STEREO MOSAICS FROM A MOVING VIDEO CAMERA FOR ENVIRONMENTAL MONITORING

Zhigang Zhu, Allen R. Hanson, Howard Schultz, Frank Stolle, Edward M. Riseman

Computer Vision Lab, Department of Computer Science
University of Massachusetts at Amherst, MA 01003, U.S.A.
{zhu, hanson, hschultz, stolle, riseman }@cs.umass.edu

Abstract

Environmental monitoring using automated analysis of high-resolution aerial video is an application of growing importance with its own set of technical challenges. A mosaic is a commonly used tool for representing the enormous amount of data generated from aerial video sequences in an easily viewable form. In contrast to the usual applications of mosaics, the environmental monitoring domain requires both geo-corrected mosaics tied to real-world coordinates and three-dimensional information. The standard techniques of generating seamless mosaics using only image data in a frame-by-frame image registration process cannot satisfy this requirement because registration errors accumulate over extended periods of time, and the mosaicing results are typically only 2D images that lose 3D information. On the other hand, 3D reconstruction of the terrain on a frame-by-frame base has proven to be both difficult and time-consuming. In this paper we will present a method for automatically and efficiently generating stereoscopic mosaics by seamless registration of optical data collected by a video camera mounted on an airborne platform. The resultant mosaics are globally correct with respect to the ground and exhibit correct 3D views when viewed stereoscopically.

1. INTRODUCTION

A critical issue among nations in the coming decades will be how to manage the use of land and natural resources. Our interdisciplinary NSF environmental monitoring project, being conducted jointly by researchers from the Department of Computer Science and the Department of Natural Resources Conservation at the University of Massachusetts at Amherst, aims at developing a methodology for estimating the standing biomass of forests. The instrumentation package mounted on an airplane consists of two video cameras (with different focal lengths), a GPS system, an INS system, and a profiling pulse laser. This paper will focus on the development of automated tools that can create video mosaics from high resolution low-altitude video sequences that can be registered with lower resolution high-altitude aerial image data (or satellite image data) as a tool for interpreting the lower resolution data. The previous manual approach used by our forestry experts utilized only a fraction of the available data due to the labor involved in hand interpreting the large amount of video data. For example, a recent project in Bolivia involved over 600 sites and more than 20 hours of video, which is prohibitive if the video is interpreted manually.

Automatically generated mosaics of forests that both are geo-referenced and support stereo viewing are of crucial importance when huge amounts of video data of the forest must be processed.

In fact, for many applications, extending the field of view (FOV) of a 2D image and introducing the third dimension (depth) is useful. Image-based rendering of an urban or natural scene and large-scale environmental monitoring are just two examples of the applications that would benefit from an extended image-based representation. Theoretically, reconstructing 3D from pairs of images, and then integrating multiple 3D maps (as well as the texture maps) into a large texture-mapped digital elevation map (DEM) would be an ideal solution. However, there are many technical issues that need to be solved, such as camera calibration, dense image correspondence, prohibitive computation, and 3D recovery consistency. The visual representation of large-scale 3D scenes is still an open problem. Recently, the construction of panoramic images and high quality mosaic images from video sequences has attracted significant attention. Many of the current successful image mosaic algorithms, however, only generate 2D mosaics from a camera rotating around its nodal point [1-5]. Creating stereo panoramas from two rotating cameras was proposed by Huang & Hung [6], and from one rotating camera by Ishiguro, Yamamoto & Tsuji [7], Peleg & Ben-Ezra [8], and Shum & Szeliski [9]. In these kinds of stereo mosaics, the viewpoint is limited within a very small area, usually along a circle of meters' diameter. A system for creating a global view for visual navigation by pasting together columns taken by a smoothly translating camera comprising only a vertical slit was proposed by Zheng & Tsuji [10]. The moving slit paradigm was used as the basis for image mosaicing [11], image-based rendering [12] and 3D reconstruction [13]. Kumar, et al [15] presented a geo-registration method that can register video mosaics to a high-attitude reference image using the geo-data. The fine geo-registration requires knowledge of a reference image (geo-referenced aerial image with broader coverage) and accompanying co-registered DEM (digital elevation map).

Our interest in the mosaic representation of a 3D scene is partially inspired by the techniques used in classical Chinese paintings[1], and by Peleg & Ben-Ezra's work [8]. In their paper, the authors mentioned that their technology for a rotating camera could also be used to create stereo panoramic images from a translating camera. However, as far as we know, there is little serious work on stereo mosaics from a translating camera, which is typically the situation of the prevalent airplane motion when performing forest surveys. Image mosaicing from a translating camera will result in an image with multiple viewpoints on a long flying path, which raises a set of different problems from that of a rotating camera, including motion model, geo-referenced image mosaic in the case of large-range translation, and epipolar geometry. Our aim is to generate seamless, continuous and stereoscopic image mosaics that are accurately tied to ground coordinate (i.e. geo-referenced), before we have a 3D DEM.

---

[1] It is commonplace to remark that the classical Chinese paintings are without a correct vanishing point and possess no exact laws of foreshortening figures [14]. But this characteristics is nothing more than a convention by which artists endeavor to represent 3D objects. In a typical Chinese painting, "the scenery changes as one walks". A concrete reflection of this multi-focal perspective (no fixed vantage point) and connotative space can be seen in the structure of horizontal landscape scrolls and their seemingly endless extension .

2. THE GEOMETRIC MODEL OF STEREOSCOPIC MOSAICS

The multi-perspective stereo model will be introduced in a ideal case, i.e., 1D translational motion. However, the main concepts and features of multi-perspective stereo mosaics are included. If the motion of the camera is a 1D translation of constant speed, 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 near the two borders of the image that are perpendicular to the motion (Fig.1a). These mosaic images are similar to multi-perspective images captured by a linear pushbroom camera [16]. In contrast to the common pushbroom aerial image, these mosaics are obtained from two different oblique angles of a single camera, one looking forward and the other looking backward, so that a stereo effect can be generated. In each multi-perspective mosaic, the sub-image (a scanline image in concept) obtained from a slit is full perspective, but successive slit images have a different viewpoint.

Let us first consider the stereo geometry of the two-slit windows in a perspective image (Fig.1). Without loss of generality, we assume that two vertical slit windows have $d_y/2$ offsets to the left and right of the center of a image respectively. Suppose that in viewpoint O_l, a curve C_l in the 3D scene can be seen through the right-side slit window, and p_l is the image of a point $P\ (X,Y,Z)$ on the curve. Now when the camera moves a certain distance B_y in the Y direction to viewpoint O_r, point P can be seen from the left-side slit window as image p_r, and on a 3D curve C_r. The depth of the point P can be calculated as

$$Z = D = F\frac{B_y}{d_y} \tag{1}$$

where F is the focal length of the camera. Now we will connect the two-slit-window geometry to the stereo mosaic images. By definition, both of the digital stereo mosaics are represented as an xy image whose two dimensions are in pixels. It is convenient that both F and d_y are also measured in pixels, and the origins of both the left and right mosaics are the same as the origin of a common reference frame (Fig. 1b). Hence the parallel-perspective projection model of the stereo mosaics is represented by the following equations

$$\begin{aligned} x_l &= x_r = F\,X/Z \\ y_r &= FY/H + (Z/H\text{-}1)\,d_y/2 \\ y_l &= FY/H - (Z/H\text{-}1)\,d_y/2 \end{aligned} \tag{2}$$

where the last two equations are derived from

$$\begin{aligned} y_r &= FT_{yr}/H - d_y/2, \quad -d_y/2 = F(Y\text{-}T_{yr})/Z \\ y_l &= FT_{yl}/H + d_y/2, \quad d_y/2 = F(Y\text{-}T_{yl})/Z \end{aligned} \tag{3}$$

In the above equations, (X,Y,Z) is the 3D coordinates of the space point P in the reference camera coordinate system, H is the average height of the terrain, and T_{yl} and T_{yr} are the Y translational components of the left and right viewpoints of the moving camera when point P can be seen through the right and left slits (so $B_y = T_{yr} - T_{yl}$). Note that Eq.(2) gives the relation between a pair of 2D points, $p_l(x_l,y_l)$ and $p_r(x_r,y_r)$, and the corresponding 3D point $P(X,Y,Z)$. In practice, image frames are captured only in discrete viewpoints (denoted by translational component T_y). In the ideal case, $T_y = V\,k$, where V (meters/frame) is the speed of the camera, and k is the frame number (k=0,1,…). Eq. (3) tells us how to generate dense left or right mosaic in the ideal case: we should take a slice of FV/H pixel wide per frame for each of the mosaics, and put them in the corresponding mosaics centered at y_l and y_r,

which are only functions of the camera displacements as well as the fixed offset of the two slits. The purpose of the scaling (*F/H*) in these equations is to balance the aspect ratio of each mosaic in *x* and *y* dimensions, so the value of H is not so critical. In fact, the mosaic can be constructed based on any plane a distance H from the camera.

By finding correspondence pair of the points, (x_l,y_l) and (x_r,y_r), in the stereo mosaics, we can compute the scaled "baseline" as $b_y = d_y+\Delta y$ where $\Delta y = y_r - y_l$ is the y displacement, and the real "baseline" as $B_y = b_y H/F$. In the ideal case, a epipolar line is the scanline in y direction, i.e. $x_r = x_l$. Substituting $B_y = b_y H/F$ into Eq.(1), we have

$$Z = H\frac{b_y}{d_y} = H(1+\frac{\Delta y}{d_y}) \qquad (4)$$

It is interesting to note that, since the selection of the two mosaic coordinate systems brings a constant shift d_y to the scaled "baseline", it produces the fixation of the stereo mosaics to a horizontal "*fixation plane*" of average height *H*, which is desirable in the stereo matching as well as stereoscopic viewing.

3. IMAGE CAPTURE AND MOSAIC GENERATION

In real applications, such as environmental monitoring from an aerial camera, it is unlikely to have an ideal 1D translational motion, with image frames being evenly and densely captured. Can we generate stereo mosaics in a more general and practical motion model? The solution to this quite challenging problem is enabled by a sophisticated aerial instrumentation package that augments the video data with 3D motion of the camera and range profile of the scene, and robust image mosaicing techniques that fill the gaps between slits from two successive frames and stitch them together. The geographical data ("geo-data") from our aerial instrumentation package includes 3D locations from the GPS, 3D rotations from the INS, and the range data along the flight path from a laser profiler.

3.1. Image Rectification

The camera motion is modeled as a matrix **R**, and a translation vector $\boldsymbol{T} = (Tx,Ty,Tz)^t$ which has a known dominant motion direction. A 3D point $\mathbf{X_k} = (X_k,Y_k,Z_k)^T$ with image coordinates $\mathbf{u} = (u_k, v_k, 1)^t$ at current frame *k* can be related to its reference coordinates $\mathbf{X} = (X, Y, Z)^T$ by

$$\mathbf{X} = \mathbf{R}\mathbf{X}_k + \mathbf{T} \qquad (5)$$

Without loss of generality, assuming that airplane's motion is mostly along the *y* axis, then we will have $T_x << T_y$, $T_z << T_y$ (if not, we can achieve this by applying an image rotation transformation), and $T_z << H$. Hence Eq. (5) can be approximated quite well by

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathbf{Q}\begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ 0 \end{pmatrix}, \quad \mathbf{Q} = \mathbf{R} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & T_z/H \end{pmatrix} \qquad (6)$$

where the average height *H* can be estimated from the range information accompanying the image sequence. Note that the only approximation in Eq. (6) is replacing T_zZ/H with T_z. Before image mosaicing, a projective transformation **A** is applied to each frame of the video using the motion parameters from geo-data. The resulting video sequence will be a "re-projected" image sequence as if it is captured by a camera undergoing 2D translations (T_x,T_y), with the T_y component dominant. The projective transformation is expressed as

$$\mathbf{u}_k^{\mathbf{p}} \cong \mathbf{A}\mathbf{u}_k \quad \mathbf{A} = \mathbf{F}\mathbf{Q}\mathbf{F}^{-1}, \quad \mathbf{F} = \begin{pmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

where $\mathbf{u}_k^{\mathbf{p}}$ is the reprojection image point of frame k.

3.2. Image Mosaicing

After image rectification, two x-slices (columns in direction x) centered at the two slit windows are taken from each frame for mosaicing. As an example, we give the equations of the left mosaic. The center of the left-view slice in frame *k* is always at $(x,y) = (0, d_y/2)$, and the widths of the slices from the center to the left and right are

$$w_1^{(k)} = F\frac{|T_y^{(k)} - T_y^{(k-1)}|}{2H}, \quad w_2^{(k)} = F\frac{|T_y^{(k+1)} - T_y^{(k)}|}{2H} \tag{8}$$

where $T_y^{(k-1)}$, $T_y^{(k)}$ and $T_y^{(k+1)}$ are the *Y* translational components of frame *k-1*, *k* and *k+1* respectively. This slice will be put in the left-view mosaic centered at

$$(t_x^{(k)},\ t_y^{(k)} + \frac{d_y}{2}) = (F\frac{T_x^{(k)}}{H},\ \ F\frac{T_y^{(k)}}{H} + \frac{d_y}{2}) \tag{9}$$

The central column of the (w_1+w_2)-wide slice is called the *fixed line*, which is fixed by the camera's motion, and the two borders of the slice are called *stitching lines*, where the current slice is stitched to the previous and the next slices.

Assuming that the intrinsic and extrinsic parameters of the camera are accurate enough, then the misalignment of the images is less than one pixel if the depth variation is less than 10% of the average height, and the interframe image displacement (s_y) is less than 10% of the distance (d_y) between two slit windows. However, the direct paste of each slit would not result in a seamless mosaic if the 3D variation is large, or if the motion of the camera is fast (i.e. s_y is large), or if the error in motion measurements cannot be neglected. Therefore a local registration and warping is applied between two successive frames. Thus, a parametric model of the image motion of the region around the left (and the right) slit window is fitted between two images. In the current implementation, a projective model (with 8 independent parameters) is used, which can be represented as a 3x3 matrix $\mathbf{W_k}$ for frame *k* and has the same form as **A** in Eq. (7). Note that in the parametric model, both the structure of the scene and the motion of the camera are included. So for the (w_1+w_2)-wide slices from the current frame (Fig. 2b), the image transformations from the first stitching line to the fixed line, and then to the second stitching line are gradually interpolated from the image match model between the current and previous frames, $\mathbf{W_k}$, to the current camera model $(t_x^{(k)},t_y^{(k)})$, and then to the image match model between the current and the next frames, $\mathbf{W_{k+1}}$, column by column. A "two-track" line-by-line transformation algorithm under a 4-parameter transformation model with more details is given in [18].

3.3. Implementation Issues

In real application, the motion parameters (geo-data) provided by the measurement instrumentation may not be accurate enough due to the measurement resolution and noises. In other circumstances, we may not have *priori* motion parameters. In both cases, the complete solution of the problem becomes difficult because it turns out to be the bundle adjustment problem [19] of multiple frames. In the following sub-sections, we will show how we can obtain a reasonably good result while greatly simplifying the problem.

Free mosaic **-** To build a free mosaic, we will still use the same camera model in section 3.1. However, we will assume that two of the rotation angles, tip (around x axis) and tilt (around y axis), are small random variables, whereas the heading angle around the z axis can be large. In addition, assume that the depth of the terrain varies only around a virtual horizontal plane - the fixation plane - which is almost true for many videos of forest scenery. That is to say, a projective transformation $\mathbf{M}_k$ can be established by matching image points between two frames (*k* and *k-1*) about the virtual fixation plane. By decomposing planar projective transformation $\mathbf{B}_k (= \mathbf{A}_{k-1}\mathbf{M}_k)$ into a warping matrix $\mathbf{A}_k$ and an interframe translational vector $\Delta\mathbf{t}_k$, we can warp the current frame using $\mathbf{A_k}$ and find the global translational vector related to the reference frame as $\mathbf{t}_k = \mathbf{t}_{k-1} + \Delta\mathbf{t}_k$.

Motion refinement **-** The free mosaic algorithm can generate a rather "realistic" pair of stereoscopic mosaics with low computational cost. However, the mosaics may not be faithful to the geo-referenced path due to the model simplification employed and error accumulation. For example, even sub-pixel errors between two successive frames in the free-mosaic may lead to a drift from the correct path of up to hundred pixels for a thousand-frame video sequence. On the other hand, the GPS system used in our previous data collection only gives us an accuracy of about one meter in the ground, which means 5 to 10 pixels in the images of our zoom camera. Obviously it is not accurate enough to be directly used for image rectification and image mosaicing. Fortunately, errors in GPS locations and INS orientations do not accumulate in time, although the absolute values of the errors may be large. Thus it is possible to combine the two different kinds of measurements from geo-data and image match with different error properties. An Extended Kalman Filter (EKF) approach [20,21] is used in our system. The basic idea is that, even if the parametric motion model from image matching is only an approximation of the real motion, it is a good prediction of the motion parameters of the current frame. Then the measurement of the instrumentation package is used to correct this prediction to prevent the warping and translational parameters from drifting further away from the geo-referenced path.

4. DEPTH FROM STEREOSCOPIC MOSAICS

In the case of general motion model in Section 3, the stereo mosaics are generated from a rectified image sequence with only 2D translation (T_x,T_y), with the T_y component dominant. The epipolar lines are no longer horizontal scanlines if T_x is not zero for every frame. Then the corresponding point (x_r,y_r) in the right-view mosaic of any point (x_l,y_l) in the left-view mosaic will be constrained to the following *epipolar curve*

$$x_r = x_l + b_x \frac{y_r - y_l}{y_r - y_l + d_y} \tag{10}$$

where $b_x = F\dfrac{T_{xr} - T_{xl}}{H} = t_{xr} - t_{xl}$, and T_{xl} and T_{xr} are the *x* translational components in frame t_l and t_r (where columns y_l of left mosaic and y_r of the right mosaic come from) respectively. Recalling the generation of stereo mosaics, we can easily find that $t_{xr}(y) = t_{xl}(y + dy)$, where t_{xr} and t_{xl} are expressed as functions of the coordinate y in each mosaic. The epipolar curve can be re-written as

$$\Delta x = [t_{xl}(y_l + d_y + \Delta y) - t_{xl}(y_l)]\frac{\Delta y}{\Delta y + d_y} = b_x(y_l, \Delta y)\frac{\Delta y}{\Delta y + d_y} \tag{11}$$

where $\Delta x = x_r - x_l$, $\Delta y = y_r - y_l$, and $b_x(y_l,\Delta y)$ is the x baseline function of y_l and Δy. Note that Δx is a function of position y_l as well as displacement Δy, which is quite different from the epipolar-line geometry of a two-view perspective stereo. Here are several interesting observations. (1) Each column of the mosaic pair has its own cluster of parallel epipolar curves (Fig.3) due to the multi-perspective geometry; (2) The epipolar curve in the right mosaic for a given point (x_l,y_l) in the left mosaic is a curve passing through this point, which means that the stereo mosaics have aligned for all the points whose depths are H; (3) If we know the depth variation range $\pm\Delta Z_m$ of the terrain, the search region will be constrained in an interval $\Delta y \in [-\frac{d_y}{H}\Delta Z_m, -\frac{d_y}{H}\Delta Z_m]$ on the epipolar curve; and (4) If the x translation t_x is zero everywhere, there will be no x displacement in the mosaic pair and the epipolar curves will become horizontal epipolar lines.

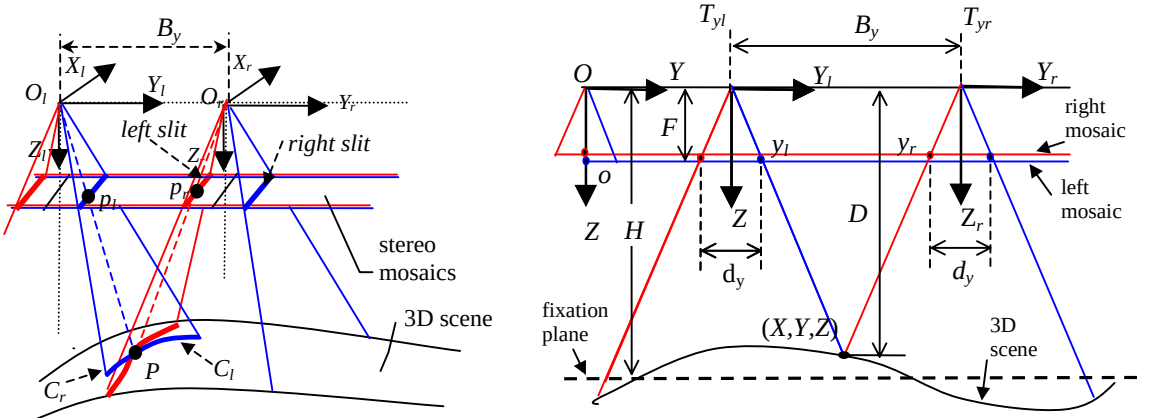
Multi-perspective stereo mosaics greatly reduce the computational burden of 3D reconstruction. Correspondence and 3D estimation are only performed on two mosaics, which are constructed using sparse yet robust image matches during image mosaicing. The disparity is fixed for every point in a scene in stereo mosaics; on the other hand, the "baseline" is proportional to the depth of a point, which means that the depth resolution is independent of the depth of the point.

Multi-perspective stereo mosaics can also serve as an image-based representation for large-scale stereoscopic viewing. Without 3D recovery, true 3D perception can be achieved by a human viewer while allowing the viewer to change viewpoint in a large scale. It should be noted that multi-perspective geometry is not full perspective. However human observers seem to have no problem seeing correct 3D from these images when viewed stereoscopically. This may be due in part to the fact that human fovea has only about 1-degree acceptance angle (the area over which the image is sharp). When viewing a stereo mosaic, the data in sharp focus is almost a true perspective image. Moreover, there are two additional advantages of the stereo mosaics over a traditional stereoscopic pair: viewing fixation and moving viewpoints, which are enabled by the inherent nature of the stereo mosaics.

5. EXPERIMENTAL RESULTS

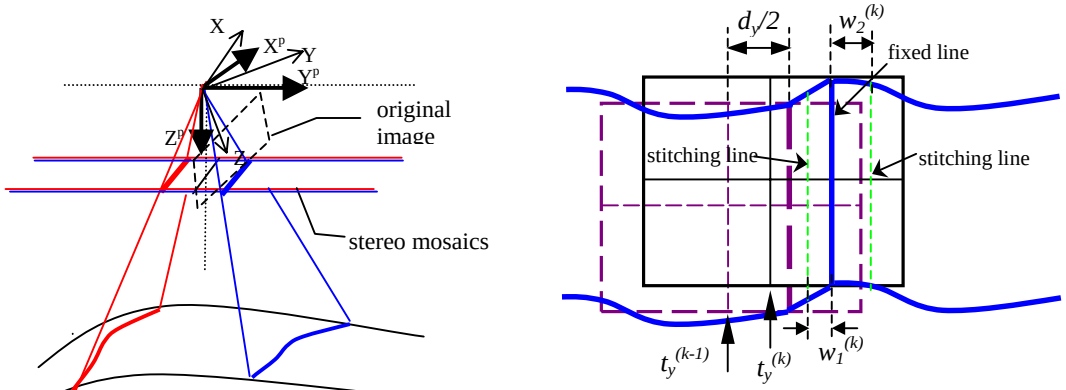
Preliminary results show that multi-perspective stereo mosaics are very useful for both 3D reconstruction and image-based rendering of large-scale terrain or urban scenes. The stereoscopically viewed mosaics of the forestry scene are both compelling and vivid to human viewing. The 3D reconstruction obtained from stereo mosaics is also very encouraging. Fig. 4 shows three frames from a 165-frame video sequence. The image resolution is 640*480. Fig. 4a and Fig. 4b show two successive images (frame 20 and frame 21) with 18.0-pixel interframe displacement. Two slit windows are 240-pixel apart. Fig.4c is a frame where scene in the right-slit window of Fig.4b can be seen in the left slit window of this image. These three images also show the baseline versus common FOV dilemma in a classic stereo/motion approach. Stereo mosaic method solves this problem by extending the FOV in the baseline's direction. From the same image sequence from a video camera, a stereo pair of large mosaics can be generated, having virtually the same FOVs in the two images and the optimal baselines for all the points. Fig. 5 shows the left view

mosaic and the right view mosaic, and the 3D map after sub-pixel image match [22]. More stereo mosaics and the depth map with high resolution can be found at
http://www.cs.umass.edu/~zhu/geomosaic.html.



(a). left(blue) and right(red) views (b). geometry of multi-perspective stereo (side view)

Fig.1. The model of stereoscopic mosaics



(a). Image rectification (b). Image mosaic

Fig. 2. Image rectification and image mosaic

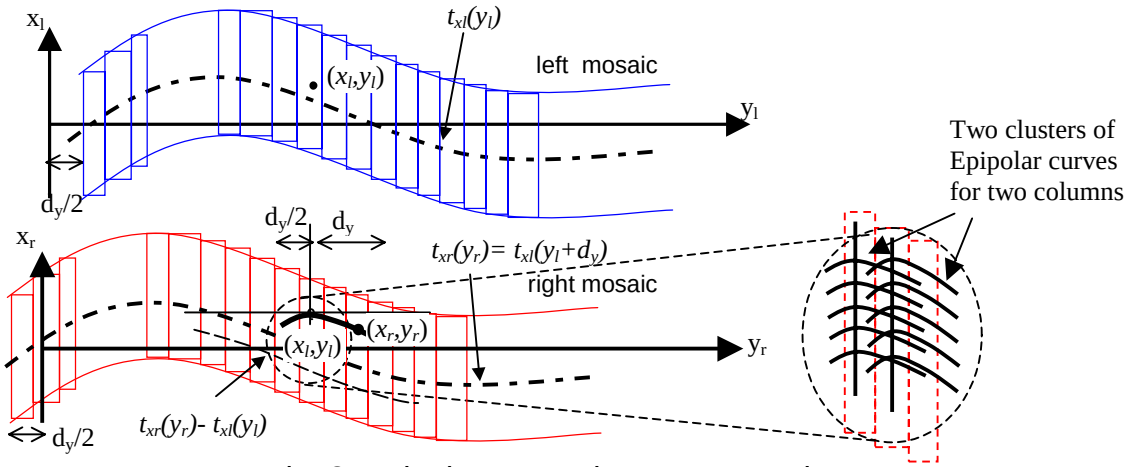
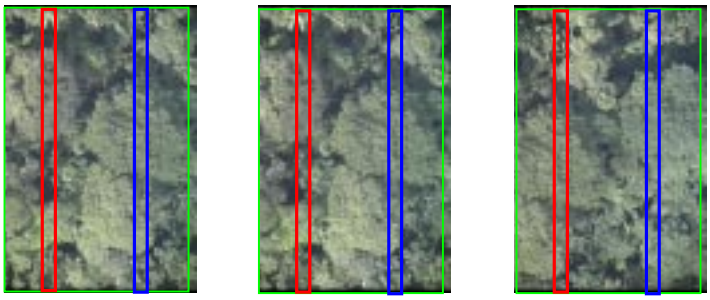
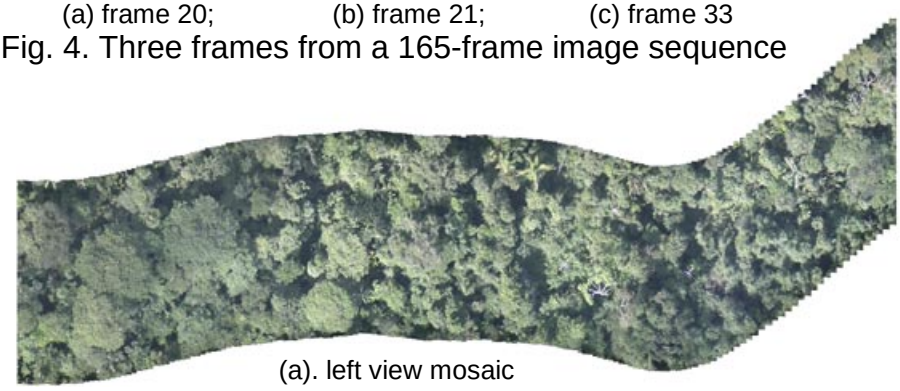


Fig. 3. Epipolar curves in stereo mosaics

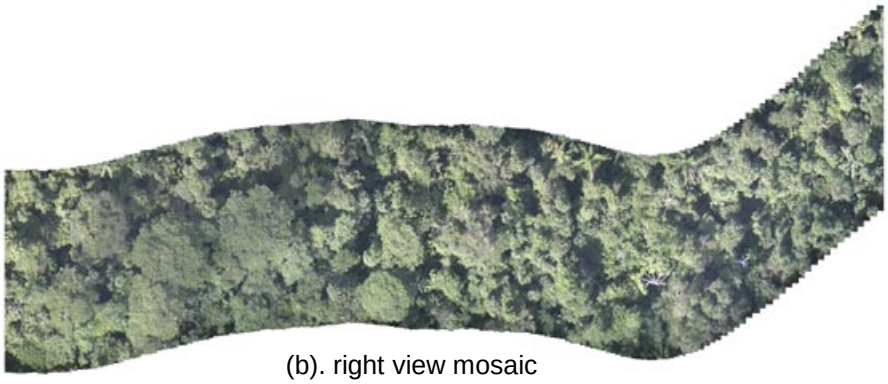


(a) frame 20; (b) frame 21; (c) frame 33

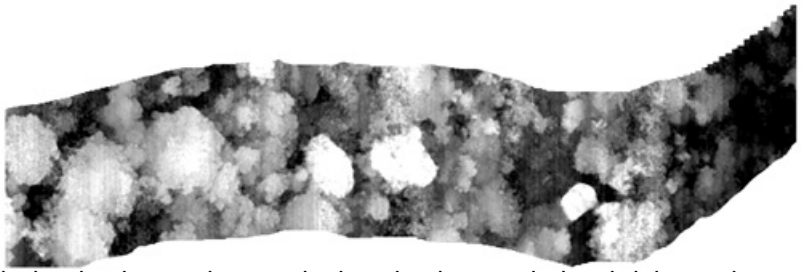
Fig. 4. Three frames from a 165-frame image sequence



(a). left view mosaic



(b). right view mosaic



(c). the depth map (cropped; elevation is encoded as brightness).

Fig. 5. Stereo mosaics and the depth map.

ACKNOWLEDGEMENTS

This work is partially supported by NSF EIA-9726401, and the NFWF #98089. Additional funding is supported by TNC in conjunction with the Winrock Corporation. Thanks are given to Kris Denio, Chris Holmes and Chris Hayward for their efforts in data collection, programming and helping evaluating results.

REFERENCES

[1]. S. E. Chen, QuickTime VR - an image based approach to virtual environment navigation, *Proc. Siggraph 95*:29-38.

[2]. H.S. Sawhney, R. Kumar, G. Gendel, J. Bergen, D.Dixon, V. Paragano, VideoBrushTM: Experiences with consumer video mosaicing, *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 1998: 56-62

[3]. H.-Y. Shum, R. Szeliski, Panoramic Image Mosaics, *Microsoft Research, Technical Report, MSR-TR-97-23, 1997*

[4]. Y. Xiong, K. Turkowski, Registration, calibration, and blending in creating high quality panoramas, *Proc. IEEE WACV'98*: 69-74.

[5]. Z. Zhu, G. Xu, E. M. Riseman and A. R. Hanson, Fast generation of dynamic and mutliresolution panorama from video sequences, *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, 1998, vol 1: 400-406.

[6]. H.-C. Huang, Y.-P. Hung, Panoramic stereo imaging system with automatic disparity warping and seaming, *Graphical Models and Image processing*, 60(3), 1998: 196-208.

[7]. H. Ishiguro, M. Yamamoto, S Tsuji, Omni-directional stereo for making global map, *Proc. IEEE Int. Conf. Computer Vision*,1990 *(ICCV'90)*, 540-547.

[8]. S. Peleg, M. Ben-Ezra, Stereo panorama with a single camera, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1999 *(CVPR'99)*: 395-401

[9]. H. Shum and R, Szeliski, Stereo reconstruction from multiperspective panoramas, *Proc. IEEE ICCV'99*, 14-21, 1999.

[10]. J. Y. Zheng, S. Tsuji, Generating Dynamic Projection Images for Scene Representation and Understanding, *Computer Vision and Image Understanding*, 72(3), Dec 1998: 237-256.

[11]. S. Peleg, J. Herman, Panoramic Mosaics by Manifold Projection. *Proc. IEEE CVPR'97*: 338-343.

[12]. P. Rademacher, G. Bishop, Multiple-center-of-projection images, *Proc. Siggraph'98*, 199-206.

[13]. Z. Zhu, G. Xu, X. Lin, Panoramic EPI Generation and Analysis of Video from a Moving Platform with Vibration, *Proc. IEEE CVPR99*: 531-537

[14]. S. Jenyns, *A Background to Chinese Paintings*, Schocken Books Inc. 1966

[15]. R. Kumar, H. Sawhney, J. Asmuth, J. Pope, S. Hsu, Registration of Video to Geo-referenced Imagery, *Proc. IAPR Int. Conf. Pattern Recognition 1998*, vol. 2: 1393-1400

[16]. R. Gupta , R. Hartley, Linear pushbroom cameras, *IEEE Trans. PAMI*, 19(9), Sep. 1997: 963-975

[17]. M. Irani, P. Anandan, S. Hsu, Mosaic based representation of video sequence and their applications, *Proc. IEEE ICCV'95*: 605-611

[18]. Z. Zhu, E. M. Riseman, A. R. Hanson, H. Schultz, Automatic Geo-Correction of Video Mosaics for Environmental Monitoring, *Technical Report, TR #99-28*, Computer Science Department, University of Massachusetts at Amherst, April, 1999.

[19]. C. C. Slama (Ed.), Manual of Photogrammetry, Fourth Edition, *Am. Soc. Photogrammetry*, 1980.

[20]. A. Gelb, *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.

[21]. T. J. Broida, R. Chellappa, Estimating the kinematics and structure of a rigid object from a sequence of monocular images, *IEEE Trans PAMI*, 13(6), 1991: 497-513.

[22]. H. Schultz, E. M. Riseman, F. Stolle, and D.-M. Woo, Error detection and DEM fusion using self-consistency, *Proc. IEEE ICCV99*, 1174 - 1181, 1999.