Chapter 1

GENERATION AND ERROR CHARACTERIZATION OF PARARELL-PERSPECTIVE STEREO MOSAICS FROM REAL VIDEO*

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

Allen R. Hanson, Howard Schultz and Edward M. Riseman
Department of Computer Science
University of Massachusetts at Amherst
Amherst, MA 01003
{hanson, hschultz, riseman}@cs.umass.edu

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, 1998, and from a single off-center rotating camera by Ishiguro, et al, 1990, Peleg & Ben-Ezra, 1999, and by Shum & Szeliski, 1999. In these kinds of stereo mosaics, however, viewpoints - therefore the parallax - are limited to images taken from a very small area. Recently our work at UMass (Zhu, et al, 1999; Zhu, et al, 2001a; Zhu, et al, 2001b) 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

*This work was partially supported by NSF EIA-9726401 and NSF CNPq EIA9970046.

the desired motion. On the contrary, the translation of a camera over a large distance is much harder to achieve in real vision applications such as robot navigation (Zheng & Tsuji, 1992) and environmental monitoring (Kumar, et al, 1995; Schultz, et al, 2000; Zhu, et al, 2001a). In an applications to environmental monitoring, we have previously shown (Zhu, et al, 1999; Zhu, et al, 2001a; Zhu, et al, 2001b) that image mosaicing from a translating camera raises a set of different problems from that of circular projections of a rotating camera. These include the choice of suitable mosaic representations, the generation of seamless image mosaics under a rather general motion with motion parallax, and epipolar geometry associated with multiple viewpoint geometry.

It has been shown independently by others (Chai & Shum, 2000) and by us (Zhu, et al, 1999; Zhu, et al, 2001a) that parallel-perspective is superior to both the conventional perspective stereo and to the recently developed multi-perspective stereo for 3D reconstruction (e.g., in Shum & Szeliski, 1999), in that the adaptive baseline inherent in the parallel-perspective geometry permits depth accuracy independent of absolute depth. However, this conclusion was obtained and verified in an ideal "simulated" case - i.e. enough samples of parallel projection rays from a "virtual camera" with ideal 1D or 2D translational motion were generated from a simulated scene model in order to compare the depth accuracy of parallel versus perspective projections (Chai & Shum, 2000). In the practice of stereo mosaicing from a real video sequence, however, we need to consider the errors in the final mosaics with respect to types of the camera motion, frame rates, focal lengths, and scene depths. The analysis of the error characteristics of 3D reconstruction from parallel-perspective stereo mosaics generated from real video sequences will be the focus of this chapter.

First we will show why an efficient "3D mosaicing" technique is important for accurate 3D reconstruction from stereo mosaics. Obviously simple use of standard 2D mosaicing techniques based on 2D image transformations such as a manifold projection (Peleg & Herman, 1997) cannot generate a seamless mosaic in the presence of large motion parallax, particularly in the case of surfaces that are highly irregular or with significantly different heights. Moreover, perspective distortion causing the geometric seams in the mosaics 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 (Kumar, et al, 1995; Sawhney, 1994; Szeliski & Kang, 1995). 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 *parallel ray interpolation for stereo mosaicing*(PRISM)(Zhu, et al, 2001b) to efficiently convert the sequence of *perspective* images with 6 DOF motion into 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 embody 3D information of the scene with optimal baseline. We have noticed that the view interpolation approach was also suggested for generating seamless 2D mosaics under motion parallax (Rousso, et al, 1998). The authors noted that in order to overcome the parallax problems, intermediate images could be synthetically generated between two original frames, and thus narrower strips used. Our work is different from theirs in two aspects. First, our approach is *direct* and much more efficient. We do not need to generate many new images between each pair of original frames. Instead we directly generate interpolated rays for the parallel-perspective mosaics from only two narrow slices of a pair of successive frames. Second, we proposed to stitch two images in the middle of the two fixed lines corresponding to the two parallel ray directions, so that views of points in the original images are as close as possible to the rays of the final mosaics, thus minimizing the occluding problems due to view changes (Zhu, et al, 2001b).

Here we further examine (1) whether the PRISM process (of image rectification followed by ray interpolation) introduces extra errors in the succeeding steps (e.g. depth recovery); and (2) whether the final "disparity equation" of the stereo mosaics, which exhibits a linear relation between depth and *stereo mosaic displacements* and which does not depend upon focal length *really* means that the depth recovery accuracy is independent of the focal length and absolute depths. Finally, to show the advantages of 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 3D reconstruction from these mosaics are given. Several important conclusions for generating and using stereo mosaics will be made based on our theoretical and experimental analysis.

2. Para-Perspective Stereo Geometry

To illustrate the fundamental geometry of the parallel-perspective (*para-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. Figure 1.1 illustrates the basic idea of the para-perspective stereo mosaics. We can generate two spatio-temporal images by extracting two scanlines of pixels (perpendicular to the motion direction) at the front and rear edges of each frame. The mosaic images thus generated are similar to *parallel-perspective* images captured by a linear pushbroom camera (Gupta & Hartley, 1997), 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 images, 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.

2.1 Parallel-perspective stereo model

Without loss of generality, we assume that two slit windows of two scanline locations have $d_y/2$ offsets to the left and right of the center of the image respectively (Figure 1.1a). 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 (Figure 1.1b; Zhu, et al, 1999; Zhu, et al, 2001a)

$$\begin{aligned} x_l = x_r &= F\frac{X}{Z} \\ y_r &= F\frac{Y}{H} + (\frac{Z}{H} - 1)\frac{d_y}{2} \\ y_l &= F\frac{Y}{H} - (\frac{Z}{H} - 1)\frac{d_y}{2} \end{aligned} \tag{1.1}$$

where F is the focal length of the camera and H is the height of a plane, for example, the average height of the terrain. Figure 1.1 gives the relation between a pair of 2D points, (x_l, y_l) and (x_r, y_r), one from each mosaic, and their corresponding 3D point $P(X, Y, Z)$. It serves a function similar to the classical pin-hole perspective camera model. A generalized model under 3D translation (Zhu, et al, 2001b) extends the parallel- perspective stereo geometry to image sequences with 3D translation, and further with six degrees of freedom (DOF) motion (i.e., rotation + translation). Here we will use the 1D translational motion case to introduce and characterize the basic parallel-perspective stereo geometry. From Equation 1.1 the depth of the point P can be computed

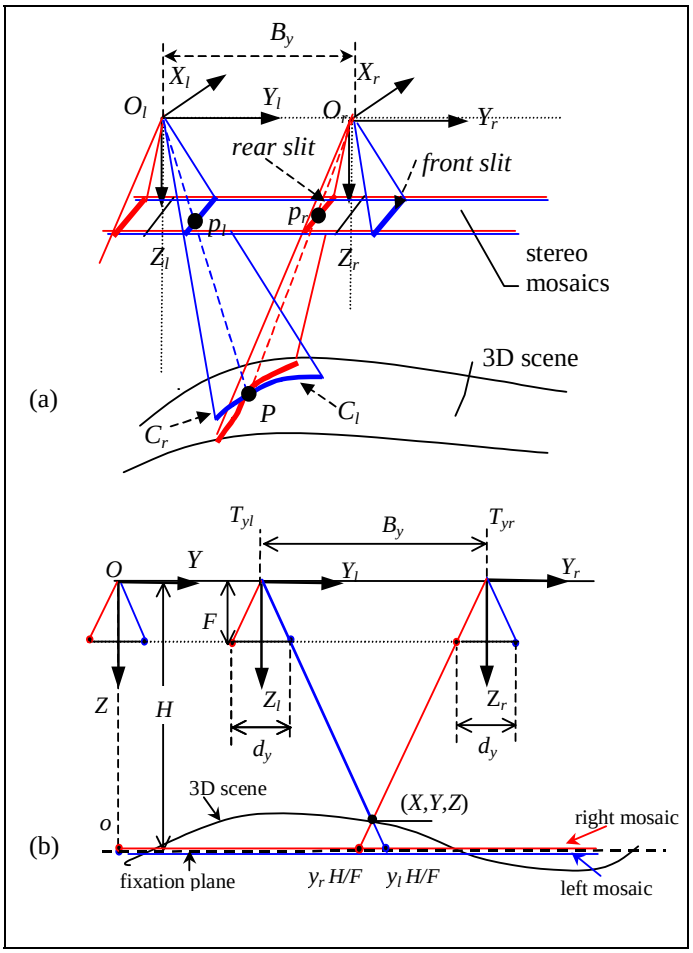


Figure 1.1. Parallel-perspective stereo geometry. (a) Two slit-window imaging geometry - perspective projection in the x direction and parallel projection in the y direction. We assume that two slit-windows have $d_y/2$ offsets to the front and rear of the image from the center respectively. From the viewpoint O_l, a curve C_l in the 3D scene can be seen through the front slit-window, and p_l is the image of a point $P(X, Y, Z)$ on the curve. When the camera moves a certain (baseline) distance B_y in the Y direction to viewpoint O_r, the point P can be seen from the rear-slit window as image p_r, and on a 3D curve C_r. (b). Stereo geometry of parallel projection in the y direction. Both mosaics are built on the fixation plane $Z = H$, but their units are in pixels - each pixel represents H/F world distances - the left mosaic pixel y_l represents a point $y_l H/F$ in the fixation plane, and the right mosaic pixel y_r represents a point $y_r H/F$ in the fixation plane.

as

$$Z = H\frac{b_y}{d_y} = H(1 + \frac{\Delta y}{d_y}) \tag{1.2}$$

where

$$b_y = d_y + \Delta y = F\frac{B_y}{H} \tag{1.3}$$

is the "scaled" version (in pixels) of the baseline B_y, and

$$\Delta y = y_r - y_l \tag{1.4}$$

is the *mosaic displacement* in the stereo mosaics. We use "displacement" instead of "disparity" since it is related to the baseline in a two-view perspective stereo system.

2.2 Parallel-perspective stereo properties

From a mathematical point of view (Equation 1.2), parallel-perspective stereo geometry has the following interesting properties.

(1)*Adaptive baseline configuration* - 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 with varying depths is created. In other words, for any point in the left mosaic, searching for the match point in the right mosaic means (in effect) finding an original frame in which the match pair has a pre-defined disparity (the distance of the two slit windows) and hence has an adaptive baseline depending on the depth of the point (Figure 1.1).

(2) *Fixation plane geometry* - The stereo displacement Δy is a function of the depth variation of the scene around a *fixation plane* with the depth H. The stereo displacements are zeros for all the points in the fixation plane, and negative (positive) for points above (below) the fixation plane.

(3) *focal lengh independency* - In parallel-perspective stereo (Equation 1.2), stereo displacements Δy are independent of the focal length of the camera used to generate the stereo mosaics. Ideally, the image resolutions in the y direction are the same no matter how far away the scene points are. The reason is that due to the parallel projection in the y direction, parallel imaging rays intersect with the 3D scene points instead of converging rays (Figure 1.2). Does this property mean that the depth accuracy is also independent of the focal lengths of the camera? We will answer this question in Section 5

(4) *constant depth resolution* - In a pair of parallel-perspective stereo mosaics, Equation 1.2 tells us that the depth Z is proportional to the im-

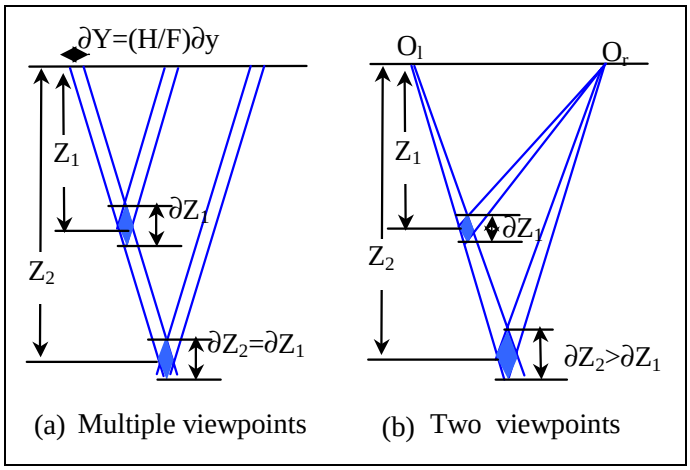


Figure 1.2. Depth resolution of stereo mosaics vs. two-view stereo.

age displacement Δy. Since the displacement Δy is measured in *discrete* mosaicing images, we assume that the image localization resolution is δy pixels (usually $\delta y \leq 1$) in the stereo mosaics, so that $\Delta y = 0, \pm\delta y, \pm 2\delta y$ and so on. In theory, *depth resolution in the parallel-perspective stereo mosaic is a constant value* since the derivative of Z with respect to Δy gives us (Figure 1.2a)

$$\delta Z = \frac{H}{d_y}\delta y = constant \tag{1.5}$$

In contrast, in a two-view perspective stereo (Figure 1.2b), the depth error of a point is proportional to the square of the depth. As a comparison to parallel-perspective stereo, we show the depth errors of a two-view perspective stereo with a long baseline $B_y = (H/F)d_y$, i.e. the disparity for the depth H is d_y. With this stereo configuration, the depth equation of the two view perspective stereo is

$$Z = H\frac{d_y}{\Delta y} \tag{1.6}$$

where Δy is the stereo disparity in the stereo pair. The depth estimation error of the perspective stereo can be derived as

$$\delta Z = \frac{Z^2}{H d_y} \tag{1.7}$$

The observation of constant depth resolution of parallel-perspective stereo has been obtained by us in aerial video mosaicing (Zhu, et al,

1999) and others in image-based modelling (Chai & Shum, 2000). However further investigation shows that it is not the entire story for parallel-perspective stereo mosaics constructed from an image sequence captured by a pin-hole camera. In the following section, we will first briefly summarize what we need to do in order to generate seamless para-perspective stereo mosaics from real videos. Then we will analyze their depth errors in Sections 4.2 and 5.

3. Stereo Mosaicing from Real Video

In real world applications, it is difficult to constrain the motion of cameras on either air or ground vehicles to 1D translation. In addition, extracting one scanline from each frame of a video sequence is not sufficient to generate a uniformly dense mosaic, due to large and possibly varying displacement between each pair of successive frames. Generally speaking, we are facing the difficult problem of structure from motion - that is estimating 3D structure of the scene as well as poses of the moving camera, which requires extensive computation in registration(matching) and reconstruction (3D estimation). In our approach for large-scale 3D scene modeling from real video, the computation of "matching" is efficiently distributed in three stages: 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 sections. Algorithms and discussions in detail can be found in our previous publications (Zhu, et al, 2001a; Zhu, et al, 2001b).

3.1 Pose estimation for image rectification

The stereo mosaicing mechanism can be generalized to the case of 3D translation if the 3D curved motion track of the camera has a dominant translational motion for generating a parallel projection in that direction (Zhu, et al, 2001b). 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. With this generalization, the motion of the camera can further be extended to a 6 DOF motion with some reasonable constraints on the values and rates of changes of motion parameters of a camera

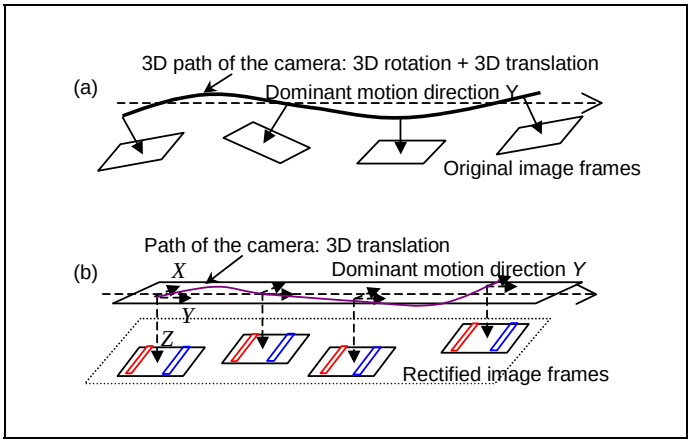


Figure 1.3. Image rectification. (a) Original and (b) rectified image sequence.

(Figure 1.3a; Zhu, et al, 2001a; Zhu, et al, 2001b), which are often satisfied by the motion of 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:

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) (Schultz, et al, 2000) and a bundle adjustment technique (Slama, 1980). The detail is out the scope of our discussion here, but the main point we want to make is that we do not need to carry out dense matching between two successive frames. Instead only sparse tie points widely distributed in the two images are needed to estimate the camera orientations.

Image rectification. A 2D projective transformation is applied to each frame in order to eliminate the rotational components(Figure 1.3b). 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 mosaic pair. 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 for stereo mosaicing

How can we generate seamless mosaics 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 - *Parallel Ray Interpolation for Stereo Mosaicing* (PRISM). 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 (Figure 1.4), 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 and can be neglected (Zhu, et al, 2001a). We take the left mosaic as an example and illustrate the geometry in Figure 1.4. 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. This fixed line of pixels can be directly copied to the corresponding mosaics. 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 (1st) frame of viewpoint (T_x, T_y) and the current (2nd) frame of viewpoint $(T_x + S_x, T_y + S_y)$ (Figure 1.4). We need to interpolate parallel rays between the two *fixed lines* of the 1st and the 2nd frames. For each point (x_l, y_1)(to the right of the 1st fixed line $y_0 = d_y/2$) in the first frame, which will contribute to the left mosaic, we can find a corresponding point (x_2, y_2) (to the left of the 2nd fixed line) in the second frame. We assume that (x_1, y_l) 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 corresponding pair can be computed as

$$T_{yi} = T_y + \frac{y_1 - d_y/2}{y_1 - y_2} S_y \tag{1.8}$$
$$T_{xi} = T_x + \frac{S_x}{S_y}(T_{yi} - T_y)$$

where T_{yi} is calculated in a synthetic IP that passes through the point

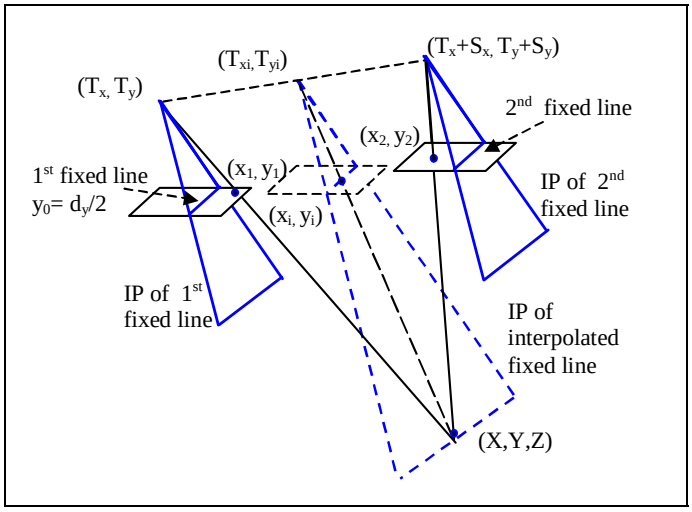


Figure 1.4. Ray interpolation by ray re-projection.

(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 such 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 (of course we can find a better fit for the motion curve rather than this linear fitting). Note that Equation 1.8 also holds for the two fixed lines: when $y_1 = d_y/2$ (the first fixed line), we have $(T_{xi}, T_{yi}) = (T_x, T_y)$; 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 $y_1 - 1 \geq d_y/2 \geq y_2 + 1$, which will be the assumption of our error analysis in the next section. Otherwise a super dense image sequence could generate a pair of stereo mosaics with super-resolution, but this will not be discussed in this paper.

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

$$(x_i, y_i) = [x_1 - \frac{S_x}{S_y}(y_1 - \frac{d_y}{2}), \frac{d_y}{2}] \tag{1.9}$$

and the mosaicing coordinates of the point is

$$(x_m, y_m) = [t_{xi} + x_1 - \frac{S_x}{S_y}(y_1 - \frac{d_y}{2}), t_{yi} + \frac{d_y}{2}] \tag{1.10}$$

where

$$t_{xi} = FT_{xi}/H, \; t_{yi} = FT_{yi}/H \tag{1.11}$$

are the "scaled" translational components of the interpolated view. Note that the interpolated rays are also parallel- perspective, with perspective projection in the x direction and parallel projection 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 rectified image sequences exhibiting 1D, 2D and 3D translation respectively, with a dominant y component(Zhu, et al, 2001b). To be consistent with the ray interpolation stage in this paper, we give the equations assuming 2D translation. With 2D camera 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*

$$\Delta x = b_x(y_l, \Delta y)\frac{\Delta y}{\Delta y + d_y} \tag{1.12}$$

where

$$\Delta x = x_r - x_l,\ \Delta y = y_r - y_l \tag{1.13}$$

are the stereo mosaic displacements in both the x and y directions, and

$$b_x(y_l, \Delta y) = t_{xl}(y_l + d_y + \Delta y) - t_{xl}(y_l) \tag{1.14}$$

is the baseline function of variables y_l and Δy. In Equation 1.14, $t_{xl}(y_l)$ is the "scaled" x translational component (as in Equation 1.3 or Equation 1.11) of the original frames corresponding to column y_l in the left mosaic. Clearly Equation 1.12 shows that *the displacement Δx in the x direction is a nonlinear function of position y_l as well as displacement Δy* (Figure 1.5), which is quite different from the epipolar geometry of a two-view perspective stereo. The reason is that image columns of different y_l in parallel-perspective mosaics are projected from different viewpoints. However, 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. In real applications, if the motion track of the camera does not deviate from the dominant motion direction too far, the epipolar curves are pretty close to horizontal epipolar lines (Zhu, et al, 2001a). In our current experiments, the depth maps of stereo mosaics were obtained by using the Terrest system designed for perspective stereo match (Schultz, 1995) without modification. The Terrest system

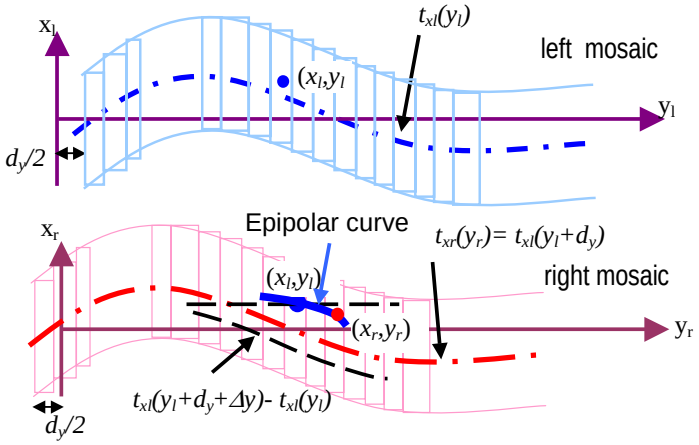


Figure 1.5. Illustration of epipolar curves in stereo mosaics.

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 incorporate the epipolar curve constraints into the search for corresponding 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 needs a three-stage matching process, i.e., interframe global matching to estimate camera poses, interframe local matching for ray interpolation, and the correspondences of the stereo mosaics to generate a depth map. While pose estimation and correspondences are the same for any stereo methods using calibrated images, our question is: does the ray interpolation step introduce extra errors? Second, the final disparity equation of the stereo mosaics does not include any information about the focal length. Does it mean 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 two sections.

4. Error Analysis of Ray Interpolation

4.1 Why ray interpolation?

First, we give the rationale why "3D mosaicing" via the PRISM approach is so important for 3D reconstruction from stereo mosaics by a real example. Figure 1.6 shows the local match and ray interpolation of a successive frame pair of a UMass campus scene, where the inter-frame motion is $(s_x, s_y) = (27, 48)$ pixels, and points on the top of a tall building (the UMass Campus Center Building) 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 (see high resolution mosaics at our web site: Zhu, 2002), the distance between the front and the rear slice windows was selected as $d_y = 192$ 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 $\Delta y = -29$ pixels. Using Equation 1.2 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 $\Delta Z = 45.31$ m. A 4-pixel misalignment in the stereo mosaics will introduce a depth (height) error of $\delta Z = 6.25$ m, even though the 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 applied as we did in our related work (Schultz, 1995).

4.1.1 A fast PRISM algorithm and its generalization. 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 (Figure 1.4). In an effort to reduce the computational complexity in our current implementation, we have designed a fast 3D mosaicing algorithm (Zhu, et al, 2001b) based on the proposed PRISM approach. It only requires matches between a set of point pairs in two successive images

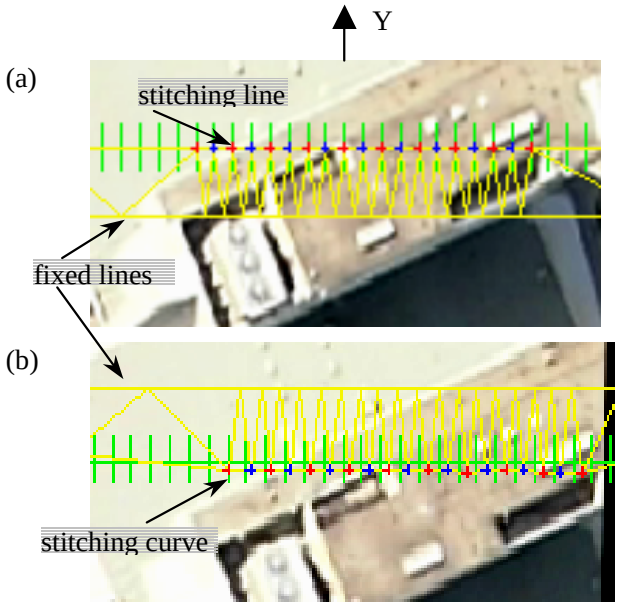


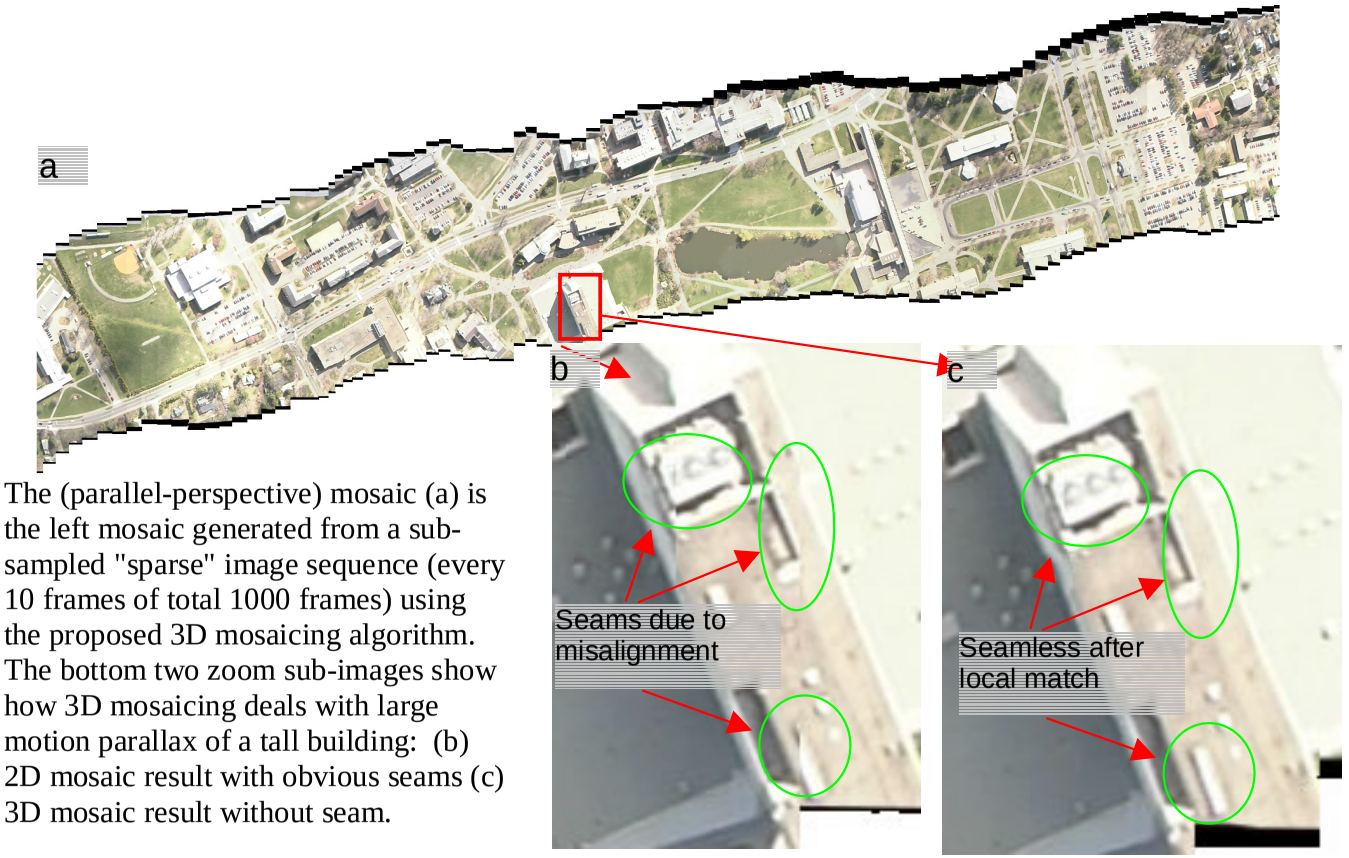
Figure 1.6. Examples of local match and triangulation for the left mosaic. Close-up windows (tiles) of (a) the previous and (b) the current frames. The green (gray in the B-W version) crosses show the initially selected points (which are evenly distributed along the ideal stitching line) in the previous frame and its initial matches in the current frame by using the global transformation. The blue and red (both dark in the B-W version) 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 in yellow(light gray in the B-W version).

around their *stitching line*, which is defined as a virtual line in the middle of the two fixed lines (see Figure 1.6). 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 (Equation:1.10). 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 (e.g., Figure 1.7), but it is good enough when the inter-frame motion is small (e.g., Figure 1.12 to Figure 1.15). 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.

4.1.2 Experimental results. While we are still working on 3D camera orientation estimation using our instrumentation and the bundle adjustments (Schultz, et al, 2000), Figure 1.7 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 building boundaries associated with depth changes in the entire 4160×1536 mosaics at our web site (Zhu, 2002). Since it is hard to see subtle errors in 2D mosaics the size of Figure 1.7a, Figures 1.7b and 1.7c show two close-up windows of the 2D and 3D mosaics side by side for portion of the scene with the tall Campus Center building. In Figure 1.7b 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 (Figure 1.7c) does not exhibit these problems.



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.

Figure 1.7. Parallel-perspective mosaics of the UMass campus scene from an airborne camera.

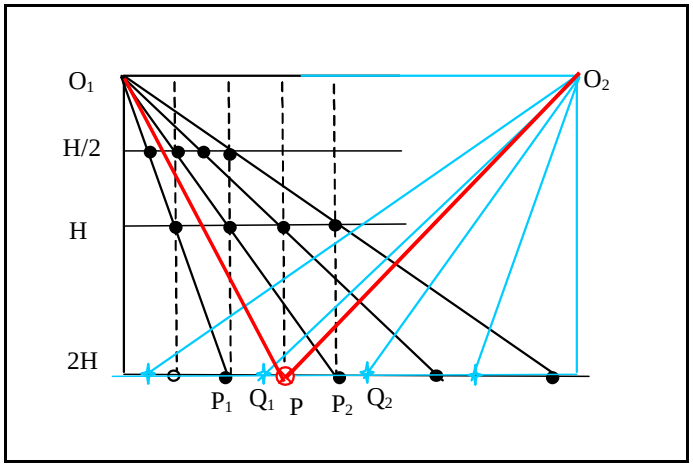


Figure 1.8. Resolution changes from perspective projection (solid dark rays from O_1) to parallel projection (dashed orthogonal rays). In a simple ray interpolation where each pixel in the mosaics is only from a single frame, resolution remains the same for the plane at the distance H, reduces to half for plane $H/2$, and could be two times (the original black dots plus 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 to the left and right oblique views of the stereo mosaics.

4.2 Errors from ray interpolation

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

4.2.1 Question 1: image resolution issues. The answer to the first question is relatively simple: A simple transformation of perspective frames to parallel-perspective mosaics does introduce resolution changes in images (Figure 1.8). 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 smaller 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 higher 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 with an average depth of all the scene points, we can still preserve the image resolution for the nearer points by a super-sampling process (e.g., double or triple the image resolutions).

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 Figure 1.8, assuming that we are performing a sub-pixel match for the ray interpolation. For example, for a sub-pixel point P that lies between two point P_1 and P_2 that are on the grids of the image O_1, we find its match between point Q_1 and Q_2 that are on the grids of the image O_2. Then the value (intensity or color) of the point P can be better interpolated by using the existing points Q_1 and P_2 since they are closer to the point P in space.

4.2.2 Question 2: ray interpolation errors. In order to answer the second question, we formulate the problem as follows (under 1D translation): Given an accurate point $y_3 = -d_y/2$ in the view O_3 that contributes to the right mosaic, we try to find its match point $y_i = +d_y/2$ in a view that contributes to the left mosaic with parallel-perspective projection (Figure 1.9). Note that we express these points in their corresponding frame coordinate systems instead of the mosaicing coordinate system for ease of notation; the mappings from these points to the mosaicing coordinates are straightforward. Usually the point y_i is reprojected from a virtual interpolated view O_i defined by a pair of correspondence points 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. The numerical analysis (Equation 1.A.12; see Appendix for detail) shows that the depth error of the stereo mosaics is proportional to the absolute depth:

$$\delta Z = \frac{Z}{d_y}\delta y \tag{1.15}$$

Comparing Equation 1.15 with Equation 1.5 , 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

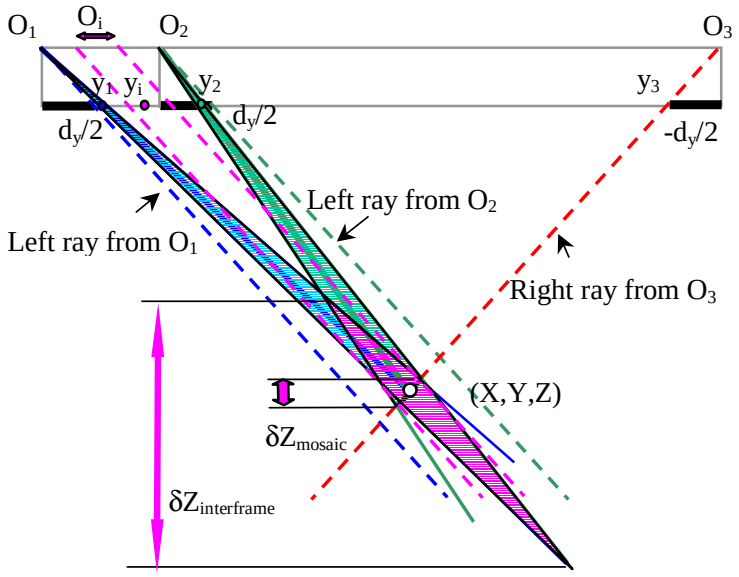


Figure 1.9. Error analysis of ray interpolation. While depth estimation for two consecutive frames is subject to large error ($\delta Z_{interframe}$), the localization error of the interpolated ray for stereo mosaics turn out to be very small and so does the depth error of stereo mosaics (δZ_{mosaic}).

depth H. Therefore, in practice the depth accuracy is not independent of absolute depth. However, this is not necessarily bad news. First we define the *relative depth error* as $|\delta Z/\delta y|$). While the relative depth error, Z/d_y, is larger than the constant number H/d_y when $Z < H$, it is in fact smaller than this constant number H/d_y when $Z > H$ (Figure 1.10). We have not yet incorporated the image resolution changes in stereo mosaics till now, but let us first make the following conclusion:

> *Conclusion 1: In theory, the depth accuracy of parallel-perspective stereo vision is independent of absolute depths; however, in practice, the depth estimation errors of parallel-perspective stereo mosaics are proportional to the absolute depths of scene points.*

How good is this linear error characterization in stereo mosaics? Analysis in the Appendix also shows that even if the depth estimation from two successive views O_1 and O_2 cannot give us good 3D information (Equation 1.A.10), as shown by the large (pink) diamond error region in Figure 1.9, the localization error of the interpolated point (i.e. the left-viewing ray from O_i) is quite small (Equation 1.A.9). Then it turns out that the depth error of stereo mosaics introduced by the ray interpolation 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 configurations as the stereo mosaics (Equation 1.A.13). This observation is summarized in Conclusion 2:

> *Conclusion 2: 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 "optimal" configurations. From the derivation of the localization error for ray interpolation, we have the following interesting conclusion with regard to the inputs of stereo mosaics (see Equation 1.A.9):

> *Conclusion 3: The ray interpolating accuracy is independent of the magnitude of the interframe motion.*

This implies 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.

4.2.3 Summary: depth accuracy versus depth. As a summary, parallel-perspective stereo mosaics provide a stereo geometry with a pre-selected and fixed disparity and adaptive baselines for all the points of different depths, even if the depth resolution is not a constant number. In fact, by incorporating the resolution changes in the mosaics as we have discussed for the first question at the beginning of this section, the depth estimation error of stereo mosaics can be improved as

$$\left|\frac{\delta Z}{\delta y}\right| = \begin{cases} \in (\frac{Z}{d_y}, \frac{H}{d_y}) & \text{if } Z \leq H \\ \in (\frac{H}{d_y}, \frac{Z}{d_y}) & \text{otherwise} \end{cases} \tag{1.16}$$

where pixel localization error δy is measured in the mosaics rather than in the original frames, as in the derivation of the error characterization in Equation 1.A.10 (and also Equation 1.15). Equation 1.16 states the fact that the relative depth error $|\delta Z/\delta y|$ of the para-perspective stereo mosaics is between a constant number H/d_y and a linear function of the depth, Z/d_y. This implies a possible depth error increase (from Z/d_y to H/d_y) due to a resolution loss (if without super-sampling) when $Z \leq H$, but a depth error decrease (from H/d_y to Z/d_y) thanks to a resolution enhancement (via sub-pixel interpolation) when $Z > H$. This leads to our fourth important conclusion for stereo mosaics:

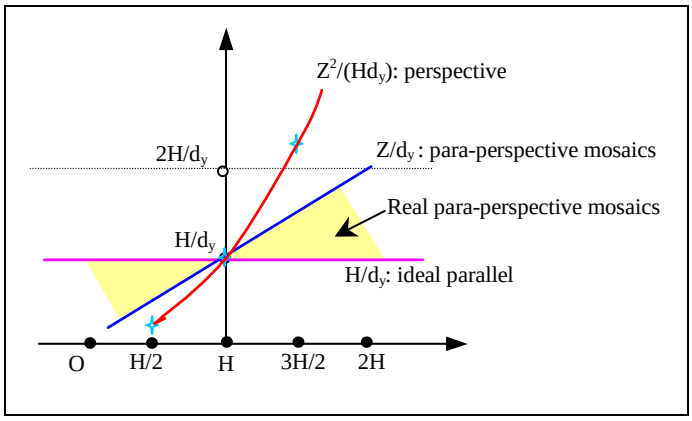


Figure 1.10. Depth errors versus depths. In the ideal case of the parallel perspective stereo, the depth error is a constant number H/d_y, independent of the absolute depths. In para-perspective stereo mosaics, however, the depth errors are a linear function of the absolute depths, Z/d_y. By incorporating super-sampling and the subpixel interpolation in stereo mosaicing, the real depth errors lie in between these two lines (shaded regions). As a comparison, the depth errors of a two view perspective stereo with the baseline $B_y = (H/F)d_y$ (i.e. the disparity for depth H is d_y) are also shown, indicating rapidly increasing errors when $H > Z$.

> *Conclusion 4: Parallel-perspective stereo mosaics provide a stereo geometry with adaptive baselines for all the points of different depths, and the depth error is between a constant function of the average depth of the scene and a linear function of absolute depth of the point. In contrast, the two-view perspective stereo has a fixed baseline, and the depth error is a second order function of absolute depth.*

This conclusion is visualized in Figure 1.10. As a comparison, the depth errors of a two view perspective stereo with a long baseline $B_y = (H/F)d_y$ (Equation 1.7) are also shown, indicating rapidly increasing errors when $H > Z$.

5. Error Analysis versus Focal Lengths

It is commonly 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 (i.e. By in Figure 1.1), 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, we must consider several factors that affect the generation of the stereo mosaics and the correspondences of the stereo mosaics, to see whether this argument is true or not.

5.1 Analyzing focal length and image resolution

First we assume that cameras using for video mosaicing have the same numbers of pixels no matter what the focal lengths (and the fields of view) are. 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 the depth estimate mainly comes from the localization error of the stereo displacement Δy, which consists of two parts: the mosaic generating error δb_1 and the stereo match error δb_2. The first part mainly comes from the baseline estimation error δB (in camera pose estimation), by the following equation:

$$\delta b_1 = \frac{F}{H}\delta B \tag{1.17}$$

where H is the depth of the fixation plane in generating the mosaics. From Equation 1.2 the part of the depth error due to the mosaicing error is

$$\delta Z_1 = \frac{H}{d_y}\delta b_1 = \frac{F}{d_y}\delta B \tag{1.18}$$

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 (Figure 1.11)

$$\delta Y = \frac{H}{F}\delta b_2 \tag{1.19}$$

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

$$\delta Z_2 = \frac{F}{d_y}\delta Y = \frac{H}{d_y}\delta b_2 \tag{1.20}$$

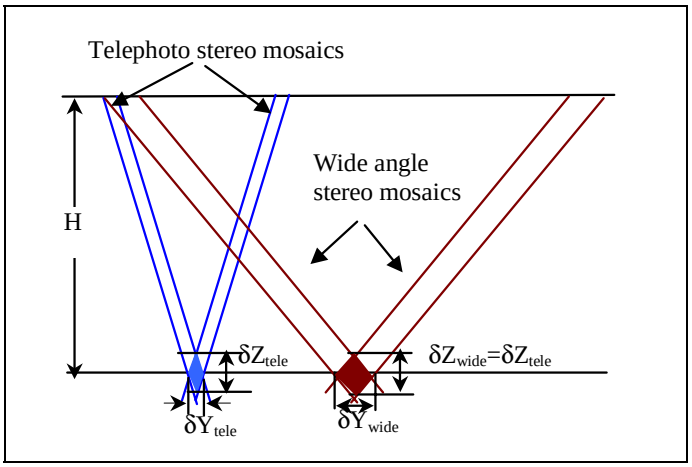


Figure 1.11. Depth error versus focal lengths (and fields of view). Note that rays are parallel due to parallel projections, which gives the same depth accuracy with different focal lengths since the larger baseline in the case of the wider field of view compensates the larger footprint on the ground. However, in practice, stereo mosaics from a telephoto camera have better depth accuracy because of better stereo match.

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 (Figure 1.11). The total depth error is

$$\delta Z = \frac{H}{d_y}(\delta b_1 + \delta b_2) \tag{1.21}$$

or

$$\delta Z = \frac{F}{d_y}\delta B + \frac{H}{d_y}\delta b_2 \tag{1.22}$$

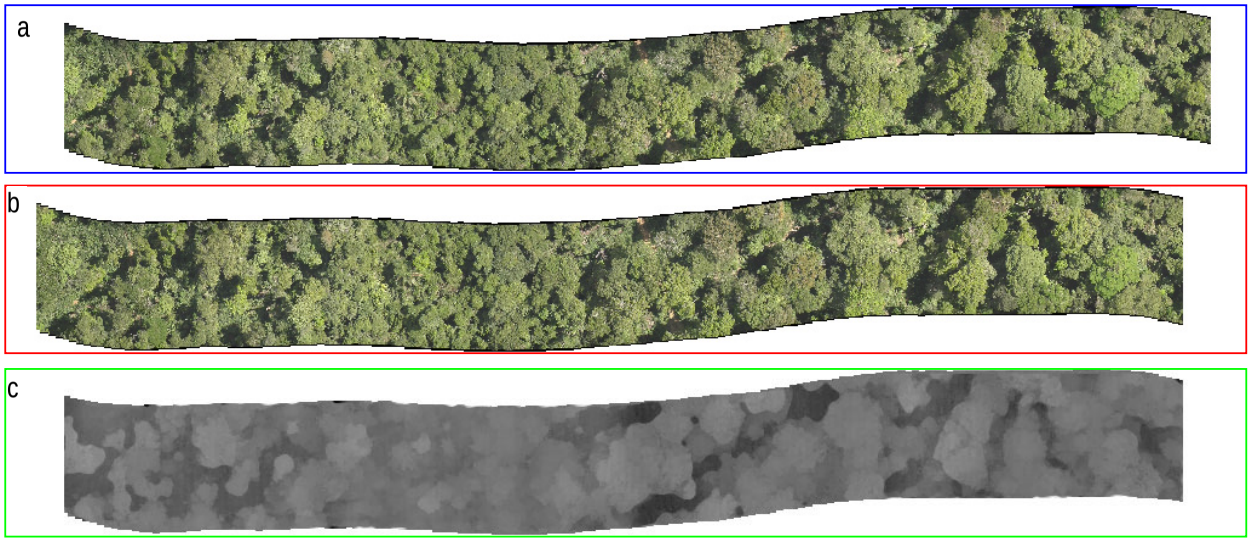
whose differences will be explained in the following:

(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 Equation 1.21 shows that depth error is independent of the focal length (Figure 1.11). 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 larger separated view angles.

(2) If the absolute camera orientation (and hence the baseline B_y in Equation 1.3) is estimated from other independent instrumentation other than image registration, the registration error in generating mosaics will be proportional to the focal length (Equation 1.17). This means the same baseline estimation error will introduce a larger mosaic registration error if a larger focal length is used. In this case, Equation 1.22 should be used to estimate the depth error, which indicates that given the baseline estimate error δB, a larger focal length will introduce a larger error in the first part due to the multiplication of F, but a 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 matching errors 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 an enlargement of the calibration error in the images. Therefore, it might be 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, depth estimation errors of stereo mosaics are independent of the focal length of the camera that generates the stereo mosaics. However, in practice a 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.*



(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 displacements; points below the fixation plane: positive displacements)

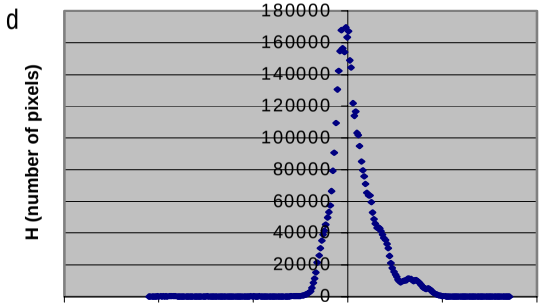


Figure 1.12. Stereo mosaics and 3D reconstruction of a 166-frame telephoto video sequence. The size of each of the original mosaics is 7056*944 pixels.

Figure 1.13. Close-up windows of the telephoto stereo mosaics in Figure 1.12 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.

5.2 Experimental analysis

To validate the above analysis, Figures 1.12 to 1.15 compare real examples of 3D recovery from two sets of stereo mosaics generated from two video sequences. The video sequences were captured simultaneously by a telephoto camera and a wide angle camera of the same forest scene (in the Amazon rain forest for estimation the standing biomass of forests). The instrumentation package (Schultz, et al, 2000) was mounted on a light airplane and consists of a GPS system, an INS system and a profiling pulse laser, as well as the two video cameras (side by side and with different focal lengths). The average height of the airplane was $H = 385$ m, and the distance between the two slit windows for both the telephoto and wide-angle stereo mosaics was selected as $dy = 160$ pixels. 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 $720(x) * 480(y)$, and the camera moved in the y direction (perpendicular to the scanlines of the cameras). 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 (Schultz, 1995). Figure 1.12c and Figure 1.14c show the derived "depth" maps (i.e., the y displacement maps) from the two pairs of telephoto and the 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 pose parameters estimated by the same dominant motion analysis as in Figure 1.7. Here, the fixation plane is a "virtual" plane with an average distance ($H = 385$ m) from the scene to the camera. Figure 1.12d and Figure 1.14d show the distributions of the mosaic displacements of the Δy components of the corresponding stereo mosaics. It can be found that the 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, which is consistent with the parallel stereo mosaic geometry (Equation 1.2), saying that the y displacements in the stereo mosaics are independent of the focal lengths used. Using Equation 1.2 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).

Figure 1.13 and Figure 1.15 show close-up windows of the stereo mosaics and the depth maps for both telephoto and wide-angle cameras respectively. 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 over-saturated) than the top part (Figure 1.16). 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.

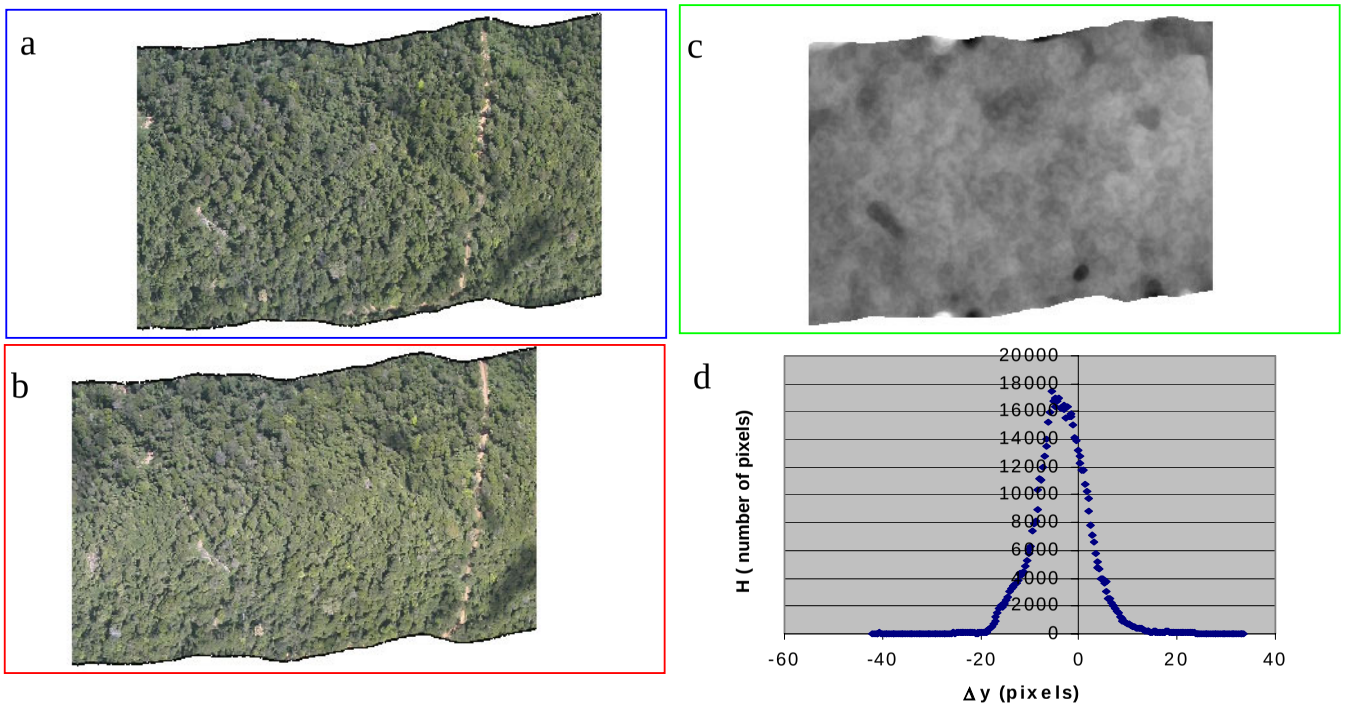


Figure 1.14. Stereo mosaics and 3D reconstruction of a 344-frame wide angle video sequence. The size of each of the original mosaics is 1680*832 pixels. (a) left mosaics (b) right mosaics (c) depth map (displacement Δy from 33 to -42 is encoded as brightness from 0 to 255) (d) depth (displacement) distribution(canopies above the fixation plane: negative displacement; points below the fixation plane: positive displacement).

Figure 1.15. Close-up windows of the wide angle stereo mosaics in Figure 1.14 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.

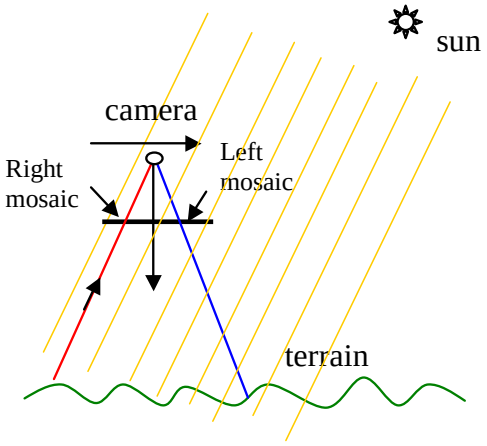


Figure 1.16. The camera moves toward the sun so the bottom part is always brighter (and sometime over- saturated) than the top part of each frame due to the sunlight reflection. It is an unusual case that you take a photo both along and against the direction of light. The right mosaic comes from the bottom part while the left mosaic comes from the top part of video frames.

6. Conclusions

In the proposed stereo mosaicing approach for large-scale 3D scene modeling, the computation of "matching" is efficiently distributed in three stages: 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 ray 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 the correspondence in a matching frame with a fixed disparity for every point in one of the mosaics.

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 three-stage matching mechanism (i.e. matching for poses, mosaicing and correspondences) 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 same 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.

As a future research topic, we want to deal with the issues of stereo mosaicing using wide FOV cameras. For example, we can extract multiple (i.e. more than 2) mosaics with small viewing angle differences between each pair of nearby mosaics (Zhu, 2002) - thus constructing a "multi-disparity" stereo mosaic system, analog to the multi-baseline stereo system (Okutomi & Kanade, 1993)). Multi-disparity stereo mosaics could be a natural solution for the problem of matching across large oblique viewing angles.

Appendix: Error Analysis of Ray Interpolation

We formulate the problem as follows: Given an accurate point $y_3 = -d_y/2$ in the 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 (Figure 1.9). The point y_i is usually from an interpolated view O_i defined by a match point pair y_1 and y_2 in the two existing consecutive views O_1 and O_2. Suppose the interframe baseline between views O_1 and O_2 is S_y, and the baseline between views 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 (with respect to depth). The depth from the pair of stereo views O_1 and O_3 is

$$Z = F\frac{B_y}{y_1 - y_3} = F\frac{B_y}{y_1 + d_y/2} \tag{1.A.1}$$

and the depth estimation error is

$$\left|\frac{\delta Z}{\delta y}\right|_{1,3} = \frac{Z}{y_1 + d_y/2} \tag{1.A.2}$$

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

$$y_1 = d_y/2 + |\delta y_1| \tag{1.A.3}$$

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

$$Z = F\frac{B_y - S_y}{y_2 - y_3} = F\frac{B_y - S_y}{y_2 + d_y/2} \tag{1.A.4}$$

and the depth estimation error is

$$\left|\frac{\delta Z}{\delta y}\right|_{2,3} = \frac{Z}{y_2 + d_y/2} \tag{1.A.5}$$

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

$$y_2 = d_y/2 + |\delta y_2| \tag{1.A.6}$$

Using Equation 1.8 we can calculate the translational component S_{yi} of the interpolated view O_i relative to the first view:

$$S_{yi} = \frac{y_1 - d_y/2}{y_1 - y_2} S_y \tag{1.A.7}$$

The localization error of the point S_{yi}, which determines the mosaicing localization accuracy using Equation 1.10, depends on the errors in matching and localizing points y_1 and y_2, which can be derived by computing the derivatives of S_{yi} with respect to both y_1 and y_2:

$$|\delta S_{yi}| = S_y[\frac{(y_1 - y_2) - (y_1 - d_y/2)}{(y_1 - y_2)^2}|\delta y_1| + \frac{y_1 - d_y/2}{(y_1 - y_2)^2}|\delta y_2|] \tag{1.A.8}$$

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

$$\left|\frac{\delta S_{yi}}{\delta y}\right| = \frac{Z}{F} \tag{1.A.9}$$

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

$$\left|\frac{\delta Z}{\delta y}\right|_{1,2} = \frac{Z}{y_1 - y_2} = \frac{Z^2}{FS_y} \tag{1.A.10}$$

Apparently smaller interframe motion S_y will introduce much larger depth estimation error (see the (pink) diamond region in Figure 1.9), and the depth error is proportional to the square of the depth. On the contrary, the depth estimation from stereo mosaics can be written as

$$Z = F\frac{B_y - S_{yi}}{y_i - y_3} = \frac{F}{d_y}(B_y - S_{yi}) \tag{1.A.11}$$

where we insert $y_3 = -d_y/2$ and $y_i = +d_y/2$ in Equation 1.A.11 . This equation is equivalent to Equation 1.2 that is used to calculate depth. Using Equation 1.A.9, the depth estimation error of stereo mosaics can be derived as

$$\left|\frac{\delta Z}{\delta y}\right|_{i,3} = \frac{Z}{d_y} \tag{1.A.12}$$

Comparing Equations 1.A.2, 1.A.3 and 1.A.12, 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.

$$\left|\frac{\delta Z}{\delta y}\right|_{1,3} \leq \left|\frac{\delta Z}{\delta y}\right|_{i,3} \leq \left|\frac{\delta Z}{\delta y}\right|_{2,3} \tag{1.A.13}$$

References

Chai, J., and H. -Y. Shum. (2000). Parallel projections for stereo reconstruction, In Proceedings of *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*: II 493-500.

Gupta, R., and R. Hartley. (1997). Linear pushbroom cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9): 963-975

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

Ishiguro, H., M. Yamamoto and S. Tsuji. (1990). Omni- directional stereo for making global map. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'90)*, 540-547.

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

Okutomi, M. and T. Kanade. (1993). A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4): 353-363.

Peleg, S., and J. Herman. (1997). Panoramic Mosaics by Manifold Projection. In Proceedings of *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*: 338-343.

Peleg, S., and M. Ben-Ezra. (1999). Stereo panorama with a single camera, In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)*: 395-401

Rousso, B., S. Peleg, I. Finci and A. Rav-Acha. (1998). Universal mosaicing using pipe projection, In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'98)*, 945-952

Sawhney, H.S. (1994). Simplifying motion and structure analysis using planar parallax and image warping. In Proceedings of *the IAPR International Conference on Pattern Recognition (ICPR'94)*: 403- 408

Schultz, H. (1995). Terrain reconstruction from widely separated images, In *SPIE*. Orlando, FL.

Schultz, H., A. R. Hanson, E. M. Riseman, F. Stolle and Z. Zhu. (2000). A system for real-time generation of geo-referenced terrain models. In Proceedings of *the SPIE Symposium on Enabling Technologies for Law Enforcement*, Boston MA, Nov 5-8.

Shum, H.-Y., and R. Szeliski. (1999). Stereo reconstruction from multiperspective panoramas. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV99)*, 14-21.

Slama, C. C. (Ed.). (1980). Manual of Photogrammetry. Fourth Edition, American Society of Photogrammetry.

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

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

Zhu, Z., A. R. Hanson, H. Schultz, F. Stolle and E. M. Riseman. (1999). Stereo mosaics from a moving video camera for environmental monitoring, In Proceedings of *the first International Workshop on Digital and Computational Video*, Tampa, Florida, 45-54.

Zhu, Z., E. M. Riseman and A. R. Hanson. (2001a). 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).

Zhu, Z., E. M. Riseman and A. R. Hanson. (2001b). Parallel- perspective stereo mosaics. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'01)*, Vancouver, Canada, July 2001, vol I: 345-352.

Zhu, Z. (2002). *PRISM: Parallel ray interpolation for stereo mosaics.* http://www-cs.engr.ccny.cuny.edu/ zhu/StereoMosaic.html