Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Web-based interactive visualization of 3D video mosaics using X3D standard

CHON Jaechoon, LEE Yang-Won[†], SHIBASAKI Ryosuke

(*Center for Spatial Information Science, the University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan*)

[†]E-mail: jwlee@iis.u-tokyo.ac.jp

Received Apr. 24, 2006; revision accepted May 16, 2006

Abstract: We present a method of 3D image mosaicing for real 3D representation of roadside buildings, and implement a Web-based interactive visualization environment for the 3D video mosaics created by 3D image mosaicing. The 3D image mosaicing technique developed in our previous work is a very powerful method for creating textured 3D-GIS data without excessive data processing like the laser or stereo system. For the Web-based open access to the 3D video mosaics, we build an interactive visualization environment using X3D, the emerging standard of Web 3D. We conduct the data preprocessing for 3D video mosaics and the X3D modeling for textured 3D data. The data preprocessing includes the conversion of each frame of 3D video mosaics into concatenated image files that can be hyperlinked on the Web. The X3D modeling handles the representation of concatenated images using necessary X3D nodes. By employing X3D as the data format for 3D image mosaics, the real 3D representation of roadside buildings is extended to the Web and mobile service systems.

Key words: 3D video mosaics, Interactive 3D visualization, Web 3D
doi:10.1631/jzus.2006.A1259 **Document code:** A **CLC number:** TP39

INTRODUCTION

The visualization of roadside buildings in virtual space using synthetic photorealistic view is one of the common methods for representing the background scene of Car Navigation Systems (CNS) and Internet map services. Since most of these background scenes are implemented by two-dimensional (2D) images, they may give somewhat monotonous impression due to fixed viewpoint position and orientation. For more interactive visualizations that enable arbitrary adjustment of the viewpoint position and orientation, the utilization of three-dimensional Geographic Information Systems (3D-GIS) data can be an alternative approach.

Two major methods for generating 3D-GIS data are (1) reconstruction of detailed 3D surfaces and (2) mosaicing of sequential image frames. The former method employs laser scanner and Charge-Coupled Device (CCD) combined with Global Positioning System (GPS) and Inertial Measurement Unit (IMU). With this complex system, detailed 3D surfaces can be acquired from any building regardless of its texture (Fruh and Zakhor, 2002; Zhao and Shibasaki, 2003). To use stereo images is another way to reconstruct detailed 3D surfaces (Pollefeys *et al*., 2000).

However, laser and stereo system may not be very effective in that they require excessive data processing to reconstruct the detailed 3D surfaces. Besides, it is difficult to apply detailed 3D surfaces to current Web or mobile systems because of the limitation of their transmission speed. In this sense, image mosaicing techniques are thought to be more appropriate method for visualizing roadside buildings on the Web and mobile systems because they can reduce the difficulties of generating textured 3D-GIS data and provide the lightweight 3D scenes suitable for human perception.

In this paper, we present a method of 3D image mosaicing for real 3D representation of roadside buildings, and implement a Web-based interactive visualization environment for the 3D video mosaics created by 3D image mosaicing for which we employ multiple projection planes to overcome the 2D-confined representation of the existing crossed-slits projection technique. Our method concatenates a sequence of vertical planar faces to create a 3D model in which the image frames are back-projected as texture.

For the Web-based interactive visualization environment, we employ eXtensible 3D (X3D), the emerging standard of Web 3D. X3D is an advanced eXtensible Markup Language (XML) version of Virtual Reality Markup Language (VRML) to enable real-time communication of 3D data across all applications and network applications. By using X3D as the data format for 3D image mosaics, the real 3D representation of roadside buildings can be extended to the Web and mobile service systems.

While the background and objectives of this study are briefly described in this section, we explore in more detail the issues on the image mosaicing in Section 2 of the paper. Optical flow detection and camera orientation setup for extracting 3D data from a sequence of images are examined in Section 3. As to the 3D image mosaicing technique, Section 4 describes the concept and procedure of multiple projection planes. Section 5 deals with the X3D standard for Web-based interactive 3D visualization. Demonstrations of the X3D-based interactive visualization of 3D video mosaics are presented in Section 6. Section 7 concludes the paper with a summary and previews of our future work.

RELATED WORK

Image mosaicing technique builds a compiled image covering a large area by connecting a series of 2D images to a plane. It is used in a variety of applications including satellite imagery mosaics, virtual reality modelling (Szeliski, 1996), medical image mosaics (Chou *et al.*, 1997), and video compression. The image mosaicing techniques are divided into two types according to the dependency on a given 3D vector data. In the first type, images of a perspective projection are registered to given 3D vector data (Hartly and Zisserman, 2000; Mikhail *et al.*, 2001; Jiang *et al.*, 2004). In the second type, images of a perspective projection are conjugated without given 3D vector data. Alternatively, the latter type includes two methods for obtaining textured 3D vector data: panoramic or spherical mosaicing using a sequence of images taken from pan/tilt cameras (Mann and Picard, 1994; Chen, 1995; McMillan and Bishop, 1995; Krishnan and Ahuja, 1996; Coorg and Teller, 2000; Shum and Szeliski, 2000) and image mosaicing using a sequence of images taken from moving cameras (Zheng and Tsuji, 1992; Peleg *et al.*, 2000; Zomet *et al.*, 2000; 2003; Roman *et al.*, 2004; Zhu *et al.*, 2004).

If images are acquired by a tilted camera, the image mosaics created by generic perspective projection and affine transformation tend to be curled. Zomet *et al.*(2000) presented a solution to this problem by warping trapezoids into rectangles while maintaining other image feature invariants. In addition, parallel-perspective mosaicing using a moving camera was proposed by Zheng and Tsuji (1992) and later developed by Peleg *et al.*(2000) and Zhu *et al.*(2004). This technique, first computes the relative position between two consecutive frames for all pairs of a sequential image set. Then, the center strips are extracted from each frame and placed in corresponding positions to build an image mosaic.

However, parallel-perspective mosaicing based on one projection plane is not appropriate for a side-looking video camera at the intersection. Zomet *et al.*(2003) developed crossed-slits projection technique to solve this problem. Roman *et al.*(2004) proposed several user-specified slits to apply the crossed-slits projection technique. This technique, however, has a disadvantage in that the image motion for each frame is limited to less than a single pixel when generating an image mosaic with the original resolution. In addition, it is difficult to calculate accurate camera orientation for well-aligned crossed-slits images. More importantly, the image mosaics based on the crossed-slits projection do not provide a 3D impression, as the roadside buildings look like standing on a flat street.

To overcome the drawbacks of the crossed-slits projection technique and to reduce the costs and difficulties of creating textured 3D data, we proposed 3D image mosaicing in our previous work (Chon *et al.*, 2004). As a novel 3D version of image mosaicing, our method provides real 3D impression for visualizing roadside buildings. The 3D video mosaics composed of 3D vector data and textured image slits are obtained by a side-looking video camera. The 3D impression of overall scene and each object is originated from the combination of 3D vector data and textured image slits as the result of our 3D image mosaicing which uses a sequence of textured vertical planar faces named "multiple projection planes". The scene geometry is approximated to multiple vertical planar faces using sparsely distributed feature points that are assigned to 3D vector data through bundle adjustments (Hartly and Zisserman, 2000; Mikhail *et al.*, 2001). The feature points are extracted and tracked by an edge tracking algorithm based on epipolar geometry (Han and Park, 2000; Hartly and Zisserman, 2000). These vertical planar faces are concatenated to create a 3D model where the image frames are back-projected as texture.

OPTICAL FLOW DETECTION AND CAMERA ORIENTATION SETUP

Since the 3D data of feature points can be calculated by using collinearity condition with camera parameters in two previous frames (Mikhail *et al.*, 2001), it is necessary for the feature points to be tracked in at least three successive frames. For the robust tracking of several feature points in the three successive frames, we employ an algorithm that tracks each pixel of the edges based on epipolar geometry. The epipolar lines are in almost horizontal direction according to the motion of the moving video camera. Tracking each pixel along the horizontal edges leads to the increase in mismatch rate because the same textures are often found in a horizontal edge. To reduce this mismatch rate, we extract only vertical edges using Canny operator (Canny, 1986).

Fig.1c shows the tracked feature points of vertical edges extracted from the previous and current frames in Figs.1a and 1b, respectively. Since the well-distributed feature points are appropriate for reducing the approximation error of camera parameters, we select best-matched feature points in an image divided into 5×5 blocks like the pattern of chessboard. Fig.1d illustrates the tracked contours (i.e., optical flows) composed of feature points that score the best-matched rate for each block. These optical flows are used as the criteria of vertical planar approximation of the scene geometry for each frame.

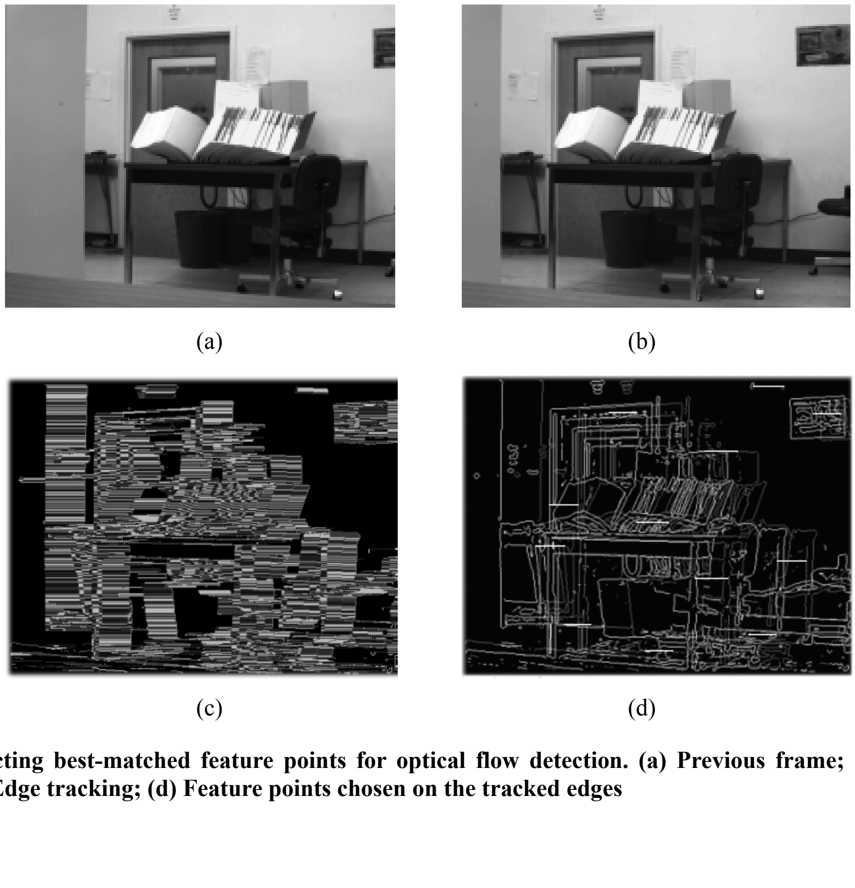

(a) (b) (c) (d)

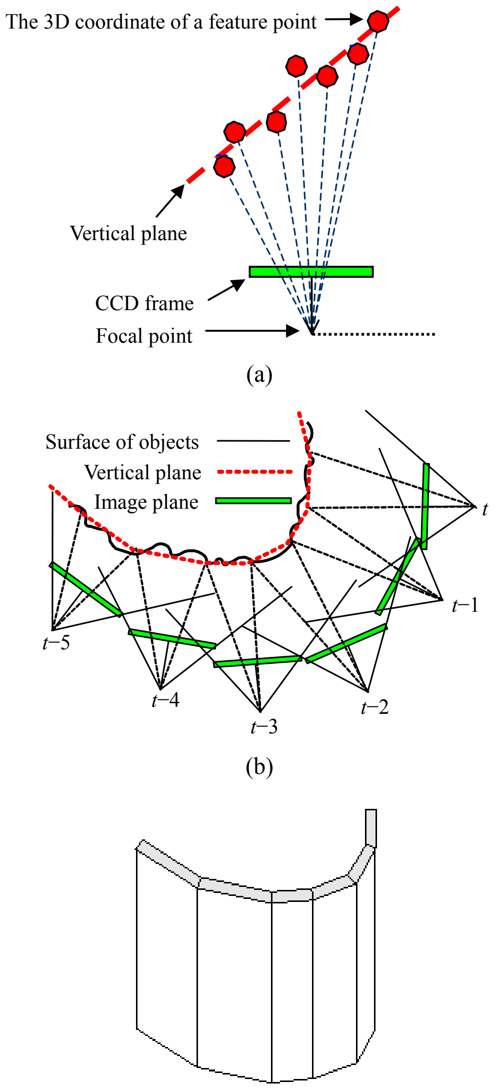
Fig.1 Selecting best-matched feature points for optical flow detection. (a) Previous frame; (b) Current frame; (c) Edge tracking; (d) Feature points chosen on the tracked edges

In order to build 3D data using the optical flows, the information of exterior and interior parameters of camera is required. Suppose the interior parameters are given, the exterior parameters for photo-triangulation should be approximated for a rigorous solution to nonlinear problems in the collinearity condition. This approximation solution can be built by a closed form space resection (Zheng and Wang, 1992) or a classical nonlinear space resection in the collinearity condition, given four or more feature points (Hartly and Zisserman, 2000).

We employ the classical nonlinear space resection to obtain the exterior parameters of the camera. Assuming that the first frame is based on an arbitrary global reference system, the exterior parameters of the second frame can be approximated in the coplanarity condition. The exterior parameters from the third to the last frame are approximated by bundle adjustment in the collinearity condition. The approximation of the second frame uses coplanarity condition instead of bundle adjustment in order to reduce divergence probability. The number of unknown parameters in the collinearity condition is at least twenty-four, whereas that of the coplanarity condition is six. The approximation from the third to the last frame requires only six unknown parameters owing to the information of 3D data of the referenced feature points.

3D IMAGE MOSAICING USING MULTIPLE PROJECTION PLANES

Seen from the sky, the building wall surface appears to be a line. Thus, we can regard each image frame representing the wall surface as a vertical planar face. This vertical planar face is approximated by using 3D data of sparsely distributed feature points in an image frame. The Least Median of Squares (LMedS) method is employed for approximating the projection planar face, with the computation of the regression line (*Z*=*aX*+*b*) as in Fig.2a (Rousseeuw, 1984). Suppose the roadside building wall surface is the thick curve, and the position of a side-looking video camera is each rectangle in Fig.2b, the multiple planar faces are represented in the dotted curve of Fig.2b. The 3D representation of these multiple projection planes is as shown in Fig.2c.



(a)

(b)

(c)

Fig.2 Concatenation of vertical planar faces generated by the sparsely distributed feature points for each frame. (a) Vertical planar face of each frame; (b) Multiple projection planes in 2D space; (c) Multiple projection planes in 3D space

X3D STANDARD FOR WEB-BASED INTERACTIVE 3D VISUALIZATION

Three components are central to designing a Web-based interactive 3D visualization (Ying *et al.*, 2004). The first one is data source and storage for the 3D video mosaics. Each image frame of the 3D video mosaics is stored in the form of an image file, so that it can be hyperlinked and concatenated on the Web.

The second component is the information technology to support open accessibility and interactivity of the Web. The third component is a 3D graphics method for Web representation. For incorporating these three components, we employ the X3D standard that ensures interactive 3D visualization on the Web.

X3D is a software standard for defining interactive Web- and broadcast-based 3D content integrated with multimedia. X3D is intended for use on a variety of hardware devices and in a broad range of application areas such as engineering and scientific visualization, multimedia presentations, entertainment and educational titles, Web pages, and shared virtual worlds. X3D is also intended to be a universal interchange format for integrated 3D graphics and multimedia. X3D is the successor to the VRML, the original ISO standard for Web-based 3D graphics. X3D improves upon VRML (Fig.3) with new features, advanced Application Programming Interfaces (API), additional data encoding formats, stricter conformance, and a componentized architecture that ensures a modular approach to related standards (Tsai *et al*., 2004).

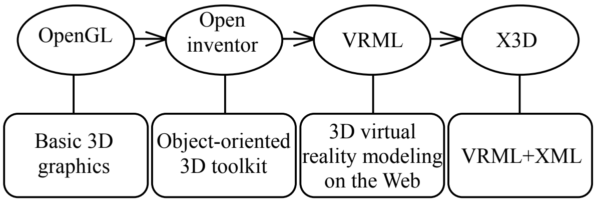

OpenGL
Open inventor
VRML
X3D
Basic 3D graphics
Object-oriented 3D toolkit
3D virtual reality modeling on the Web
VRML+XML

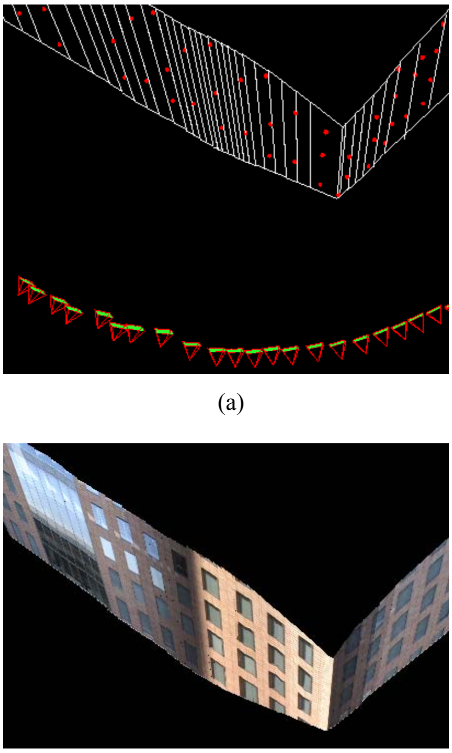
Fig.3 Evolution of 3D graphics

Main goal of X3D scene model is to provide a precise and rich representation framework for geometric and environmental features. An X3D scene graph is a directed acyclic graph. Nodes can contain specific fields with one or more children nodes that participate in the hierarchy (Bilasco *et al*., 2005). Using the nodes, 3D functionalities such as polygonal geometry, parametric geometry, hierarchical transformations, lighting, materials, multi-pass/multi-stage texture mapping, etc. are implemented.

WEB DEMONSTRATION OF 3D VIDEO MOSAICS IN X3D

Our Web-based visualization environment for 3D video mosaics is composed of two parts: the creation of 3D video mosaics and the generation of X3D document.

First, for the creation of 3D image mosaics, we conduct the optical flow detection, camera orientation setup, and 3D image mosaicing with multiple projection planes, using a sequence of image frames acquired by side-looking video camera (Fig.4).



(a)

(b)

Fig.4 Creation of 3D video mosaics. (a) Best-matched feature points and multiple projection planes; (b) Textured planar faces for 3D vector data

Second, for the generation of X3D document, we conduct the data preprocessing for 3D video mosaics and the X3D modelling for textured 3D data. The data preprocessing includes the conversion of each frame of 3D video mosaics into concatenated image files that can be hyperlinked on the Web. The X3D modelling handles the representation of concatenated images using necessary X3D nodes. <Shape> node includes the information of 3D vector data and texture image. <IndexedFaceSet> node, a child node of <Shape>, defines a 3D vector surface model based on the polygons derived from irregularly distributed height points (Gelautz *et al*., 2004; Farrimond and Hetherington, 2005). <Appearance> node, which is also a child node of <Shape>, defines a hyperlink to the texture image assigned to corresponding 3D vector data. As a client-side user interface, we employ Octaga Player (http://www.octaga.com) for an ActiveX plug-in of Web browsers. Fig.5 illustrates the Web-based interactive visualization of 3D video mosaics in X3D, as the result of our work. This X3D document provides the arbitrary adjustment of viewpoint position and orientation in addition to the real 3D representation of roadside buildings as a textured 3D-GIS data on the Web.

CONCLUSION

We presented a method of 3D image mosaicing for real 3D representation of roadside buildings, and implemented a Web-based interactive visualization environment for the 3D video mosaics created by 3D image mosaicing. The 3D image mosaicing technique developed in our previous work is a very powerful method for creating textured 3D-GIS data without excessive data processing like the laser or stereo system. For the Web-based open access to the 3D video mosaics, we built an interactive visualization environment using X3D, the emerging standard of Web 3D. We conducted the data preprocessing for 3D video mosaics and the X3D modelling for textured 3D data. The data preprocessing includes the conversion of each frame of 3D video mosaics into concatenated image files that can be hyperlinked on the Web. The X3D modelling handles the representation of concatenated images using necessary X3D nodes. By employing X3D as the data format for 3D image mosaics, the real 3D representation of roadside buildings is extended to the Web and mobile service systems.

As the result of our work, the X3D document composed of 3D video mosaics created by 3D image mosaicing provides the arbitrary adjustment of viewpoint position and orientation in addition to the real 3D representation of roadside buildings as textured 3D-GIS data on the Web. One of the advantages of X3D is the interoperability with Moving Picture Experts Group Layer 4 (MPEG-4), the multimedia standard for wired and wireless Internet. An animated

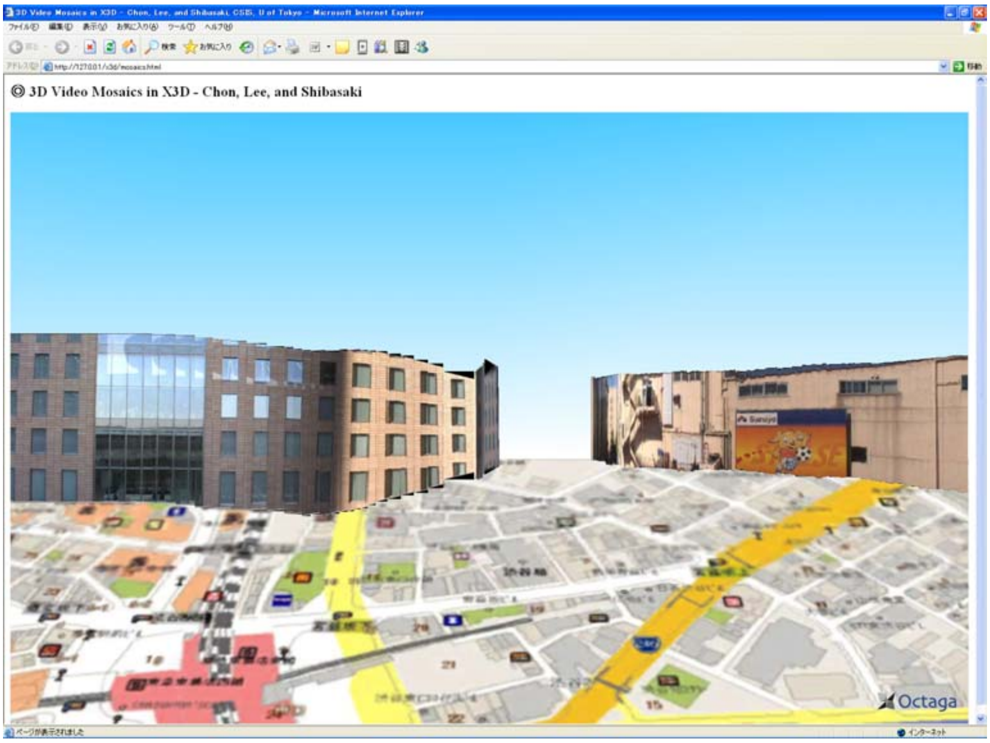

3D Video Mosaics in X3D - Chon, Lee, and Shibasaki, CSIS, U of Tokyo - Microsoft Internet Explorer
3D Video Mosaics in X3D - Chon, Lee, and Shibasaki
Octaga

Fig.5 Web demonstration of 3D video mosaics in X3D

X3D document is converted into MPEG-4 format so that it can be broadcasted on the Web and mobile systems, with the streaming service for stable real-time multimedia transmission.

References

Bilasco, I.M., Gensel, J., Villanova-Oliver, M., Martin, H., 2005. 3DSEAM: A Model for Annotating 3D Scenes using MPEG-7. Proceedings of the 7th IEEE International Symposium on Multimedia, p.310-319.

Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **8**(6):679-698.

Chen, S.E., 1995. Quicktime VR—An Image-based Approach to Virtual Environment Navigation. Proceedings of ACM SIGGRAPH'95, p.29-38.

Chon, J., Fuse, T., Shimizu, E., 2004. Urban visualization through video mosaics based on 3D multi-baselines, the International Archives of the Photogrammetry. *Remote Sensing and Spatial Information Science*, **35**(B3): 727-731.

Chou, J.S., Qian, J., Wu, Z., Schramm, H., 1997. Automatic Mosaic and Display from a Sequence of Peripheral Angiographic Images. Proceedings of SPIE Medical Imaging, **3034**:1077-1087.

Coorg, S., Teller, S., 2000. Spherical mosaics with quaternions and dense correlation. *International Journal of Computer Vision*, **37**(3):259-273. [doi:10.1023/A:1008184124789]

Farrimond, B., Hetherington, R., 2005. Compiling 3D Models of European Heritage from User Domain XML. Proceedings of the 9th International Conference on Information Visualisation, p.163-171.

Fruh, C., Zakhor, A., 2002. Data Processing Algorithms for Generating Textured 3D Building Facade Meshes from Laser Scans and Camera Images. Proceedings of International Symposium on 3D Data Processing, Visualization, and Transmission, p.834-847.

Gelautz, M., Brandejski, M., Kilzer, F., Amelung, F., 2004. Web-based Visualization and Animation of Geospatial Data using X3D. Proceedings of 2004 IEEE International Geoscience and Remote Sensing Symposium, **7**:4773-4775.

Han, J.H., Park, J.S., 2000. Contour matching using epipolar geometry. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **22**(4):358-370. [doi:10.1109/34.845378]

Hartly, R., Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Jiang, B., You, S., Neumann, U., 2004. A Robust Tracking System for Outdoor Augmented Reality. Proceedings of IEEE Virtual Reality 2004, p.3-10.

Krishnan, A., Ahuja, N., 1996. Panoramic Image Acquisition. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition 1996, p.379-384. [doi:10.1109/CVPR.1996.517100]

Mann, S., Picard, R., 1994. Virtual Bellows: Constructing High Quality Stills from Video. Proceedings of the 1st IEEE International Conference on Image Processing, **1**:363-367. [doi:10.1109/ICIP.1994.413336]

McMillan, L., Bishop, G., 1995. Plenoptic Modelling: An Image Based Rendering System. Proceedings of ACM SIGGRAPH'95, p.39-46.

Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. Introduction to Modern Photogrammetry. John Wiley & Sons.

Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A., 2000. Mosaicing on adaptive manifolds. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **22**(10):1144-1154. [doi:10.1109/34.879794]

Pollefeys, M., Koch, R., Vergauwen, M., van Gool, L., 2000. Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **55**(4):251-267. [doi:10.1016/S0924-2716(00)00023-X]

Roman, A., Garg, G., Levoy, M., 2004. Interactive Design of Multi-perspective Images for Visualizing Urban Landscapes. Proceedings of IEEE Visualization 2004.

Rousseeuw, P.J., 1984. Least median of squares regression. *Journal of American Statistical Association*, **79**(388): 871-880. [doi:10.2307/2288718]

Shum, H.Y., Szeliski, R., 2000. Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, **36**(2):101-130. [doi:10.1023/A:1008195814169]

Szeliski, R., 1996. Video mosaic for virtual environments. *IEEE Computer Graphics and Applications*, **16**(2):22-30. [doi:10.1109/38.486677]

Tsai, J.F., Kouh, J.S., Chen, L., 2004. Constructing the Simulation Examples for the Courses of Dynamics and Fluid Mechanics by X3D. Proceedings of MTS/IEEE TECHNO-OCEAN'04, **1**:573-577.

Ying, J., Gracanin, D., Lu, C.T., 2004. Web Visualization of Geo-spatial Data using SVG and VRML/X3D. Proceedings of the 3rd International Conference on Image and Graphics, p.497-500. [doi:10.1109/ICIG.2004.147]

Zhao, H., Shibasaki, R., 2003. A vehicle-borne urban 3D acquisition system using single-row laser range scanners. *IEEE Trans. on SMC Part B: Cybernetics*, **33**(4):658-666. [doi:10.1109/TSMCB.2003.814280]

Zheng, J.Y., Tsuji, S., 1992. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, **9**(1):55-76. [doi:10.1007/BF00163583]

Zheng, Z., Wang, X., 1992. A general solution of a closed-form space resection. *Photogrammetric Engineering & Remote Sensing Journal*, **58**(3):327-338.

Zhu, Z., Hanson, A.R., Riseman, E.M., 2004. Generalized parallel-perspective stereo mosaics from airborne video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **26**(2):226-237. [doi:10.1109/TPAMI.2004.1262190]

Zomet, A., Peleg, S., Arora, C., 2000. Rectified Mosaicing: Mosaics without the Curl. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, **II**:459-465.

Zomet, A., Feldman, D., Peleg, S., Weinshall, D., 2003. Mosaicing new views: the crossed-slits projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **25**(6):741-754. [doi:10.1109/TPAMI.2003.1201823]