guericke Univ. Magdeburg (Germany); Mario Hildebrandt, Otto-von-Guericke University Magdeburg (Germany); Stefan Kiltz, Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

The application of contact-less optical 3D sensing techniques yielding digital data for the acquisition of toolmarks on forensic ballistic specimens found at crime scenes, as well as the development of computer-aided, semi-automated firearm identification systems that are using 3D information, are currently emerging fields of research with rising importance. Traditionally, the examination of forensic ballistic specimen is done manually by highly skilled forensic experts using comparison microscopes. A partly automation of the comparison task promises examination results that are less dependent on subjective expertise and furthermore a reduction of the manual work needed. One necessary requirement for the examination of forensic ballistic specimens is a reliable circle-detection and segmentation of cartridge bottoms. This information is later used for example for alignment and registration tasks, determination of regions of interest, and locally restricted application of complex feature-extraction algorithms. In this work we are using a Keyence VK-X 105 laser-scanning confocal microscope to acquire a very high detail topography image, a laser-intensity image, and a color image of the assessed cartridge bottoms simultaneously. The work is focused on a comparison of Hough-Transform and geometric-shape determination for circle-detection on cartridge bottoms using 3D as well as 2D information. We compare the pre-processing complexity, the required processing time, and the ability for a reliable detection of all desired circles. We assume that the utilization of geometric-shape detection can reduce the required processing time due to a less complex processing. For application of shape determination as well as for Hough-Transform we expect a more reliable circle-detection when using additional 3D information. Our first experimental evaluation, using 100 cartridges shot from 3 different firearms shows positive tendency to verify these suppositions.

8665-15, Session 4

Photocopier forensics based on arbitrary text characters

Changyou Wang, Xiangwei Kong, Shize Shang, Dalian Univ. of Technology (China); Xin'gang You, Beijing Institute of Electronic Technology and Application (China)

A photocopied document can provide characteristics of the photocopier so as to distinguish which photocopier it comes from. While how to extract the optimal intrinsic features is critical in photocopier forensics. In this paper, character texture features are

extracted as the intrinsic features and a photocopier forensics method based on the texture analysis of arbitrary text characters is proposed. Firstly, image preprocessing process is practiced to get individual character images. Next three sets of texture features are extracted from each individual character, including the gray level features, the gradient differential matrix (GDM) features and the gray level-gradient co-occurrence matrix (GLGCM) features. Then each individual character in a document is classified using a Fisher classifier and a majority vote scheme is performed on the character classification results to identify the source photocopier. Experimental results on seven photocopiers prove the effectiveness of the proposed method and an average character classification accuracy of 88.47% can be achieved.

8665-16, Session 4

Accelerating video carving from unallocated space

Hari Kalva, Florida Atlantic Univ. (United States); Anish Parikh, Nirma Univ. of Science & Technology (India); Avinash Srinivasan, George Mason Univ. (United States)

Video carving, the process of recovering video from unallocated disk space, has become an essential tool in digital forensics [1, 2, 3]. There has been some prior work on video carving including open source tools such as Defraser. Hard disk capacity has continued to increase and 1 TB disks are now common on consumer grade laptops and PCs. As the disk capacity grows, the complexity of video carving grows as the amount of unallocated space also grows. In this paper we present algorithms to accelerate video carving.

8665-17, Session IPI

A histogram shifting based RDH scheme for H.264/AVC with controllable drift

Zafar Shahid, LIRMM Labs (France); William Puech, Univ. Montpellier 2 (France)

This paper presents an efficient method for reversible data-hiding (RDH) in H.264/AVC intra bitstream in an open-loop fashion. The proposed scheme has a minimal controllable drift, which is proportional to the payload. The data is reversibly embedded in quantized transform coefficients (QTCs) using histogram shifting outside the reconstruction loop. In general histogram shifting based RDH methods for QTCs, there is a mismatch on encoder and decoder side because of modification of boundary pixel which are used for spatial intra prediction. It results in a severe degradation of the visual quality of watermarked video. In the proposed scheme, we have solved a system of linear equations which makes it possible to perform embedding in the QTCs without changing the value of the boundary pixels. It makes it possible to embed data in the transform domain without intra prediction drift. Moreover, the distortion is proposed to only those bits of payload which are '0'. Experimental result shows that the proposed scheme has a minimal increase in bitrate and guarantees the complete recovery of the original video content after extraction of the embedded content.

8665-19, Session 5

On the role of side information in steganography in empirical covers

Jessica Fridrich, Binghamton Univ. (United States)

In an attempt to alleviate the negative impact of unavailable cover model, some steganographic schemes utilize the knowledge of the so-called "precover" when embedding secret data. The precover is typically a higher-resolution (unquantized) representation of the cover, such as the raw sensor output before it is converted to an 8-bit per channel color image. The precover object is only available to the sender but not to the Warden, which seems to give a fundamental advantage to the sender. In this paper, we provide theoretical insight for why side-informed embedding schemes for empirical covers might

provide high level of security. By adopting a piece-wise polynomial model corrupted by AWGN for the content, we prove that when the cover is sufficiently non-stationary, embedding by minimizing distortion w.r.t. the precover is more secure than by preserving a model estimated from the cover (the so-called model-based steganography). Moreover, the side-informed embedding enjoys four times lower steganographic Fisher information than LSB matching.

8665-20, Session 5

Embedding operations and locations for steganography in H.264 video

Andreas Neufeld, Andrew D. Ker, Univ. of Oxford (United Kingdom)

This work studies the fundamental building blocks for steganography in H.264 compressed video: the embedding operation and the choice of embedding locations. Our aim is to inform the design of better video steganography, a topic on which there has been relatively little publication so far. We determine the best embedding option, from a small menu of embedding operations and locations, as benchmarked by an empirical estimate of Maximum Mean Discrepancy (MMD) for first- and second-order features extracted from a video corpus. A highly-stable estimate of MMD can be formed because of the large sample size. The best embedding operation (so-called F5) is identical to that found by a recent study of still compressed image steganography, but in video the options for embedding location are richer and the optimal choice involves unequal use of luma and chroma channels.

8665-21, Session 5

Video steganography with multi-path motion estimation

Yun Cao, Xianfeng Zhao, Fenghua Li, Institute of Information Engineering of Chinese Academy of Sciences (China); Nenghai Yu, Univ. of Science and Technology of China (China)

This paper proposes a novel video steganography during motion estimation. Compared with existing schemes, the new approach has been enhanced from two aspects to improve steganographic security. First, to reduce the single change distortion, multiple quick search functions are employed to generate optimized alternates for motion vector replacement. Secondly, to improve both the embedding and computational efficiencies, a flexible embedding structure is designed to perform matrix embedding.

8665-22, Session 6

Random projections of residuals as an alternative to co-occurrences in steganalysis

Vojtech Holub, Jessica Fridrich, Tomas Denemark, Binghamton Univ. (United States)

Today, the most reliable detection of steganography in empirical cover sources, such as digital images, has been achieved using machine-learning by representing images with co-occurrence matrices of neighboring noise residual samples computed using local pixel predictors. In this paper, we propose an alternative statistical description by binning random projections of residuals from a local neighborhood. The size and shape of the neighborhood allow the steganalyst to further diversify the statistical description and thus improve steganalysis detection. Other key advantages of this approach include the possibility to model longer-range dependencies among pixels and making use of information that was previously underutilized in the marginals of co-occurrences. Moreover, the proposed approach is much more flexible, allowing the steganalyst to better balance performance and feature dimensionality (complexity). We demonstrate the usefulness of the proposed approach on highly adaptive embedding schemes, such as HUGO and WOW.

8665-23, Session 6

The challenges of rich features in universal steganalysis

Tomas Pevny, Czech Technical Univ. in Prague (Czech Republic); Andrew D. Ker, Univ. of Oxford (United Kingdom)

No Abstract Available

8665-24, Session 6

Exploring multitask learning for steganalysis

Julie M. Makelberge, Andrew D. Ker, Univ. of Oxford (United Kingdom)

This paper introduces a new technique for multi-actor steganalysis. In conventional settings, it is unusual for one actor to generate enough data to be able to train a personalized classifier. On the other hand, in a network there will be many actors, between them generating large amounts of data. Prior work has pooled the training data, and then tried to deal with its heterogeneity. In this work, we use multitask learning to account for differences between actors' image sources, while still sharing domain (globally-applicable) information. We tackle the problem by learning separate feature weights for each actor, and sharing information between the actors through the regularization. This way, the domain information that is obtained by considering all actors at the same time is not disregarded, but the weights are nevertheless personalized. This paper explores whether multitask learning improves accuracy of detection, by benchmarking new multitask learners against previous work.

8665-25, Session 7

A cost-effective decision tree based approach to steganalysis

Liyun Li, Polytechnic Institute of New York Univ. (United States); Husrev T. Sencar, TOBB Univ. of Economics and Technology (Turkey); Nasir Memon, Polytechnic Institute of New York Univ. (United States)

An important issue concerning real-world deployment of steganalysis systems is the computational cost of ac- quiring features used in building steganalyzers. Conventional approach to steganalyzer design crucially assumes that all features required for steganalysis have to be computed in advance. However, as the number of features used by typical steganalyzers grow into thousands and timing constraints are imposed on how fast a decision has to be made, this approach becomes impractical. To address this problem, we focus on machine learning aspect of steganalyzer design and introduce a decision tree based approach to steganalysis. The proposed steganalyzer system can minimize the average computational cost for making a steganalysis decision while still maintaining the detection accuracy. To demonstrate the potential of this approach, a series of experiments are performed on well known steganography and steganalysis techniques.

8665-26, Session 7

Quantitative steganalysis using rich models

Jan Kodovsky, Jessica Fridrich, Binghamton Univ. (United States)

In this paper, we propose a regression framework for steganalysis of digital images that utilizes the recently proposed rich models -- high-dimensional statistical image descriptors that have been shown to substantially improve classical (binary) steganalysis. Our proposed system is based on gradient boosting and utilizes a steganalysis-specific variant of regression trees as base learners. The conducted experiments confirm that the proposed system outperforms prior quantitative steganalysis (both structural and feature-based) across a wide range of steganographic schemes: HUGO, LSB replacement, nsF5, BCHopt, and MME3.

8665-27, Session 8

Stegatone performance characterization

Yung-Yao Chen, Purdue Univ. (United States); Robert Ulichney, Hewlett-Packard Labs. (United States); Jan P. Allebach, Purdue Univ. (United States); Matthew Gaubatz, Hewlett-Packard Labs. (United States); Stephen Pollard, Hewlett-Packard Labs. (United Kingdom)

Embedding data in hard copy is in widespread use for applications that include pointing the reader to on-line content by means of a URL, tracing the source of a document, labeling and packaging. Most solutions involve placing overt marks on the page. The most common are 1D, 2D, and 3D (color) barcodes. However, while barcodes are a popular means for encoding information for printed matter, they add unsightly overt content.

In order to avoid such overt content, Stegatones are clustered-dot halftones that encode a data payload by single-pixel shifts of selected dot-clusters. In Stegatone, we can embed information in images or graphics – not in the image file as is done in traditional watermarking, but in the halftone on the printed page. However, the recovery performance of stegatones is not well understood across a wide variety of printing technologies, models, and resolutions, along with variations of scanning resolution. It would thus be very useful to have a tool to quantify stegatone performance under these variables. The results would then be used to better calibrate the encoding system. We conduct three test procedures to characterize Stegatone performance. The experimental results characterize Stegatone performance for a number of printers, scanners, and resolutions.

8665-28, Session 8

Image tampering localization via estimating the non-aligned double JPEG compression

Lanying Wu, Xiangwei Kong, Bo Wang, Shize Shang, Dalian Univ. of Technology (China)

In this paper, we present an efficient method to localize the forged parts in a tampered JPEG image. The forged region usually undergoes a different JPEG compression with the background region in JPEG image forgeries. When a JPEG image is cropped to another host JPEG image and resaved in JPEG format, the JPEG block grid of the tampered region often mismatches the JPEG block grid of the host image with a certain shift. This phenomenon is called non-aligned double JPEG compression (NA-DJPEG). In this paper, we identify different JPEG compression forms by estimating the shift of NA-DJPEG compression. Our shift estimating approach is based on the percentage of non zeros of JPEG coefficients in different situations. Compared to previous work, our tampering localization method (i) performances better for dealing with small image size, (ii) is robust to common tampering processing such as resizing, rotating, feathering and so on, (iii) doesn't need a image database to train a machine learning based classifier or to get a proper threshold.

Conference 8666: Visual Information Processing and Communication IV

Wednesday - Thursday 6 –7 February 2013

Part of Proceedings of SPIE Vol. 8666 Visual Information Processing and Communication IV

8666-1, Session 1

Fairness issues in resource allocation schemes for wireless visual sensor networks

Katerina Pandremmenou, Lisimachos P. Kondi, Konstantinos E. Parsopoulos, Univ. of Ioannina (Greece)

This work addresses the problem of fairness and efficiency evaluation of various resource allocation schemes for wireless visual sensor networks (VSNs). These schemes are used to optimally allocate the source coding rates, channel coding rates, and power levels among the nodes of a wireless direct sequence code division multiple access (DS-CDMA) VSN. All of the considered schemes optimize a function of the video qualities of each node. However, there is no single scheme that maximizes the video quality of each node simultaneously. In fact, all presented schemes are able to provide a Pareto-optimal solution, meaning that there is no other solution that is simultaneously preferred by all nodes. Thus, it is not clear which scheme results in the best resource allocation for the whole network. To handle the resulting tradeoffs, in this study, a metric that considers both fairness and performance issues is used for the qualitative evaluation of the results of each scheme. Ideally, a desirable scheme could achieve high total utility while being equally fair to all nodes and requiring low amounts of power. For this reason, we also evaluated the schemes in terms of the total consumed power relative with the total achieved utility.

8666-2, Session 1

Discussion on information theoretic and simulation analysis of linear shift-invariant edge detection operators

Bo Jiang, National Institute of Aerospace (United States)

Generally, the designs of digital image processing algorithms and image gathering devices remain separate. However, experiments show that the image gathering process profoundly impacts the performance of digital image processing and the quality of the resulting images. We proposed an end-to-end information theory based system to assess linear shift-invariant edge detection algorithms, where the different parts, such as scene, image gathering, and processing, are assessed in an integrated manner using Shannon's information theory. We evaluated the performance of the different algorithms as a function of the characteristics of the scene and the parameters, such as sampling, additive noise etc., that define the image gathering system. The edge detection algorithm is regarded as having high performance only if the information rate from the scene to the edge image approaches its maximum possible. This goal can be achieved only by jointly optimizing all processes. To validate our information theoretical conclusions, a series of experiments simulated the whole image acquisition process were conducted. After comparison and discussion between theoretic analysis and simulation analysis, we can draw a conclusion that the proposed information-theoretic assessment provides a new tool which allows us to compare different linear shift-invariant edge detectors in a common environment.

8666-3, Session 1

Analysis of uncoded image communication over noisy channels with unequal protection

Ruiqin Xiong, Peking Univ. (China); Xiaopeng Fan, Harbin Institute of Technology (China); Feng Wu, Microsoft Research Asia (China); Wen Gao, Peking University (China)

For image communication over noisy channels, conventional coding schemes generally suffer from threshold effect when channel quality falls under a certain threshold. An uncoded analog transmission (UAT) scheme named SoftCast, on the other hand, has recently shown to provide graceful quality transition and achieve superior performance to existing state-of-the-art digital methods, for channels with wide instantaneous signal-to-noise ratio range. However, the mechanism of UAT and the factors that influence the distortion-power efficiency of UAT are not investigated thoroughly in literatures. In this paper, we present a theoretical analysis for uncoded analog transmission. The contributions are three-fold. Firstly, a formulation for the end-to-end performance of UAT with unequal noise protection is developed. A metric called "data activity" is introduced to measure the complexity of a source in the context of UAT. Secondly, we show that, for signals with strong correlation, significant performance gain can be provided by promoting energy diversity in data using a decorrelation transform. Thirdly, we show that the transform gain can be realized only if the energy diversity in data is known by both the encoder and the decoder and properly exploited for unequal noise protection. Specifically, the granularity for describing the data energy diversity knowledge shared by the encoder and the decoder controls the actual achieved performance gain. The analyses presented in the paper are verified by experimental results.

8666-20, Session IPI

Block-layer bit allocation for quality constrained video encoding based on constant perceptual quality

Chao Wang, Xuanqin Mou, Wei Hong, Xi'an Jiaotong Univ. (China); Lei Zhang, The Hong Kong Polytechnic Univ. (Hong Kong, China)

In lossy image/video encoding, there is a compromise between the number of bits (rate) and the extent of distortion. Bits need to be properly allocated to different sources, such as frames or macro blocks (MBs). Since the human eyes are more sensitive to the difference than the absolute value of signals, the MINMAX criterion suggests to minimizing the maximum distortion of the sources to limit the quality fluctuation. There are many works aimed to such constant quality encoding, however, almost all of them focus on the frame layer bit allocation, and use PSNR as the quality index. We suggest that the bit allocation for MBs should also aim at the constant quality, and furthermore, perceptual quality indices should be used instead of PSNR. Based on this idea, we propose a multi-pass block-layer bit allocation scheme for quality constrained encoding. The experimental results show that the proposed method can achieve much better encoding performance.

8666-21, Session IPI

A bilateral hole filling algorithm for time-of-flight depth camera

Seung-Won Jung, Samsung Advanced Insitute of Technology (Korea, Republic of)

Time-of-flight (ToF) depth camera provides absolute depth values by measuring the phase delay between the emitted and received light signals. Since the phase delay is calculated by comparing electric

charges accumulated at the photodiode, the saturation occurs when the amount of electric charges exceeds the capacity of the photodiode. Such a saturation phenomenon is particularly problematic in the ToF depth camera since a long exposure time is required for accurate depth acquisition. In this paper, we present an effective algorithm to compensate for a saturated region, i.e. a hole region, in depth images acquired by the ToF camera. The proposed algorithm recovers the hole region by taking into account contour pixels surrounding the hole region. In particular, eight contour pixels are selected and then grouped into four pairs according to the four representative directions, i.e. horizontal, vertical, and two diagonal directions. The four pairs of contour pixels are then combined via a modified bilateral filtering framework in which the filter coefficients are obtained by considering the depth value difference between two depth pixels in each pair and the geometric distance between the hole pixel and the pixel in the contour. Experimental results demonstrate the effectiveness of the proposed algorithm.

8666-22, Session IPI

Utilization of reverse super-resolution techniques to reduce image size for transmission

Jeremy Straub, The Univ. of North Dakota (United States)

The reduction of file size for transmission of imagery over a constrained communications path provides significant benefit in a multitude of bandwidth-constrained applications including transmission from interplanetary spacecraft to Earth, transmission from orbit to Earth, and bandwidth reduction for various commercial and entertainment applications. Common methods utilize various techniques ranging from the identification and symbolic replacement of common patterns and various loss-causing techniques, which replace regions with abstractions. This paper presents an approach which utilizes a shared database and super-resolution techniques to downsize and re-enhance data to facilitate low-date-rate transmission of image data.

8666-5, Session 2

Improved motion prediction for scalable extensions of HEVC *(Invited Paper)*

Xiaoyu Xiu, InterDigital, Ltée (Canada); Yong He, Yuwen He, Yan Ye, InterDigital, Inc. (United States)

Due to the emergence and maturity of new advanced video compression technologies, a new video coding standard, High Efficiency Video Coding (HEVC), is under development by the Joint Collaborative Team on Video Coding (JCT-VC) of the ITU-T VCEG and the ISO/IEC MPEG, which has already demonstrated significant improvement in video compression efficiency compared to the existing video coding standards such as H.264/AVC. As the first version of single-layer HEVC standard comes close to completion, there are great interests to extend the standard with scalable capabilities. In this paper, several advanced coding techniques are presented to improve the motion prediction process for scalable extensions. Experimental results are presented to validate their efficiency.

8666-6, Session 2

An HEVC extension for spatial and quality scalable video coding *(Invited Paper)*

Tobias Hinz, Philipp Helle, Haricharan Lakshman, Mischa Siekmann, Jan Stegemann, Heiko Schwarz, Detlev Marpe, Thomas Wiegand, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

No Abstract Available

8666-7, Session 2

Towards a next generation open-source video codec *(Invited Paper)*

Jim Bankoski, Ronald Bultje, Adrian Grange, John Koleszar, Debargha Mukherjee, Paul Wilkins, Yaowu Xu, Google (United States)

Google has recently started developing a next generation open-source video codec - called VPNext, as part of the experimental branch of libvpx included in the WebM project (http://www.webmproject.org/). Starting from the VP8 video codec released by Google in 2010 as the baseline, a number of enhancements and new tools have been added to improve the coding efficiency. So far, at least 25% bit-rate reduction has been achieved over VP8 on a wide range of test sets. This paper will present a technical overview of the current status of this project along with comparisons made to VP8 and other mainstream state-of-the-art video codecs.

The new tools that have been added so far include: more modes for INTRA prediction, various forms of compound INTER prediction, ?-pel motion vectors and 8-tap switchable subpel interpolation filters, filtered prediction, improved motion reference generation and motion vector coding, improved entropy coding and frame-level adaptive entropy coding for various symbols, improved loop filtering, incorporation of larger 32x32 super-blocks and coding modes, frame level segmentation to group similar areas together, and inclusion of larger 8x8 and 16x16 DCTs, as well as 4x4 and 8x8 Asymmetric DSTs. Other tools are being actively worked on as well.

8666-8, Session 3

Scalable extensions of HEVC for next generation services *(Invited Paper)*

Louis J. Kerofsky, Jane Zhao, Kiran Misra, Andrew Segall, Sharp Labs. of America, Inc. (United States)

No Abstract Available

8666-9, Session 3

An improved hypothetical reference decoder for HEVC *(Invited Paper)*

Sachin G. Deshpande, Sharp Labs. of America, Inc. (United States); Miska Hannuksela, Nokia Research Ctr. (Finland); Kimihiko Kazui, Fujitsu Labs., Ltd. (Japan); Thomas Schierl, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany); Gary J. Sullivan, Microsoft Corp. (United States); Ye-Kui Wang, Qualcomm Inc. (United States)

Hypothetical Reference Decoder (HRD) describes a hypothetical decoder model that specifies constraints on a conforming bitstream that an encoding process may produce. HEVC builds upon and improves the design of the generalized hypothetical reference decoder of H.264/ AVC. This paper describes some of the main improvements of hypothetical reference decoder for HEVC.

8666-10, Session 3

On lossless coding for HEVC *(Invited Paper)*

Wen Gao, Minqiang Jiang, Haoping Yu, Huawei Technologies Co., Ltd. (United States)

No Abstract Available

8666-11, Session 3

Edge adaptive intra field de-interlacing of video images

Vladimir Lachine, Gregory Smith, Louie Lee, Qualcomm Inc. (Canada)

Expanding image by an arbitrary scale factor and thereby creating an enlarged image is a crucial image processing operation. Deinterlacing is an example of such operation where a video field is enlarged in vertical direction with 1:2 scale factor. The most advanced deinterlacing algorithms use a few consequent input fields to generate one output frame. In order to save hardware resources in video processors, missing lines in each field may be generated without reference to the other fields. Line doubling, known as "bobbing", is the simplest intra field deinterlacing method. However, it generates a lot of visual artifacts. Even interpolation of an inserted line from a few neighboring lines by vertical filter may produce such visual artifacts as "jaggies".

In this work we present edge adaptive image up-scaling algorithm, which can produce "jaggies" free video output frames. As a first step, an edge and its parameters in each interpolated pixel are detected from gradient squared tensor based on local signal variances. Then, according to the edge parameters including orientation, anisotropy and variance strength, the algorithm determines footprint and frequency response of two-dimensional interpolation filter for the output pixel. Filter's coefficients are defined by edge parameters, so that quality of the output frame is controlled by local content. The proposed algorithm of the edge adaptive image up-scaling has been hardware implemented in video display processor for intra field deinterlacing of video images.

8666-12, Session 3

On the efficiency of image completion methods for intra prediction in video coding with large block structures

Dimitar Doshkov, Oscar Jottrand, Thomas Wiegand, Patrick Ndjiki-Nya, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

Intra prediction is a fundamental tool in video coding with hybrid block-based architecture. Recent investigations have shown that one of the most beneficial elements for a higher compression performance in high-resolution videos is the incorporation of larger block structures. In this work, we investigate the performance of novel intra prediction modes based on different image completion techniques in a new video coding scheme with large block structures. Image completion methods exploit the fact that high frequency image regions yield high coding costs when using classical H.264/AVC prediction modes. This problem is tackled by investigating the incorporation of several intra predictors using the concept of Laplace partial differential equation (PDE), Least Square (LS) based linear prediction and the Auto Regressive model. A major aspect of this article is the evaluation of the coding performance in a qualitative (i.e. coding efficiency) manner. Experimental results show significant improvements in compression (up to 7.41 %) by integrating the LS-based linear intra prediction.

8666-13, Session 4

Depth-layer-based multiview image synthesis and coding for interactive z- and x-dimension view switching

Yu Mao, The Graduate Univ. for Advanced Studies (Japan); Gene Cheung, Yusheng Ji, National Institute of Informatics (Japan)

No Abstract Available

8666-14, Session 4

Wyner-Ziv coding of depth maps exploiting color motion information

Matteo Salmistraro, Marco Zamarin, Søren O. Forchhammer, Technical Univ. of Denmark (Denmark)

Distributed coding of multi-view data and depth maps is an interesting and challenging research field, whose interest is growing thanks to the recent advances in depth estimation and the development of affordable devices able to acquire depth information. In applications like video surveillance and object tracking, the availability of depth data can be beneficial and allow for more accurate processing. In these scenarios, the encoding complexity is typically limited and therefore distributed coding approaches are desirable.

In this paper a novel algorithm for distributed compression of depth maps exploiting corresponding color information is proposed. Thanks to the high correlation of the motion in color and depth videos, motion information from the decoded color signal can effectively be exploited to generate accurate side information for the depth signal, allowing for higher rate-distortion performance without increasing the delay at the decoder side. The proposed scheme has been evaluated against state-of-the-art distributed video coding techniques applied on depth data. Experimental results show that the proposed algorithm can provide PSNR improvement between 2.18 dB and 3.40 dB on depth data compared to the reference decoder.

8666-15, Session 4

Multimodal image registration by iteratively searching keypoint correspondences

Yong Li, Robert L. Stevenson, Univ. of Notre Dame (United States)

This paper proposes a multimodal image registration algorithm through searching the best matched keypoints by employing the global information. Keypoints are detected from images from both the reference and test images. For each test keypoint, a certain number of reference keypoints are chosen as mapping candidates. A triplet of keypoint mappings determine an affine transformation, and then it is evaluated with the similarity metric between the reference image and the transformed test image by the determined transformation. An iterative process is conducted on triplets of keypoint mappings, and for every test keypoint updates and stores its best matched reference keypoint. The similarity metric is defined to be the number of overlapped edge pixels over entire images, allowing for global information being incorporated in evaluating triplets of mappings. Experimental results show that the proposed algorithm can provide more accurate registration than existing methods on EO-IR images.

8666-16, Session 4

A spatially varying PSF model for Seidel aberrations and defocus

Jonathan D. Simpkins, Robert L. Stevenson, Univ. of Notre Dame (United States)

Contrary to common assumptions in the literature, the blur kernel corresponding to lens-effect blur has been demonstrated to be spatially-varying across the image plane. Existing models for the corresponding point spread function (PSF) are either parameterized and spatially-invariant, or spatially-varying but ad-hoc and discretely-defined.

In this paper, we develop and present a novel, spatially-varying, parameterized PSF model that accounts for Seidel aberrations and defocus in an imaging system. We also demonstrate that the parameters of this model can easily be determined from a set of discretely-defined PSF observations, and that the model accurately describes the spatial variation of the PSF from a test camera.

8666-17, Session 5

Efficient determination of intra predictability in H.264/AVC and similar codecs

Seyfullah H. Oguz, Qualcomm Inc. (United States)

Intra prediction as introduced by H.264/MPEG-4 AVC is a very potent image and video compression tool to exploit spatial correlation within a picture prior to transform coding. In an extended form it has been also adopted to the HEVC standard. The significantly increased intra coding efficiency provided by this tool compared to earlier coding standards comes at the expense of considerably increased computational requirements. With increasing spatio-temporal resolutions for video signals and the constraints imposed by real-time encoding, the significance of computational requirements becomes more pronounced.

In this paper, a novel, computationally efficient algorithm to assess the intra predictability of an image block is introduced. In the proposed algorithm, the input image block size is chosen to be larger than the smallest block size used for intra prediction for example 8x8 and 4x4 respectively. The larger size of the image block input and analyzed enables capturing the prominent directional structure present in the image block within a support region larger than the (smallest) block size used for intra prediction i.e. at a larger scale, and hence contributes to the overall efficiency of the algorithm. The proposed algorithm subsamples the pixels of the input image block on two or more 2D grids, the axes of which are rotated in the plane. For example, for H.264/MPEG-4 AVC, two 4x4 subsampling grids are used, the axes of which are rotated by pi/4 radians relative to each other. The pixels of the input image block e.g. 8x8 sized, are subsampled on these two different 4x4 grids, and 4x4 Hadamard transforms – already in H.264/MPEG-4 AVC tool set – are applied to both subsampled pixel sets. Analysis of the transform coefficients resulting from both grids' transforms provides accurate, concise information about the presence and nature of any prominent smooth or directional structure e.g. horizontal, vertical, oblique (the two different diagonals), present within the input image block. Hence, in addition to inferring the intra predictability of the input image block at a large scale, the proposed algorithm also points to the potential intra prediction direction(s) in effect. Features of the algorithm such as the use of Hadamard transforms and simple sums of coefficient magnitudes based comparisons facilitate its fast computation.

RD efficiency comparisons with another fast intra prediction algorithm and future work considerations are provided.

8666-18, Session 5

Cubic-panorama image dataset analysis for storage and transmission

Saeed Salehi, Eric Dubois, Univ. of Ottawa (Canada)

In this work we address the problem of cubic-panorama image dataset analysis for storage and transmission with emphasis on disparity estimation and disparity vector transcoding. Based on our earlier works on cubic-panorama image datasets and by using properties of the Epipolar geometry, a novel disparity estimation and disparity vector transcoding method suitable for cubic-panorama image datasets is presented. Our aim is to achieve increased rate-distortion performance with equivalent computational complexity for storage as compared to the classic solutions. Comparing the computational complexity, search range, and search direction, it is shown that our method is more intelligent. Performance improvement is shown both subjectively and objectively in disparity estimation error using PSNR (Peak Signal to Noise Ratio) measure. Furthermore, obtained results are used in the transmission stage of the project. Our disparity vector transcoding method can be used to convert cubic-panoramas of IPIP format to rectangular views of similar prediction structure very efficiently. Similar approach can be used in case of IBIB structure. Subsequently, in applications where at the transmission stage IPPP sequence is preferred, our method can be followed by a video transcoding scheme to convert the sequence from rectangular IBIB format to rectangular IPPP format.

8666-19, Session 5

Efficient streaming of stereoscopic depth-based 3D videos

Dogancan Temel, Mohammed Aabed, Georgia Institute of Technology (United States); Mashhour Solh, Texas Instruments Inc. (United States); Ghaassan Alregib, Georgia Institute of Technology (United States)

In this paper, we summarize our current work on reconstructing the depth map from monocular cues and depth map statistics. We first analyze a ground truth depth map to extract a set of depth cues or statistics. Then, based on these depth cues, we process the colored reference video and generate an estimate of the depth map per monocular cue. The monocular cues we consider in this work are motion, texture, and intensity. The processing of each channel in the Y CRCB color space is conducted separately. We tested our approach on different video sequences with different monocular properties. The results show that the extracted depth maps generate a 3D video with quality close to the video rendered using the ground truth depth map. We report objective results using 3VQM and PSNR. Furthermore, we analyze the savings in bitrate as a consequence of eliminating the need for two video codecs, one for the reference color video and one for the depth map. In this case, only the depth cues are sent as a side information with the color reference video and this can lead to bit-rate savings up to 40%.

Conference 8667A: Mobile Imaging System Design and Image Quality

Tuesday 5 –5 February 2013

Part of Proceedings of SPIE Vol. 8667 Multimedia Content and Mobile Devices

8667-43, Session 7

Mobile-based text detection and recognition for visually impaired persons

Mustafa I. Jaber, Jeremi Sudol, Bing Song, Ipplex Holdings Corp. (United States)

In this paper, we propose a system for detecting and localizing text regions in images and videos capturing printed page, books, magazine, mail envelope, and receipt in real-time using smartphone camera. The system includes stages for i) identifying text regions from low-resolution video frames, ii) generating audio feedback to guide the visually impaired personal to capture the entire text region in the scene, iii) triggering the camera to capture a high-resolution still-image of the same scene, iv) recognizing the text regions using off-the-shelf optical character recognition tool that runs on the mobile device or in the cloud, and v) pronouncing the recognized text using text-to-speech module. The originality of the proposed algorithm stands in its real-time audio guided feedback to capture an acceptable image for the OCR engine. Methods for corner detection, connected component analysis, and paragraph structure test are used in the proposed text detection module. The algorithm has been tested on iPhone devise where enhanced performance was achieved. The usage simplicity and availability of the application on smartphones gave its advantage over traditional scanner-based OCR systems.

8667-45, Session 7

Determination of sensor oversize for stereo-pair mismatch compensation and image stabilization

Prajit Kulkarni, Aptina Imaging LLC (United States)

Stereoscopic cameras consist of two sensor modules that in theory are mounted parallel to each other at a fixed distance along a single plane. Practical tolerances in the manufacturing and assembly process can, however, cause mismatches in the relative orientation of the modules. One solution to this problem is to design sensors that image a larger field-of-view than is necessary to meet system specifications. This requires the computation of the sensor oversize needed to compensate for the various types of mismatch. This work introduces a mathematical framework to compute these oversize values for mismatch along each of the six degrees of freedom. The basis of the mathematical treatment is that the original field-of-view for each sensor will need to be maintained despite any relative displacement between the sensor modules. One module is considered as the reference and the extreme rays of the field-of-view of the second sensor are traced in order to derive equations for the required horizontal and vertical oversize. As a further application, by modeling user hand-shake as the displacement of the sensor from its intended position, these deterministic equations could be used to estimate the sensor oversize required to stabilize images that are captured using cell phones.

8667-46, Session 8

Image quality evaluation using moving targets

Uwe Artmann, Image Engineering GmbH & Co. KG (Germany)

The image quality is highly influenced by the signal processing for noise and resolution and the processing is the main reason for the loss of low contrast, fine details, the so called "texture blur". We present an approach to describe the image processing in more detail.

All standardized test methods use a defined chart and require, that the chart and the camera are not moved in any way during test. In this paper, we present our results investigating the influence of chart movement during the test.

Different charts, optimized for different aspects of image quality evaluation, are moved with a defined speed during the capturing process. The chart movement will change the input for the signal processing depending on the speed of the target during the test.

The basic theoretical changes in the image will be the introduction of motion blur. With the known speed and the measured exposure time, we can calculate the theoretical motion blur. We compare the theoretical influence of the motion blur with the measured results. We use different methods to evaluate image quality parameter vs. motion speed of the chart. For this test series we use a D-SLR camera as a reference and different mobile phone cameras.

8667-47, Session 8

Multiple-field approach for aberration correction in miniature imaging systems based on wafer-level production

Eric Logean, Toralf Scharf, Nicolas Bongard, Hans Peter Herzig, Ecole Polytechnique Federale de Lausanne (Switzerland); Markus Rossi, Heptagon (Switzerland)

Mobile imaging systems are often limited by the objective lens. Here we present an intermediate approach between the costly traditional objectives and the low-resolution objectives inspired by the insect eyes. Our multi-field approach uses a small number of optical channels each imaging a portion of the desired field of view. The full-field image is obtained by digital reconstruction. Each optical channel is of adequate dimension to obtain images with an angular resolution close to the resolution of the human eye, and simultaneously, it is kept simple for easy mass production using wafer-level technology.

We present the design and fabrication of two prototypes, each using 9 plano-convex lenses (9 channels) to image a field of +/-30 degrees with field curvature correction. The first prototype uses glass lenses from Edmund Optics glued on a wafer to form an multi-field imaging system with an f-number of 3. The second prototype uses lenses fabricated by the reflow technique. It has an f-number of 5. Experimental results including simple image reconstructions are presented.

In conclusion, we have presented objectives fabricated with micro-optics technology and using the multi-field concept which are thin, simple to mount, robust, and easily replicated.

8667-48, Session 8

Color correction using multi-cycle feedback system for adaptive color image quality in mobile phone camera

Setaek Oh, Jaemin Joo, Jin Wook Kwon, Chang Chun Goh, Sang Jo Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

Recently, Camera module system in mobile phone is steadily developed by the miniature of chip size and optimized lens design for high end level. However it has definite advantages such as useful size, optimized weight and efficient power control in CMOS image sensor, it is difficult to make it implement high quality image on account of skillful limitation of image processing. Color correction is important in vision applications and process to improve color constancy of

image. Colors from CMOS or CCD image sensor do not match colors of the photographed object due to output characteristic's difference compared to standard color. Therefore, we need color correction to complement that weakness similar to ideal camera image. In the conventional method, color correction is implemented by empirical method in terms of lots of iterative experiments. In this paper, we propose faithful to the original colors adaptive color correction using multi cycle feedback system for high quality image. Multi cycle feedback system is the adaptive method that color matrix at color space (R-G-B-Y-M-C) is separated and each color's value reproduces repetitively. Also, we can evaluate a proposed color correction method quantitatively using color matching method. The acquired experiment results show the proposed color correction is close to the quantitative value of ideal or target camera. In addition, we can confirm the dramatic improvement of the image quality for real taken picture applicability.

8667-49, Session 8

Auto-focus algorithm based on statistical blur estimation

Prajit Kulkarni, Aptina Imaging LLC (United States)

Conventional auto-focus techniques in movable-lens camera systems use a measure of sharpness to determine the lens position that brings the scene into focus. This paper presents a novel wavelet-domain approach to determine the position of best focus. In contrast to current techniques, the proposed algorithm estimates the level of blur in the captured image at each lens position. The degree of blur in an image is quantified using a measure of Gaussianity in the high-pass filtered version of the image. The algorithm overcomes several shortcomings of the sharpness-based approach. Sharpness estimation requires the application of band-pass filters and is sensitive to image noise. This leads to larger filters and the use of data-dependent thresholds which necessitates tuning and calibration for different imaging conditions. The proposed blur-based method works on the statistics computed from a filtered version of the input image. Since noise has no impact on performance, the algorithm works with a short filter and is devoid of thresholds and parameter tuning. These advantages, along with the optimization presented in the paper, make the proposed very attractive for hardware implementation on cell phones. Experiments prove that the algorithm performs very well on noisy images and under resolution and data scaling.

8667-77, Session 8

Nokia PureView oversampling technology
(Invited Paper)

Tero Vuori, Juha Alakarhu, Eero Salmelin, Ari Partinen, Nokia Corp. (Finland)

No Abstract Available

8667-50, Session 9

Low light performance of digital still cameras

Dietmar Wüller, Image Engineering GmbH & Co. KG (Germany)

Looking at the latest developments in cameras the pixel race is still going on and pixel pitches close to 1 micron have become reality. Especially in mobile cameras sensor size and the related module size matters. With shrinking pixel sizes the sensitivity of sensors decreases. Modern technologies try to compensate for this. On the one hand back side illumination increases the quantum efficiency of the signal itself. On the other hand image processing is used to increase the quality of the amplified signal afterwards.

One major aspect of image processing is noise reduction that decreases the amount of noise in the images. Unfortunately the processing algorithms face the problem to differentiate between

fine detail in the images and noise. Therefore the denoising also affects low contrast fine detail and reduces resolution in the images. Noise reduction is typically also higher in color channels than on the luminance part of the signal resulting a decrease of color fidelity.

While lowering the illumination level some experiments look at the different image quality aspects for a huge variety of cameras. The paper will discuss and summarize the results of these experiments and will be an input for the discussion of the standards committee on how to proceed forward to create a low light performance measurement standard.

8667-51, Session 10

Noise evaluation standard of image sensor using visual spatio-temporal frequency characteristics

Takeyuki Fujii, Shoichi Suzuki, Shinichiro Saito, Sony Corp. (Japan)

Regarding the noise evaluation of image sensor, it is important to establish the visual noise standard, which is a noise evaluation metric using visual characteristics.

The visual noise level can vary depending on the viewing distance, spatial frequency, color and viewing conditions and frame rate. A method of measuring the visual noise level is provided in ISO 15739. Visual characteristics depend on contrast and frame rate; however, the method doesn't consider that. We propose solutions to solve two problems using visual spatio-temporal frequency characteristics.

Firstly, we investigated visual spatial frequency characteristics that depend on contrast and propose a new evaluation method. It shows that the image sensor with large pixels count is effective in noise reduction.

Secondly, we investigated visual temporal frequency characteristics and propose a new evaluation method for the moving image. It shows that the image sensor with high frame rate is effective in noise reduction.

Finally, by combining two proposed methods, we show the method in which noise evaluation is possible in both a still image and moving image sequences. We applied the proposal method to moving image sequences acquired by the image sensor and investigated the validity. The proposed method showed good performance in the evaluation experiment.

Conference 8667B: Mobile Computational Photography

Monday - Tuesday 4 –5 February 2013

Part of Proceedings of SPIE Vol. 8667 Multimedia Content and Mobile Devices

8667-52, Session 11

Image deblurring in smartphone devices using built-in inertial measurement sensors

Ondrej Sindelar, Charles Univ. in Prague (Czech Republic); Filip Sroubek, Institute of Information Theory and Automation (Czech Republic)

Long exposure hand-held photography is degraded with blur, which is difficult to remove without prior information about the camera motion. In this work, we utilize inertial sensors (accelerometers and gyroscopes) in modern smartphones to detect exact motion trajectory of the smartphone camera during exposure and remove blur from the resulting photography based on the recorded motion data. The whole system is implemented on Android platform and embedded in the smartphone device resulting in a close-to-real-time deblurring algorithm. To our knowledge this is the first attempt in this direction and renders the method particularly appealing for end users. The performance of the proposed system is demonstrated in real-life scenarios. We have justified the space-invariant simplification for certain camera motions, but simultaneously we have uncovered intrinsic sources of space-variant blur, such as rolling shutter. The space-variant implementation of the deblurring algorithm, which would solve some of the current issues, is in theory possible, but the computational cost on the smartphone may be too high. It will be a topic of our future research to find out whether this is viable.

8667-53, Session 11

Compensating specular highlights for non-Lambertian projection surfaces

Chen-Tai Kao, Tai-Hsiang Huang, Homer H. Chen, Hua Lee, National Taiwan Univ. (Taiwan)

We have presented a method for radiometric compensation of specular highlight of a non-Lambertian projection surface. An attractive feature of the method is that it relaxes the limitation on the viewer's position with respect to the camera. Our method is able to automatically reconstruct specular highlight response of the projection surface and dynamically compensate the image for viewer at an arbitrary viewing angle. We implemented this algorithm on the radiometric compensation framework proposed in [1]. As a preprocessing module of the whole procam system, the proposed algorithm is simple and efficient. Only one calibration image is needed. Experimental results show that our algorithm can effectively predict the specular light and eliminate the photometric distortion seen at any arbitrary viewing angle. The projection quality is greatly improved. Since the algorithm is designed for a procam system, it is well applicable for mobile devices with projector and camera bound together and using a nearby wall as the ad hoc screen for projection.

8667-54, Session 12

On plenoptic camera resolution

Andrew Lumsdaine, Indiana Univ. (United States); Todor G. Georgiev, Qualcomm Inc. (United States); Lilie Lin, Indiana Univ. (United States)

Plenoptic cameras capture the 4D radiance of a scene -- two dimensions of spatial information and two dimensions of angular information. Final images are rendered from the 4D data using integral projection, often in conjunction with a shearing transformation used to effect "refocusing" of the final image. The various transformations applied to the radiance complicate the question of what is the achievable resolution of the final rendered image. In this paper we analyze the geometry of discrete plenoptic function capture to determine theoretical limits of plenoptic camera resolution and provide experimental results to verify our analyses.

8667-55, Session 12

Wave analysis of a plenoptic system and its applications

Sapna A. Shroff, Kathrin Berkner, Ricoh Innovations, Inc. (United States)

Traditional imaging systems directly image a 2D object plane on to the sensor. Plenoptic imaging systems contain a lenslet array at the conventional image plane and a sensor at the back focal plane of the lenslet array. In this configuration each lenslet effectively images the aperture of the main imaging lens at the sensor. Therefore the data captured at the sensor retains angular light-field information. If a filter array is placed at the pupil aperture of the main imaging lens, each lenslet images the filter array on to the sensor. This enables the separation of multiple filter modalities giving single snapshot, multi-modal images. Plenoptic systems are increasingly being used in recent times for imaging in various applications. As the application of these systems moves towards microscopes and other complex systems and as pixel sizes become smaller, consideration of diffraction effects in these systems becomes increasingly important. We discuss a plenoptic system and its wave propagation analysis for both the coherent and incoherent cases. We simulate a system response using our analysis and discuss various applications of the system response pertaining to plenoptic system design, implementation and calibration.

8667-56, Session 12

Fourier analysis of the focused plenoptic camera

Andrew Lumsdaine, Lilie Lin, Indiana Univ. (United States); Todor G. Georgiev, Qualcomm Inc. (United States)

The focused plenoptic camera is a recently developed approach to lightfield capture that uses the microlens array as an imaging system focused on the focal plane of the main camera lens. Since lightfields can be captured with significantly higher spatial resolution than with the traditional approach, images can be rendered at resolutions that meet the expectations of modern photographers. The focused plenoptic camera captures lightfields with a different tradeoff between spatial and angular information than with the traditional approach. To more rigorously characterize these tradeoffs, including the limits of this new approach, this paper presents a Fourier analysis of the focused plenoptic camera. Based on this analysis, we also present an extended Fourier-slice rendering algorithm that can be used to render high-resolution images from lightfields.

8667-57, Session 13

Design of user interfaces for selective editing of digital photos on touchscreen devices

Thomas Binder, Meikel Steiding, Manuel Wille, Nils Kokemohr, Nik Software GmbH (Germany)

When editing images it is often desirable to apply a filter with a spatially varying strength. With the usual selection tools like gradient, lasso, or

Conference 8667B

brush tools, creating masks containing such spatially varying strength values is time-consuming and cumbersome. We present an interactive filtering approach which allows to process photos selectively without the intermediate step of creating a mask containing strength values. In using this approach, the user only needs to place reference points (called control points) on the image and adjusts the spatial influence and filter strength for each control point. The filter is then applied selectively to the image, with strength values interpolated for each pixel between control points. The interpolation is based on a mixture of distances in space, luminance, and color; it is therefore a low-level operation. Since the main goal of the approach is to make selective image editing intuitive, easy and playful, emphasis is put on the user interface: We describe the evolution process of developing an existing mouse-driven user interface into a touchscreen-driven one. Many question needed to be answered anew, such as how to present a slider widget on a touchscreen. Several variants are discussed and compared.

8667-58, Session 13

Touch HDR: photograph enhancement by user controlled wide dynamic range adaptation

Steve M. Verrall, Qualcomm Inc. (United States); Hasib Siddiqui, Qualcomm Inc (United States); Kalin Atanassov, Sergio R. Goma, Vikas Ramachandra, Qualcomm Inc. (United States)

High Dynamic Range (HDR) technology typically consists of (1) acquiring a wide dynamic range image and (2) adapting the wide dynamic range image to fit to the display range of the device. The first part can be achieved in multiple ways, two common examples are by using a special sensor or by combining two or more images with same or different exposures. The second part locally adapts the tone mapping function thus effectively re-using the available range. In this paper we address the second part and we describe a photo editing technique, referred to as "Touch HDR" that enables the user to create a new image by selectively enabling and specifying the aggressiveness of the adaptation.

8667-59, Session 13

Image and video denoising and enhancement through frame stacking and alignment

Kalin Atanassov, James Nash, Sergio R. Goma, Vikas Ramachandra, Qualcomm Inc. (United States); Hasib Siddiqui, Qualcomm Inc (United States)

Due to the current cell-phone cameras pixel size, there is a gap in the low-light and high dynamic performance of those cameras compared with traditional cameras (DSC/DSLRs). Most of the systems connected to those cameras have significantly more processing resources than traditional cameras. Mobile computational photography is an emerging field that tries to leverage those resources and compensate the lack in performance. In this paper we propose an algorithm that reduces the noise and increases the dynamic range of a final image by "stacking" sequentially captured frames. The frame stacking is done by frame alignment subject to a projective transform and temporal anisotropic diffusion. Our approach effectively accomplishes an average of images in which we reduce the quantization and dark current noise up to the limit of the fixed pattern noise. Further image improvement is accomplished by fixed pattern noise characterization and removal. To enhance the result we apply adaptive contrast enhancement, and show 10-20 times dynamic range improvement for typical scenes

8667-60, Session 13

Accelerating defocus blur magnification

Florian Kriener, Thomas Binder, Manuel Wille, Nik Software GmbH (Germany)

A shallow depth of field is often used as a creative element, for example in portrait photography. This, however, comes at the cost of expensive and heavy camera equipment, such as large sensor DSLR bodies and fast lenses. In contrast, cheap small-sensor cameras with fixed lenses usually exhibit a larger depth of field than is sometimes desirable. In this case a computational solution is suggesting, since a shallow depth of field cannot be achieved by optical means. One possibility is to algorithmically magnify the defocus blur already present in the image. Yet, existing algorithmic solutions in this direction suffer from poor performance due to the ill-posedness of the problem: The amount of defocus blur can be estimated at edges only; homogeneous areas do not contain such information. In order to enhance the defocus blur, these gaps need to be closed which requires to solve a very large optimization problem.

We propose a faster way to propagate the amount of blur from the edges to the entire image by solving the optimization problem on a coarse scale, followed by edge-aware upsampling using the original image as guide. The resulting approximate defocus map can be used to synthesize images with shallow depth of field with a quality comparable to the original approach. This is demonstrated by experimental results.

8667-62, Session 14

Adaptive DOF for plenoptic cameras

Alexander Oberdörster, Fraunhofer-Institut für Angewandte Optik und Feinmechanik (Germany); Hendrik P. A. Lensch, Eberhard Karls Univ. Tübingen (Germany)

Plenoptic cameras promise to provide arbitrary re-focusing through a scene after the capture. In practice, however, the refocusing range is limited by the depth of field (DOF) of the plenoptic camera. For the focused plenoptic camera, this range is given by the range of object distances for which the microlens images are in focus.

We propose a technique of recording light fields with an adaptive depth of focus. Between multiple exposures -- or multiple recording of the light field -- the microlens array of the plenoptic camera is moved. The depth and quality of focus is adjusted by changing the number of exposures and the spacing of the MLA movements.

In contrast to traditional cameras, extending the DOF does not force an all-in-focus image. Instead, the refocus range is extended. There is full creative control about the focus depth; images with shallow or selective focus can be generated.

8667-63, Session 14

Plenoptic depth map in the case of occlusions

Zhan Yu, Jingyi Yu, Univ. of Delaware (United States); Andrew Lumsdaine, Indiana Univ. (United States); Todor G. Georgiev, Qualcomm Inc. (United States)

Recent realization of hand-held plenoptic cameras has given rise to previously unexplored effects in photography. Designing a mobile phone plenoptic camera is becoming feasible with the significant increase of computing power of mobile devices and the introduction of System on Chip. However, capturing high number of views is still impractical due to special requirements such as ultra-thin camera and low costs. In this paper, we analyze a mobile plenoptic camera solution with small number of views. Such camera can produce refocusable high resolution final image if a depth map is generated for every pixel in the sparse set of views.

With the captured multi-view images, the obstacle to recover a high-resolution depth is occlusions. To robustly resolve these, we first analyze the behavior of pixels in such situations. We show that even under severe occlusion, one can still distinguish different depth layers based on statistics. We estimate the depth of each pixel on by discretizing the space in the scene and conducting plane sweeping. Specifically, for each given depth, we gather all corresponding pixels from other views and model the in-focus pixels as a Gaussian distribution. We show how it is possible to distinguish occlusion pixels, and in-focus pixels in order to find the depths. Final depth maps are computed in real scenes captured by a mobile plenoptic camera.

8667-64, Session 14

Reduced depth of field using multi-image fusion

Boris Ajdin, Univ. of Ulm (Germany); Timo Ahonen, Nokia Research Ctr. (United States)

This paper presents a novel multi-image fusion method for artificially reducing depth of field in phone camera photographs. In addition to two high-resolution images, one focused on the object of interest selected by the user using a single tap, and another one with the maximum background blur (achieved by focusing at the near focus plane), the system captures a VGA resolution focal stack of images with variable focus settings. The focal stack is then used to guide a graph cut based segmentation of the object of interest from the sharp high-resolution image by comparing per pixel the sharpness profile across the stack with the sharpness profile of the point of interest. Afterwards the segmented object is blended into the background from the blurry high-resolution image, resulting in a visually pleasing image with a shallow depth of field effect.

8667-65, Session 14

Optimizing depth of field based on a range map and a wavelet transform

Mike Wellner, Pattern Recognition Lab. (Germany); Thomas Käster, Pattern Recognition Co. GmbH (Germany); Thomas Martinetz, Erhardt Barth, Univ. zu Lübeck (Germany)

The imaging properties of small cameras in mobile devices exclude restricted depth of field and distance-dependent blur with a sensation of depth. Algorithmic solutions to this problem usually fail because high-quality, dense range maps are hard to obtain, especially with a mobile device. However, methods like stereo, shape from focus stacks, and the use of flashlights may yield coarse and sparse range maps. A standard procedure is to regularize such range maps to make them dense and more accurate. In most cases, regularization leads to insufficient localization and sharp edges in depth cannot be handled well. In a wavelet basis, an image is defined by its significant wavelet coefficients, only these need to be encoded. If we wish to perform range-dependent image processing, we only need to know the range for the significant wavelet coefficients. We therefore propose a method that determines a sparse range map only for significant wavelet coefficients, and then weights the wavelet coefficients depending on the associate range information. The image reconstructed from the resulting wavelet representation exhibits space-variant, range-dependent blur. We present different results, e.g. based on images and range maps obtained with the stereo camera of the mobile phone LG P920.

8667-66, Session 15

Resolution and sensitivity of wafer-level multi-aperture cameras

Alexander Oberdörster, Fraunhofer-Institut für Angewandte Optik und Feinmechanik (Germany); Hendrik P. A. Lensch, Eberhard Karls Univ. Tübingen (Germany)

The scaling limits of multi-aperture systems have been widely discussed from an information-theoretical standpoint. While these arguments are valid as an upper limit, the real-world performance of mass-market systems for mobile devices is still restricted by optical aberrations. We argue that aberrations can be more easily controlled with certain architectures of multi-aperture systems, especially those manufactured on wafer scale (WLO). We complement our analysis with measurements of one single- and one multi-aperture WLO camera, showing that multi-aperture systems can indeed deliver superior performance. We examine both sharpness and sensitivity, giving measurements of MTF and SNR.

8667-67, Session 15

Refinement of depth maps by fusion of multiple estimates

Balaji Krishnamurthy, Anubha Rastogi, ADOBE Systems India Private Ltd. (India)

No Abstract Available

8667-68, Session 15

On the application of the plenoptic camera to mobile phones

Iciar Montilla Garcia, Instituto de Astrofísica de Canarias (Spain); Marta Puga, Jose G. Marichal-Hernandez, Jonas Philipp Lüke, José Manuel Rodríguez-Ramos, Univ. de La Laguna (Spain)

The plenoptic camera was originally created to allow the capture of the Light Field, a four-variable volume representation of all rays and their directions, that allows the creation by synthesis of a 3D image of the observed object. This method has several advantages with regard to 3D capture systems based on stereo cameras, since it does note need frame synchronization or geometric and color calibration. And it has many applications, from 3DTV to medical imaging. A plenoptic camera uses a microlens array to measure the radiance and direction of all the light rays in a scene. The array is placed at the focal plane of the objective lens, and the sensor is at the focal plane of the microlenses. In this paper we study the application of our super resolution algorithm to mobile phones cameras. With a commercial camera, it is already possible to obtain images of good resolution and enough number of refocused planes, just placing a microlens array in front of the detector.

8667-69, Session 15

A novel 3D model for rotation/zoom motion deblur

Zhihui Li, Silong Peng, XiYuan Hu, Institute of Automation (China); Xiaojing Xu, Institute of Forensic Science (China)

Spatially non-uniform image blur caused by camera shake is a common problem for domestic digital cameras. Rota- tion/Zoom motion blur is a special type of spatially non- uniform image blur. However, most of the existing ap- proaches for solving spatially variant blur use too complex models and suffer from high computational cost. And there have seldom paper discussion the estimation methods of zoom/rotation parameters. On the contrary, based on the real image capture process, we propose a novel time spatially invariant 3-D convolution model for rotation-blur and zoom- blur. Under this model, rotation-blur and zoom-blur restora- tion become quite simple and a fast and efficient algorithm can be used to solve such kind of spatially variant blur. And we can estimate the rotation/zoom parameters base the 3-D model. Further more, we present the blind deblurring method of zoom/rotation blur. The experimental results demonstrate the effective and accuracy of our methods.

8667-70, Session 16

Lytro camera technology: theory, algorithms, performance analysis *(Keynote Presentation)*

Sergio R. Goma, Todor G. Georgiev, Qualcomm Inc. (United States); Andrew Lumsdaine, Indiana Univ. (United States)

No Abstract Available

8667-71, Session 17

A new fusion-based low light still-shot stabilization

Young-Su Moon, Samsung Advanced Institute of Technology (Korea, Republic of) and SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Shi-Hwa Lee, Samsung Advanced Institute of Technology (Korea, Republic of) and SAMSUNG Electronics, Ltd. (Korea, Republic of)

Digital cameras under a dark illumination reveals some artifacts like motion blur in a long-exposed shot or salient noise corruption in a (High ISO) short-exposed shot. To prevent such artifacts, multi-frame fusion of either differently exposed images or multiple short-exposed images has been studied actively, and as a practical still-shot stabilization, the approach using multiple short-exposure images has begun to be applied to consumer cameras. However, it requires highly complicated and time-consuming computational procedures to achieve effective motion compensated noise filtering and brightness restoration in a difficult situation of strong noise corruption, low brightness and distorted color.

In this paper, we propose a new fusion-based low-light stabilization approach, which inputs both short-exposure images and one proper (or long)-exposure (blurry) image. First, a coarse-to-fine motion compensated temporal noise filtering is suggested for getting a clean image from the multiple short-exposure noisy images. More specifically, global motion estimation, block-wise local motion estimation, and motion compensated temporal noise filtering are iterated in a hierarchical framework to acquire better denoising and its speed-up. Then, online image restoration to the denoised image is followed to obtain a visually good (properly bright and colored) result. Our test results show the effectiveness of the proposed algorithm.

8667-72, Session 17

Real-time skeleton tracking for embedded systems

Foti Coleca, Univ. zu Lübeck (Germany) and gestigon GmbH (Germany); Sascha Klement, gestigon GmbH (Germany); Thomas Martinetz, Erhardt Barth, Univ. zu Lübeck (Germany)

Touch-free gesture technology is beginning to become more popular with consumers and may have a significant future impact on interfaces for digital photography. However, almost every commercial software framework for gesture and pose detection is aimed at either desktop PCs or high-powered GPUs, making mobile implementations for gesture recognition an attractive area for research and development. In this paper we present an algorithm for hand skeleton tracking and gesture recognition that runs on an ARM-based platform (Pandaboard ES, OMAP 4460 architecture). The algorithm uses self-organizing maps to fit a given topology (skeleton) into a 3D point cloud. This is a novel way of approaching the problem of pose recognition as it does not employ complex optimization techniques or data-based learning. After an initial background segmentation step, the algorithm is ran in parallel with heuristics which detect and correct artifacts arising from insufficient or erroneous input data. We then optimize the algorithm for the ARM platform using fixed-point computation and the NEON SIMD architecture the OMAP4460 provides. We tested the algorithm with two different depth-sensing devices (Microsoft Kinect, PMD Camboard). For both input devices we were able to accurately track the skeleton at the native framerate of the cameras.

8667-74, Session IPI

Reconstruction of the image on the Cartesian lattice from a finite number of projections in computed-tomographic imaging

Nan Du, The Univ. of Texas at San Antonio (United States); Yusheng Feng, Univ of Texas at San Antonio (United States); Artyom M. Grigoryan, The Univ. of Texas at San Antonio (United States)

In this paper, we discuss a new approach of reconstructing the image from a finite number of projections using the tensor representation of the image, after processing the projections by solving the special systems of equations. In the discrete space, all components of the tensor transform are defined as the ray-sums of the image in the Cartesian lattice, which can exactly be calculated from the ray-integrals. To describe this transformation of integrals from the image plane to the lattice, we introduce the concept of shifted geometrical rays, or G-rays that allow for each projection to calculate the splitting-signal by solving a system of algebraic equations described by the binary Toeplitz matrices. Then, the discrete image is calculated by the inverse tensor transform or through the inverse 2-D DFT. The proposed approach is presented for the continuous model, when the image $f(x,y)$ is on the unit square $[0,1]$ x $[0,1]$ and consists of N2 cells of constant intensity each on the Cartesian lattice NxN, where N is a power of two or a prime. The proposed method was implemented in MATLAB and C++, and the experimental results are illustrated, for the images of size up to 1024 x 1024.

8667-75, Session IPI

Method of G-particles for image reconstruction from a finite number of projections

Artyom M. Grigoryan, Nan Du, The Univ. of Texas at San Antonio (United States)

Novel method of summation of projections is proposed, which uses the tensor representation of the image as a set of splitting-signals defined by sums of parallel rays passing through the Cartesian lattice. These signals define the direction components of the image and can be calculated from the ray-integrals of sets of shifted parallel rays, which are defined in a unique way for each projection. This set of rays is called the geometrical, or G-rays. We introduce the concept of the point map of projections, when each image element is considered as a particle described by a field function in tensor representation. The map of each particle on the Cartesian lattice is considered in the form of a matrix describing all G-rays passing through the particle. The consideration of field functions for G-particles leads to a field transformation of the image with following reconstruction. The set of projections is defined by the tensor representation. However, other sets of projections can also be considered; the sets defined by a small number of angles, symmetric and non-symmetric sets of angles, as well as random sets of angles. Experimental results of the proposed reconstruction are given and compared with the filtered backprojection.

Conference 8667C: Multimedia on Mobile Devices 2013

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8667 Multimedia Content and Mobile Devices

8667-17, Session 4

Enabling customer self service through image processing on mobile devices

Joern Kreutel, Fachhochschule Brandenburg (Germany); Sascha Hellmann, Brandenburg University of Applied Sciences (Germany); Jörn Kreutel, Beuth University of Applied Sciences (Germany)

We describe an innovative use of augmented reality technology in a mobile application that is distributed to end customers of an enterprise. Using the application, customers will be able to solve issues related to the enterprise's products without requiring assistance by a human customer service agent.

8667-18, Session 4

Cognitive styles and visual quality

Satu Jumisko-Pyykkö, Tampere Univ. of Technology (Finland)

Assessors are the main measurement instruments in subjective quality of experience evaluation. Although the perceptual opportunities and constrains are influenced by multiple demographic and psychographic factors, they are typically disregarded as a part of quality assessment (overview Jumisko-Pyykkö, 2011). Cognitive styles refer to individual's consistent approaches to organize and represent information (Riding and Rayner, 1998). Goal of this study is to explore influence of cognitive styles on visual quality requirements. The data-collection is conducted using the Style of Processing (SOP) questionnaire (Childers et al., 1985) in three visual video quality experiments with a total of 80 participants. All participants were categorized into four groups according to sensorial preferences in information processing (visual, auditive, bimodal – high processing, and bimodal - low processing). The experiments were conducted in the controlled circumstances when varying depth in video quality with several content types on the mobile device. The results showed variation in quality requirements between these groups. Finally, these results also indicate that sensorial processing styles are essential to explore for sample characterization in quality assessment and for exploring more user-aware quality adjustments in future services and products.

8667-19, Session 4

Subjective evaluation of HEVC in mobile devices

Ray Garcia, General Dynamics Itronix Corp. (United States); Hari Kalva, Florida Atlantic Univ. (United States)

Mobile compute environments provide a unique set of user needs and expectations that designers must consider. With increase multimedia use in mobile environments, video encoding methods within the smart phone market segment are key factors that contribute to positive user experience. Currently available display resolutions and expected cellular bandwidth are major factors the designer must consider when determining which encoding methods should be supported. The desired goal is to maximize the user experience, reduce cost, and reduce time to market

Subjective comparisons are made between H.264/AVC and HEVC encoding standards in accordance with Double-stimulus impairment scale (DSIS) as defined by ITU-R BT.500-13. Test environment are based on smart phone LCD resolutions and expected cellular bit rates, such as 200kbps and 400kbps.

Subjective feedback shows both encoding methods are adequate at 400kbps constant bit rate. However, a noticeable user experience gap was observed for 200 kbps constant bit rate. Significantly less H.264 subjective quality is noticed with video sequences that have multiple objects moving and no single point of visual attraction. Video sequences with single points of visual attraction or few moving objects tended to have higher H.264 subjective quality.

8667-20, Session 5

Location-based tracking using long-range passive RFID and ultrawideband (UWB) communications *(Invited Paper)*

Faranak Nekoogar, Lawrence Livermore National Lab. (United States)

The ability to locate and track first responders inside buildings is an important problem for which there are currently no good solutions. Indoor localization and navigation poses many challenges for search and rescue teams (i.e. firefighters) such as inability to determine their exact location and communicate with the incident commander outside the building. Although RF navigation and tracking systems have many advantages over other technologies, the harsh indoor RF environment demands new ways of developing and using RF sensor and communication systems. In this paper, we introduce an integrated system for reliable indoor navigation that includes (i) long-range passive RF tags with read distance of over 300 feet, and (ii) ultra-wideband (UWB) communications link for sending reader data from inside the building to outside in order to locate and track first responders. The mobile platform architecture described in this paper make optimal use of long-range passive tags, and takes advantage of the frequency diversity of UWB communication systems for a reliable, robust and yet low-cost design. Further, the UWB communication allows transmission of both voice and data over the same harsh RF environment.

8667-21, Session 5

Real-time content-aware video retargeting for tunnel vision assistance

Thomas A. Knack, Andreas E. Savakis, Rochester Institute of Technology (United States)

Image and video retargeting technologies have become an effective means of resizing media for aspect ratio constrained applications and are useful for mobile smartphone systems. In this paper, a novel real-time video retargeting model is proposed for mobile phone implementation to assist individuals with tunnel vision. Spatial and temporal improvements to the baseline seam carving algorithm provide the basis for our model. Seam carving is a content-aware retargeting operator which defines 8-connected horizontal or vertical paths, or seams of pixels. The optimality of these seams is based on a specific energy function. Seam removal permits changes in the aspect ratio while simultaneously preserving important regions. This work introduces a video retargeting model for tunnel vision assistance which incorporates spatial and temporal considerations to preserve visual integrity. Face detection masks and salience maps are provided to achieve more comprehensive results. Additionally, formulation of a novel temporal coherence measure is presented that allows for retargeting of streaming video. Integration of the proposed model with an Android mobile platform demonstrates its portability and potential for use in mobile low vision assistance systems.

Conference 8667C

8667-22, Session 5

Human movement activity classification approaches that use wearable sensors and mobile devices

Sahak I. Kaghyan, Yerevan State Univ. (Armenia); Hakob G. Sarukhanyan, Institute for Informatics and Automation Problems (Armenia); David Akopian, The Univ. of Texas at San Antonio (United States)

This paper reviews different methods and approaches of solving human activity classification problem. To classify activity current work reviews following data collection methods: 1) raw data retrieving from mobile devices' sensors (such as GPS sensor, accelerometer and gyroscope), and 2) signal values' collecting process and transferring to server simultaneously from several biaxial accelerometers, attached to different parts of human body. Generally, signals were uninterruptedly transferred from mentioned sensors or from smartphone via wireless network to server, located nearby, for raw data further analysis and recognition. Current work also notes different algorithms which were used in classification during incoming data analyzing process. From comparing these results with each other it will be clear which approach of activity classification will be more efficient to use depending on limitations that may be and means, available to solve activity classification problem.

8667-23, Session 5

Concept for practical exercises for studying autonomous flying robots in a university environment

Ricardo Band, Johann-Sebastian Pleban, Stefan Schön, Reiner Creutzburg, Arno Fischer, Fachhochschule Brandenburg (Germany)

The aim of this paper is to demonstrate the usefulness of a concept of practical exercises for studying autonomous flying robots for computer science students in a university environment.

It is shown how the students can assemble, program, fly, network and apply autonomous flying robots (drones, quadrocopters, hexacopters, octocopters, helicopters, helicams, bugbots, ...) in different exercises and improve their skills and theoretical and practical knowledge in different aspects.

8667-24, Session 5

Applications of multimedia technology on autonomous flying robots for university technology transfer projects

Ricardo Band, Johann-Sebastian Pleban, Stefan Schön, Reiner Creutzburg, Arno Fischer, Fachhochschule Brandenburg (Germany)

The aim of this paper is to give an overview of the wide area of possible applications of multimedia technology on autonomous flying robots for university technology transfer projects.

In particular we summarize the usefulness of different imaging and video technology applications in order to strengthen the links between industrial and administrative partners and the university.

8667-25, Session 6

Mobile variable data pages: from apps to Cloud printing

Nathan Moroney, Kok-Wei Koh, Hewlett-Packard Labs. (United States)

Variable data pages or printing is the dynamic generation of partially or completely customized content for print production. In a commercial print production context, these pages are often generated using centralized tools, databases and print engines. One challenge in a centralized production environment is managing the variable data, which can include personal information or vendor data. Mobile computing presents an interesting alternative for generation of mobile variable data pages. In a mobile context, individual apps can create highly customized content making use of distributed processing, data and printers. We present results for an experimental application, to be described in greater detail elsewhere, which makes use of the PDF generation tools included with the developer SDK to create personalized letter-formatted pages which are then printed using HP's ePrint cloud printing technology. In this way, mobile devices generate unique data which is then formatted by the app as a print-ready document which is then printed via a distributed network of printers, without the need for an intermediate driver or dedicated general purpose formatting engine. This capability is a general purpose one that scales in two directions. First, the app and mobile device are used to generate and manage the variable data content. If a processing module can be implemented given the limitations of the target device then the generation of variable data pages will scale directly with the number of mobile computing devices. Second, use of a distributed print production process means that printing will also scale directly with the number of printers making use of the cloud printing infrastructure. This paper will describe the specifics of the creation of the variable data pages using a specific developer SDK and our experiences using the ePrint cloud printing technology. The paper will also include a provocative example in which use of the PDF generation tools in the mobile environment allowed the creation of a highly specialized printed swatch-book in manner that was competitive with using highly specialized desktop tools.

8667-26, Session 6

Digitized forensics: retaining a link between physical and digital crime scene traces using QR-codes

Mario Hildebrandt, Stefan Kiltz, Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

The digitization of physical traces from crime scenes in forensic investigations entrains the challenge of creating a link between the two or more representations of the same trace. In order to be forensically sound, especially the two security aspects of integrity and authenticity need to be maintained at all times. Especially the adherence to the authenticity using technical means proves to be a challenge at the boundary between the physical object and its digital representations. In this article we propose a new method of linking physical objects with its digital counterparts using two-dimensional barcodes and additional meta-data accompanying the acquired data for integration in the conventional bagging and tagging process. Using the exemplary chosen QR-code as particular implementation of a barcode and a model of the forensic process, we also supply a means to integrate our suggested approach into forensically sound proceedings.

We use the example of the digital dactyloscopy as a forensic discipline, where currently progress is being made by digitizing some of the processing steps. We show an exemplary prototypic implementation of the suggested approach using a smartphone as a mobile device for the verification of the physical trace to extend the chain-of-custody from the physical to the digital domain.

8667-27, Session 6

Smart apps for applied machine learning on mobile devices - the MOMO project

Stefan Edlich, Mathias Vogler, Beuth Hochschule für Technik Berlin (Germany)

The MOMO project is a multimillion € research project supported by EFRE run at Beuth University of Technology Berlin (App.Sc.). It consists of two parts: Eco Mobility and Mobile computing. The latter goal is to develop new Smartphone and Tablet applications for mobile computing

Conference 8667C

(as NFC, indoor navigation, etc.). A sub-project of MOMO is smart apps and adressing machine learning. In the first part we developed a smart app for amusement parks as an example. The users were tracked and machine learning algorithms have been applied thereafter to predict visitor behaviour in these parks (or buildings, concerts, airports, etc.). This experience lead us to more research in terms of a) machine learning (ML) result visualization, b) ML implementation and c) ML configuration and execution on mobile devices. In this paper we show examples of implementations and realizations for these three areas. Furthermore we show how such applications can itself be used by users / visitors - and not only scientists / park operators - to create new (visual) experiences.

8667-28, Session 6

Real-time volume rendering of digital medical images on an iOS device

Christian Noon, Dreamworks Animation SKG, Inc. (United States); Joseph Holub, Iowa State Univ (United States); Eliot Winer, Iowa State Univ. (United States)

Performing high quality 3D visualizations on mobile devices, while tantalizingly close in many areas, is still a quite difficult task. This is especially true for 3D volume rendering of digital medical images. Allowing this would empower medical personnel a powerful tool to diagnose and treat patients and train the next generation of physicians. This research focuses on performing real time volume rendering of digital medical images on iOS devices using custom developed GPU shaders for orthogonal texture slicing. An interactive volume renderer was designed and developed with several new features including dynamic modification of render resolutions, an incremental render loop, a shader-based clipping algorithm to support OpenGL ES 2.0, and an internal backface culling algorithm for properly sorting rendered geometry with alpha blending. The application was developed using several application programming interfaces (APIs) such as OpenSceneGraph (OSG) as the primary graphics renderer coupled with iOS Cocoa Touch for user interaction, and DCMTK for handling of DICOM files. The developed application is able to render volume datasets over 450 slices up to 50-60 frames per second, depending on the specific device and model. All rendering is done locally on the device so no Internet connection is required.

8667-29, Session 6

MessageSpace: a messaging system for health research

Rodrigo D. Escobar, David Akopian, The Univ. of Texas at San Antonio (United States); Deborah Parra-Medina, Laura A. Esparza, The Univ. of Texas Health Science Ctr. at San Antonio (United States)

Most developed countries have been increasing their health care expenditures over the years. Increases of around 30% and 50% of the per capita income of those countries have been noticed in the last eight years. Disease prevalence and medical treatments represents around two-thirds of these raisings. This trend reveals the importance of disease prevention. Basically, the more people are aware of how to prevent diseases, the more expenses can be cut, thus making an impact to the whole society. Short Message Service (SMS) messages have been used to empower health care and health promotion, but existing SMS based health promotion systems do not completely address health promotion research needs. In order to empower health care promotion and education research, we present a system whose main goal is to fulfill the needs of health care researchers. We do so by integrating and extending existing technologies including J2EE, databases, mobile messaging, and Smartphones applications among others. The SMS is used as the key technology to send messages and polls to research projects' participants. The integration of the mentioned technologies along with the Short Message Service unique characteristics and usage patterns lead the way towards providing a promising system for supporting health-promotion research.

8667-30, Session 6

Multi-resolution edge detection with edge pattern analysis

Bo Jiang, National Institute of Aerospace (United States)

Edge detection is defined as the process of detecting and representing the presence of and locations of image signal discontinuities. Generally, the edge detection operation has two main steps: filtering, and detection and localization. In the first step, finding an optimal scale of the filter is an ill-posed problem, especially when a single (global) scale is used over the entire image. Multi-resolution description of the image which can represent the image features occurring in a range of scales is used, where a combination of Gaussian filters with different scales can ameliorate the single scale issue. In the second step, often edge detectors have been designed to capture simple ideal step functions in image data, but real image signal discontinuities deviate from this ideal form. Another three types of deviations from the step function which relate to real distortions occurring in natural images are examined. These types are impulse, ramp, and sigmoid functions which respectively represent narrow line signals, simplified blur effects, and more accurate blur modeling. General rules for edge detection based upon the classification of edge types into four categories- ramp, impulse, step, and sigmoid are developed from this analysis. The performance analysis on experiments supports that the proposed algorithm does lead to more effective edge detection and localization with improved accuracies.

8667-31, Session IPI

Clientside Skype forensics: an overview

Tina Meißner, Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

Skype communication over the Internet is very popular nowadays. The aim of this paper is to give an overview on the clientside Skype Forensics and to explain what data is stored on the user's computer and how these data can be read out in a forensic investigation. Both, the manual analysis and the analysis by open source tools and commercial tools are described and the corresponding problems are addressed.

8667-32, Session IPI

Real-time human activity classification using tri-axial accelerometer of mobile device and SQLite database engine

Sahak I. Kaghyan, Armenian-Russian (Slavonic) Univ. (Armenia); David Akopian, The Univ. of Texas at San Antonio (United States); Hakob G. Sarukhanyan, Institute for Informatics and Automation Problems (Armenia)

The movement of telemedicine from desktop platforms to wireless and mobile configurations may have a significant impact on future health care. Physical activity is recognized as one of the most important dimensions of human health. Available sensors in mobile devices allow for activity recognition and, thus, physical health assessment. During recent years there have been many publications on activity classification. These approaches typically exploit wearable sensors attached to patient's body to collect necessary information concerning activity. In this paper, our focus is on activity recognition using sensors built in in smartphones. Particularly tri-axial accelerometers on Android systems are used. Raw sensor signal sequences are collected at programmed sampling rates. The processing is performed using SQLite relational database management system in smartphone memory. As portable database allows simultaneously read and write data, so mobile application does data collection and acquired information analyzing at the same time in order to do real-time activity classification straight on smartphone.

8667-33, Session IPI

Gradient-based fusion of infrared and visual face images using support vector machine for human face identification

Priya Saha, Mrinal K. Bhowmik, Tripura Univ. (India); Debotosh Bhattacharjee, Jadavpur Univ. (India); Barin K. De, Tripura Univ. (India); Mita Nasipuri, Jadavpur Univ. (India)

Human Face Identification has achieved a lot of attention from research community during past several years. But, performance of face identification system degrades in uncontrolled environment when random lighting sources from different directions change visual representations significantly. In this paper, we have proposed gradient based fusion method of gradient visual and corresponding infrared face images to overcome the problem of illumination varying conditions. This technique mainly extracts illumination insensitive features under different conditions for effective face recognition purpose. The gradient image computed from a visible light image, the subject's face is compared with the gradient image extracted from a thermal face image of the same subject. The image fusion of infrared image and corresponding visual gradient image is done in wavelet domain by taking the maximum information of approximation and detailed coefficients. These fused images are taken for dimension reduction using Independent Component Analysis (ICA). The reduced 1?49 face images are taken for training and testing purposes from different classes of different datasets of IRIS face database. We have taken SVM multiclass strategy 'one-vs.-all' in our experiment. Linear kernel and Polynomial kernel with degree 3 are used in our experiment as kernel functions. The experiment results reveal that our method can perform well for the face images under different lighting conditions.

8667-34, Session IPI

Future mobile access for open-data platforms and the BBC-DaaS system

Stefan Edlich, Sonam Singh, Ingo Pfennigstorf, Beuth Hochschule für Technik Berlin (Germany)

We develop a platform which provides a marketplace for data which can act as base for applications or information seekers relying on this data. We examine various methods to facilitate automatic augmentation of metadata for effective querying related tasks in a scalable manner for the platform. Data is automatically preprocessed, analyzed, visualized, transformed, filtered and made accessible to interested parties through access points like REST, Multimedia devices or directly accessible through web application. We examine strategies for automatic pricing of data usage using various parameters e.g. the amount of data, bandwidth used, CPU and memory costs or license costs specified by the provider.

Data as a Service (DaaS) provide state of the art services for transformations, visualizations, filtering, sorting of data and analytics, pricing as core and an extensible framework for adapting to changing requirements of different types of data , data providers and data consumers(mostly applications).

The key issues of this poster are advanced ways of mobile access to open data platforms like BBC-DaaS and open-data platforms in general.

8667-35, Session IPI

HIPAA compliance for messaging systems: MessageSpace case study

Anuvrath Joshi, Rodrigo D. Escobar, David Akopian, The Univ. of Texas at San Antonio (United States); Deborah Parra-Medina, Laura A. Esparza, The Univ. of Texas Health Science Ctr. at San Antonio (United States)

No Abstract Available

8667-36, Session IPI

Location tracking forensics on GPS handhelds, mobile navigation devices, and smartphones

Stefan Sack, Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

The spread of navigation devices has increased significantly over the last 10 years. With the help of the current development of even smaller Navigation receiver units it is possible to navigate with almost any current Smartphone. Due to the increased use of navigation devices their relevance to forensic investigations has risen rapidly. Navigation data is forensically interesting because by the residence of the devices in most cases the location and the travelled path of the owner can be reconstructed. In this work practices for forensic analysis of Navigation devices are developed. Different devices will be analyzed and it is attempted, by means of forensic procedures to restore the traveled path of the device. For analysis of the various devices different software and hardware is used. There will be presented common procedures for securing and testing of mobile devices and some specialties in the investigation of each device. The different classes considered are GPS handhelds, mobile navigation devices and smartphones. It will be attempted, to read all data of the device. This is realized by the usage of current forensic software e.g. TomTology or Oxygen Forensic Suite. To limit the practical work the data extraction is focused on the frequently used device of a class, as the procedure for many groups of devices can be identical. In the present work a GPS handheld, a TomTom GO 700, an iPhone 3G and a HTC Desire (Android) is used because they have a wide circulation.

8667-37, Session IPI

Conception of a course for professional training and education in the field of computer and mobile forensics - Part II: Android Forensics

Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

The growth of Android in the mobile sector and the interest to investigate these devices from a forensic point of view has rapidly increased. Many companies have security problems with mobile devices in their own IT infrastructure. To respond to these incidents, it is important to have professional trained staff. It is necessary to further train the own employees in the field of mobile forensics because a lot of companies work with very sensitive data. Motivated by these facts, this paper shows proven practical examples for training approaches and practical exercises to investigate the forensically most important tasks of Android mobile devices.

8667-38, Session IPI

Possibilities and modification of the forensic investigation process of solid state drives

Frank Irmler, Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

This paper deals with the possibilities of the forensic investigation of solid-state drives. The aims of this study are to clarify the information gained by a forensic analysis of these media and to explain the diffenerences to conventional forensic examinations of conventional hard disk drives. For each test series, a variety of hardware and software was used. One interesting result is that the build-in trim function of the SSD has great impact on the result of the forensic investigation.

8667-40, Session IPI

Mobile learning in medicine

Sabri Serkan Gulluoglu, Istanbul Arel Univ. (Turkey)

This paper outlines the main infrastructure for implicating mobile learning in medicine and present a sample mobile learning application for medical learning within the framework of mobile learning systems.

Mobile technology is developing nowadays. In this case it will be useful to develop different learning environmets using these innovations in internet based distance education. M-learning makes the most of being on location, providing immediate access, being connected, and acknowledges learning that occurs beyond formal learning settings, in places such as the workplace, home, and outdoors. Central to m-learning is the principle that it is the learner who is mobile rather than the device used to deliver m learning. The integration of mobile technologies into training has made learning more accessible and portable. Mobile technologies make it possible for a learner to have access to a computer and subsequently learning material and activities; at any time and in any place. Mobile devices can include: mobile phone, personal digital assistants (PDAs), personal digital media players (eg iPods, MP3 players), portable digital media players, portable digital multimedia players.

Mobile learning (m-learning) is particularly important in medical education, and the major users of mobile devices are in the field of medicine. The contexts and environment in which learning occurs necessitates m-learning. Medical students are placed in hospital/clinical settings very early in training and require access to course information and to record and reflect on their experiences while on the move.

The main titles will be mobile learning, medical education, sub-branches of medicine, suitable education branches for mobility and, main differences of m-learning and normal learning systems, sample m-learning interface model for medicine for tablet computer and mobile phone systems, Conclusions and advises to the future.

As a result of this paper, this paper strives to compare and contrast mobile learning with normal learning in medicine from various perspectives and give insights and advises into the essential characteristics of both for sustaining medical education.

8667-41, Session IPI

Overview and forensic investigation approaches of the gaming console Sony PlayStation Portable

Ralph Schön, Stephan Schön, Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

This paper deals with forensically interesting features of the Sony Playstation Portable game console. The construction and the internal structure are analyzed more precisely. Interesting forensic features of the operating system and the file system are presented.

8667-42, Session IPI

An efficient and fast iris location algorithm

Guangyuan Jiang, Changchun Normal Univ. (China)

The accuracy and speed of iris localization affect recognition system performance in the iris recognition system. Based on analyzing some prevailing iris localization algorithms, in this paper the edge information of pupil is extracted by the least-square method, and the iris outer circle is extracted by the improved Canny Operator plus Hough Transform, and the experimental result shows that this localization method is fast and of high precision. The noise from iris region includes eyelid, eyelash, eyelid shadow and specular reflections. The segmental-secondary linear localization method adopting edge detection and Radon Transform is proposed to remove the noise from eyelid on the eyelid localization, the eyelash noise and eyelid shadow are removed by threshold method, and the experimental result shows that the algorithm is efficient and accurate.

Conference 8667D: Multimedia Content Access: Algorithms and Systems VII

Monday - Wednesday 4 –6 February 2013

Part of Proceedings of SPIE Vol. 8667 Multimedia Content and Mobile Devices

8667-1, Session K1

The three R's of computer vision: recognition, reconstruction and reorganization *(Keynote Presentation)*

Jitendra Malik, Univ. of California, Berkeley (United States)

No Abstract Available

8667-2, Session 1

Video structure extraction *(Invited Paper)*

Remi Trichet, Ram Nevatia, The Univ. of Southern California (United States)

Action recognition is a challenging computer vision task that has a plethora of applications in nowadays society of pervasive technology, such as semantic indexing, or sport sequences skimming. Approaches can broadly be splintered in 2 categories: structural and statistical based strategies. As the former tries to match the data to some pre-defined model, the latter classifies events according to feature distributions.

But in every case, the utilized techniques are an attempt to build higher semantic information from raw, low-level features. This talk will focus on some common difficulties that the community is facing in this scope and the ploys used to tackle these challenges.

8667-3, Session 1

Taming the wild: acoustic segmentation on consumer-produced videos

Gerald Friedland, Benjamin Elizalde, International Computer Science Institute (United States)

In the last decade, the number of consumer-produced videos in the Internet has increased exponentially. For example, one of the most popular websites for consumer-produced videos, Youtube, claims that 72 hours of video are uploaded every minute, resulting in nearly 14 years of content uploaded every day. This results in a massive demand for techniques that can index the data for retrieval. A first step in automatic indexing is usually segmentation. Segmentation is the process of identifying the boundaries between classes. In speech processing, examples of these classes are phones, words, sounds, speech, music, and silence. While high accuracies have been reached for traditional, corpus-based, supervised segmentation tasks, segmentation approaches on ''wild'' videos are still a major challenge. In consumer-produced media, one cannot rely on any single characteristic to draw boundaries between classes and it is difficult to pre-train models because of the high variance in the data. Consumer-produced audio may contain low quality audio, noisy environments, singing, music, overlapping situations, and other unexpected circumstances. This talk presents our approach to a hybrid segmentation system for consumerproduced videos, which consists of both unsupervised clustering (using so-called audio percepts) and supervised classification based on an event space.

8667-4, Session 1

Sparse conditional mixture model: late fusion with missing scores for multimedia event detection

Ramesh M. Nallapati, Eric Yeh, SRI International (United States)

In this work, we present the Sparse Conditional Mixture Model (SCMM) which addresses the problem of late fusion in multimedia event detection in the presence of missing detection scores from one or more modalities. Our experiments prove that the SCMM model significantly outperforms the traditional Conditional Mixture model and several other baselines on the TREC task of Multimedia Event Detection, by dynamically adapting the model to capture only the observed modalities for a given video clip.

8667-5, Session 1

Can object detectors aid Internet video event retrieval?

Davide Modolo, Cees Snoek, Univ. van Amsterdam (Netherlands)

The problem of event representation for automatic event detection in Internet videos is acquiring an increasing importance, due to their applicability to a large number of applications. Existing methods focus on representing events in terms of either low-level descriptors or domain-specific models suited for a limited class of video only, ignoring the high-level meaning of the events. Ultimately aiming for a more robust and meaningful representation, in this paper we question whether object detectors can aid video event retrieval. We propose an experimental study that investigates the utility of present-day local and global object detectors for video event search. By evaluating object detectors optimized for high-quality photographs on low-quality Internet video, we establish that present-day detectors can successfully be used for recognizing objects in web videos. We use an object-based representation to re-rank the results of an appearance-based event detector. Results on the challenging TRECVID multimedia event detection corpus demonstrate that objects can indeed aid event retrieval. While much remains to be studied, we believe that our experimental study is a first step towards revealing the potential of object-based event representations.

8667-6, Session 1

Multimedia event detection using visual concept signatures

Ehsan Younessian, Michael Quinn, Teruko Mitamura, Alex Hauptmann, Carnegie Mellon Univ. (United States)

Multimedia Event Detection (MED) is a multimedia retrieval task with the goal of finding videos of a particular event in a large-scale Internet video archive, given example videos and text descriptions. In this paper, we mainly focus on an 'ad-hoc' scenario in MED where we do not use any example video. We aim to retrieve test videos based on their visual semantics using a Visual Concept Signature (VCS) generated for each event only derived from the event description provided as the query. Visual semantics are described using the Semantic INdexing (SIN) feature which represents the likelihood of predefined visual concepts in a video. To generate a VCS for an event, we project the given event description to a visual concept list using the proposed textual semantic similarity. Exploring SIN feature

properties, we harmonize the generated visual concept signature and the SIN feature to improve retrieval performance. We conduct different experiments to assess the quality of generated visual concept signatures with respect to human expectation and in the context of the MED task to retrieve the SIN feature of videos in the test dataset when we have no or only very few training videos.

8667-7, Session 2

A fast approach for integrating ORB descriptors in the bag of words model

Costantino Grana, Daniele Borghesani, Rita Cucchiara, Univ. degli Studi di Modena e Reggio Emilia (Italy)

In this paper we propose to integrate the recently introduces ORB descriptors in the currently favored approach for image classification, that is the Bag of Words model. In particular the problem to be solved is to provide a clustering method able to deal with the binary string nature of the ORB descriptors. We suggest to use a k-means like approach, called k-majority, substituting Euclidean distance with Hamming distance and majority selected vector as the new cluster center. Results combining this new approach with other features are provided over the ImageCLEF 2011 dataset.

8667-8, Session 2

Video-based analysis of motion skills in simulation-based surgical training

Qiang Zhang, Lin Chen, Qiongjie Tian, Baoxin Li, Arizona State Univ. (United States)

Analysis of motion expertise is an important problem in many domains including sports and surgery. In recent years, surgical simulation has emerged at the forefront of new technologies for improving the education and training of surgical residents. In simulation-based surgical training, a key task is to rate the performance of the operators, which is done currently by senior surgeons. This paper introduces a novel solution to this problem through employing vision-based techniques. We develop an automatic, video-based approach to analyzing the motion skills of a surgeon in simulation-based surgical training, where a surgical action is captured by multiple video cameras with little or no calibration, resulting in multiple video streams of heterogeneous properties. Typical multiple-view vision techniques are inadequate for processing such data. We propose a novel approach that employs both canonical correlation analysis (CCA) and the bag-of-words model to classify the expertise level of the subject based on the heterogeneous video streams capturing both the motion of the subject's hands and the resultant motion of the tools. Experiments were designed and performed to validate the proposed approach using realistic data captured from resident surgeons in local hospitals. The results suggest that the proposed approach may provide a promising practical solution to the real world problem evaluating motion skills in simulation-based surgical training.

8667-9, Session 2

Adaptive segmentation grids for human action recognition in videos

Nicolas Ballas, Bertran Delezoide, Commissariat à l'Énergie Atomique (France)

In this paper, we introduce an adaptive segmentation grid scheme for action recognition in unconstrained videos. State-of-arts solutions focusing on action recognition embeds low-level features spatio-temporal context in a Bag-of-Words model through statically defined segmentation grids. While providing a coarse localization of low-level features, those approaches tend to be limited by their grids rigidity and regular geometry. To adress this issue we propose to learn spatio-temporal grids directly from the video data. We propose new segmentation grids, named Adaptive Grid (AG), which are learnt and

adapted to a given training dataset. Our adapaive grids are then exploited by a Bag-of-Words model at the aggregation step for action recognition. Our proposal is evaluated on 4 publicly available datasets showing an improvement over the state-of-art.

8667-10, Session 2

Exploiting visual search theory to infer social interactions

Paolo Rota, Duc-Tien Dang-Nguyen, Nicola Conci, Nicu Sebe, Univ. degli Studi di Trento (Italy)

In this paper we propose a new method to infer human social interactions using typical techniques adopted in literature for visual search and information retrieval. The main piece of information we use to discriminate among different types of interactions is provided by proxemics cues acquired by a tracker, and used to distinguish between intentional and casual interactions. The proxemics information has been acquired through two different metrics: the first one is related to the current distance between subjects, while the second one measures the O-space synergy between subjects. The obtained values are taken at every time step in a temporal sliding window and then uncorrelated using the DFT transformation. The obtained features are eventually merged into an unique array. All the possible DFT tracks of the available dataset are clustered via K-means algorithm. The clusters are reorganized using a second temporal window into a Bag Of Words framework, so as to build the feature vector that will feed the SVM classifier.

8667-11, Session 3

Machine perception for content discovery at YouTube *(Keynote Presentation)*

Paul Natsev, Google (United States)

YouTube's mission is for YOU to discover and shape the world through video. At the heart of this mission is content discovery, or the problem of finding interesting content relevant to a given topic or user. This problem is particularly challenging given the variety and volume of YouTube videos: more than an hour of video is uploaded to YouTube every second (that's more than ten years worth of content every day). In this talk, I will give an overview of some work in the machine perception department at Google Research aiming to improve content discovery at YouTube. Specifically, I will present several case studies of applying machine perception and machine learning at YouTube scale to tackle problems such as automatically identifying and labeling celebrities and tourist landmarks in video, tagging videos semantically, discovering talent on YouTube, and using games to crowdsource video discovery on YouTube.

8667-12, Session 3

Presentation video retrieval using automatically recovered slide and spoken text

Matthew L. Cooper, FX Palo Alto Lab. (United States)

Video is becoming a prevalent medium for e-learning. Lecture videos contain text information in both the visual and aural channels: the presentation slides and lecturer's speech. This paper examines the relative utility of automatically recovered text from these sources for lecture video retrieval. To extract the visual information, we apply video content analysis to detect slides and optical character recognition to obtain their text. Automatic speech recognition is used similarly to extract spoken text from the recorded audio. We perform controlled experiments with manually created ground truth for both the slide and spoken text from more than 60 hours of lecture video. We compare the automatically extracted slide and spoken text in terms of accuracy relative to ground truth, overlap with one another, and utility for video retrieval. Results reveal that automatically recovered slide text and

spoken text contain different content with varying error profiles. Experiments demonstrate that automatically extracted slide text enables higher precision video retrieval than automatically recovered spoken text.

8667-13, Session 3

VidCat: an image and video analysis service for personal media management

Lee Begeja, Eric Zavesky, Zhu Liu, David Gibbon, Raghuraman Gopalan, Behzad Shahraray, AT&T Labs. Research (United States)

This paper presents a system for storage, organization and consumption of consumer media. Photo and video capture for consumers has never been easier, so the need for an enterprise-scale media service that does not burden the user during upload, share, and organization is critical. Cloud-based storage and consumption of personal photos and videos provides increased accessibility, functionality, and satisfaction for mobile users. One cloud service frontier that is recently growing is that of personal media management. In this work, a system called VidCat is presented that assists users in the tagging, organization, and retrieval of their personal media by faces and visual content similarity in addition to time and date information. It couples a low-level analysis engine with network-based media services for traditional organization and retrieval tasks. Algorithms that perform automated content segmentation, intelligent keyframe selection, near-duplicate detection, and face detection and similarity are described and are presented with evaluations on internationally recognized datasets. Design decisions for harmonious coupling of the VidCat back-end media service and a tablet-based client prototype with four powerful consumer use cases are also discussed.

8667-14, Session IPI

Audio stream classification for multimedia database search

Maria Teresa Artese, Consiglio Nazionale delle Ricerche (Italy); Simone Bianco, Univ. degli Studi di Milano-Bicocca (Italy); Isabella Gagliardi, Consiglio Nazionale delle Ricerche (Italy); Francesca Gasparini, Univ. degli Studi di Milano-Bicocca (Italy)

Search and retrieval of huge archives of Multimedia data is a challenging task. A classification step is often used to reduce the number of entries on which to perform the subsequent search. In particular, when new entries of the database are continuously added, a fast classification based on simple threshold evaluation is desirable.

In this work we present a CART-based (Classification And Regression Tree]) classification framework for audio streams belonging to multimedia databases. The database considered is the Archive of Ethnography and Social History (AESS), which is mainly composed of popular songs and other audio records describing the popular traditions handed down generation by generation, such as traditional fairs, and customs.

The peculiarities of this database are that it is continuously updated; the audio recordings are acquired in unconstrained environment; and for the non-expert human user is difficult to create the ground truth labels.

In our experiments, half of all the available audio files have been randomly extracted and used as training set. The remaining ones have been used as test set. The classifier has been trained to distinguish among three different classes: speech, music, and song. All the audio files in the dataset have been previously manually labeled into the three classes above defined by domain experts.

8667-15, Session IPI

Structuring a sharded image retrieval database

Eric Liang, Avideh Zakhor, Univ. of California, Berkeley (United States)

In previous work we described an approach to localization based in image retrieval. Specifically, we assume coarse localization based on GPS or cell tower and refine it by matching a user generated image query to a geotagged image database. We partition the image dataset into overlapping cells, each of which contains its own approximate nearest-neighbors search structure. By combining search results from multiple cells as specified by coarse localization, we have demonstrated superior retrieval accuracy on a large image database covering downtown Berkeley. In this paper, we investigate how to select the parameters of such a system e.g. size and spacing of the cells, and show how the combination of many cells outperforms a single search structure over a large region. We use two datasets from Berkeley and Oakland to demonstrate our results. We conclude that (a) cell radius should be chosen such that the number of features is a few million rather than a few tens of million; (b) even when cells of 20 million features are used, it is preferable to combine their results rather than use a single large cell; (c) a cell density where the radius of cell is equal to the distance between the center of cells offers the best accuracy and complexity performance.

8667-16, Session IPI

Diversification of visual media retrieval results using saliency detection

Oleg Muratov, Giulia Boato, Francesco G. B. De Natale, Univ. degli Studi di Trento (Italy)

Diversification of retrieval results allows for better and faster search.

Recently there has been proposed different methods for diversification of image retrieval results mainly utilizing text information and techniques imported from natural language processing domain. However, images contain visual information that is impossible to describe in text and the use of visual features is inevitable. Visual saliency is information about the main object of an image implicitly included by humans while creating visual content. For this reason it is naturally to exploit this information for the task of diversi cation of the content.

In this work we study whether visual saliency can be used for the task of diversifi cation and propose a method for re-ranking image retrieval results using saliency. The evaluation has shown that the use of saliency information results in higher diversity of retrieval results.

2014 Electronic Imaging

SCIENCE AND TECHNOLOGY

Technologies for digital imaging systems, 3D display, image quality, multimedia, and mobile applications

We're moving to San Francisco in 2014!

Mark your calendar

www.electronicimaging.org

Conferences and Courses
2–6 February 2014

Location
Hilton San Francisco, Union Square
San Francisco, California, USA

Technologies

- 3D Imaging, Interaction, and Metrology
- Visualization, Perception and Color
- Image Processing
- Image Capture
- Multimedia Processing and Applications
- Computer Vision

IS&T
imaging.org

SPIE®