Error Characteristics of Parallel-Perspective Stereo Mosaics

Zhigang Zhu, Allen R. Hanson, Howard Schultz, Edward M. Riseman
Department of Computer Science, University of Massachusetts at Amherst, MA 01003
E-mail: zhu@cs.umass.edu

Abstract

This paper analyzes different aspects of the error characteristics of parallel-perspective stereo mosaics generated from an airborne video camera moving through a complex three-dimensional scene. First, we show that theoretically a stereo pair of parallel-perspective mosaics is a good representation for an extended scene, and the adaptive baseline inherent to the geometry permits depth accuracy independent of absolute depth. Second, in practice, we have proposed a 3D mosaicing technique PRISM (parallel-ray interpolation for stereo mosaicing) that uses interframe match to interpolate the camera position between the original exposure centers of video frames taken at discrete spatial steps. By analyzing the errors introduced by a 2D mosaicing method, we explain why the "3D mosaicing" solution is important to the problem of generating smooth and accurate mosaics while preserving stereoscopic information. We further examine whether this ray interpolation step introduces extra errors in depth recover from stereo mosaics by comparing to the typical perspective stereo formulation. Third, the error characteristics of parallel stereo mosaics from cameras with different configurations of focal lengths and image resolutions are analyzed. Results for mosaic construction from aerial video data of real scenes are shown and for 3D reconstruction from these mosaics are given. We conclude that (1) stereo mosaics generated with the PRISM method have significantly less errors in 3D recovery (even if not depth independent) due to the adaptive baseline geometry; and (2) longer focal length is better since stereo matching becomes more accurate.

1. Introduction

There have been attempts in a variety of applications to add 3D information into an image-based mosaic representation. Creating stereo mosaics from two rotating cameras was proposed by Huang & Hung [1], and from a single off-center rotating camera by Ishiguro, et al [2], Peleg & Ben-Ezra [3], and Shum & Szeliski [4]. In these kinds of stereo mosaics, however, the viewpoint -- therefore the parallax -- is limited to images taken from a very small area. Recently our work [5,6,7] has been focused on parallel-perspective stereo mosaics from a dominantly translating camera, which is the typical prevalent sensor motion during aerial surveys. A rotating camera can be easily controlled to achieve the desired motion. On the contrary, the translation of a camera over a large distance is much hard to control in real vision applications such as robot navigation [8] and environmental monitoring [6, 9]. We have previously shown [5-7] that image mosaicing from a translating camera raises a set of different problems from that of circular projections of a rotating camera. These include suitable mosaic representations, the generation of a seamless image mosaic under a rather general motion with motion parallax, and epipolar geometry associated with multiple viewpoint geometry.

In this paper we will give a thorough analysis on various aspects of the error characteristics of 3D reconstruction from parallel-perspective stereo mosaics generated from real video sequences. It has been shown independently by Chai and Shum [10] and by Zhu, et al [5,6] that parallel-perspective is superior to both the conventional perspective stereo and the recently developed multi-perspective stereo for 3D reconstruction, in that the adaptive baseline inherent to the parallel-perspective geometry permits depth accuracy independent of absolute depth. However, this conclusion is obtained in an ideal case – i.e. enough samples of parallel projection rays from a "virtual camera" with ideal 1D or 2D motion can be generated from a complete scene model. In the practice of stereo mosaicing from a real video sequence, however, we need to consider the errors in the final mosaics versus camera motion types, frame rates, focal lengths, and scene depths. The analysis of the error characteristics of 3D construction from real stereo mosaics will be the focus of this paper.

First we will show why an efficient "3D mosaicing" techniques are important for accurate 3D reconstruction from stereo mosaics. Obviously use of standard 2D mosaicing techniques based on 2D image transformations such as a manifold projection [11] cannot generate a seamless mosaic in the presence of large motion parallax, particularly in the case of surfaces that are highly irregular or with large different heights. Moreover, perspective distortion causing the geometric seams will introduce errors in 3D reconstruction using the parallel-perspective geometry of stereo mosaics. In generating image mosaics

with parallax, several techniques have been proposed to explicitly estimate the camera motion and residual parallax [9,12,13]. These approaches, however, are computationally intense, and since a final mosaic is represented in a reference perspective view, there could be serious occlusion problems due to large viewpoint differences between a single reference view and the rest of the views in the image sequence.

We have proposed a novel "3D mosaicing" technique called PRISM (parallel ray interpolation for stereo mosaicing) [7] to efficiently convert the sequence of *perspective* images with 6 DOF motion into the parallel-perspective stereo mosaics. In the PRISM approach, global image rectification eliminates rotation effects, followed by a fine local transformation that accounts for the interframe motion parallax due to 3D structure of the scene, resulting in a stereo pair of mosaics that embodies 3D information of the scene with optimal baseline. This paper further examines (1) whether the PRISM process of image rectification followed by ray interpolation introduces extra errors in the following step of depth recovery; and (2) whether the final disparity equation of the stereo mosaics really means that the depth recovery accuracy is independent of the focal length and absolute depths. To show the advantages of the stereo mosaics, depth recovery accuracy is analyzed and compared to the typical perspective stereo formulation. Results for mosaic construction from aerial video data of real scenes are shown and for 3D reconstruction from these mosaics are given in the paper. Several important conclusions for generating and using stereo mosaics will be made based on our theoretical and experimental analysis.

2. Parallel-Perspective Stereo Geometry

Fig. 1 illustrates the basic idea of the parallel-perspective stereo mosaics. Let us first 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. 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. The mosaic images thus generated are similar to *parallel-perspective* images captured by a linear pushbroom camera [14], which has *perspective projection in the direction perpendicular to the motion and parallel projection in the motion direction*. In contrast to the common pushbroom aerial image, these mosaics are obtained from two different oblique viewing angles of a single camera's field of view, one set of rays looking forward and the other set of rays looking backward, so that a stereo pair of left and right mosaics can be generated as the sensor moves forward, capturing the inherent 3D information.

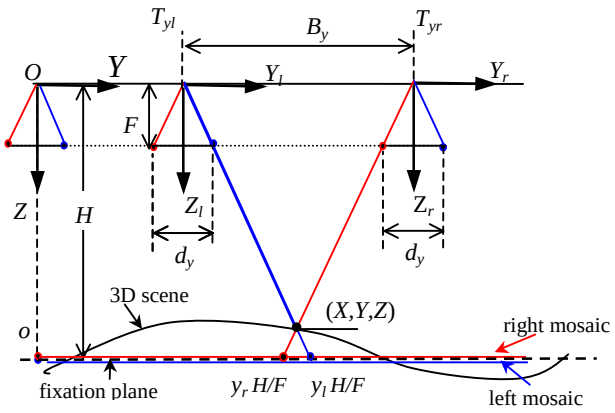


Fig. 1. Parallel-perspective stereo geometry. Both mosaics are built on the fixation plane, but their unit is in pixel – each pixel represents H/F world distances.

2.1. Parallel-perspective stereo model

Without loss of generality, we assume that two vertical 1-column slit windows have $d_y/2$ offsets to the left and right of the center of the image respectively (Fig. 1). The "left eye" view (left mosaic) is generated from the front slit window, while the "right eye" view (right mosaic) is generated from the rear slit window. The *parallel-perspective projection model* of the stereo mosaics thus generated can be represented by the following equations [6]

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

where F is the focal length of the camera, H is the height of a *fixation plane* (e.g., average height of the terrain). Eq.(1) gives the relation between a pair of 2D points (one from each mosaic), (x_l,y_l) and (x_r,y_r), and the corresponding 3D point (X,Y,Z). It serves a function similar to the classical pin-hole perspective camera model. A generalized model under 3D translation is given in [7]. The depth can be computed as (from Eq. (1))

$$Z = H\frac{b_y}{d_y} = H(1+\frac{\Delta y}{d_y}) \qquad (2)$$

where

$$b_y = d_y + \Delta y = FB_y/H \qquad (3)$$

is the "scaled" version of the baseline B_y, $\Delta y = y_r - y_l$ is the "mosaic displacement"[1] in the stereo mosaics. Displacement Δy is a function of the depth variation of the scene around the fixation plane H. Since a fixed angle between the two viewing rays is selected for generating the stereo mosaics, the "disparities" (d_y) of all points are fixed; instead a geometry of optimal/adaptive baselines (b_y) for all the points is created. In other words, for any

[1] We use "displacement" instead of "disparity" since it is related to the baseline in a two view-perspective stereo system.

point in the left mosaic, searching for the match point in the right mosaic means finding an original frame in which this match pair has a pre-defined disparity (by the distance of the two slit windows) and hence has an adaptive baseline depending on the depth of the point (Fig. 1).

2.2. Depth resolution of stereo mosaics

In a pair of parallel-perspective stereo mosaics, depth is proportional to the image displacement Δy (Eq.(2)). Since Δy is measured in discrete images, we assume that the image localization resolution is ∂y pixels (usually $\partial y <= 1$) in the stereo mosaics, so that $\Delta y = 0, \pm \partial y$, $\pm 2\partial y$, …. *The depth resolution in the parallel-perspective stereo is a constant value* (Fig. 2)

$$\partial Z = \frac{H}{d_y}\partial y = \text{constant} \qquad (4)$$

which is a contrast to the two-view perspective stereo where the depth error of a point is proportional to the square of the depth (Eq. 9a-6) in Appendix).

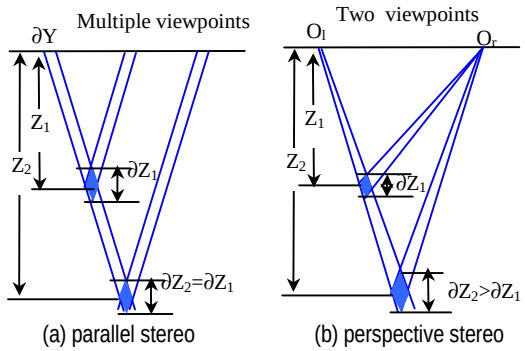


Fig. 2 Depth resolution of stereo mosaics

Ideally, for parallel-perspective stereo, depth resolution is independent of absolute depths of scene points and of the focal length of the camera used to generate the stereo mosaics. In addition, the image resolutions in the y direction are the same no matter how far scene points are. The reason is that due to the parallel projection in the y direction, parallel rays intersect in the 3D scene points instead of converging rays (Fig. 2).

3. Stereo Mosaicing from Real Video

In the PRISM approach for large-scale 3D scene modeling from real video, the computation of "match" is efficiently distributed in three steps: camera pose estimation, image mosaicing and 3D reconstruction. In estimating camera poses (for image rectification), only sparse tie points widely distributed in the two images are needed. In generating stereo mosaics, matches are only performed for parallel-perspective rays between small overlapping regions of successive frames. In using stereo mosaics for 3D recovery, matches are only carried out between the two final mosaics. This section gives a brief summary of the techniques in the three steps, as the base for the error analysis in the following section. Algorithms and discussions in detail can be found in [6,7].

3.1. Image rectification

The stereo mosaicing mechanism can be generalized to the case of 3D translation if the 3D curved motion track has a dominant translational motion for generating a parallel projection in that direction [7]. Under 3D translation, seamless stereo mosaics can be generated in the same way as in the case of 1D translation. The only difference is that viewpoints of the mosaics form a 3D curve instead of a 1D straight line. Further, the motion of the camera can be generalized to a 6 DOF motion with some reasonable constraints on the values and rates of changes of motion parameters of a camera [6,7] (Fig. 3a), which are satisfied by a sensor mounted in a light aircraft with normal turbulence. There are two steps necessary to generate a rectified image sequence that exhibits only 3D translation, from which we can generate seamless mosaics:

1) *Camera orientation estimation*. Assuming an internally pre-calibrated camera, the extrinsic camera parameters (camera orientations) can be determined from our aerial instrumentation system (GPS, INS and a laser profiler)[15] and a bundle adjustment technique [16]. The detail is out the scope of this paper, but the main point here is that we do not need to carry out dense match between two successive frames. Instead only sparse tie points widely distributed in the two images are needed to estimate the camera orientations.

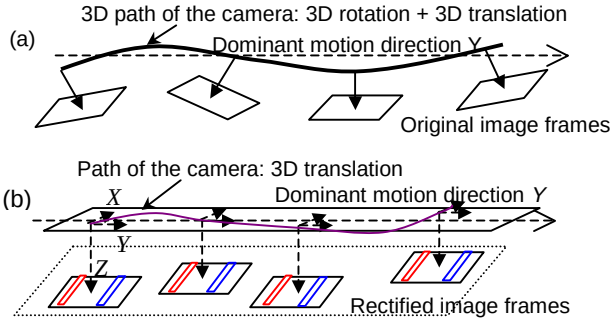


Fig. 3. Image rectification. (a) Original and (b) rectified image sequence.

2) *Image rectification*. A 2D projective transformation is applied to each frame in order to eliminate the rotational components(Fig. 3b). In fact we only need to do this kind of transformation on two narrow slices in each frame that will contribute incrementally to each of the stereo mosaics. The 3D motion track formed by the viewpoints of the moving camera will have a dominant motion direction (*Y*) that is perpendicular to the optical axis of the "rectified" images.

3.2. Ray interpolation

How can we generate seamless mosaic from video of a translating camera in a computational effective way? The key to our approach lies in the parallel-perspective representation and an interframe ray interpolation approach. For each of the left and right mosaics, we only need to take a front (or rear) slice of a certain width (determined by interframe motion) from each frame, and perform local registration between the overlapping slices of successive frames (Fig. 4), then generate parallel *interpolated rays* between two known discrete perspective views for the left (or right) mosaic.

Since we will use the mathematical model of the ray interpolation in the following error analysis, let us examine this idea more rigorously in the case of 2D translation after image rectification when the translational components in the Z direction is small [6]. We take the left mosaic as an example (Fig. 4). First we define the central column of the front (or rear) mosaicing slice in each frame as a *fixed line*, which has been determined by the camera's location of each frame and the pre-selection of the front (or rear) slice window (Fig. 4, Fig. 5). An interpretation plane (IP) of the fixed line is a plane passing through the nodal point and the fixed line. By the definition of parallel-perspective stereo mosaics, the IPs of fixed lines for the left (or right) mosaic are parallel to each other. Suppose that (S_x, S_y) is the translational vector of the camera between the previous (1^{st}) frame of viewpoint (T_x, T_y) and the current (2^{nd}) frame of view point (T_x+S_x, T_y+S_y) (Fig. 4). We need to interpolate parallel rays between the two *fixed lines* of the 1st and the 2nd frames. For each point (x_1, y_1)(to the right of the 1^{st} fixed line $y_0=d_y/2$) in frame(T_x, T_y), which will contribute to the left mosaic, we can find a corresponding point (x_2, y_2) (to the left of the 2^{nd} fixed line) in frame (T_x+S_x, T_y+S_y). We assume that (x_1, y_1) and (x_2, y_2) are represented in their own frame coordinate systems, and intersect at a 3D point (X,Y,Z). Then the parallel reprojected viewpoint (T_{xi}, T_{yi}) of the correspondence pair can be computed as

$$T_{yi} = T_y + \frac{(y_1 - d_y/2)}{y_1 - y_2} S_y,\ T_{xi} = T_x + \frac{S_x}{S_y}(T_{yi} - T_y) \qquad (5)$$

where T_{yi} is calculated in a synthetic IP that passes through the point (X,Y,Z) and is parallel to the IPs of the fixed lines of the first and second frames, and T_{xi} is calculated in a way that all the viewpoints between (T_x, T_y) and (T_x+S_x, T_y+S_y) lie in a straight line. Note that Eq. (6) also holds for the two fixed lines such that when $y_1 = d_y/2$ (the first fixed line), we have $(T_{xi}, T_{yi})=(T_x, T_y)$, and when $y_2 = d_y/2$ (the second fixed line), we have $(T_{xi}, T_{yi})=(T_x+S_x, T_y+S_y)$. We assume that normally the interframe motion is large enough to have $y_1-1 \geq d_y/2 \geq y_2+1$. A super dense image sequence could generate a pair of stereo mosaics with super-resolution, but this will not be discussed in this paper.

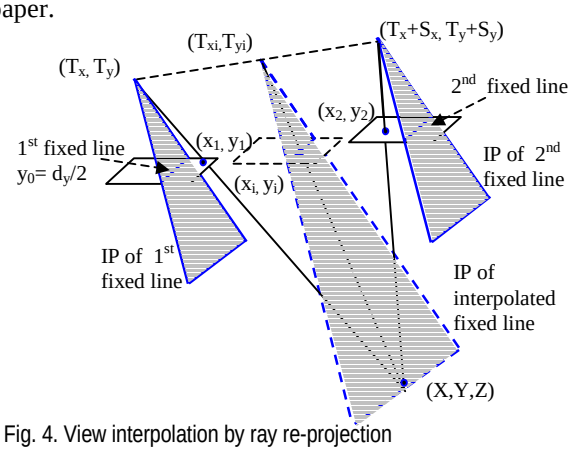


Fig. 4. View interpolation by ray re-projection

The reprojected ray of the point (X,Y,Z) from the interpolated viewpoint (T_{xi}, T_{yi}) is

$$(x_i,\ y_i) = [x_1 - \frac{S_x}{S_y}(y_1 - \frac{d_y}{2}),\ \frac{d_y}{2}] \qquad (6)$$

and the mosaicing coordinates of this point is

$$(x_m,\ y_m) = [t_{xi} + x_1 - \frac{S_x}{S_y}(y_1 - \frac{d_y}{2}),\ t_{yi} + \frac{d_y}{2}] \qquad (7)$$

where

$$t_{xi}=F\,T_{xi}/H\,,\ \ t_{yi}=F\,T_{yi}/H. \qquad (8)$$

are the "scaled" translational components of the interpolated view. Note that the interpolated rays are also parallel-perspective, with perspective in the x direction and parallel in the y direction.

3.3. 3D reconstruction from stereo mosaics

In the general case, the viewpoints of both left and right mosaics will be on the same smooth 3D motion track. Therefore the corresponding point in the right mosaic of any point in the left mosaic will be on an epipolar curve determined by the coordinates of the left point and the 3D motion track. We have derived the epipolar geometry of the stereo mosaics generated from a rectified image sequence exhibiting 3D translation with the y component dominant [7]. Under 2D translation (T_x,T_y), 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 an *epipolar curve*

$$\Delta x = b_x(y_l, \Delta y)\frac{\Delta y}{\Delta y + d_y}, \qquad (9)$$

$$\Delta x = x_r - x_l\,,\ \ \Delta y = y_r - y_l$$

where $b_x(y_l, \Delta y) = [t_{xl}(y_l + d_y + \Delta y) - t_{xl}(y_l)]$ is the baseline function of y_l and Δy, and $t_{xl}(y_l)$ is the "scaled" x translational component (as in Eq. (3) or (8)) of the original frames corresponding to column y_l in the left

mosaic. Hence Δx is a nonlinear function of position y_1 as well as displacement Δy, which is quite different from the epipolar geometry of a two-view perspective stereo. The reason is that image columns of different y_1 in parallel-perspective mosaics are projected from different viewpoints. In the ideal case where the viewpoints of stereo mosaics form a 1D straight line, the epipolar curves will turn out to be horizontal lines.

The depth maps of stereo mosaics were obtained by using the Terrest system designed for perspective stereo match[17] without modification. The Terrest system was designed to account the illumination differences and perspective distortion of stereo images with largely separated views by using normalized correlation and multi-resolution un-warping. Further work is needed to apply the epipolar curve constraints into the search of correspondence points in the Terrest to speedup the match process. Currently we perform matches with 2D search regions estimated from the motion track and the maximum depth variations of a scene.

In parallel-perspective stereo mosaics, since a fixed angle between the two sets of viewing rays is selected, the disparities of all points are pre-selected (by mosaicing) and fixed; instead the geometry of optimal/adaptive baselines for all the points is created. From the parallel-perspective stereo geometry, the depth accuracy is independent to the depth of a point and the image resolution. However, there are two classes of issues that need to be carefully studied in stereo mosaics from real video sequences. First, 3D recovery from stereo mosaics need a two-step matches, i.e., interframe matching (and ray interpolation) to generate the mosaics, and the correspondences of the stereo mosaics to generate a depth map. Does the ray interpolation step introduce extra errors? Second, the final disparity equation of the stereo mosaics does not include any focal length. Does it means that the depth recovery accuracy from stereo mosaics is really independent of the focal length of the camera that captures the original video? We will discuss these two issues in the following sections.

4. Error Analysis of Ray Interpolation

4.1. Comparison of 3D vs. 2D mosaicing

First, we give the rationale why "3D mosaicing" is so important for 3D reconstruction from stereo mosaics by a real example. Fig. 5 shows the local match and ray interpolation of a successive frame pair of a UMass campus scene, where the interframe motion is (s_x, s_y) = (27, 48) pixels, and points on the top of a tall building (the Campus Center) have about 4 pixels of additional motion parallax. As we will see next, these geometric misalignments, especially of linear structures, will be clearly visible to human eyes. Moreover, perspective distortion causing the geometric seams will introduce errors in 3D reconstruction using the parallel-perspective geometry of stereo mosaics. In the example of stereo mosaics of the UMass campus scene [18], the distance between the front and the rear slice windows is d_y = 192 pixels, and the average height of the aerial camera from the ground is H = 300 meters (m). The relative y displacement of the building roof (to the ground) in the stereo mosaics is about Δy = -29 pixels. Using Eq. (2) we can compute that the "absolute" depth of the roof from the camera is Z = 254.68 m, and the "relative" height of the roof to the ground is ΔZ = 45.31 m. A 4-pixel misalignment in the stereo mosaics will introduce a depth (height) error of δZ = 6.25 m, though stereo mosaics have rather large "disparity" (d_y =192). While the relative error of the "absolute" depth of the roof ($\delta Z/Z$) is only about 2.45%, the relative error of its "relative" height ($\delta Z/\Delta Z$) is as high as 13.8%. This clearly shows that geometric-seamless mosaicing is very important for accurate 3D estimation as well as good visual appearance. It is especially true when sub-pixel accuracy in depth recovery is needed [17].

In principle, we need to match all the points between the two fixed lines of the successive frames to generate a complete parallel-perspective mosaic. In an effort to reduce the computational complexity, we have designed a fast *3D mosaicing* algorithm [7] based on the proposed PRISM method. It only requires matches between a set of point pairs in two successive images around their *stitching line*, which is defined as a virtual line in the middle of the two fixed lines (see Fig. 5). The pair of *matching curves* in the two frames is then mapped into the mosaic as a *stitching curve* by using the ray interpolation equation (7). The rest of the points are generated by warping a set of triangulated regions defined by the control points on the matching curve (that correspond to the stitching curve) and the fixed line in each of the two frames. Here we assume that each triangle is small enough to be treated as a planar region.

Using sparse control points and image warping, the proposed 3D mosaicing algorithm only approximates the parallel-perspective geometry in stereo mosaics (Fig. 9), but it is good enough when the interframe motion is small (e.g., Fig. 10 and Fig. 11). Moreover, the proposed 3D mosaicing 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.

While we are still working on 3D camera orientation estimation using our instrumentation and the bundle

adjustments [15], Fig. 9 shows mosaic results where camera orientations were estimated by registering the planar ground surface of the scene via dominant motion analysis. However the effect of seamless mosaicing is clearly shown in this example. Please compare the results of 3D mosaicing (parallel-perspective mosaicing) vs. 2D mosaicing (multi-perspective mosaicing) by looking along many building boundaries associating with depth changes in the entire 4160x1536 mosaics at our web site [18]. Since it is hard to see subtle errors in the 2D mosaics of the size of Fig. 9a, Fig. 9b and Fig. 9c show close-up windows of the 2D and 3D mosaics for the portion of the scene with the tall Campus Center building. In Fig. 9b 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. 9c) does not exhibit these problems.

4.2. Errors from ray interpolation

In theory, the adaptive baseline inherent to the parallel-perspective geometry permits depth accuracy independent of absolute depth. However, in practice, two questions need to be answered related to local match and ray interpolation. First, since we use motion parallax between two successive frames, will the small baseline between frames introduce large errors in ray interpolation? Second, is there any resolution gain or loss due to the change of the perspective projection of original frames to the parallel projection of the stereo mosaics for different depths?

The answer to the second question is relatively simple: A simple transformation of perspective frames to parallel-perspective mosaics does introduce resolution changes in images (Fig. 6). Recall that we build the mosaics on a fixation plane of the depth H. It means that the image resolution in the stereo mosaics are the same as the original frames only for points on plane H. However, for regions whose depths are less than H, a simple parallel ray re-sampling process will result in resolution loss. On the other hand, regions whose depths are larger than H could have their resolution enhanced by sub-pixel interpolation. This tells us that if we select the fixation plane above all the scene points, we can make full use of the image resolution of the original video frames. However, if we still want to keep the fixation plane between the scene points, we can still preserve the image resolution for the nearer points by a super-sampling process. For the points below the fixation plane, resolution could be better enhanced by using sub-pixel interpolation between a pair of frames as illustrated in Fig. 6, assuming that we are performing a sub-pixel match for the ray interpolation. For example, for a point P that lies between two point P_1 and P_2 on the grids of the image O_1, we find its match between point Q_1 and Q_2 on the grids of the image O_2. Then the color of the point P can be better interpolated by using points Q_1 and P_2 since they are closer to the point P in space.

In order to answer the first question, we formulate the problem as follows (under 1D translation): Given an accurate point $y_3 = -d_y/2$ in view O_3 that contributes to the right mosaic, we try to find a match point $y_i=+d_y/2$ in a view that contributes to the left mosaic with parallel-perspective projection (Fig. 7). Note that we express these points in their corresponding frame coordinate systems instead of the mosaicing coordinate system for easy notations; the mappings from these points to the mosaicing coordinates are straightforward. The point y_i is usually reprojected from an interpolated view O_i generated from a match point pair y_1 and y_2 in two existing consecutive views O_1 and O_2. The localization error of the point y_i depends on the errors in matching and localizing points y_1 and y_2. Analysis (see Appendix) shows that even if the depth from two successive views O_1 and O_2 cannot give us good 3D information (as shown by the large pink error region in Fig. 7, Eq. (a-6)), the localization error of the interpolated point (i.e. the left ray of O_i) is quite small (Eq. (a-5). It turns out that such depth error of stereo mosaics is bounded by the errors of two pairs of stereo views O_1+O_3 and O_2+O_3, both with almost the same "optimal" baseline configuration as the stereo mosaics. Using Eq. (a-6), and also considering the resolution changes in the mosaics discussed above, the depth estimation error of stereo mosaics can be derived as

$$\left|\frac{\partial Z}{\partial y}\right| = \begin{cases} = \dfrac{H}{d_y}, & \text{if } Z \le H \\ \in \left(\dfrac{H}{d_y}, \dfrac{Z}{d_y}\right), & \text{if } Z > H \end{cases} \qquad (9)$$

where pixel localization error ∂y is measured in the mosaics rather than in the original frames as in Eq. (a-6). So the resolution lose ($Z<H$) and enhancement ($Z>H$) are reflected in Eq. (9). Comparing Eq. (9) with Eq. (4), it can be seen that the depth error of the "real" stereo mosaics generated by ray interpolation is related to the actual depth (Z) of the point instead of just the average depth H. Therefore, in practice the depth accuracy is not independent of absolute depth. Nevertheless, parallel-perspective stereo mosaics still provide a stereo geometry with a pre-selected and fixed disparity and adaptive baselines for all the points of different depths. Here are four conclusions that are very important to the generation and applications of the stereo mosaics (refer to the equations in Appendix):

Conclusion 1. In theory, the depth accuracy of parallel-perspective stereo is independent of absolute depths;

however, in practice, the depth error of the stereo mosaics is roughly proportional to the absolute depth of a point.

Conclusion 2. Parallel-perspective stereo is apparently more accurate than two-view perspective stereo. Parallel-perspective stereo mosaics provide a stereo geometry with adaptive baselines for all the points of different depths, and depth error is better than a linear function of absolute depth. In contrast, the two-view perspective stereo has a fixed baseline, and the depth error is a second order function of absolute depth.

Conclusion 3. Ray interpolation does not introduce extra errors to depth estimation from parallel-perspective stereo mosaics. The accuracy of depth estimation using stereo mosaics via ray interpolation is comparable to the case of two-view perspective stereo with the same "adaptive" baseline configurations (if possible). Obviously, stereo mosaics provide a nice way to achieve such configurations.

Conclusion 4. The ray interpolating accuracy is independent of the magnitude of the interframe motion. This means that stereo mosaics with the same degree of accuracy can be generated from sparse image sequences, as well as dense ones, given that the interframe matches are correct.

5. Error Analysis versus Focal Lengths

5.1. Selecting focal length and image resolution

It is well known that in stereo vision, a large baseline will give us better 3D accuracy in 3D recovery. The geometric property of the parallel-perspective stereo mosaics also indicates that a larger angle between the two sets of rays of the stereo mosaics will give us larger baselines (B_y in Fig. 1), hence better 3D accuracy. It seems to tell us that a wide-angle lens (with shorter focal length) could give us larger baselines and hence better stereo mosaic geometry than a tele-photo lens (with longer focal length). However, one must consider several factors that affect the generation of the stereo mosaics and the correspondence of the stereo mosaics, to see this argument is not necessarily true.

First we assume that the camera has the same number of pixels no matter what the focal length (and the field of view) is. A simple fact is that wider field of view (FOV), i.e., shorter focal length always means lower image resolution (which is defined as the *number of pixels per meter length of the footprint on terrain*). Our question is: given the same distance of the two slit windows, d_y(in pixels), what kind of focal length gives us better depth resolution, the wide angle lens or the telephoto lens?

In stereo mosaics, the error in depth estimate comes from the localization error of the stereo displacement Δy, which consists of two parts: the mosaic registration error δb_1 and the stereo match error δb_2. The first part mainly comes from the baseline estimation (i.e. "calibration") error δB, by the following equation:

$$\delta b_1 = \frac{F}{H}\delta B \tag{10}$$

where H is the depth of the fixation plane in generating the mosaics. From Eq. (2) the depth error part due to the mosaic error is

$$\delta Z_1 = \frac{H}{d_y}\delta b_1 = \frac{F}{d_y}\delta B \tag{11}$$

Second, the depth estimation error due to the stereo match error δb_2 depends on how big a δb_2-pixel footprint is on the ground. Since the image resolution of a point of depth H in the image of the focal length F is F/H (pixels/meter), the size of the footprint on the ground will be (Fig. 8)

$$\delta Y = \frac{H}{F}\delta b_2 \tag{12}$$

Obviously shorter focal lengths produce larger footprints., hence lower spatial resolution This part of the depth error can be expressed as

$$\delta Z_2 = \frac{F}{d_y}\delta Y = \frac{H}{d_y}\delta b_2 \tag{13}$$

Note that the same depth accuracy in terms of stereo matching is achieved for different focal lengths since the larger baseline in the case of the wider field of view exactly compensates for the larger footprint on the ground with parallel projections (Fig. 8). The total depth error is

$$\delta Z = \frac{H}{dy}(\delta b_1 + \delta b_2) \tag{14}$$

or

$$\delta Z = \frac{F}{dy}\delta B + \frac{H}{dy}\delta b_2 \tag{15}$$

whose differences will be explained in the following:

(1). If the registration error in generating mosaics is independent of the focal length, which could be the case when the relative camera orientation is directly estimated from interframe image registration and bundle adjustments, then Eq. (14) shows that depth error is independent of the focal length (Fig. 8). However, since a smaller focal length (wide FOV) means a larger angle between the two set of left and right rays of the stereo mosaics (given the same distance of the slit windows), it will introduce larger matching error δb_2 due to occlusion, perspective distortion and illumination changes of large separated view angles.

(2) If the absolute camera orientation (and hence the baseline B_y) is estimated from other instrumentation other than image registration, the registration error in generating mosaics will be proportional to the focal length (Eq. (10)). This means the same baseline error will introduce larger

mosaic registration error if larger focal length is used. In this case, Eq. (15) should be used to estimate the depth error, which indicates that given the baseline estimate error δB, larger focal length will introduce larger error in the first part due to the multiplication of F, and smaller error in the second part due to the smaller stereo match error δb_2. As it is hard to give an explicit function of the stereo match error versus focal length (and view difference), it is roughly true that the second part is dominant using a normal focal length. In this case, a shorter focal length (and wider view direction difference) in generating stereo mosaics will introduce larger match error due to lower image resolution, significantly larger occlusion and more obvious illumination differences. On the other hand, too long a focal length will result in too short baselines, hence too big enlargement of the calibration error in the images. Therefore, it is possible to find an optimal focal length if we can specify a stereo matching error function versus the focal length (field of view), considering the texture and depth variation of the terrain and the size of the stereo match primitives in stereo images. Quantitatively, we have the following conclusion:

Conclusion 5. Ideally, estimating depth error of stereo mosaic is independent to the focal length of the camera that generates the stereo mosaics. However, in practice longer focal length will give better 3D reconstruction from the stereo mosaics, due to the finer image resolution, less occlusion and fewer lighting problems if a reasonably good baseline geometry can be constructed.

5.2. Experimental analysis

As an example, Fig. 10 an Fig. 11 compare the real examples of 3D recovery from stereo mosaics generated from a telephoto camera and a wide angle camera for the same forest scene. The average height of the airplane is H = 385 m, and the distance between the two slit windows for both the telephoto and wide-angle stereo mosaics is d_y = 160. The focal length of the telephoto camera is F_{tele} = 2946 pixels and that of the wide angle camera is F_{wide} = 461 pixels (which were estimated by a simple calibration using the GPS/INS/laser range information with the camera, and the results from image registration). In both cases, the size of the original frames are 720 (x)*480(y) where the camera moved in the vertical (y) direction. By a simple calculation, the image resolution of the telephoto camera is 7.65 pixels/meter and that of the wide-angle camera is 1.20 pixels/meter.

The depth maps of stereo mosaics were obtained by using the Terrest system based on a hierarchical sub-pixel dense correlation method [17]. Fig. 10c and Fig. 11c show the derived "depth" maps (i.e., displacement maps) from the pairs of telephoto and wide angle parallel-perspective stereo mosaics of the forest scene. In the depth maps, mosaic displacements are encoded as brightness so that higher elevations (i.e. closer to the camera) are brighter. It should be noted here that the parallel-perspective stereo mosaics were created by the proposed 3D mosaicing algorithm, with the camera orientation parameters estimated by the same dominant motion analysis as in Fig. 9. Here, the fixation plane is a "virtual" plane with an average distance (H=385 m) from the scene to the camera. Fig. 10d and Fig. 11d show the distributions of the mosaic displacements of the real stereo mosaics in Fig. 10 and Fig. 11. It can be found that the Δy displacement distribution of the telephoto stereo mosaics has almost a zero mean, which indicates that the numbers of points above and below the virtual fixation plane are very close. In the depth map of the wide-angle mosaics, more points on tree canopies can be seen. For both cases, most of the pixels have displacements within -10.0 pixels to +10.0 pixels. Using Eq. (2) we can estimate that the range of depth variations of the forest scene (from the fixation plane) is from -24.0 m (tree canopy) to 24.0 m (the ground).

Fig. 12 and Fig. 13 show close-up windows of the stereo mosaics and the depth maps for both telephoto and wide-angle cameras. By comparison, the telephoto stereo mosaics have much better spatial resolutions of the trees and the ground, and have rather similar appearance in the left and right views. In contrast, the left and right wide angle stereo mosaics have much large differences in illumination and occlusion, as well as much lower spatial resolution. The large illumination differences in the wide-angle video are due to the sunlight direction that always made the bottom part of a frame brighter (and sometime oversaturated) than the top part (Fig. 13d). From the experimental results, we can see that better 3D results are obtained from the telephoto stereo mosaics than from the wide-angle stereo mosaics.

6. Conclusions

In the proposed stereo mosaicing approach for large-scale 3D scene modeling, the computation of "match" is efficiently distributed in three steps: camera pose estimation, image mosaicing and 3D reconstruction. In estimating camera poses, only sparse tie points widely distributed in the two images are needed. In generating stereo mosaics, matches are only performed for view interpolation between small overlapping regions of successive frames. In using stereo mosaics for 3D recovery, matches are only carried out between the two final mosaics, which is equivalent to finding a matching frame for every point in one of the mosaics with a fixed disparity.

In terms of depth recovery accuracy, parallel-perspective stereo mosaics provide adaptive baselines and fixed

disparity. We have obtained several important conclusions. Ray interpolation between two successive views is actually very similar to image rectification, thus the accuracy of two-step match mechanism (mosaicing + stereo match) for 3D recovery from stereo mosaics is comparable to that of a perspective stereo with the same adaptive/optimal baseline configurations. We also show that the ray interpolation approach works equally well for both dense and sparse image sequences in terms of accuracy in depth estimation. Finally, given the number of pixels in the original frames, the errors of depth reconstruction is somewhat related to the focal length (and the image resolution) of the camera that captures the video frames. Although further study is needed to investigate what is the best focal length for a certain spatial relation of the camera and the terrain, it seems that stereo mosaics using a telephoto lens (with narrower FOV, higher image resolution and less perspective distortion) gives better 3D reconstruction results than those of a wide angle lens. In fact we can extract multiple (>2) mosaics with small viewing angle differences between each pair of nearby mosaics [18]. Multi-disparity stereo mosaics could be a natural solution for the problem of matching across large oblique viewing angles.

Acknowledgements

This work is partially supported by NSF EIA-9726401, and NSF CNPq EIA9970046. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

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

[2]. H. Ishiguro, M. Yamamoto, and Tsuji, Omni-directional stereo for making global map, *ICCV'90*, 540-547.

[3]. S. Peleg, M. Ben-Ezra, Stereo panorama with a single camera, *CVPR'99*: 395-401

[4]. H. Shum and R, Szeliski, Stereo reconstruction from multiperspective panoramas, *Proc. IEEE ICCV99*, 14-21, 1999.

[5]. Z. Zhu, A. R. Hanson, H. Schultz, F. Stolle, E. M. Riseman, Stereo mosaics from a moving video camera for environmental monitoring, *Int. Workshop on Digital and Computational Video*, 1999, Tampa, Florida, pp 45-54.

[6]. Z. Zhu, E. M. Riseman, A. R. Hanson, Theory and practice in making seamless stereo mosaics from airborne video, *Technical Report #01-01*, CS Dept., UMass-Amherst, Jan. 2001 (http://www.cs.umass.edu/~zhu/UM-CS-2001-001.pdf).

[7]. Z. Zhu, E. M. Riseman, A. R. Hanson, Parallel-perspective stereo mosaics, In *ICCV'01*, Vancouver, Canada, July 2001.

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

[9]. R. Kumar, P. Anandan, M. Irani, J. Bergen and K. Hanna, Representation of scenes from collections of images, In *IEEE Workshop on Presentation of Visual Scenes*, 1995: 10-17.

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

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

[12]. H.S. Sawhney, Simplifying motion and structure analysis using planar parallax and image warping. ICPR'94: 403- 408

[13]. R. Szeliski and S. B. Kang, Direct methods for visual scene reconstruction, In *IEEE Workshop on Presentation of Visual Scenes*, 1995: 26-33

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

[15].Schultz, H., Hanson, A., Riseman, E., Stolle, F., Zhu. Z., A system for real-time generation of geo-referenced terrain models, *SPIE Symposium on Enabling Technologies for Law Enforcement*, Boston MA, Nov 5-8, 2000

[16].C. C. Slama (Ed.), Manual of Photogrammetry, Fourth Edition, *American Society of Photogrammetry*, 1980.

[17].H. Schultz. Terrain Reconstruction from Widely Separated Images, In *SPIE*. Orlando, FL, 1995.

[18].Z. Zhu, PRISM: Parallel ray interpolation for stereo mosaics, http://www.cs.umass.edu/~zhu/StereoMosaic.html

Appendix: Error analysis of ray interpolation

We formulate the problem as follows: Given an accurate point $y_3 = -d_y/2$ in view O_3 that contribute to the right mosaic, we try to find a match point $y_i=+d_y/2$ in a view that contributes to the left mosaic with parallel-perspective projection (Fig. 7). The point y_i is usually from an interpolated view O_i generated from a match point pair y_1 and y_2 in existing consecutive views O_1 and O_2. Suppose the interframe baseline between O_1 and O_2 is S_y, and the baseline between O_1 and O_3 is B_y. First we can write out equations of the depth errors by two view stereos O_1+O_3 and O_2+O_3 , both with almost the same baseline configurations as the adaptive baseline between O_i and O_3. The depth from the pair of stereo views O_1 and O_3 is

$$Z = F\frac{B_y}{y_1 - y_3} = F\frac{B_y}{y_1 + d_y/2}$$

and the depth estimation error is

$$\left|\frac{\partial Z}{\partial y}\right|_{1,3} = \frac{Z}{y_1 + d_y/2} \qquad \text{(a-1)}$$

where y_1 is slightly greater than $d_y/2$ by a small value δy_1:

$$y_1 = d_y/2 + |\delta y_1| \qquad \text{(a-2)}$$

Similarly, The depth from the pair of stereo views O_2 and O_3 is

$$Z = F\frac{B_y - S_y}{y_2 - y_3} = F\frac{B_y - S_y}{y_2 + d_y/2}$$

and the depth estimation error is

$$\left|\frac{\partial Z}{\partial y}\right|_{2,3} = \frac{Z}{y_2 + d_y/2} \qquad \text{(a-3)}$$

where y_2 is slightly smaller than $d_y/2$ by a small value δy_2:

$$y_2 = d_y/2 - |\delta y_2| \qquad \text{(a-4)}$$

Using Eq. (5) we can calculate the relative translational component S_{yi} of the interpolated view O_i to the first view:

$$S_{yi} = \frac{(y_1 - d_y/2)}{y_1 - y_2} S_y \qquad \text{(a-4)}$$

The localization error of the point S_{yi}, which determines the mosaicing accuracy(Eq. (7)), depends on the errors in matching and localizing points y_1 and y_2, which can be derived by differentialing Eq. (a-4) by both y_1 and y_2:

$$|\partial S_{yi}| = S_y[\frac{(y_1 - y_2) - (y_1 - d_y/2)}{(y_1 - y_2)^2}|\partial y_1| + \frac{(y_1 - d_y/2)}{(y_1 - y_2)^2}|\partial y_2|]$$

By assuming that $\partial y_1 = \partial y_2 \equiv \partial y$, and using $Z = FS_y/(y_1 - y_2)$, we can conclude that

$$\left|\frac{\partial S_{yi}}{\partial y}\right| = \frac{Z}{F} \qquad \text{(a-5)}$$

where F is the focal length. It is interesting to note that interpolating accuracy is independent of the magnitude of the interframe motion S_y. For comparison, the depth error from the two consecutive frames O_1+O_2 is

$$\left|\frac{\partial Z}{\partial y}\right|_{1,2} = \frac{Z}{y_1 - y2} = \frac{Z^2}{FS_y} \qquad \text{(a-6)}$$

Apparently smaller interframe motion will introduce much larger depth estimating error (see the pink region in Fig. 7). The depth estimation from stereo mosaics can be written as

$$Z = F\frac{B_y - S_{yi}}{y_i - y_3} = \frac{F}{d_y}(B_y - S_{yi})$$

where we insert $y_3 = -d_y/2$ and $y_i = +d_y/2$. This equation is equivalent to Eq. (2). Using Eq. (a-5), the depth estimation error of stereo mosaics can be expressed by

$$\left|\frac{\partial Z}{\partial y}\right|_{i,3} = \frac{Z}{d_y} \qquad \text{(a-7)}$$

It turns out that the depth error of stereo mosaics is bounded by the errors of two view stereos O_1+O_3 and O_2+O_3, both with almost the same adaptive baselines as the stereo mosaics, i.e.

$$\left|\frac{\partial Z}{\partial y}\right|_{1,3} \le \left|\frac{\partial Z}{\partial y}\right|_{i,3} \le \left|\frac{\partial Z}{\partial y}\right|_{2,3} \qquad \text{(a-8)}$$

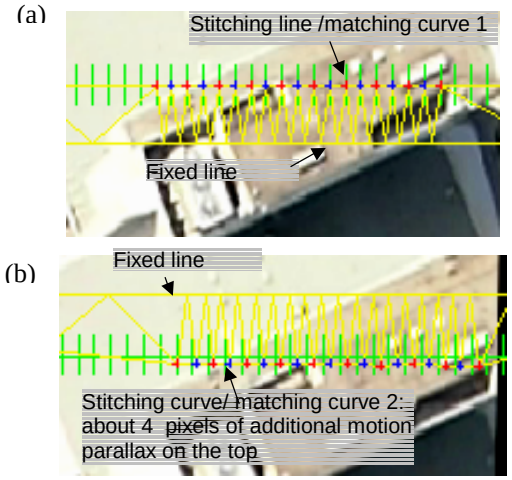


Fig. 5. Examples of local match and triangulation for the left mosaic. Close-up windows of (a)the previous and (b) the current frame. The green crosses show the initially selected points (which are evenly distributed along the ideal stitching line) in the previous frame and its initial matches in the current frame by using the global transformation. The blue and red crosses show the correct match pairs by feature selection and correlation (red matches red, blue matches blue). The fixed lines, stitching lines/curves and the triangulation results are shown as yellow.

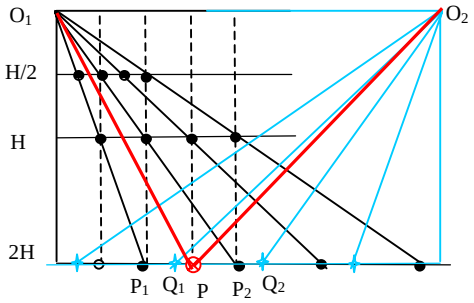


Fig. 6. Resolution changes from perspective projection (solid rays from O_1) to parallel projection (dashed rays). In a simple ray interpolation where each pixel in the mosaics is only from a single frame, resolution remains the same for plane H, reduces to half (gray dots) for plane H/2, and could be two times (the original black dots pus the interpolated white dots) for plane 2H. With image interpolation from more than one frames, image resolution can be better enhanced by sub-pixel interpolation (see text).This figure shows the case where parallel rays are perpendicular to the motion. However, same principle applies for the left and right views of the stereo mosaics.

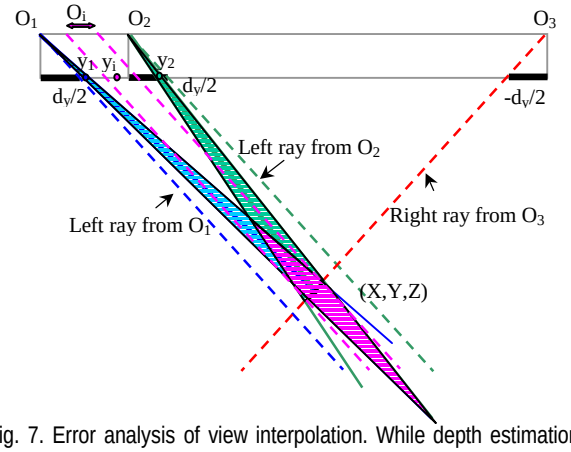


Fig. 7. Error analysis of view interpolation. While depth estimation for two consecutive frames is subject to large error, the localization error of the interpolated ray for stereo mosaics turn out to be very small

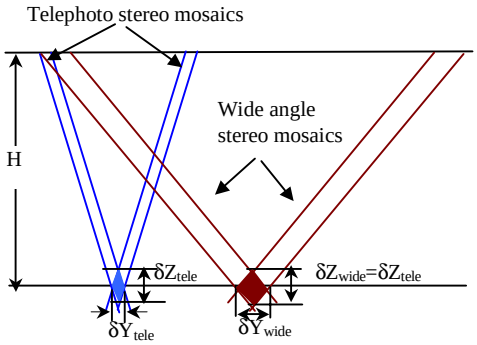


Fig. 8. Depth error versus focal lengths (and fields of view). Note that rays are parallel due to parallel projections, which gives the same depth accuracy with different focal lengths since the larger baseline in the case of the wider field of view compensates the larger footprint on the ground. However, in practice, stereo mosaics from a telephoto camera have better depth accuracy because of better stereo match.

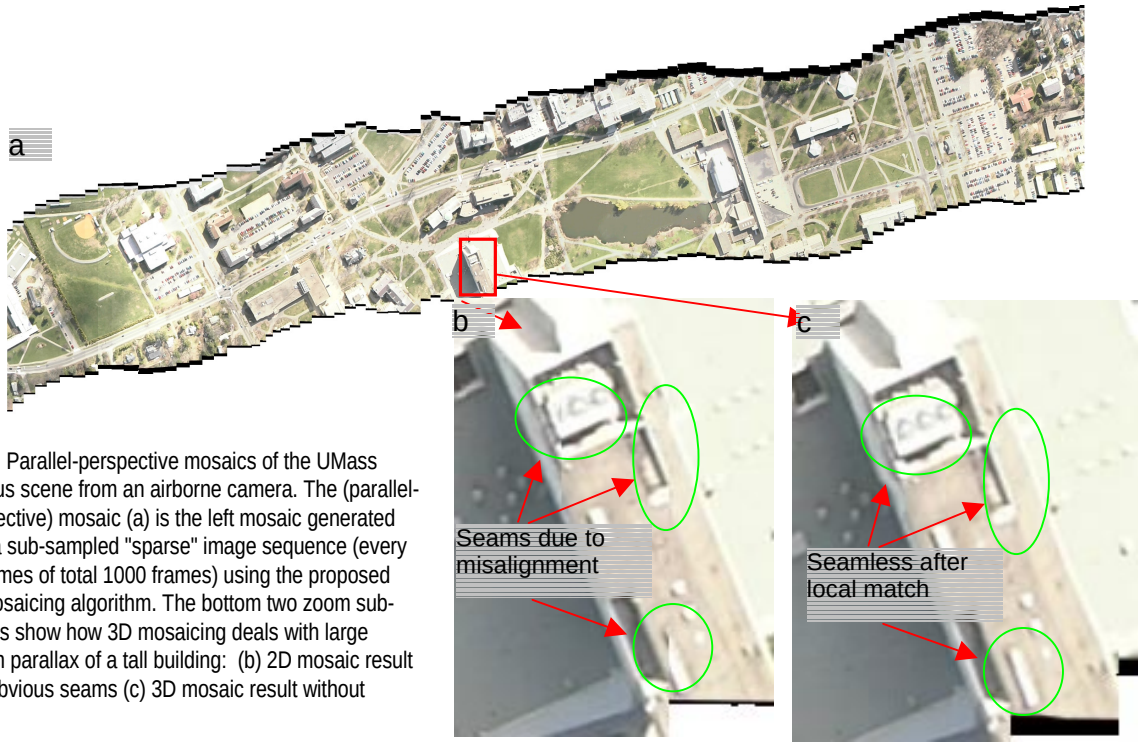
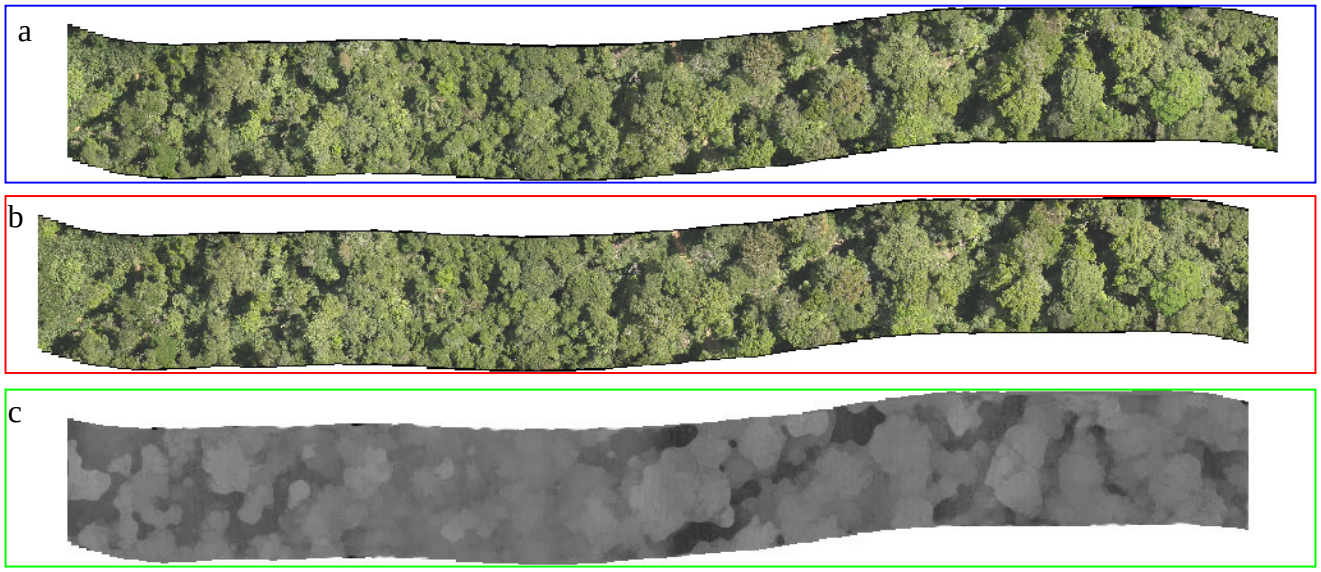


Fig. 9. Parallel-perspective mosaics of the UMass campus scene from an airborne camera. The (parallel-perspective) mosaic (a) is the left mosaic generated from a sub-sampled "sparse" image sequence (every 10 frames of total 1000 frames) using the proposed 3D mosaicing algorithm. The bottom two zoom sub-images 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.



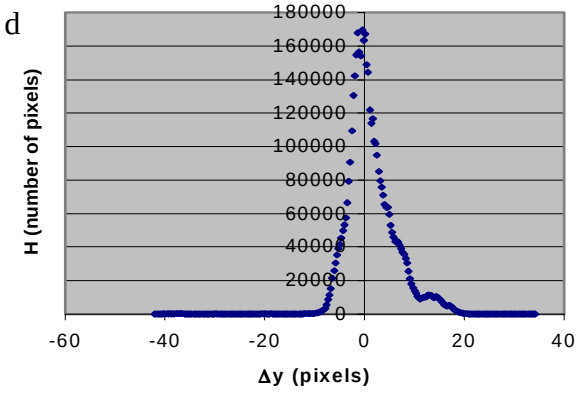


Fig. 10. Stereo mosaics and 3D reconstruction of a 166-frame telephoto video sequence. The size of each of the original mosaics is 7056*944 pixels. (a) left mosaics (b) right mosaics (c) depth map (displacement Δy from 33 to -42 is encoded as brightness from 0 to 255) (d) depth (displacement) distribution (canopies above the fixation plane: negative displacement; points below the fixation plane: positive displacement)

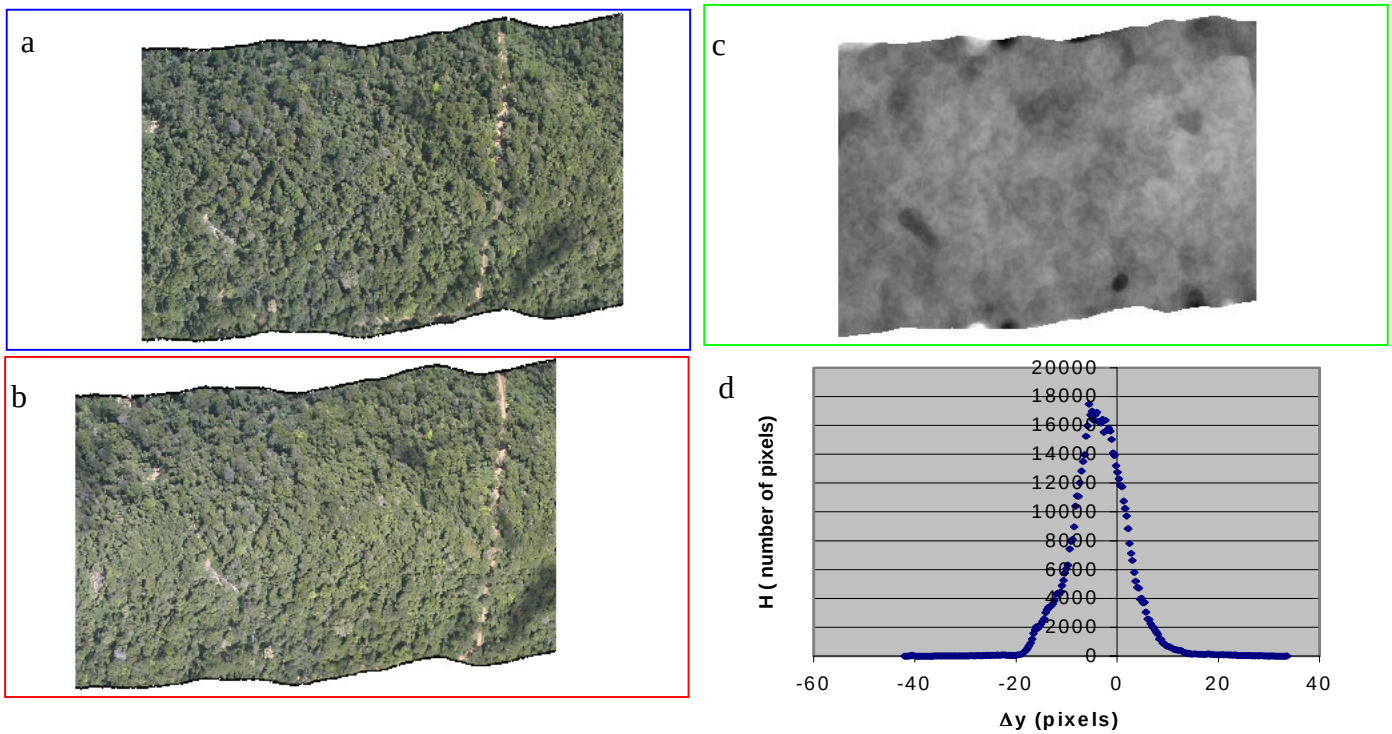


Fig. 11. Stereo mosaics and 3D reconstruction of a 344-frame wide angle video sequence. The size of each of the original mosaics is 1680*832 pixels. (a) left mosaics (b) right mosaics (c) depth map (displacement Δy from 33 to -42 is encoded as brightness from 0 to 255) (d) depth (displacement) distribution(canopies above the fixation plane: negative displacement; points below the fixation plane: positive displacement)

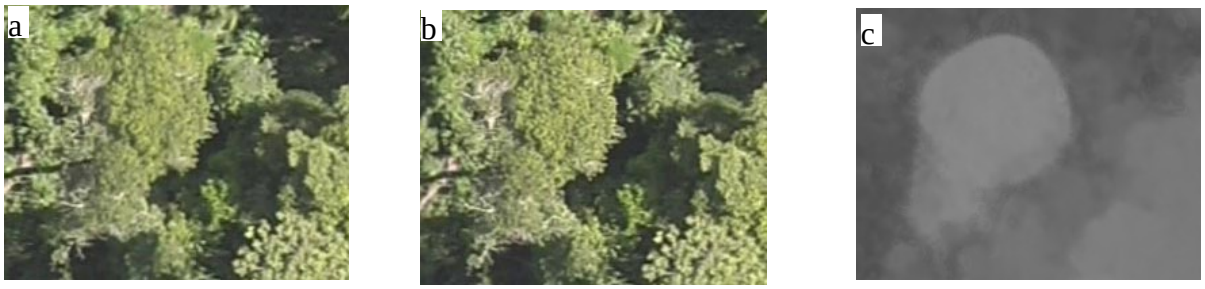


Fig. 12. Zoom regions of the telephoto stereo mosaics in Fig. 10 show high resolution of the trees and good appearance similarity in (a) the left and (b) the right mosaics, and hence produce (c) good 3D results.

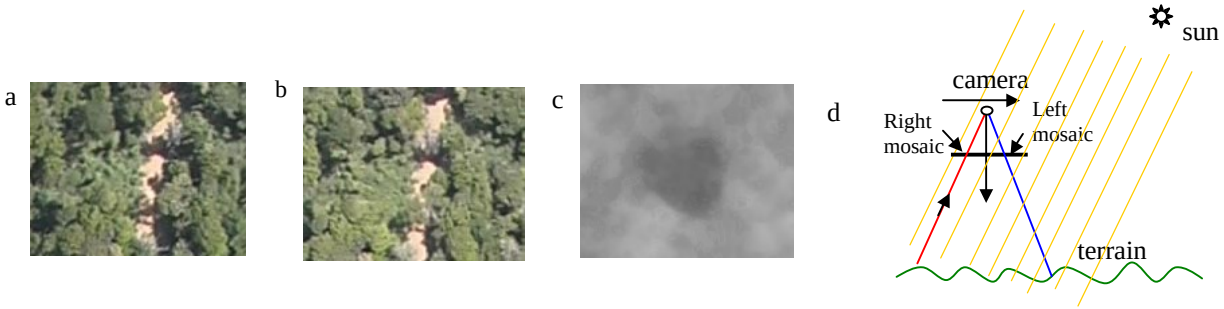


Fig. 13. Zoom regions of the wide angle stereo mosaics in Fig. 11 show much lower resolution of trees and largely different illuminations, perspective distortions and occlusions in (a) the left and (b) the right mosaics, and hence produce (c) less accurate 3D results. (d) The camera moves toward the sun so the bottom part is always brighter (and sometime over-saturated) than the top part of each frame due to the sunlight reflection. It is an unusual case that you take a photo both along and against the direction of light. The right mosaic comes from the bottom part while the left mosaic comes from the top part of video frames.