Presented at AIPR 2003: Imagery and Data Fusion, October 15-17 in Washington, DC, USA

Stereo Mosaics with Slanting Parallel Projections from Many Cameras or a Moving Camera

Zhigang Zhu
Department of Computer Science, The City College of New York, New York, NY 10031
zhu@cs.ccny.cuny.edu

Abstract

This paper presents an approach of fusing images from many video cameras or a moving video camera with external orientation data (e.g. GPS and INS data) into a few mosaiced images that preserve 3D information. In both cases, a virtual 2D array of cameras with FOV overlaps is formed to generate the whole coverage of a scene (or an object). We propose a representation that can re-organize the original perspective images into a set of parallel projections with different slanting viewing angles. In addition to providing a wide field of view, there are two more benefits of such a representation. First, mosaics with different slanting views represent occlusions encountered in a usual nadir view. Second, stereo pair can be formed from a pair of slanting parallel mosaics thus image-based 3D viewing can be achieved. This representation can be used as both an advanced video interface for surveillance or a pre-processing for 3D reconstruction.

1. Introduction

It is a commonplace to generate a 2D panoramic mosaic of the 3D scene from a moving camera, with a single multiple-viewpoint viewing direction [1,2], but 3D information and/or surface information from other viewing directions is lost in such a representation. A digital elevation map (DEM) generated from aerial photometry consists of a sampled array of elevations (depths) for a number of ground positions at regularly spaced intervals [3]. It usually only has a nadir viewing direction, hence the surfaces from other viewing directions cannot be represented. However, in some applications such as surveillance and security inspection, a scene or an object (e.g. a vehicle) needs to be observed from many viewing directions to reveal the abnormal areas hidden in unusual views. Stereo panoramas [4,5] have been presented to obtain the best 3D information from an off-center rotating camera. In the case of a translating camera, various layered representations [6-8] have been proposed to represent both 3D information and occlusions, but such representations need 3D reconstructions.

This paper presents an approach of fusing images from many spatially distributed video cameras or a moving video camera with external orientation data (e.g. GPS and INS data) into a few mosaiced images that preserve 3D information. In both cases, a virtual 2D array of cameras with FOV overlaps is formed to generate the whole coverage of a scene (or an object). As a matter of fact, many viewing directions are included in the original camera views. X-slit mosaics with non-parallel rays [9] have been proposed using this property to generate mosaics for image-based rendering. In this paper we propose a representation that can re-organize the original perspective images into a set of parallel projections with different slanting viewing angles (in both the x and the y directions of the 2D images). Such representations provide a wide field of view, 3D information for stereo viewing and reconstruction, and the capability to represent occlusions. This representation can be used as both an advanced video interface for surveillance or a pre-processing for 3D reconstruction and scene representation.

As the organization of this paper, we first present the stereo mosaicing representations with a set of slanting parallel projections in both directions. Second we show how to construct a 2D (virtual) camera array in three different cases. Third we present the Parallel Ray Interpolation for Stereo Mosaicing (PRISM) approach for generate mosaics under real camera setups and for arbitrary 3D scenes. Then we analyze the advantages of such representations in stereo viewing and 3D reconstruction. Finally experimental results are given for two important applications – aerial video surveillance and under vehicle inspection. In the aerial video case, a moving camera is accompanied by GPS and INS measurements to provide orientation data. In the under vehicle inspection system, a pre-calibrated 1D array of cameras are used to scan the bottom of a vehicle when the vehicle is driven over the camera array. Finally a brief summary is given.

2. 2D Slanting Parallel Projection

A normal perspective camera has a single viewpoint, which means all the rays pass through a common nodal point. On the other hand, an orthogonal image with parallel projections in both the x and y directions has all the rays parallel to each other. Imagining that we have a sensor with parallel projections, we could rotate the sensor to capture images with different *slanting* angles (including nadir and oblique angles) in both directions so that we can create many pair of parallel stereo images with two different slanting angles, and observe surfaces that could be occluded in a nadir view.

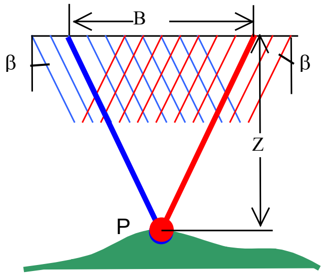


Fig. 1. Depth from parallel stereo with multiple viewpoints: 1D case.

Fig. 1 shows the parallel stereo in a 1D case, where two slanting (oblique) angles are chosen to construct stereo geometry. The depth of a point can be calculated as (Fig. 1)

$$Z = \frac{B}{2tg\beta} \qquad (1)$$

where 2β is the angle between the two viewing directions, and B is the adaptive baseline between the two viewpoints that construct the triangulation relation. It has been shown by others [10] and by us [11, 12] that parallel stereo is superior to both conventional perspective stereo and to the recently developed multi-perspective stereo with concentric mosaics for 3D reconstruction (e.g., in [5]), in that the adaptive baseline inherent in the parallel-perspective geometry permits depth accuracy independent of absolute depth in theory [10,11] and as a linear function of depth in stereo mosaics from perspective image sequences [12].

We can make two extensions to this parallel stereo. First, we can select various slanting angles for constructing multiple parallel projections. By doing so we can observe various degrees of occlusions and can construct stereo pairs with different depth resolution via the changes of baselines. Second, we can extend this 1D parallel projection to 2D (Fig. 2): we can obtain a mosaiced image that has a nadir view (Fig. 2a), slanting angle(s) only in one direction (Fig. 2b and c) or in both the x and the y directions (Fig. 2d).

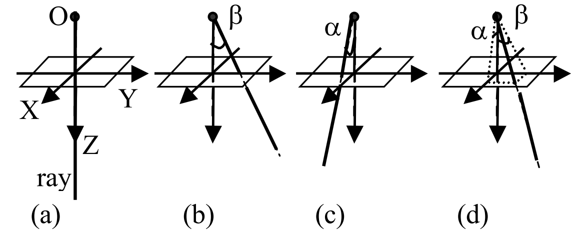


Fig. 2. Parallel projections with two slanting angles α and β (in the *x* and *y* directions). (a) Nadir view (α=β=0); (b) y-slanting view (α=0, β≠0); (c) x-slanting view (α≠0, β=0) and (d) dual-slanting view (α≠0, β≠0). Parallel mosaics can be formed by populating the single selected ray in each case in both the *x* and *y* directions.

3. 2D (Virtual) Array of Cameras

It is impractical to use a single sensor to capture orthogonal images with full parallel projections in both x and y dimensions for a large scene, and with various oblique directions. However we could have at least three practical approaches in generate images with slanting parallel projections with existing sensors: a 2D sensor array of many spatially distributed cameras, a "scanner" with a 1D array of cameras, and a single perspective camera that moves (Fig. 3).

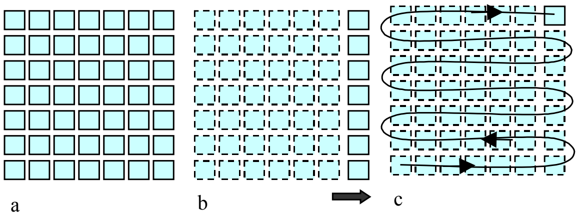


Fig. 3. Parallel mosaics from 2D bed of cameras. (a) 2D array; (b) 1D scanning array and (c) a single scanning camera.

With a 2D array of many perspective cameras (Fig. 3a), we first assume that the optical axes of all the cameras point to the same directions (inside the paper in Fig 3a), and the viewpoints of all cameras are on a plane perpendicular to their optical axes. Then we can reorganize the perspective images into mosaiced images with any slanting viewing angles by extract rays from the original perspective images with the same viewing directions, one ray from each image. If the camera array is dense enough, then we can generate densely mosaiced images.

If we only have a 1D linear array of perspective cameras (Fig. 3b), we can translate the camera to scan over the scene to synthesize a virtual 2D bed of camera array. Then we can still generate stereo mosaic pairs with slanting parallel projections, given that we can accurately control the translation of the camera array. We have actually used this approach in an Under Vehicle Inspection System (UVIS) [13, 14, 18].

Even if we just use a single camera, we can still generate a 2D virtual bed of cameras by moving the cameras in two dimensions, along a 2D scanning path shown in Fig. 3c. This is the case for aerial video mosaics [11, 12, 15, 17].

4. PRISM: Video Mosaicing Algorithm

In real applications, there are two challenging issues. First, it is difficult to have all cameras point to the same directions, with their viewpoints in a plane. Second, it is impractical to have such a dense camera array (or such a dense scan) to generate dense parallel mosaics. However, we can still generate dense parallel mosaics after we solve the following two main problems.

The first problem is camera orientation estimation (calibration). It is well known that camera calibration is a hard problem, especially for a moving camera. In our previous study of aerial video application, we used external orientation instruments, i.e., GPS, INS and laser profiler to ease the problem of camera orientation estimation [11, 12]. In this paper, we assume that the extrinsic and intrinsic camera parameters are known at each camera location.

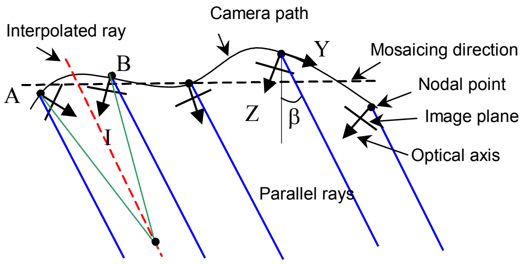


Fig. 4. Ray interpolation for parallel mosaicing from an arbitrary camera array.

The second problem is to generate dense parallel mosaics with a sparse, uneven, camera array, and for a complicated 3D scene. To solve this problem, a Parallel Ray Interpolation for Stereo Mosaics (PRISM) approach has been proposed [11]. While the PRISM algorithm was originally designed to generate parallel-perspective stereo mosaics (parallel projection in one direction and perspective projection in the other), the core idea of *ray interpolation* could be used for generating mosaics with full parallel projection of any slanting angles.

Fig. 4 shows how the PRISM works for 1D images. The 1D camera has two axes – the optical axis (Z) and the Y-axis. Given the known camera orientation at each camera location, one ray with a given slanting angle β can be chosen from the image at each camera location to contribute to the parallel mosaic with the same slanting angle β. The slanting angle is defined against the direction perpendicular to the *mosaicing direction*, which is the mean direction of the camera path (Fig. 4). But the problem is that the "mosaiced" image with only those existing rays will be sparse and uneven since the camera array cannot be regular and very dense. Therefore interpolated parallel rays between a pair of parallel rays from two successive images should be generated by performing local matching between these two images, or other additional images. The assumption is that we can found at least two images to generate the parallel ray. Such an interpolated ray is shown in Fig 4, where Ray I is interpolated from Image A and Image B.

One interesting property of the parallel mosaics is that all the (virtual) viewpoints are in infinite. Therefore, even if the original camera path has large deviation in the direction perpendicular to the mosaicing direction, we can still generate full parallel mosaics. However, we should note that in practice, too large deviation in the perpendicular direction will result in a captured image sequence with rather different image resolutions, hence the resulted mosaics will have an uneven spatial resolution.

The extension to 2D images (particular to the X direction of the cameras) of the above approach is straightforward, and a similar region triangulation strategy as in [11] can be applied here to deal with 2D images. However, one practical issue here is the selection of neighborhood images of each image for ray interpolation. For example, with a 1D scan sequence of a single camera, it is hard to generate full parallel projection in the X direction perpendicular to the motion of the camera, since the interpolated parallel rays far off the center of the images in the x direction have to use rays with rather different oblique angles in the original perspective images.

5. Stereo Viewing and 3D Reconstruction

Parallel mosaics with various slanting angles represent scenes from the corresponding viewing angles with parallel rays, with virtually endless fields of view. There are two obvious applications of such representation. First, a human can perceive the 3D scene with a pair of mosaics with different slanting angles (e.g. using polarized glasses) without any 3D recovery. If we have mosaics with various slanting angles in both the x and the y direction, we can generate a virtual fly/walk-through – the translation in the *xy* plane can be simulated by shifting the current displayed mosaic pair, the rotations around the X and the Y axes can be simulated by selecting different pairs of mosaics, and the rotation around the optical axis only needs to rotate the pair of mosaics in their image planes.

Second, for 3D recovery, matches are only performed on a pair of mosaics, not on individual video frames. Stereo mosaic methods also solve the baseline versus field-of-view (FOV) dilemma efficiently by extending the FOV in the directions of mosaicing – in both the *x* and *y* directions. More important, the parallel stereo mosaics have fixed "disparities" and optimal/adaptive baselines for all the points, which leads to uniform depth resolution in theory and linear depth resolution in practice. For 3D reconstruction, epipolar geometry is rather simple due to the full parallel projections in the mosaic pair.

6. Experimental Examples

We present results of stereo mosaics for two applications: airborne videography for aerial surveillance, and 1D video array for under-vehicle inspection.

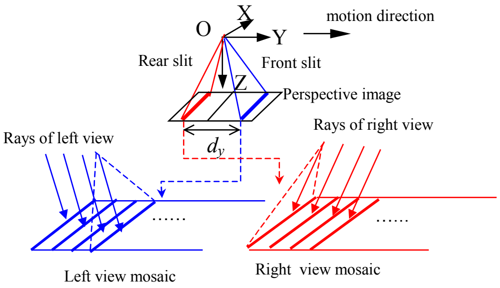


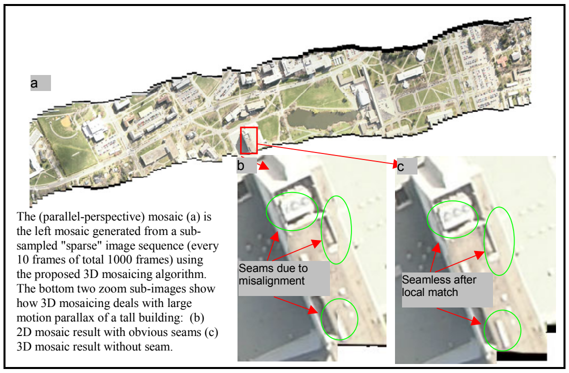
Fig. 5. Parallel-perspective stereo mosaics with a 1D scan path of camera motion.

6.1. Video mosaics from aerial video

First we assume the motion of a camera is an ideal 1D translation, the optical axis is perpendicular to the motion, and the frames are dense enough. Then, we can generate two spatio-temporal images by extracting two columns of pixels (perpendicular to the motion) at the front and rear edges of each frame in motion (Fig. 5). The mosaic images thus generated are *parallel-perspective*, which have perspective projection in the direction perpendicular to the motion and parallel projection in the motion direction. In addition, these mosaics are obtained from two different oblique viewing angles of a single camera's field of view, so that a stereo pair of left and right mosaics captures the inherent 3D information. Note that we do not generate parallel projection in the *x* direction for this 1D scan case due to the difficulty mentioned in Section 4.

In the aerial video application, a single camera is mounted in a small aircraft undergoing 6 DOF motion, together with a GPS, INS and laser profiler to measure the moving camera locations and the distances of the terrain [11, 12]. So we can generate seamless stereo parallel-perspective video mosaic strips from image sequences with a 1D scan path, but with a rather general motion model, using the proposed parallel ray interpolation for stereo mosaicing (PRISM) approach [11]. In the PRISM approach for large-scale 3D scene modeling, the computation is efficiently distributed in three steps: camera pose estimation via the external measurement units, image mosaicing via ray interpolation, and 3D reconstruction from a pair of stereo mosaics.

In principle, we need to match all the points between the two overlapping slices of the successive frames to generate a complete parallel-perspective mosaic. In an effort to reduce the computational complexity, we have designed a fast PRISM algorithm [11] based on the proposed PRISM method. It only requires matches between a set of point pairs in two successive images, and the rest of the points are generated by warping a set of triangulated regions defined by the control points in each of the two images. The proposed fast PRISM algorithm can be easily extended to use more feature points (thus smaller triangles) in the overlapping slices so that each triangle really covers a planar patch or a patch that is visually indistinguishable from a planar patch, or to perform pixel-wise dense matches to achieve true parallel-perspective geometry.



The (parallel-perspective) mosaic (a) is the left mosaic generated from a sub-sampled "sparse" image sequence (every 10 frames of total 1000 frames) using the proposed 3D mosaicing algorithm. The bottom two zoom sub-images show how 3D mosaicing deals with large motion parallax of a tall building: (b) 2D mosaic result with obvious seams (c) 3D mosaic result without seam.

Fig. 6. Parallel-perspective mosaics of a campus scene from an airborne camera.

Fig. 6 shows mosaic results from an aerial video sequence of a cultural scene. Please compare the results of parallel-perspective mosaicing via the

PRISM approach [11] vs. 2D mosaicing via similar approach as the manifold mosaicing [2], by looking along many building boundaries associating with depth changes in the entire 4160x1536 mosaics at our web site [15]. Since it is hard to see subtle errors in the 2D mosaics of the size of Fig. 6a, Fig. 6b and Fig. 6c show close-up windows of the 2D and 3D mosaics for the same portion of the scene with the tall Campus Center building. In Fig. 6b the multi-perspective mosaic via 2D mosaicing has obvious seams along the stitching boundaries between two frames. It can be observed by looking at the region indicated by circles where some fine structures (parts of a white blob and two rectangles) are missing due to misalignments. As expected, the parallel-perspective mosaic via 3D mosaicing (Fig. 6c) does not exhibit these problems.

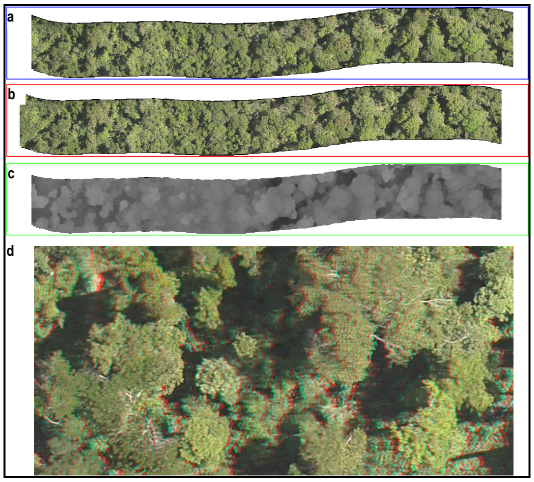


Fig. 7. Stereo mosaics and 3D reconstruction of a 166-frame telephoto video sequence. (a) Left mosaics (b) right mosaics (c) depth map and (d) stereoscopic viewing using left-blue/right-red glasses.

As another example, Fig. 7 shows a real example of stereo mosaics (with two y-slanting angles) generated from a telephoto camera and 3D recovery for a forest scene in Amazon rain forest. The average height of the airplane is $H = 385$ m, and the distance between the two slit windows is selected as 160 pixels (in the y direction) with images of 720 (x) *480 (y) pixels. The image resolution is about 7.65 pixels/meter. The depth map of stereo mosaics in Fig. 7c was obtained by using a hierarchical sub-pixel dense correlation method [16], where the range of depth variations of the forest scene (from a stereo fixation plane) is from -24.0 m (tree canopy) to 24.0 m (the ground). Even before any 3D recovery, a human observer can perceive the 3D structure of the scene with a stereo pair (Fig. 7d).

We have used the same instrumentation package (GPS/INS/Video camera) to generate multiple slanting parallel-perspective mosaics, each of them has parallel projection (with a slanting angle) in the direction of the camera path and perspective projection perpendicular to that direction. Multiple slanting parallel-perspective mosaics can be used for image-based rendering as discussed in Section 5. A mosaic-based fly-through demo may be found at [17], which uses 9 slanting mosaics generated from real video sequence of the UMass campus. This demo shows *parallax, occlusion and moving objects* in multiple parallel-perspective mosaics. We note that the rendering shows parallel-perspective rather than true perspective perception. However, a true perspective fly-through will be enabled by 3D reconstruction from the multiple mosaics.

6.2. Video mosaics for under-vehicle inspection

The slanting parallel projection has been also applied to a 2D (virtual) camera array where viewpoints of the original images are distributed in a 2D array, which will further extend the FOV in two spatial directions, with parallel projections in both the x and y directions.

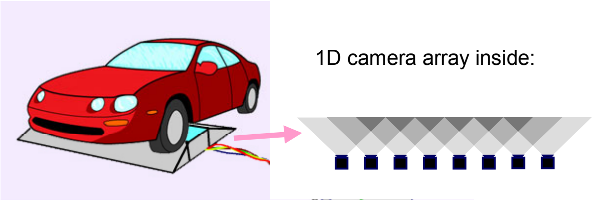


Figure 8. 1D camera array for under-vehicle inspection [13, 14]

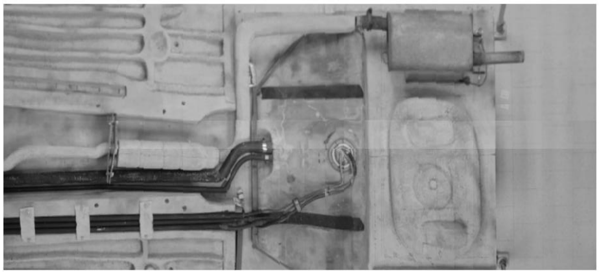


Figure 9. Full parallel projection mosaics with a bed of 2D array of cameras

As one of the real applications of full parallel stereo mosaics, we have generated an approximate version of mosaics with full parallel projection from a virtual bed of 2D camera arrays by driving a car over a 1D array of cameras in an under-vehicle inspection system

(UVIS) [13, 14, 18]. The Under Vehicle Inspection System (UVIS) is a system designed for checkpoints such as those at borders, embassy's, large sporting events, etc. It is an example of generating mosaics in very short-range video so a 2D array of camera is necessary for full coverage of the objects (under-vehicles). Fig. 8 illustrates the system setup where a array of cameras is housed in a platform. When a car drives over the platform, several mosaics with different slanting angles of the underside of a car are created. The mosaics can then be viewed by an inspector to thoroughly examine the underside of the vehicle from different angles. Figure 6 shows such a mosaic generated from four overlapping image sequences taken by four moving cameras side by side – which simulates the motion of the vehicle. A PPT demo of five oblique parallel views of the mosaics can be found at [18] where different “occluded” regions under a pipe can be observed by changing to different mosaics.

In the case of 1D camera array, we can pre-calibrated the fixed cameras, and correct the geometric and photometric distortions of those wide FOV cameras. However challenges remain since (1) the distance between cameras are large compared to the very short viewing distances to the bottom of the car; and (2) without the assistance of GPS/INS for pose estimation, we need to determine the car’s motion by other means, e.g. tracking line features on the car. The proposed ray interpolation approach needs to take these two factors into consideration.

7. Conclusions

This paper presents an approach of fusing images from many video cameras or a moving video camera with external orientation data into a few mosaiced images with slanting parallel projections. In both cases, a virtual 2D array of cameras with FOV overlaps is formed to generate the whole coverage of a scene (or an object). The proposed representation provides wide FOV, preserves 3D information, and represents occlusions. This representation can be used as both an advanced video interface for surveillance or a pre-processing for 3D reconstruction. We have shown real examples of stereo mosaics for two important applications - aerial video surveillance and under-vehicle inspection.

8. Acknowledgements

This work is partially supported by AFRL through a grant on multimedia display in surveillance (Award No. F33615-03-1-63-83). The experimental examples are mainly based on some previous work in the Computer Vision Lab at the University of Massachusetts Amherst, under a NSF environmental monitoring using aerial videography (Grant Number EIA- 9726401), and an industrial sub-contract Under Vehicle Inspection System (UVIS). The supports and collaborations of Prof. Edward Riseman and Prof. Allen Hanson and their UMass research team are greatly appreciated. The author also thanks Prof. George Wolberg and Prof. Izidor Gertner at the City College of New York for their discussions with applications in multiple ONR research projects for image registration, target detection and human-computer interaction.

9. References

[1] J. Y. Zheng and S. Tsuji, Panoramic representation for route recognition by a mobile robot, IJCV, 9(1), 1992: 55-76

[2] S. Peleg, B. Rousso, A. Rav-Akha and A. Zomet, Mosaicing on adaptive manifolds, *PAMI*, 22(10), October 2000: 1144-1154.

[3] USGS Digital elevation Models (DEM), http://data.geocomm.com/dem/

[4] S. Peleg, M. Ben-Ezra and Y. Pritch, OmniStereo: panoramic stereo imaging, *PAMI*, March 2001:279-290.

[5] H.-Y. Shum and R. Szeliski, Stereo reconstruction from multiperspective panoramas, *ICCV’99*: 14-21.

[6] S. Baker, R. Szeliski and P. Anandan, A layered approach to stereo reconstruction. *CVPR'98:* 434-441

[7] J. Shade, S. Gortler, L. He. and R. Szeliski, Layered depth image. *SIGGRAPH'98:* 231-242

[8] Z. Zhu and A. R. Hanson, 3D LAMP: a New Layered Panoramic Representation, *ICCV’01:* II 723-730.

[9] A. Zomet, D. Feldman, S. Peleg and D. Weinshall, Mosaicing new views: the crossed-slits projection, *PAMI* 25(6), June 2003.

[10] J. Chai and H. -Y. Shum, Parallel projections for stereo reconstruction, CVPR'00: II 493-500.

[11]. Z. Zhu, E. Riseman, A. Hanson, Parallel-perspective stereo mosaics, *ICCV’01*, vol I: 345-352.

[12] Z. Zhu, A. Hanson, H. Schultz and E. M. Riseman, Generation and error characteristics of parallel-perspective stereo mosaics from real video. In *Video Registration*, M. Shah and R. Kumar (Eds.), Kluwer, 2003: 72-105.

[13]. P. Dickson, J. Li, Z. Zhu, A. Hanson, E. Riseman, H. Sabrin, H. Schultz and G. Whitten, Mosaic generation for under-vehicle inspection. *WACV’02*:251-256

[14] http://vis-www.cs.umass.edu/projects/uvis/index.html

[15] http://www-cs.engr.ccny.cuny.edu/~zhu/StereoMosaic.html

[16]H. Schultz, Terrain reconstruction from widely separated images, SPIE 2486, April 1995: 113-123.

[17] http://www-cs.engr.ccny.cuny.edu/~zhu/CampusVirtualFly.avi

[18] http://www-cs.engr.ccny.cuny.edu/~zhu/mosaic4uvis.html