

Content-based dynamic 3D mosaics

Zhigang Zhu

The static and dynamic objects of a large-scale scene from an airborne camera can be automatically detected, measured, and visualized as multiple panoramic images.

In an emergency evacuation involving a large metropolitan area, a light airplane or unmanned aerial vehicle is sent to fly over the scene. Several hours of video streams are generated every time such a mobile platform performs a data collection task, resulting in 100GB of data per hour for standard 640×480 raw color images. This huge amount of video data not only poses difficulties in data archiving but also for users to digest and build up a coherent map. It would therefore be useful to able to generate a synopsis of the entire region on the fly. A large field map that can be interactively viewed and dynamically updated, with all important 3D facilities and moving objects labeled on the map, would help make it possible to create more efficient evacuation plans and save more lives.

In the past, mosaic approaches^{1,2}—where smaller images are ‘stitched together’ to create a panoramic view—have been proposed for representation and compression of video sequences. However most of the work done involves panning, or rotating cameras instead of translating them, as happens in airborne surveillance and monitoring. Nevertheless, some work has been done in 3D reconstruction of panoramic mosaics,^{3,4} usually resulting in 3D depth maps of static scenes instead of high-level 3D representations for both static and dynamic targets.

The goal of the research on video processing and representation at the City College of New York is to rapidly acquire panoramic maps with 3D and moving targets when a light aerial vehicle equipped with a video camera flies over a known or unknown urban area. We have developed a content-based 3D mosaic representation (CB3M) for such long video sequences. The motion of the camera has a dominant direction of motion, but six degrees-of-freedom (DOF) motion is allowed. The potential applications in transportation and surveillance are significant. For example, by flying a light airplane with a video camera over New York City, we could

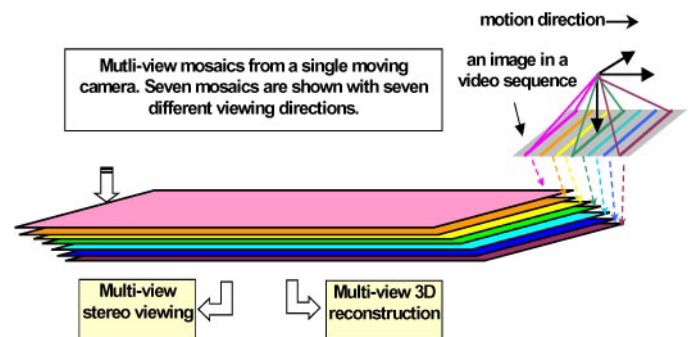


Figure 1. Mosaics: from many narrow field-of-view (FOV) images to a few wide FOV mosaics

obtain a 3D representation of the city and its traffic. User-friendly visualization (the panoramic stereo mosaics) and automatic analysis (the content-based target representations) could be provided for automatic traffic monitoring and anomaly detection.

There are two consecutive steps in constructing a CB3M representation: stereo mosaicking and 3D motion extraction. In the first step, a set of parallel-perspective mosaics⁵—panoramic images combining all the video images—is generated to capture both the 3D and dynamic aspects of the scene under the camera coverage. This step turns hundreds and thousands of images of a video sequence into just a few large field-of-view (FOV) mosaics. Multiple wide FOV mosaics are generated from a single camera, but the results are much like using multiple line-scan cameras with different oblique angles to scan through the entire scene (see Figure 1). Because of the multiple scanning angles, occluded regions in one mosaic can be seen from the others. Moving objects show up in each mosaic, and by switching to different ones, the dynamic aspects can be viewed as well. This feature leads to a very efficient multi-view stereo viewing approach.⁶ The translation and rotation of the virtual camera for a virtual fly-through are implemented by simply sliding a window across the mosaic(s) and switching between different mosaics.

Continued on next page

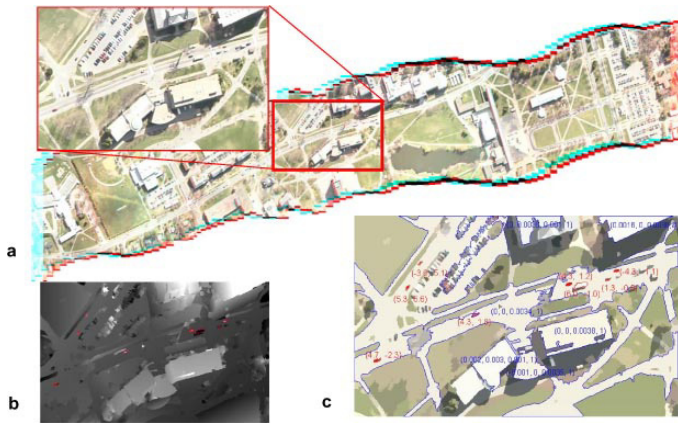


Figure 2. Content-based 3D mosaic representation of an aerial video sequence. (a) A pair of stereo mosaics from the total nine mosaics and a close-up window; (b) the height map of the objects inside that window; (c) the CB3M representation with some of the regions labeled by their boundaries and plane parameters (in blue) and the detected moving targets marked by their boundaries and motion vectors (in red).

In the second step, a segmentation-based stereo matching algorithm⁷ is applied to extract parametric representation of the color, structure, and motion of the dynamic and/or 3D objects in urban scenes. Multiple pairs of stereo mosaics are used for facilitating reliable stereo matching, occlusion handling, accurate 3D reconstruction, and robust moving target detection. Finally a CB3M representation⁸ is created: a highly compressed visual representation for very long video sequences of a dynamic 3D scene. In the CB3M representation, the panoramic mosaics are segmented into planar regions, which are the primitives for content representations. Each region is represented by its mean color, region boundary, plane normal, and distance, as well as by motion, direction, and speed, if it is a dynamic object. Relations of each region with its neighbors are also built for further object recognition and representations, such as buildings and road networks.

Figure 2 shows an example of CB3M from a real video sequence of a campus scene when the airplane was about 300m above the ground. For this example, the original image sequence has 1,000 frames of 640×480 color images. With the CB3M representation, a compression ratio of more than 10,000 is achieved. More importantly, it has object contents.

The CB3M construction and representation provide benefits for many applications, such as urban transportation planning, aerial surveillance, and urban modeling. The panoramic mosaics provide a synopsis of the scene with all the 3D objects and dynamic objects in a single view. The 3D contents of the CB3M representation make further object recognition and higher-level feature extraction possible. The motion contents of the CB3M representation provide dynamic measurements

of moving targets in the large-scale scene. Finally, the CB3M representation is highly compressed. Usually a compression ratio of thousands to ten thousands can be achieved. This saves space when archiving a lot of data for a large area.

This work is supported by the Air Force Research Laboratory under Award No. FA8650-05-1-1853. It is also partially supported by the National Science Foundation (NSF) under Grant No. CNS-0551598, Army Research Office through Grant No. W911NF-05-1-0011, the New York Institute for Advanced Studies, and by the Professional Staff Congress—City University of New York. The image data in Figure 2 were captured in the UMass Computer Vision Lab under NSF grant (No EIA-9726401).

Author Information

Zhigang Zhu

The City College of New York
New York, New York
The CUNY Graduate Center
New York, New York

Zhigang Zhu is an Associate Professor in the Department of Computer Science, City College of the City University of New York, where he directs the City College Visual Computing Laboratory and co-directs the Center for Perceptual Robotics, Intelligent Sensors, and Machines (PRISM). His research interests include 3D computer vision, human-computer interaction, augmented reality, and various applications using machine vision. He has published more than 100 technical papers in related fields and is a senior member of the IEEE and a member of the ACM.

References

1. M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, *Mosaic representations of video sequences and their applications*, **Signal Processing: Image Communication** 8 (4), May 1996.
2. W. H. Leung and T. Chen, *Compression with mosaic prediction for image-based rendering applications*, **IEEE Intl. Conf. Multimedia & Expo.**, New York, July 2000.
3. Y. Li, H.-Y. Shum, C.-K. Tang, and R. Szeliski, *Stereo reconstruction from multiperspective panoramas*, **IEEE Trans. on PAMI** 26 (1), pp. 45–62, 2004.
4. C. Sun and S. Peleg, *Fast panoramic stereo matching using cylindrical maximum surfaces*, **IEEE Trans. SMC Part B** 34, pp. 760–765, Feb 2004.
5. Z. Zhu, E. Riseman, and A. Hanson, *Generalized parallel-perspective stereo mosaics from airborne videos*, **IEEE Trans. PAMI** 26 (2), pp. 226–237, Feb 2004.
6. Z. Zhu and A. R. Hanson, *Mosaic-based 3D scene representation and rendering*, **the IEEE Eleventh International Conference on Image Processing**, pp. I 633–636, Genova, Italy, 11–14 September 2005.
7. H. Tang, Z. Zhu, G. Wolberg, and J. R. Layne, *Dynamic 3D urban scene modeling using multiple pushbroom mosaics*, **the Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006)**, University of North Carolina, Chapel Hill, USA, 14–16 June 2006.
8. Z. Zhu, H. Tang, G. Wolberg, and J. R. Layne, *Content-based 3D mosaics for dynamic urban 3D scenes*, **SPIE Defense and Security Symposium 2006**, Orlando, Florida, USA, 17–21 April 2006.