Mosaic-based 3D scene representation and rendering of circular aerial video

Edgardo Molina[a,b], Zhigang Zhu[a,b], Olga Mendoza-Schrock[c]
[a]The City College of New York, 138th Street and Convent Ave, New York, NY, USA 10031;
[b]The CUNY Graduate Center, 365 Fifth Avenue, New York, NY, USA 10016;
[c]Air Force Research Laboratory, WPAFB, Dayton, Ohio, USA

ABSTRACT

Circular aerial video provides a persistent view over a scene and generates a large amount of imagery, much of which is redundant. The interesting features of the scene are the 3D structural data, moving objects, and scenery changes. Mosaic-based scene representations work well in detecting and modeling these features while greatly reducing the amount of storage required to store a scene. In the past, mosaic-based methods have worked well for video sequences with straight camera paths in a dominant motion direction[11]. Here we expand on this method to handle circular camera motion. By using a polar transformation about the center of the scene, we are able to transform circular motion into an approximate linear motion. This allows us to employ proven 3D reconstruction and moving object detection methods that we have previously developed. Once features are found, they only need to be transformed back to the Cartesian space from the polar coordinate system.

1. INTRODUCTION

Aerial video over a large-scale scene produces a large amount of imagery. Much of that imagery may be of static and unchanging structures and background scenery. In surveillance applications a scene will be imaged continually with the purpose of detecting changes and moving objects. For aerial video taken at high-altitudes, each frame will represent a largely unchanging scene with changes and moving objects confined to small windows on the frames.

Mosaicing is a common technique used to represent a complete picture of a scene by registering all of the imagery and combining and blending it into a seamless mosaic. A single mosaic can only provide a snapshot view of the scene, losing 3D depth information and moving object information. Previously, work has been done to employ a mosaic-based approach that preserves depth and moving object information. In the past, the method from Zhu, et al.[10] has worked well for aerial video on a camera path with a dominant motion direction. This method generates a set of aligned multi-view "pushbroom" stereo mosaics of an aerial video taken with a single camera (or any standard 2D frame sensor with a projective lens). These mosaics are then used for 3D depth reconstruction and moving object detection.

Here we present an extension to the multi-view mosaicing technique to handle aerial video taken on a circular flight path. We show that the resultant circular stereo mosaics produce the same scene representation benefits (preserve depth and object motion information) that were shown for aerial video on a close to linear camera path. As an additional benefit, multi-view stereo mosaics allow direct 3D viewing and rendering of the imagery before the 3D depth reconstruction is undertaken.

Section 2 of this paper will present a brief overview of related works. Section 3 describes the ideal geometry for circular stereo mosaics. In section 4 we present how to get stereo correspondences and the geometry for 3D reconstruction from circular stereo mosaics. Section 5 describes the method for generating circular stereo mosaics. In section 6 we show results for an ideal simulated scene as well as a real scene, with concluding remarks in section 7.

2. RELATED WORK

Mosaics have become common for representing a set of images gathered by one or more (moving) cameras. Creating multi-perspective stereo panoramas with circular projections from two rotating cameras has been proposed by Huang &

Hung[4] (1998), from one rotating camera off its nodal point by Peleg, et al[7] (2001) and Shum & Szeliski[9] (1999), and from a single translating camera by Zhu, et al[10] (2001), and Chai & Shum[2] (2000). Recent advanced topics in video mosaicing include: geo-referenced mosaicing[5], alignment of two video streams without overlapping but with a common viewpoint[3], acquisition of super-resolution[12] and super-dynamic-range[1,8] in 2D panoramic mosaicing. However, existing techniques for 2D image/video registration using global transformations such as rigid, affine or projective transformations fail to work for 3D aerial videography under parallax.

Scene modeling and representations of large-scale 3D urban scenes, both dynamic and static, are critical in many military and civilian applications, such as mapping, surveillance, transportation, development planning, archeology and architecture. Research in this area has attracted increasing attention from both academia and industry. Quite a few research groups have major research projects in large-scale 3D scene modeling.

3. GEOMETRIC MODEL

First we present the ideal geometry to generate circular mosaics from a pushbroom sensor on a circular camera path. In section 4 we show how to build circular stereo mosaics and reconstruct depth.

A 1D camera with a single column off the center of an angle β, moves along a circular path with a center C and a radius R. The camera's optical axis Z is perpendicular to the circular path. A circular "panoramic" image is generated by this scanning camera (Fig. 1). The scanlines of the circular images are circles. Such an image is represented by I(α,r), where α is the angle of the pixel along the circle measured from a starting point, and r is the distance along the column direction.

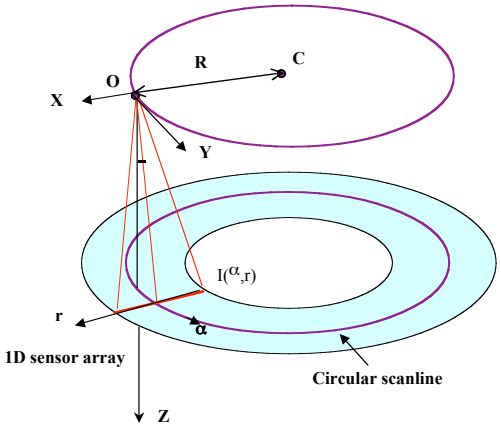


Figure 1. Circular imaging geometry

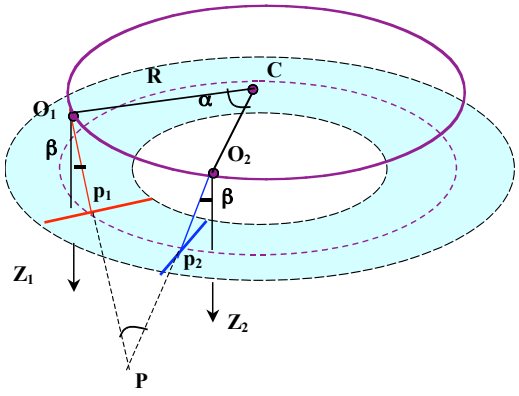


Figure 2. Circular stereo geometry

4. STEREO CORRESPONDENCE AND 3D RECONSTRUCTION

Given two such circular scanning cameras moving on the same circular path with a center C and a radius R, both with viewing angle β, one looking forward (O1) and the other looking backward (O2), a pair of circular stereo panoramas can be generated (Fig. 2). For any 3D point P, its correspondences, p1(α1,r1) and p2(α2,r2), in the two panoramas are (approximately) along a circular scanline. Therefore, we have r1 = r2, and the angular disparity

$$\alpha=\alpha 2 - \alpha 1. \quad (1)$$

The baseline B between the two views O1 and O2 can be calculated as

$$B = 2R \sin(\alpha/2) = R\alpha, \quad (2)$$

where α is measured in radians, and the radius is much larger than the arc length B. Then the distance from each view (O1 or O2) to the 3D point can be calculated as

$$D = R \sin(\alpha/2) / \sin\beta = R\alpha /(2\beta) \quad (3)$$

where β is also measured in radians. Hence the Z coordinate of the point P can be computed as

$$Z = D \cos\beta = R \sin(\alpha/2)/\tan\beta = 2R\alpha /\tan\beta \quad (4)$$

We want to note here that pushbroom stereo mosaics under a circular motion path are different from multi-perspective stereo panoramas with circular projections[3,4]. In stereo panoramas with circular projections, the optical axis of the camera points to the center of the circular motion, while in pushbroom stereo mosaics, the optical axis of the camera is perpendicular to the circular motion path. In fact, in all the cases where the optical axis is not pointing to the center of the circular path, pushbroom stereo mosaics can be generated by applying image rectification before mosaicing.

Second, in circular pushbroom stereo mosaics, the depth error is independent of the depth in theory (Eq. 4), which is the same as linear pushbroom stereo mosaics. Therefore, the two types of pushbroom stereo mosaics can be combined into one model for a more general motion, in that the motion is characterized as piecewise linear and circular. Then if a camera moves on a more general path, a generalized pushbroom mosaic can be built along that path, in which the projection is perspective perpendicular to the direction of the motion. If the rates of changes of motion directions are slow, we can fit the motion parameters onto a smooth path that is piecewise linear and circular (with large radii), so that locally the epipolar geometry is still along "scanlines".

5. CIRCULAR STEREO MOSAICING

In the following, we discuss how we can generate such circular stereo mosaics from an aerial video sequence when a camera undergoes a circular path. In theory, a pair of stereo mosaics can be created by extracting two columns, with the same viewing angle β, one looking forward and the other looking backward. However, there are several practical issues to consider. We assume the camera moves on a perfect circular path, or that the circle is large enough to compensate the small drifts of the camera from the viewing circle by 2D image translations. The camera may not have a nadir-viewing angle, so an orthorectification step is carried out before mosaicing. Finally, the camera views may not be dense enough to generate a dense and seamless circular mosaic if only one column is extracted.

We will describe the steps to create one such mosaic, with a viewing angle β. A pair of stereo mosaics can be generated in the same way.

Step 1. Camera orientation estimation.
First we need to estimate the camera positions and orientations for each frame relative to a frame of reference (the first image can serve as a frame of reference when no geo-location information exists). This can be done by using either INS/GPS or bundle adjustment techniques. We realize this is an important step, but here we will mainly focus on the geometric models and the stereo mosaicing algorithms. There is a large body of literature discussing this issue.

Step 2. Camera path generation and image rectification.
By fitting the viewpoints of the moving camera, a circular path (with center C and radius R) can be found. Then the images in the sequence are transformed (by rotation and translations) so that each rectified image has a viewpoint on the circular path, and has an optical axis perpendicular to the circular path. Then the rectified sequence is ready for use in creating mosaics.

Step 3. Circular image transformation
By using the information of the circular path, each image should be transformed from its rectilinear representation to a circular representation (Fig. 3). This step can be combined with Step 2.

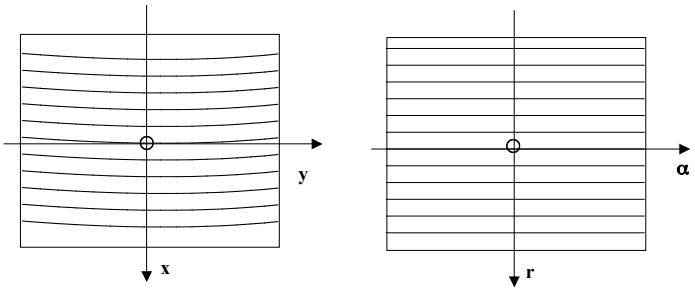


Figure 3. From rectilinear to circular images

Step 4. Circular panorama generation.
Given two views, O1 and O2, two rays (with circular angles α1 and α2 determined by the two view locations) can be directly cut from column b in each image, and pasted into the mosaics. The circular angle between them is α. Between these two columns, finding the correspondences β1 and β2 of a 3D point P, in the two images respectively, such that the distance of the point P can be calculated as

$$Z = R\,\alpha / (\beta 1 - \beta 2) \tag{5}$$

Then the circular angle α_i of the reprojected ray to the viewing direction of β is

$$\alpha_i = Z\,(\beta 1 - \beta) = R\,\alpha\,(\beta 1 - \beta) / (\beta 1 - \beta 2) \tag{6}$$

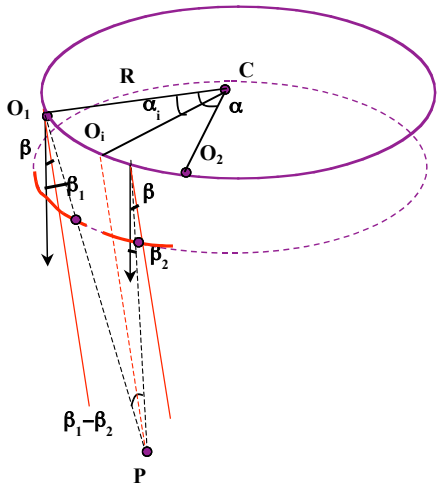


Figure 4. Circular mosaic generation

Note that Eq. (6) indicated that this procedure is strikingly similar to the parallel ray interpolation for stereo mosaicing (PRISM) we have proposed[11], therefore the same PRISM algorithm with slight modification can be directly applied for both linear pushbroom mosaicing and circular pushbroom mosaic, and then to the generalized pushbroom mosaicing. A dense 2D mosaic can be generated by image area triangulation and interpolation[11]. Furthermore, linear pushbroom mosaics are a special case of circular pushbroom mosaics when the radius of the circular path is infinite.

6. RESULTS

Two data sets have been used to evaluate our circular mosaic algorithm. The first set is a simple simulation with perfect circular camera motion. The camera has a nadir view of the scene at an altitude of 600 meters moving in a circle with radius 156.4 meters. The camera's FOV is 15.189 degrees. The second set of data was taken from the AFRL publicly released CLIF 2007 dataset[6]. Sequences of 360 degree circular paths were extracted from camera 0; in particular they are frames '000000-100305' to '000000-100535', a total of 231 frames.

Mosaics were generated in two ways. The first was using the procedure outlined in section 5. After the camera motion parameters were estimated, a center point was calculated so that individual frames could be warped into circular images and these were used to build the circular stereo mosaics. This method works better if the path of a camera is approximately a circle. In section 6.1 we present results for the simulation and the CLIF 2007 data. The second method also followed the same procedure, but in this case, a circular warping was not applied to the frames. Instead a local rectification between frames was used and mosaics were built based on the distance between consecutive frames. This method works even when the path of the camera is not a circle. Mosaics thus created will follow a smoothed version of the camera path. Section 6.2 shows those results.

6.1 Experiment 1 – with circular warping

6.1.1 Simulation Scene

In this simulation the center of the camera path is known and is the same for all frames. We go directly to Step 3 (circular image transformation) in our algorithm. Fig. 5 shows the original camera's image and it transformed into a circular projection image using a Cartesian to Polar coordinate transformation where the center is at (320,-386) in pixels. (Note: -386 in the Y-axis refers to center being 386 pixels above the image.)

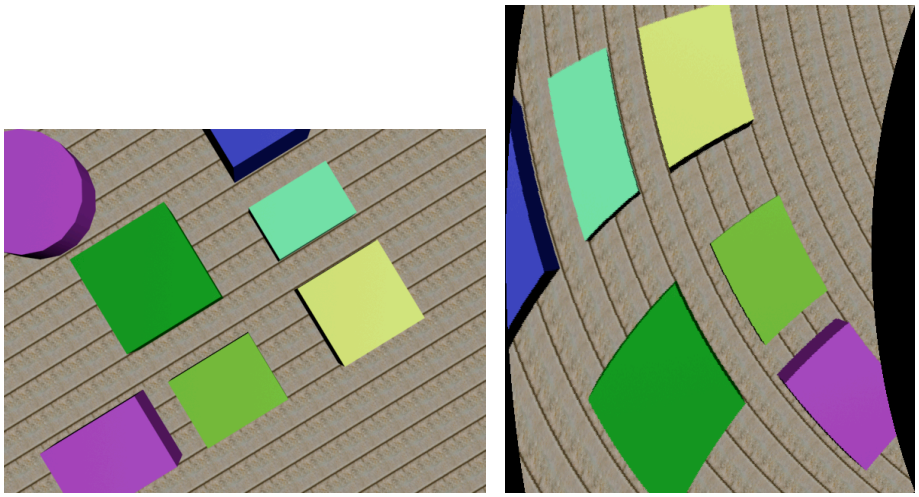
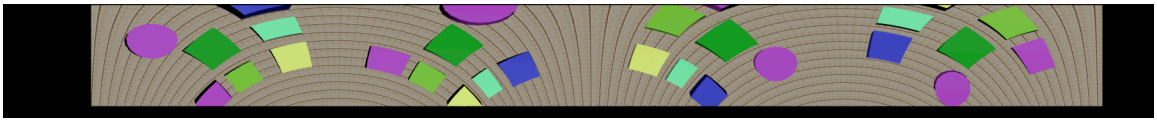


Figure 5. On the left is the original frame in Cartesian X,Y coordinates;
On the right is the transformed image in Polar R,θ coordinates.

In the original sequence the camera motion was circular and counter-clockwise in Cartesian coordinates. After step 3 the sequences radii of the images are registered and the motion is only in θ; in the transformed sequence this is a linear motion. We can do step 4 (circular panorama generation) on the transformed sequence using our pushbroom mosaicing algorithm. In addition this allows us to generate multiview pushbroom mosaics; Fig. 6 shows a left and right view of the simulated scene, which can be used to generate stereoscopic views of the 3D scene.



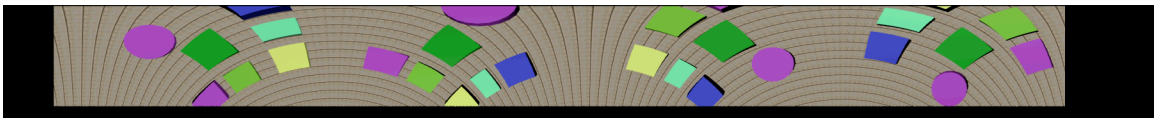


Figure 6. The top/bottom views show left/right mosaics of the scene. The Y-axis represents radius, and the X-axis represents degrees. Both mosaics show 360 degrees of coverage.

Fig. 7 shows one of the left view mosaics transformed back to Cartesian coordinates, which shows an undistorted global view of the original scene.

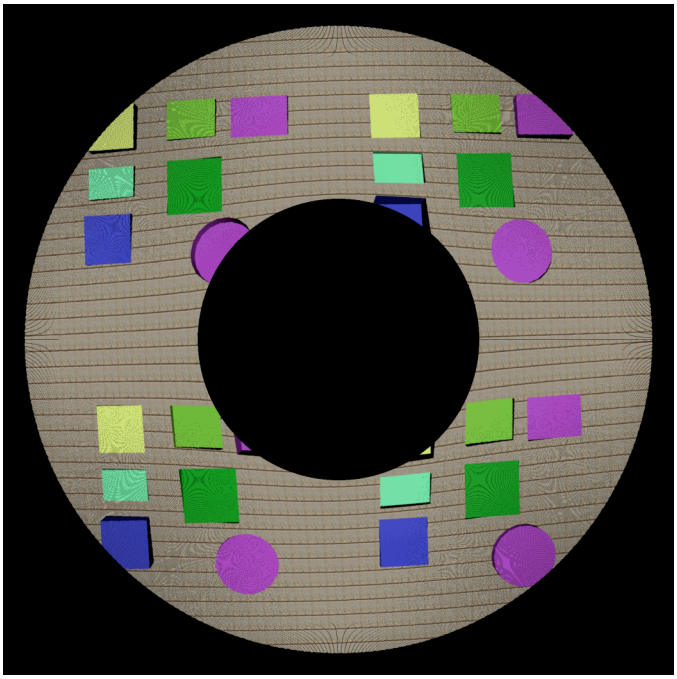


Figure 7. Left stereo mosaic warped back from circular coordinates to Cartesian coordinates.

6.1.2 CLIF 2007 Scene

These results are for a circular camera path sequence in camera 0 of the CLIF 2007 dataset. These results were attained without any image rectification prior to step 3. Only an image-based registration is applied for step 4 on the transformed imagery. A camera path center was calculated from the camera path that was estimated in step 1 of the procedure. Fig. 8 shows one of the mosaics that are generated for multiple-view 3D and motion viewing. Fig. 9 shows a close-up of a small portion of the mosaic.

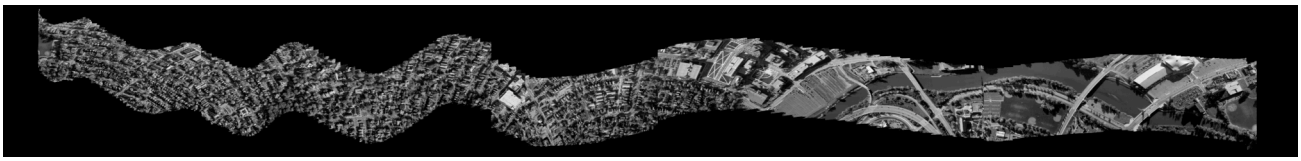


Figure 8. A full 360 degrees circular mosaic.

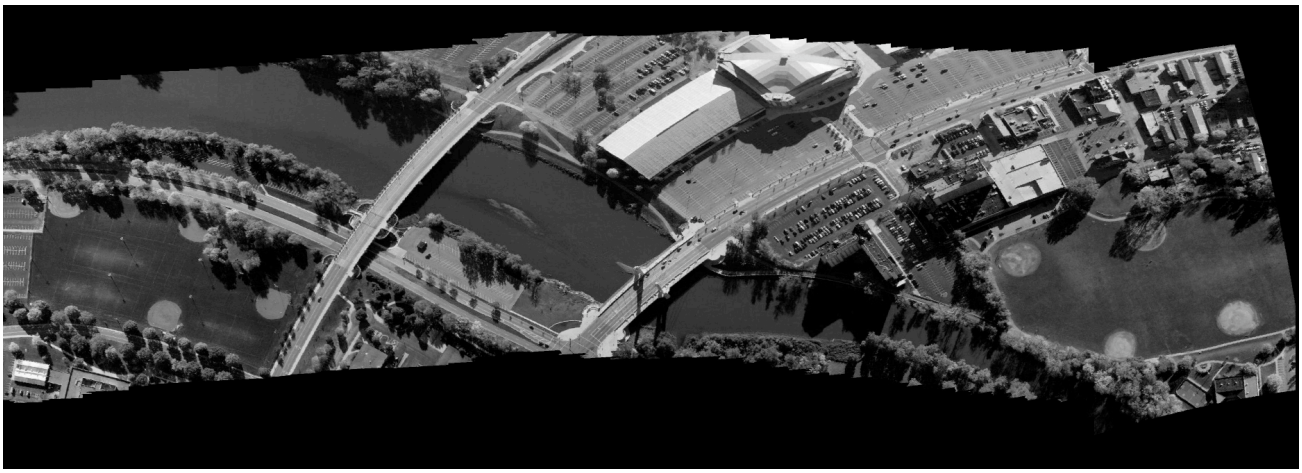


Figure 9. A cropped portion of the right end of the scene in Fig 8.

6.2 Experiment 2 – with local rectification

In this experiment we do not apply a circular warping to the frames. Instead we do a local rectification among frames. The local rectification transformation for each image is estimated as the difference of the smooth version of the estimated original camera path and the original camera path. The mosaics are generated on the smoothed camera path, taking into consideration the distance and rotation from consecutive frames. This is the first step we take at trying to generalize the multi-view mosaicing method to handle more general camera motion. Below are the results for comparison to Experiment 1. The most noticeable effect, interframe edge discontinuity, is seen in the mosaic close up of Figure 11 compared to Figure 9. Future work is to improve the local recification and the alignment for generating seamless mosaics.

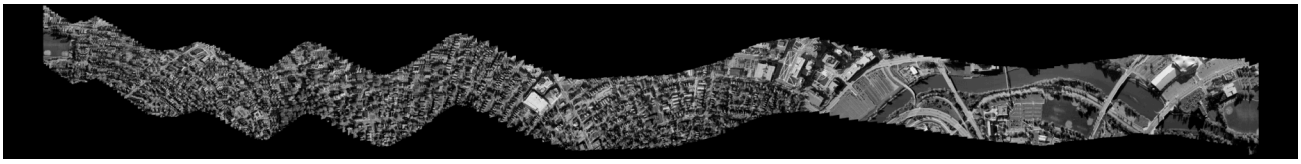


Figure 10. Full circular mosaic without circular warping

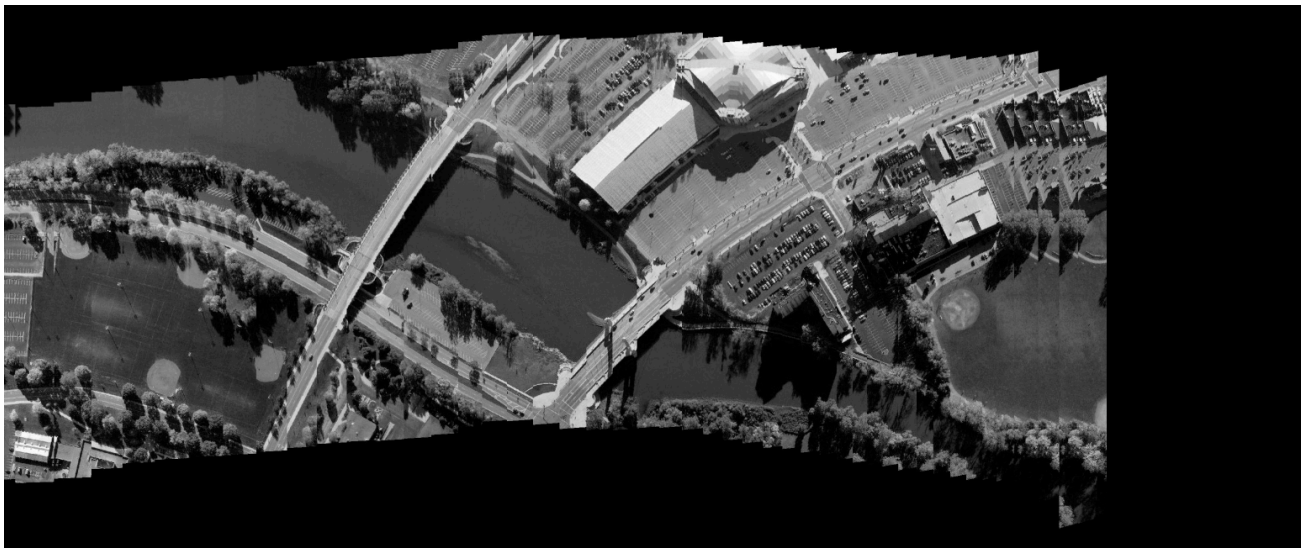


Figure 11. Shows a cropped portion of the right end of the scene in Fig 10.

7. CONCLUSION

In this paper we have shown how to extend the linear pushbroom stereo mosaic approach to work on aerial video taken with a camera on a circular path. The resultant circular stereo mosaics, like the linear pushbroom mosaics, provide us with a compact scene representation that preserves 3D structure information, as well as moving object information. The circular stereo mosaics can be used directly to view the scene in 3D using a stereoscopic display, and they can also be used to reconstruct 3D structures of the scene. We have also begun to set the framework to develop a geometric model for a mosaic-based representation that can handle more general camera motion.

ACKNOWLEDGEMENTS

This work is supported by the Air Force Research Laboratory (AFRL/RYAT) through Grant Award No. FA8650-05-1-1853. The first and second authors are also partially supported by NSF under Grant No. CNS-0551598, by AFOSR (Award #FA9550-08-1-0199), and by ARO (Grant No W911NF-05-1-0011). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. However, the views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government.

REFERENCES

[1] Aggarwal, M. and N. Ahuja, 2001. High dynamic range panoramic imaging. *ICCV'01*, 82-89.
[2] Chai, J. and H –Y. Shum, 2000. Parallel projections for stereo reconstruction. *CVPR'00*: II 493-500.
[3] Caspi, Y. and M. Irani, 2001. Alignment of non-overlapping sequences. *ICCV'01*.
[4] Huang H-C and Hung Y-P., 1998. Panoramic stereo imaging system with automatic disparity warping and seaming. *Graphical Models and Image Process*, 60(3): 196-208.
[5] Kumar, R., et al, 1998. Registration of video to geo-referenced imagery, *ICPR'98*, vol. 2: 1393-1400
[6] Mendoza-Schrock, O. L, Patrick, J. A., Garing, M. "Exploring image registration techniques for layered sensing" Evolutionary and Bio-Inspired Computation: Theory and Applications III. Edited by Teresa H. O'Donnell, Misty Blowers, and Kevin L. Priddy. Proceedings of the SPIE, May, (2009)
[7] Peleg S, M Ben-Ezra and Y Pritch, 2001. Omnistereo: panoramic stereo imaging, *IEEE Trans. PAMI*, 23(3): 279-290.
[8] Schechner, Y. and S. Nayar, 2001. Generalized mosaicing, *ICCV'01*.
[9] Shum, H.-Y., R. Szeliski, 1999. Stereo reconstruction from multiperspective panoramas. *ICCV'99*: 14-21.
[10] Zhu, Z., E M Riseman, A R Hanson, 2001. Parallel-perspective stereo mosaics. *ICCV'01*, vol I: 345-352.
[11] Zhu, Z., E. M. Riseman, A. R. Hanson, 2004. Generalized parallel-perspective stereo mosaics from airborne videos, *IEEE Trans. PAMI*, vol. 26, no. 2, Feb 2004, pp 226-237.
[12] Zomet, A., and S. Peleg, 2001. Super-resolution from multiple images having arbitrary mutual motion, in S. Chaudhuri (ed.), *Super-Resolution Imaging*, Kluwer Academic, September.

(This document is approved for public release via 88 ABW-10-1314.)