COPYRIGHT NOTICE

Generalized Parallel-Perspective Stereo Mosaics from Airborne Video

Zhigang Zhu, *Member*, *IEEE*, Allen R. Hanson, *Member*, *IEEE*, and Edward M. Riseman, *Senior Member*, *IEEE*

Abstract—In this paper, we present a new method for automatically and efficiently generating stereoscopic mosaics by seamless registration of images collected by a video camera mounted on an airborne platform. Using a parallel-perspective representation, a pair of geometrically registered stereo mosaics can be precisely constructed under quite general motion. A novel parallel ray interpolation for stereo mosaicing (PRISM) approach is proposed to make stereo mosaics seamless in the presence of obvious motion parallax and for rather arbitrary scenes. Parallel-perspective stereo mosaics generated with the PRISM method have better depth resolution than perspective stereo due to the adaptive baseline geometry. Moreover, unlike previous results showing that parallel-perspective stereo has a constant depth error, we conclude that the depth estimation error of stereo mosaics is in fact a linear function of the absolute depths of a scene. Experimental results on long video sequences are given.

Index Terms—Mosaicing, stereo vision, visual representation, epipolar geometry, image registration, view interpolation, airborne video analysis.

1 Introduction

Recently, 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 and Hung [3]. The generation of stereo mosaics from a single off-center rotating camera was proposed and fully studied by Peleg and Ben-Ezra [13] and Shum and Szeliski [20] for image-based rendering applications. In our previous work for environmental monitoring using aerial video, we proposed to create stereo mosaics from a single camera with dominant translational motion [26]. In fact, the idea of generating stereo panoramas for either an off-center rotating camera or a translating camera can be traced back to the earlier work in robot vision applications by Ishiguro, et al. [6] and Zheng and Tsuji [25]. The attraction of the recent studies on off-center rotating cameras lies in how to make stereo mosaics with good epipolar geometry and high image quality for image-based rendering [13], [14], [19], [20]. However, in stereo mosaics with a rotating camera, the viewers are constrained to rotationally viewing the stereo representations. Translational motion, on the other hand, is the typical prevalent sensor motion during ground vehicle navigation [25], [27] or aerial surveys [8], [26]. In [13], the authors mentioned that the same techniques developed for the stereoscopic circular projection of a rotating camera could be applied to a translating camera, but it turns out that there has been little serious work on this topic. A rotating camera can be easily controlled to achieve the desired circular motion, and to exhibit certain desirable geometric properties. In contrast, the translation of a camera over a large distance is much harder to control, and introduces rather different and often difficult geometric properties.

Clearly, the use of standard 2D mosaicing techniques based on 2D image transformations such as a manifold projection [12] cannot generate seamless mosaics in the presence of obvious motion parallax. Rectified mosaicing methods have been proposed for generating 2D mosaics "without the curl," with a panning camera that is not perfectly horizontal [32], [17] or with a translating camera facing a tilted planar surface [32]. However these methods are all based on global parametric transformations between successive frames, and cannot apply to a translating camera viewing surfaces that are highly irregular or with large differences in heights, resulting in significantly different motion parallax. In generating seamless 2D mosaics from a hand-held rotating camera, Shum and Szeliski [21] used a local alignment (deghosting) technique to compensate for the small amount of motion parallax introduced by small translations of the camera. For 2D rectified mosaics under more general motion cases, non-straight stitching curves have been proposed (for example, [9]) to generate seamless mosaics for aerial and satellite images. More recently, Rousso et al. [15] proposed universal mosaicing using a "pipe projection." To deal with motion with parallax, they suggested that a 2D orthogonal projection could be generated by taking a collection of strips, each with a width of one pixel, from interpolated camera views in between the original camera positions, but details were not provided. Moreover, for stereo mosaics, an accurate mathematical model is required for precise 3D reconstruction. Kumar et al. [8] dealt with the geo-registration problem by utilizing an available geo-registered aerial image with broader coverage, as well as an accompanying coregistered digital elevation map. In more general cases for generating image mosaics with parallax, several techniques have been proposed to explicitly estimate the camera motion and residual parallax by recovering a projective depth value for each pixel [7], [16], [24]. A

- *Z. Zhu is with the Department of Computer Science, City College of New York, Convent Ave. and 138th Street, New York, NY 10031. E-mail: zhu@cs.ccny.cuny.edu.*
- *A.R. Hanson and E.M. Riseman are with the Department of Computer Science, University of Massachusetts at Amherst, Amherst, MA 01003. E-mail: {hanson, riseman}@cs.umass.edu.*

Manuscript received 25 July 2002; revised 27 Apr. 2003; accepted 18 Sept. 2003. Recommended for acceptance by M. Irani.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117016.

general yet efficient approach and a suitable representation is highly desired for generating seamless stereo mosaics for long image sequences under obvious motion parallax, preferably before 3D reconstruction.

Another interesting issue is how well the parallel-perspective stereo behaves in terms of depth resolution. It has been shown independently by others [1] and by us [26], [29] that parallel-perspective stereo is superior to both conventional perspective stereo and to the recently developed multi-perspective stereo with concentric mosaics for 3D reconstruction (e.g., in [20]), in that the adaptive baseline inherent in the parallel-perspective geometry permits depth accuracy *independent of absolute depth*. But, this conclusion was obtained and verified under ideal conditions (e.g., in the study of [1]). In practice, however, a serious consideration of stereo mosaicing from a real video sequence is the degree of error in the final mosaics when using real sensors. Real video cameras have central projection, undergo a more general motion, are subject to limited frame rates, and view scenes with depth changes.

As a real application of our work, our interdisciplinary NSF environmental monitoring project aims at developing techniques for estimating the standing biomass of forests, monitoring land use changes, habitat destruction, etc., using high resolution low-altitude video sequences [23], [26]. An instrumentation package, mounted on a small airplane, consists of two digital video cameras (telephoto and wide-angle), a Global Positioning System (GPS), an Inertial Navigation System (INS), and a profiling pulse laser [26]. The previous manual approach [23] used by our forestry experts utilized only a fraction of the available data due to the labor involved in manual interpretation of the large amount of video data. For example, recent projects in Bolivia involved more than 20 hours of video over 600 sites, and in Brazil over 120 hours (10 terra bytes), which is prohibitive if the video is interpreted manually. A more compact representation and more flexible interactive 3D visualization interface are clearly necessary in such aerial video applications; in fact, for many applications dealing with large-scale natural or urban scenes, extending the field of view (FOV) of a 2D image, and then introducing the third dimension of depth would be of great utility. Video surveillance [8], environmental monitoring [26], [29], image-based rendering [13], [20], compact video representation [4], [5], and robot navigation [25], [27] are just a few examples of the applications that would benefit from an extended and 3D-enhanced image-based representation.

In this paper, we will address the problem of creating seamless and geometrically registered 3D mosaics from a moving camera, undergoing a rather general motion and allowing viewpoints to change over a large distance. There are three significant contributions in this paper. First, a precise mathematical model of generalized parallel-perspective stereo is proposed, which not only supports seamless mosaicing under quite general motion, but also captures inherent 3D information of the scene in a pair of stereo mosaics. Second, we propose a novel technique called PRISM (*parallel ray interpolation for stereo mosaicing*) to efficiently convert the sequence of perspective images with large motion parallax into the parallel-perspective stereo mosaics. We note that the PRISM approach can be generalized to mosaics with other types of projections (such as circular projection and full parallel projection). Third, we further examine 1) whether the PRISM process (of image rectification followed by ray interpolation) introduces additional 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 depths and stereo mosaic displacements, *really* means that the recovered depth accuracy is independent of absolute depth. Results for mosaic construction from aerial video data of real scenes are shown and 3D reconstructions from these mosaics are given.

This paper is organized as follows: Section 2 gives the representation of generalized parallel-perspective stereo and its epipolar curve geometry under 3D translation. In Section 3, we discuss how image sequences with rather arbitrary, but dominant translational motion (i.e., constrained 6 DOF), can be used as input to develop stereo mosaics. In Section 4, we propose the novel ray interpolation approach, PRISM, to general stereo mosaic from video with obvious motion parallax under translational motion. Section 5 gives a thorough error analysis of ray interpolation in stereo mosaicing. Several important conclusions are made regarding the conditions for generating effective stereo mosaics by the PRISM approach and subsequent 3D reconstruction. Finally, we summarize the main points of this paper and discuss directions of future work.

2 Generalized Parallel-Perspective Stereo

The basic idea of the parallel-perspective stereo can be explained as the following under 1D translation [26], [1]. Assume the camera motion is an ideal 1D translation, the optical axis is perpendicular to the motion, and the frames are dense enough. Then, we can generate two spatio-temporal *mosaic* images by extracting two scanlines of pixels (perpendicular to the motion) at the front and rear edges of each frame in motion (Fig. 1a). Each mosaic image thus generated is similar to a parallel-perspective image generated by a linear pushbroom camera [2], which has perspective projection in the direction perpendicular to the motion and parallel projection in the motion direction. In addition, these mosaics are obtained from two different oblique viewing angles (of a single camera's field of view), so that a stereo pair of left and right mosaics captures the inherent 3D information.

To cope with the real motion of an airborne camera, we will generalize (next section) the stereo mosaicing mechanism to deal with constrained 6 DOF motion—a rather general motion with a dominant translation motion direction. Since rotation effects could be removed by image rectification, here we will first show how to represent stereo mosaics under 3D translation (i.e., without camera rotation). We assume that the 3D curve collecting the moving viewpoints has a dominant translational motion (e.g., the Y direction in Fig. 1b) so that a parallel projection can be generated in that direction. Under 3D translation, parallel stereo mosaics can be generated in the same way as in the case of 1D translation. The main difference is that viewpoints of the mosaics form a 3D curve instead of a 1D straight line.

2.1 Mathematical Model

Without loss of generality, we assume that two horizontal 1D-scanline slit windows (the rear slit and the front slit) have $d_y/2$ offsets to the left and right of the center of the image, respectively (Fig. 1a). The stereo mosaics are formed by the following two steps: *scaling and then translating*. We define the *scaled* vector of a camera position $\boldsymbol{T} = (T_x, T_y, T_z)^t$ (related to a common reference frame—the first frame in Fig. 1) as $\boldsymbol{t} = (t_x, t_y, t_z)^t = F\boldsymbol{T}/H$ in the mosaicing coordinates, where F is the focal length of the camera and H is the height of a *fixation*

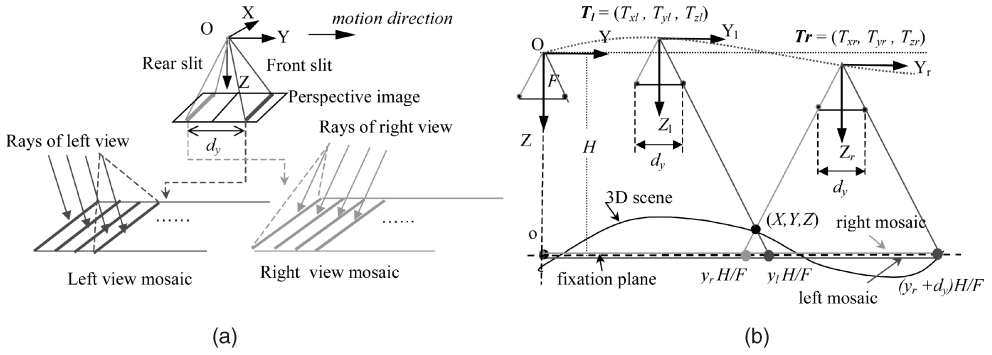


Fig. 1. Parallel-perspective stereo geometry. (a) Illustration of dominant translational motion direction, two slit windows and two parallel-perspective mosaics. (b)The stereo geometry of the generalized parallel-perspective projection under 3D translation (the X axis and the slit windows are perpendicular to the plane of the figure).

plane (e.g., average height of the terrain) on which one pixel in the y direction of the mosaics corresponds to H/F world distance. Then, from the frame in camera position (t_x, t_y, t_z), the front slit will be *translated* to $(t_x, t_y + d_y/2)$ in the "left eye" mosaic, while the rear slit will be *translated* to $(t_x, t_y - d_y/2)$ in the "right eye" mosaic; therefore, both mosaics share the same origin o. The above treatment (namely scaling and translating) has two benefits. By scaling, an appropriate parallel sampling in the y direction is maintained; therefore, a good aspect ratio of the mosaics is kept in the x and the y directions. This is the best choice for parallel sampling under parallel-perspective projection, especially for object points close to the fixation plane. By translating, appearance distortion of the mosaics is minimized (especially in the X direction) since the image slits are placed according to the camera locations of the corresponding image frames.

Suppose the corresponding pair of the 2D points (one from each mosaic), (x_l, y_l) and (x_r, y_r), of a 3D point (X, Y, Z), is generated from original frames in the camera positions (T_{xl}, T_{yl}, T_{zl}) and (T_{xr}, T_{yr}, T_{zr}), respectively. The mathematical model of the *generalized parallel-perspective stereo mosaics* can be represented by the following equations

$$\begin{cases} (x_l, y_l) = \left(F\frac{X-T_{xl}}{Z-T_{zl}} + F\frac{T_{xl}}{H}, F\frac{Y}{H} - \left(\frac{Z-T_{zl}}{H} - 1\right)\frac{d_y}{2} \right) \\ (x_r, y_r) = \left(F\frac{X-T_{xr}}{Z-T_{zr}} + F\frac{T_{xr}}{H}, F\frac{Y}{H} + \left(\frac{Z-T_{zr}}{H} - 1\right)\frac{d_y}{2} \right). \end{cases} \quad (1)$$

It should be noted that generation of stereo mosaics requires only knowledge of the camera pose information, but not the 3D structure of the scene. Under 3D translation, the image scales (especially in the x direction) of the same scene regions in the left and right mosaics could be different due to different Z translational components T_{zl} and T_{zr} in (1). However, when the translation in the Z direction is very small compared to the height H, as in our aerial video applications, the scale difference of the same regions in the left and the right mosaics of the stereo pair are small, which aids stereo matching (both by computers for 3D reconstruction and by humans during stereo viewing).

2.2 Disparity, Baseline, and Epipolar Geometry

Because of the way the stereo mosaics are generated, the viewpoints of both are on the same smooth 3D motion track. The "scaled" camera position $\boldsymbol{t_r}$ corresponding to column y in the right mosaic is exactly the camera position $\boldsymbol{t_l}$ corresponding to column $y + d_y$ in the left mosaic (e.g., the right-view point $y_r H/F$ and the left-view point $(y_r + d_y)H/F$ on the fixation plane in Fig. 1b are from the same view $\boldsymbol{T_r}$), i.e.,

$$\boldsymbol{t_r}(y) = \boldsymbol{t_l}(y + d_y) \quad (2)$$

both of which are only functions of the y coordinate. Let us define the *mosaic displacement* vector between a pair of corresponding points (x_l, y_l) and (x_r, y_r) in the stereo mosaics as

$$(\Delta x, \Delta y) = (x_r - x_l, y_r - y_l). \quad (3)$$

In the general case of 3D translation, the depth of the point can be derived from (1) as

$$Z = H\frac{b_y}{d_y} + \bar{T}_z = H\left(1 + \frac{\Delta y}{d_y}\right) + \bar{T}_z, \quad (4)$$

where

$$b_y = \frac{F}{H}B_y = d_y + \Delta y \quad (5)$$

is defined as the *scaled baseline* in the y direction, which is the scaled version of the baseline ($B_y = T_{yl} - T_{yr}$) between the two viewpoints (T_{xl}, T_{yl}, T_{zl}) and (T_{xr}, T_{yr}, T_{zr}) that generate the point pair, and

$$\bar{T}_z = \frac{T_{zl} + T_{zr}}{2} \quad (6)$$

is defined as the *average camera height deviation* of the stereo point pair related to the origin of the reference frame. Equation (4) represents the depth-baseline-disparity relation of the parallel-perspective stereo: The *disparity* of any corresponding point pair is a *constant* number d_y, but the *mosaic displacement* Δy varies with the depth of the 3D point and represents the *adaptive baseline* b_y for that point. Note that, in the depth equation, we adopt the same notations of disparity and baseline as in the two-view perspective stereo. From (4), we have

$$\Delta y = \frac{d_y}{H}(\Delta Z - \bar{T}_z), \quad (7)$$

which means that under 3D translation, the mosaic displacement (Δy) of a 3D point is proportional to the *relative depth deviation* of the point, which is the real depth deviation from the fixation plane ($\Delta Z = Z - H$), less the average camera height deviation ($\bar{T}_z$). When the motion of the camera is constrained to a 2D translation in the XY plane (i.e., $T_z = 0$,

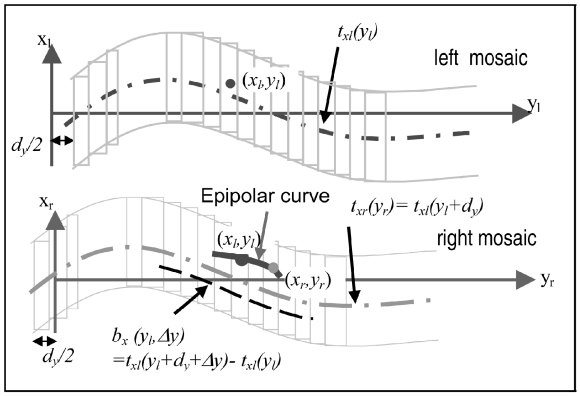


Fig. 2. Epipolar curves in stereo mosaics under 2D translation. Given a left point (x_l, y_l), the baseline function $b_x(y_l, \Delta y)$ is a shifted version of the motion track $t_{xr}(y_r)$ by a constant $t_{xl}(y_l)$, which results in the epipolar curve in the right mosaic using (10).

thus $\bar{T}_z = 0$), the stereo mosaic displacement of a 3D point is exactly proportional to the depth deviation of the point around the fixation plane H. It is also interesting to note that the selection of the two mosaic coordinate systems brings a constant shift d_y to the scaled *baseline* (b_y) and produces the fixation of the stereo mosaics to a horizontal *fixation plane* of an average height H. This is highly desirable for both stereo matching and stereoscopic viewing.

For stereo matching between two such mosaics, we need to know the epipolar geometry under general 3D translation. From (1) and (4), the corresponding point in the right mosaic of any point in the left mosaic will be on an *epipolar curve* determined by the left point and the 3D motion track, i.e.,

$$\Delta x = \frac{b_x \Delta y + b_z d_y (x_l - \frac{t_{xr}+t_{xl}}{2})/F}{\Delta y + d_y + b_z d_y/(2F)}, \tag{8}$$

where $b_x \equiv b_x(y_l, \Delta y) = t_{xl}(y_l + d_y + \Delta y) - t_{xl}(y_l)$ and $b_z \equiv b_z(y_l, \Delta y) = t_{zl}(y_l + d_y + \Delta y) - t_{zl}(y_l)$ are the "scaled" baseline functions in the x and z directions of variables y_l and Δy. Here, we use the same "scaled" notation as the baseline b_y in (5) and apply the relation in (2). Hence, Δx is a nonlinear function of the position (x_l, y_l) as well as the displacement Δy. This is quite different from the epipolar geometry of two-view perspective stereo because: 1) image columns with different y_l coordinates in parallel-perspective mosaics are projected from different viewpoints, which are reflected in the baseline function $b_x(y_l, \Delta y)$, and 2) Δx is also a function of x_l due to the nonzero Z translation and therefore nonzero b_z term in (7). We have the following conclusions for the epipolar geometry of parallel-perspective stereo:

1. *In the general case of 3D translation*, if we know the range of depth deviation (plus an average camera height deviation) from (7), i.e.,

$$\pm \Delta Z_m = \pm(|\Delta Z|_{max} + |\bar{T}_z|_{max}) \tag{9}$$

the search region for the corresponding point in the right mosaic is

$$\Delta y \in \left[-\frac{d_y}{H} \Delta Z_m, +\frac{d_y}{H} \Delta Z_m \right],$$

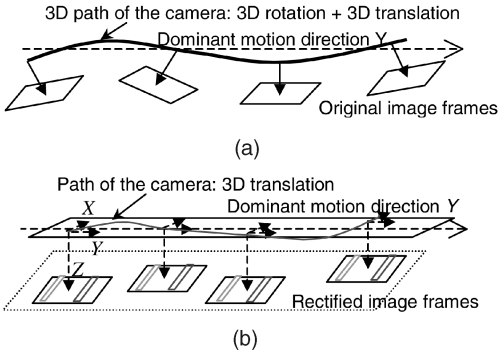


Fig. 3. Image rectification. (a) Original and (b) rectified image sequence.

and along an epipolar curve(using (8)), which is different for every point (x_l, y_l) in general.

2. *In the case of 2D translation* (i.e., $b_z = 0$), the epipolar curve for a given point (x_l, y_l) in the left mosaic passes through the location (x_l, y_l) in the right mosaic (Fig. 2):

$$\Delta x = b_x(y_l, \Delta y) \frac{\Delta y}{\Delta y + d_y}, \tag{10}$$

which implies that the stereo mosaics are aligned for all the points whose depths are H. The same epipolar curve function (of y_l and Δy) is applied to all the points in the left mosaic with the same y_l coordinate.

3. *In the ideal case* where the viewpoints of stereo mosaics lie in a 1D straight line (i.e., $b_x = b_z = 0$), the epipolar curves will turn out to be horizontal lines ($\Delta x = 0$). Therefore, we can apply most of the existing stereo match algorithms for rectified perspective stereo with little modification.

3 Mosaicing under Realistic 6 DOF Motion

This section discusses how to generate stereo mosaics under a more general motion model, with constrained 6 DOF. To generate meaningful and seamless stereo mosaics, we need to impose some constraints on the camera motion (Fig. 3a). First, the motion must have a dominant direction. Second, the angular orientation of the camera is constrained to a range that precludes it turning more than 180 degrees. Third, the rate of change of the angular orientation parameters must be slow enough to allow sufficient overlap of successive images for stereo mosaicing. These constraints are all reasonable and are satisfied by a sensor mounted in a light aircraft with normal turbulence. Within these constraints, the camera can realistically undergo six DOF motion. There are two steps necessary to generate a rectified image sequence that exhibits only (known) 3D translation, and from which we can subsequently generate seamless mosaics:

Step 1: Camera orientation estimation. Using an internally precalibrated camera, the extrinsic camera parameters (camera orientation) can be determined from an aerial instrumentation system (GPS, INS, and a laser profiler) [26] and bundle adjustment techniques [22]. The main point here is that we do not need to carry out a dense match between two successive frames. Instead, only sparse tie points widely distributed in the two images are needed to estimate the camera orientations.

Step 2: Image rectification. An image rotation transformation is applied to each frame in order to eliminate the

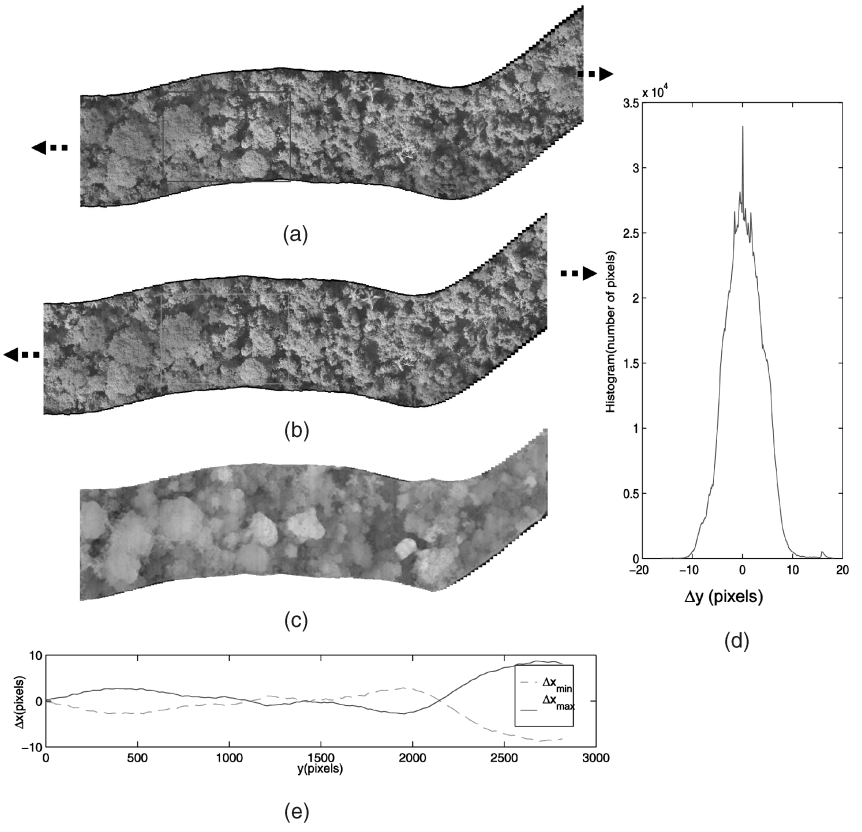


Fig. 4. Stereo mosaics and the epipolar geometry. (a) Left view mosaic. (b) Right view mosaic. (c) The y displacement map: mosaic displacement Δy (proportional to the relative depth ΔZ) is encoded as brightness (brightness is from 0 when $\Delta y = 18.3$ pixels, to 255 when $\Delta y = -16.2$ pixels), so higher elevation (i.e., closer to the camera) is brighter. (d) The histogram of the y displacements. (e) Illustration of searching ranges of the x displacements given a searching range of the y displacements [-10, +10] in the stereo mosaics, using (10).

rotational components (Fig. 3b). In fact, we only need to do this kind of transformation to two narrow slices in each frame that will contribute incrementally to a pair of mosaics. In our motion model, the 3D rotation is represented by a rotation matrix $\boldsymbol{R}$, and the 3D translation is denoted by a vector $\boldsymbol{T} = (T_x, T_y, T_z)^t$. A 3D point $\boldsymbol{X_k} = (X_k, Y_k, Z_k)^T$ with image coordinates $\boldsymbol{u_k} = (u_k, v_k, 1)^t$ at current frame k can be related to its reference coordinates $\boldsymbol{X} = (X, Y, Z)^T$ by $\boldsymbol{X} = \boldsymbol{R_k X_k} + \boldsymbol{T_k}$, where $\boldsymbol{R_k}$ and $\boldsymbol{T_k}$ are the rotation matrix and the translation vector of the kth frame related to the reference frame (e.g., the first frame). In the image rectification stage, a projective transformation $\boldsymbol{A_k}$ is applied to the kth frame of the video using the motion parameters obtained from the camera orientation estimation step:

$$\boldsymbol{u_k^p} \cong \boldsymbol{A_k u_k} \tag{11}$$

with

$$\boldsymbol{A_k} = \boldsymbol{F R_k F^{-1}}, \quad \boldsymbol{F} = \begin{pmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{12}$$

where $\boldsymbol{u_k^p}$ is the reprojected image point of the kth frame, and F is the camera's focal length. The resulting video sequence will be a rectified image sequence as if it was captured by a "virtual" camera undergoing 3D translation (T_x, T_y, T_z). We assume that vehicle's motion is primarily along the Y axis after eliminating the rotation, which implies that the mosaic will be produced along the Y direction.

3.1 A Real Example

While a full calibration of parameters in all camera positions is a very difficult task in a long video sequence and is not the focus of this paper, we point out here that two different practical treatments with near real-time implementations are applied in our experiments [29]: unconstrained image mosaics using a dominant plane fitting technique without camera calibration and geo-registered mosaics with a practical method for camera orientation estimation using the GPS/INS measurements. An underlying assumption in the practical treatments is that, if the translational component in the Z direction is much smaller than the distance itself, we use a constant scaling factor in the interframe motion estimation and image rectification for each frame to compensate for the Z translation. We show a real example of stereo mosaics from a 165-frame video sequence in Fig. 4, collected as part of a project [23], [26] with The Nature Conservancy (TNC) for determining biomass for preservation of a tropical forest in Bolivia. This example shows the creation of stereo mosaics from video of a real world scene, the epipolar geometry, and the 3D reconstruction and stereo viewing properties. Figs. 4a and 4b show a pair of stereo mosaics generated by using an unconstrained image mosaicing approach (referred to as "free mosaicing" in [29]) with the slit-window distance $d_y = 224$ (in pixels; the original image was 720*480). It is obvious from the mosaics that, after the compensation of the small rotation angles and the small Z components by image rectification, the rectified image sequence has significant translations in the x direction, as well as in the dominant y direction.

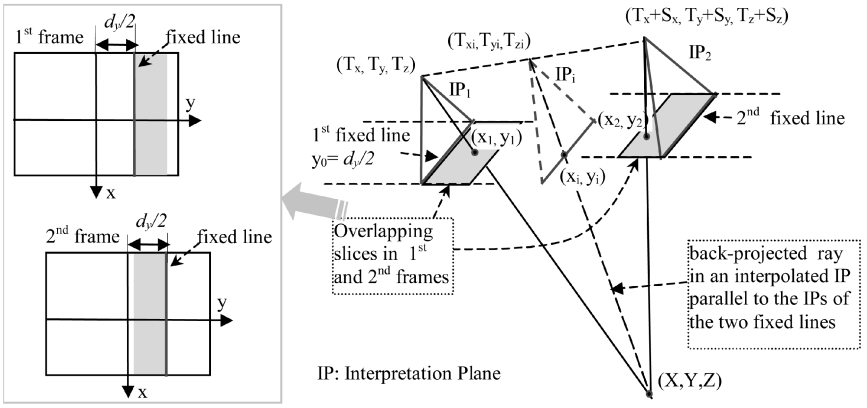


Fig. 5. Ray interpolation by ray reprojection. For any pair of points between the two fixed lines of the two successive frames, a new interpolated ray is back-projected parallel to the interpretation planes of the fixed lines, from the intersection of the two corresponding rays cast from the viewpoints of the two existing frames.

There are two benefits of generating a seamless stereo mosaic pair. First, a human can perceive the 3D scene in a stereo mosaic pair (e.g., using a pair of polarized glasses, or even red/blue-green anaglyph glasses) without any explicit 3D recovery [28], [29]. Since stereo mosaicing can be generated in real-time, this leads to a real-time stereoscopic viewing of the 3D scene. Experience of both forestry experts and laymen have shown that the stereoscopically viewed mosaics of trees are both compelling and vivid to the viewers. Results of high resolution stereo mosaics can be found at our Web sites [28]. Second, after stereo mosaicing, matches are only performed on the stereo mosaics for 3D recovery, not on individual pairs of video frames, resulting in tremendous reduction in storage and computation. Fig. 4c shows the derived y "displacement" map (where displacement Δy is proportional to the relative depth ΔZ) from the pair of parallel-perspective stereo mosaics. This displacement map is obtained using a hierarchical subpixel dense correlation method [18]. Fig. 4d shows the histogram of the y displacements of this pair of stereo mosaics, indicating most of the pixels have displacements within -10.0 pixels to $+10.0$ pixels. Using (10), we derive the search range of the x displacements, $[\Delta x_{min}, \Delta x_{max}]$, at each column of the left mosaic, if the search range of the y displacements in the stereo mosaics is $[-10, +10]$ (Fig. 4e). It can be seen that for the large part of the mosaics, the search ranges in the x direction are within ± 3 pixels, except the tail with large x motion components.

4 A Ray Interpolation Approach for Stereo Mosaicing

After image rectification, we obtain a translational motion sequence with the rotational effect removed. However, the translational sequence exhibits obvious motion parallax. How can we generate seamless mosaics in a computationally effective way from this sequence? The key to our approach lies in the parallel-perspective representation and our novel PRISM (*parallel ray interpolation for stereo mosaicing*) approach [30]. For the left (or right) mosaic, we only need to take a front (or rear) slice of a certain width (determined by the interframe motion) from each frame and perform local registration between the overlapping slices of successive frames. We then directly generate parallel-perspective interpolated rays between two known discrete perspective views for the left (or right) mosaic so that the geometry of the mosaic generated exhibits true parallel projection in the direction of the dominant motion and, therefore, the mosaic is without geometric distortions.

4.1 PRISM: Parallel Ray Interpolation for Stereo Mosaicing

Let us examine the PRISM approach more rigorously in the general case of 3D translation (after image rectification). We take the left mosaic as an example and illustrate the geometry in Fig. 5. First, we define the *fixed line* of the front mosaicing slice in each frame as a scanline that is $d_y/2$ distance from the center of the frame. We use the term "fixed line" to indicate that pixels on that line can be directly copied to the corresponding location in the left mosaic. The width of the slice used for ray interpolation are determined by the camera's locations of both frames and the depths of the points seen by the two frames. An *interpretation plane* (IP) of the fixed line is a plane passing through the nodal point and the fixed line of the frame. By the definition of parallel-perspective stereo mosaics under pure translation, all the IPs of fixed lines for the left mosaic are parallel to each other. Suppose that (S_x, S_y, S_z) is the translational vector of the camera between the previous (1^{st}) frame of viewpoint (T_x, T_y, T_z) and the current (2^{nd}) frame of viewpoint $(T_x + S_x, T_y + S_y, T_z + S_z)$ (Fig. 5). We need to interpolate parallel-perspective rays between the two *fixed lines* of the 1^{st} and the 2^{nd} frames.

For each point (x_1, y_1) (to the right of the first 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 second fixed line) in the second frame. We assume that (x_1, y_1) and (x_2, y_2), represented in their own frame coordinate systems, intersect at a 3D point (X, Y, Z). Then, the desired parallel-reprojected viewpoint (T_{xi}, T_{yi}, T_{zi}) for the corresponding pair can be computed as

$$\begin{aligned} T_{yi} &= T_y + \frac{\left(y_1 - \frac{d_y}{2}\right)(FS_y - y_2 S_z)}{(y_1 - y_2)\left(FS_y - \frac{d_y}{2} S_z\right)} S_y, \\ T_{xi} &= T_x + \frac{S_x}{S_y}(T_{yi} - T_y), \\ T_{zi} &= T_z + \frac{S_z}{S_y}(T_{yi} - T_y), \end{aligned} \tag{13}$$

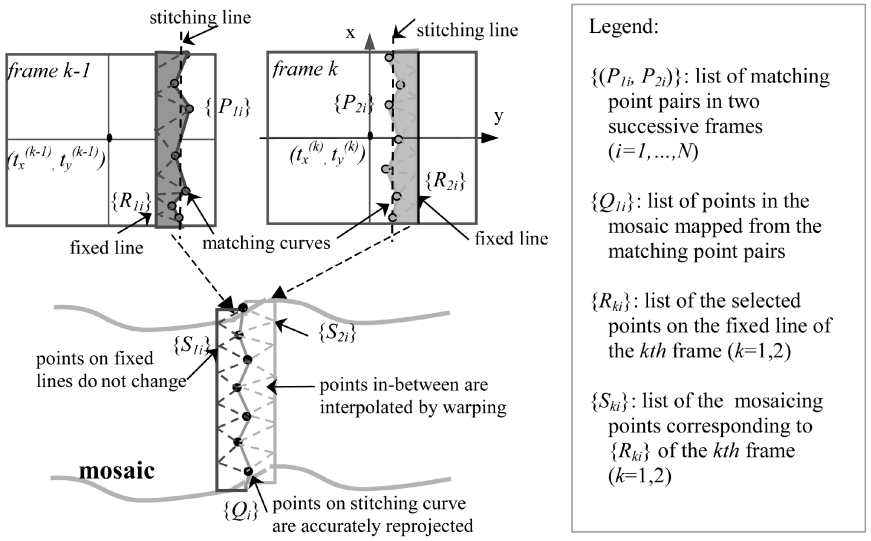


Fig 6. A fast PRISM algorithm: sparse matching, region triangulation, and image warping.

where T_{yi} is calculated in a "virtually" interpolated 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} and T_{zi} are calculated in such a way that all the viewpoints between (T_x, T_y, T_z) and $(T_x + S_x, T_y + S_y, T_z + S_z)$ lie in a straight line. (Of course we can find a better fit for the motion curve rather than this linear local fit.) The reprojected "image" (x_i, y_i) of the point (X, Y, Z) from the interpolated viewpoint (T_{xi}, T_{yi}, T_{zi}) is given by

$$(x_i, y_i) = \left[\frac{x_1 - F\frac{S_x}{Z_i}}{1 - \frac{S_z}{Z_i}}, \frac{d_y}{2}\right], \quad Z_i = \frac{FS_y - \frac{d_y}{2}S_z}{y_1 - \frac{d_y}{2}}, \tag{14}$$

where Z_i is introduced here to simplify the represenation (it is loosely related to depth but is not actually the depth measured from the interpolated viewpoint). Note that the calculation of the x coordinate in the above equation indicates perspective projection in the x direction, and the constant y coordinate $(= d_y/2)$ indicates that the point is on the fixed line of the virtually interpolated view (and, hence, the interpolated projection ray is parallel to the IPs of the fixed lines).

In our aerial video application, the actual motion of the aircraft was mostly in the xy plane; therefore, the 3D translation is reduced to 2D translation (with $T_z = 0$ and, hence, $S_z = 0$) as shown in [30]. In this case, (13) and (14) can be greatly simplified. The items in the two equations that need to be changed are

$$T_{yi} = T_y + \frac{y_1 - d_y/2}{y_1 - y_2}S_y, \quad x_i = x_1 - \frac{S_x}{S_y}\left(y_1 - \frac{d_y}{2}\right). \tag{15}$$

Knowing the interpolated view point (T_{xi}, T_{yi}, T_{zi}) and the point coordinates (x_i, y_i) in the virtually interpolated view, the left mosaicing coordinates (x_l, y_l) of the point can be calculated as

$$(x_l, y_l) = \left(t_{xi} + x_i, t_{yi} + \frac{d_y}{2}\right), \tag{16}$$

where $t_{xi} = FT_{xi}/H$ and $t_{yi} = FT_{yi}/H$ are the "scaled" translational components of the interpolated view.

4.1.1 Generalization and Discussions

The core idea of the PRISM algorithm is ray interpolation, which uses explicit *motion parallax* information between successive video frames to create better mosaics. However, the idea of ray interpolation is not limited to parallel-perspective projection. It can be generalized to other kinds of projection geometry, such as circular projection [13], [14] multiperspective panoramas [20], linear pushbroom cameras [2], full parallel projection [1], etc.

We also note that view interpolation has been suggested by others for generating seamless 2D mosaics under motion parallax [15]. Our work is different from theirs in three aspects, mainly due to explicitly employing the mosaicing geometry, namely, the parallel-perspective stereo representation, in the PRISM algorithm. 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 parallel rays for the parallel-perspective mosaics. Second, we propose to stitch two images in the middle of the two fixed lines to minimize the occlusion problem since views of the points thus selected in the original images are as close as possible to the rays of the final mosaics. Last but not least, we use an accurate geometric model to maintain precise stereo geometry for 3D reconstruction.

4.2 Implementation and Experimental Analysis

4.2.1 A Fast PRISM Implementation

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 and to handle textureless regions, a fast PRISM algorithm has been implemented ([29], [30]). As a summary, the fast PRISM algorithm consists of the following four steps, taking the left mosaic as an example (Fig. 6):

Step 1: Slice determination. Determine the fixed lines in the current frame k and the previous frame $k-1$ by the left slit window distance $d_y/2$, and "ideal" straight stitching lines by their 2D scaled translational parameters $(t_x^{(k)}, t_y^{(k)})$ and $(t_x^{(k-1)}, t_y^{(k-1)})$. The locations of the stitching lines are in the middle of the two fixed lines, i.e, at $d_y/2 + (t_y^{(k)} - t_y^{(k-1)})/2$ in the $(k-1)$th frame and $d_y/2 - (t_y^{(k)} - t_y^{(k-1)})/2$ in the kth

frame. Thus, we have two overlapping slices in the kth and $(k-1)$th frames, each of which starts from the fixed line and ends a small distance away from the stitching lines (to ensure overlap), in opposite directions.

Step 2: Match and ray interpolation. Match a set of corresponding points as control point pairs in the two successive overlapping slices, $\{(P_{1i}, P_{2i}), i = 1, 2, \ldots, N\}$, in a given small region along epipolar lines, around the straight stitching line. We use a correlation-based method to find a pair of piecewise linear *matching curves* passing through the control points in the two frames. The control point pairs are selected by measuring both the gradient magnitudes and the correlation values in matching. Then, a *stitching curve* is obtained which runs through destination locations $Q_i (i = 1, \ldots, N)$ of the interpolated rays in the mosaic computed for each corresponding pair (P_{1i}, P_{2i}) using (16).

Step 3: Triangulation. Select two sets of control points $\{R_{mi} (m = 1, 2; i = 1, \ldots, N-1)\}$ on the fixed lines in the two successive frames, whose y coordinates are determined by the fixed lines and whose x coordinates are the averages of the consecutive control points on the matching curves, P_{mi} and $P_{m,i+1} (m = 1, 2; i = 1, \ldots, N-1)$, for appropriate triangulation. Mapping of R_{1i} and R_{2i} into the corresponding mosaic coordinates results in S_{1i} and $S_{2i} (i = 1, \ldots, N)$, this time by solely using interframe translations $(t_x^{(k)}, t_y^{(k)})$ and $(t_x^{(k-1)}, t_y^{(k-1)})$. For the kth frame, we generate two sets of corresponding triangles (Fig. 6): The source triangles by point sets $\{P_{2i}\}$ and $\{R_{2i}\}$, and the destination triangles by point sets $\{Q_i\}$ and $\{S_{2i}\}$. Do the same triangulation for the $(k-1)^{st}$ frame.

Step 4: Warping. For each of the two frames, warp each *source triangle* into the corresponding *destination triangle*, under the assumption that the region within each triangle is a planar surface given small interframe displacements. Since the two sets of destination triangles in the mosaic have the same control points on the stitching curve, the two slices will be naturally stitched in the mosaic.

4.2.2 Motion Parallax and Misalignment Analysis

We present some experimental results of real video mosaicing to show why ray interpolation is important for both stereo viewing and 3D reconstruction. Fig. 7 shows real examples of local match and ray interpolation results for two pairs of successive images for a UMass campus scene. In constructing stereo mosaics, the distance between the front and the rear slice windows is $d_y = 192$ (in pixels; original image 720*480), and the range profiler tells us that the average height of the aerial camera from the ground is about $H = 300$ m. The quantitative analysis of 3D estimation error with such misalignments for both pairs is summarized in Table 1. As an example, we will explain the details for the Fine Arts Center case in Fig. 7 Example 1. By manually selecting corresponding points on the ground and the top of the ridge of the Fine Art Building in the stereo mosaics generated by the 3D mosaicing method (the fast PRISM algorithm), we find that the relative y displacement of the building top with respect to the ground is about $\Delta y = -13$ pixels. A 1-pixel misalignment (δy in Table 1) in stereo mosaics, when using a 2D mosaicing method without ray interpolation, will introduce a depth (and height) error of $\delta Z = 1.56$ m, using (4). While the relative error of the depth estimation of the roof (i.e., $\delta Z/Z$) is only about 0.56 percent, the relative error in height estimation (i.e., $\delta Z/\Delta Z$) is as high as 7.7 percent. It can be seen that even though the several pixels of interframe motion parallax are not sufficient for 3D estimation using interframe stereo, it is significant in improving the overall depth accuracy in stereo mosaics.

Fig. 7g shows mosaiced results where camera orientation parameters were estimated by registering the planar ground surface of the scene via dominant motion analysis [29]. The readers can also compare the results of 3D mosaicing (with parallel-perspective projection) versus 2D mosaicing (with manifold projection) by examining the building boundaries associating with depth changes in the full $4,160 \times 1,536$ mosaics at our Web sites [28]. Clearly geometric-seamless mosaicing by ray interpolation is very important for accurate 3D estimation as well as good visual appearance.

4.3 Discussions: Better Triangulation and Occlusion Handling

The locations of the stitching curves in the fast PRISM algorithm enable us to use the closest existing views to generate parallel-perspective rays. Using sparse control points on stitching curves and image warping, the fast PRISM algorithm only approximates the parallel-perspective geometry in stereo mosaics. However, the proposed PRISM technique can be implemented to use more feature points (thus, smaller triangles) in the overlapping slices rather than just those around a single stitching curve so that each triangle really covers a planar patch or a patch that is visually indistinguishable from a planar patch. Therefore, one of the critical issues is to robustly pick up and match the control points and to perform better triangulation that is necessary to generate a geometrically corrected and seamless mosaic. Morris and Kanade [10] have discussed the best triangulation given a set of 3D points of an object, based on its consistency with a set of images of the 3D object. The method proposed in their paper could be applied to our PRISM algorithm for better triangulation, which is an important topic that deserves further study.

Another important issue is occlusion handling in ray interpolation. In a perspective image, scene regions in different image locations have varying degrees of occlusion (Fig. 8a). In contrast, in a parallel-perspective image, the occlusion relations are always the same in the direction of the parallel projection (Fig. 8b). Now, the question is: When we transform a sequence of perspective images to a left-view (or right-view) parallel-perspective mosaic using parallel ray interpolation, how can we deal with these different spatial occlusion relations in perspective views? While occlusion is known to be a particularly difficult problem in computer vision, our analysis shows that, in our case of stereo mosaics, we only need to deal with occlusion where we detect significant *right-side* depth boundaries for generating left mosaics, or *left-side* for right mosaics. We show the principle with the *left-view* parallel-perspective mosaic, working with the 1D intersection in the direction of the parallel projection. Let us consider a pair of successive frames (with viewpoints O_1 and O_2) of an image sequence (Fig. 8c). We define an *occlusion viewpoint* O_x as a viewpoint from which the left parallel ray intersects an occluding boundary (B_x) of an object (the box). It can be easily verified that the condition to avoid the occlusion problem is when both viewpoints O_1 and O_2 are on one side of the occlusion viewpoint O_x. Otherwise, we face the occlusion problem. In such a case, a region on the more

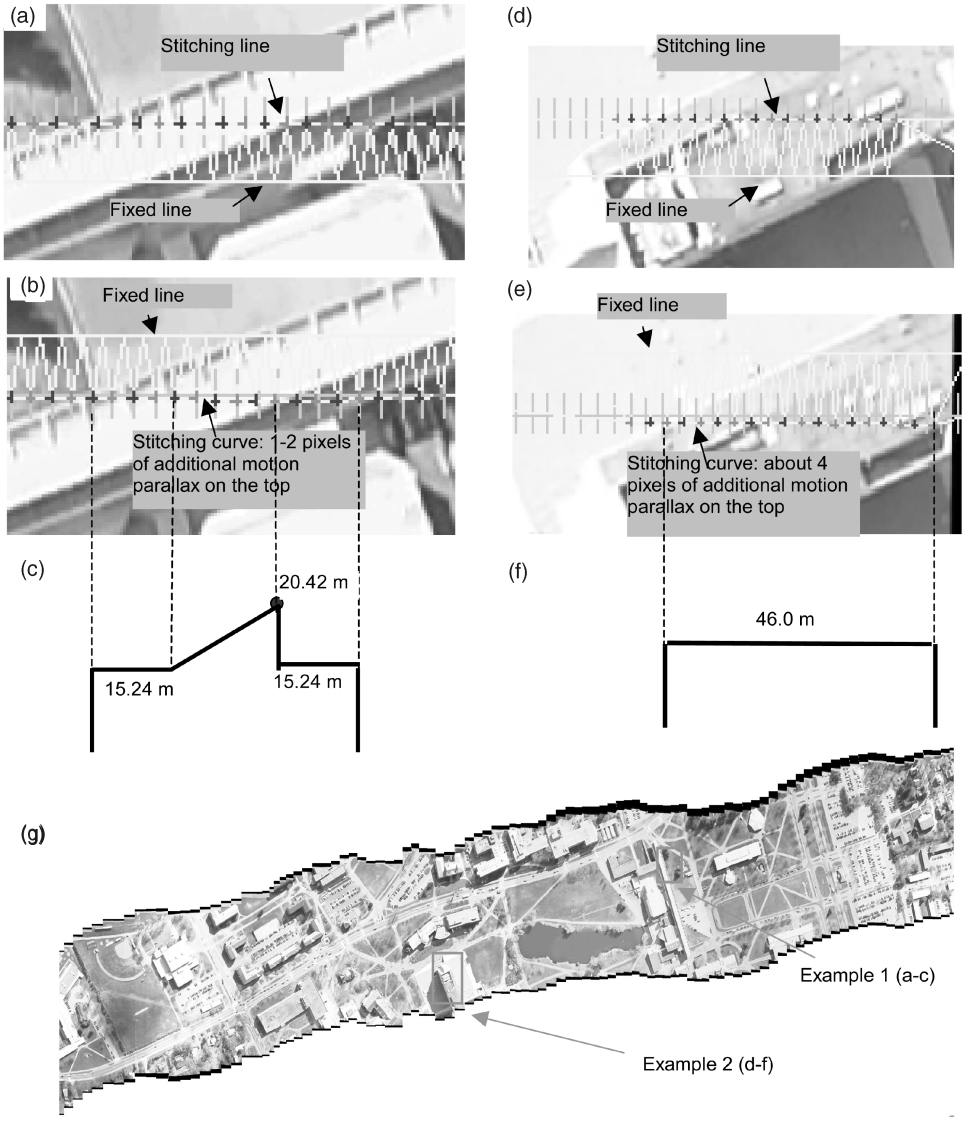


Fig. 7. Two examples of local match and triangulation for the left mosaic. In both examples (the first and the second columns), enlarged views of small windows of the following are shown: (a), (d) the previous frame; (b), (e) the current frame, and (c), (f) ground truth values of heights along the stitching lines (measured from the ground plane). The large gray crosses in (a), (b), (d), and (e) 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 dark and light small crosses show the correct match pairs by feature selection and correlation (light matches light, dark matches dark). The fixed lines, stitching lines/curves, and the triangulation results are shown as lighter lines. The local match results show that points on the tops of the narrow building (the Fine Arts Center Building) in the first example and the tall building (the Campus Center Building) in the second example have larger motion parallax than ground points. The left-view mosaic of all the frames in the sequence using the fast PRISM algorithm are shown in (g).

distant surface (the ground) can only be seen from the second view for generating the left mosaic. Part of this region, which should be visible in the generated left mosaic, is bounded by the rays O_xB_x and O_1B_x. If we are not dealing with very complicated occluding scene, the third view that follows the pair of frames under consideration can usually also see this portion of region, so that matching points in the second and a third view (O_3) for this region and back-projecting the rays to the desired left ray direction will solve this problem.

5 Depth Error Characterization of Stereo Mosaics

In theory, the adaptive baseline inherent in the parallel-perspective geometry permits depth accuracy independent of absolute depth, as shown in [26], [1]. However, in practice, an important question needs to be answered: Since the motion parallax information between two successive perspective frames is used for making stereo mosaics, will the small baseline between frames introduce large errors in ray interpolation, as it does for direct depth estimation from successive frames?

In order to answer the question, we need to reobserve the ray interpolation process. In making parallel-perspective stereo mosaics, the disparities (d_y) of all points are constant since a fixed angle between the two viewing rays is selected for generating the stereo mosaics. As a consequence, for any point in the right mosaic, searching for the match point in the left mosaic indicates a process of finding an original frame in which this match pair has a predefined *constant disparity* (by the distance of the two slit windows) but with an *adaptive baseline* depending on the depth of the point. Therefore, we

TABLE 1
Error Analysis of 2D Mosaics of a Campus Scene (d_y = 192 Pixels, H = 300 m)

Measurements ($\Delta Z = H\Delta y/d_y$, $\delta Z = H\delta y/d_y$, $Z = H + \Delta Z$)	Fine Arts Center Building (Figure 7 Example 1)	Campus Center Building (Figure 7 Example 2)
Mosaic displacements Δy	-13 pixels	-29 pixels
Object depth from camera Z	279.69 m	254.68 m
Object height to ground ΔZ	20.31 m	45.31 m
Ground truth of height ($\pm$0.15m)	20.42 m	46.00 m
Interframe motion s_y	36 pixels	48 pixels
Interframe misalignment δy	1-2 pixels	4 pixels
Depth (or height) error δZ	**1.56 - 3.13 m**	**6.25 m**
Relative depth error ($\delta Z/Z$)	**0.56% - 1.1%**	**2.45%**
Relative height error ($\delta Z/\Delta Z$)	**7.7%-15.5%**	**13.8%**

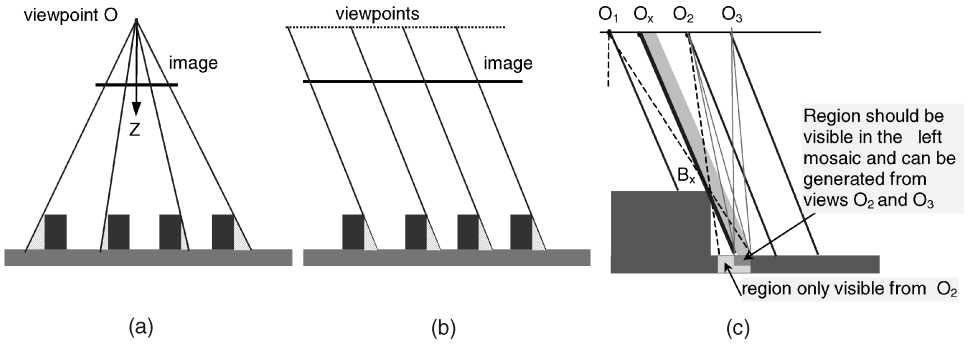


Fig. 8. Illustrations of the differences in occlusions in (a) a perspective image and (b) a parallel image. The shaded regions are occluded. (c) illustrates how to perform ray interpolation with occlusion for a left-view mosaic. A symmetric relation holds for a right-view mosaic.

formulate the problem as follows under 1D translation (Fig. 9). Let us for the moment assume that we accurately identify a point $y_3 = -d_y/2$ right in the center of the rear slit window in a view O_3 (of the original video sequence), which will contribute to the right mosaic. Now we are hoping to find its match point $y_i = +d_y/2$, in the center of the front slit window of a certain view in the sequence, which could contribute to the left mosaic with parallel-perspective projection. Usually, we will not have the exact view O_i; instead, the point y_i is reprojected (i.e., *interpolated*) from a virtual interpolated view O_i determined by a pair of correspondence points y_1 and y_2 in two existing consecutive views O_1 and O_2 in the original perspective image sequence. The localization error of the point y_i depends on the errors in matching and localizing points y_1 and y_2 (and also in camera pose estimation). After some tedious mathematical deduction [31], we obtain an important conclusion: *The depth error of the real stereo mosaics is proportional to the absolute depth*:

$$\delta Z_{mosaic} = \frac{Z}{d_y}\delta y, \tag{17}$$

where δy represents the spatial localization error of the corresponding point pair in the original perspective images.

In the beginning of this discussion, we assumed an accurate right (backward-looking) parallel ray, but we can incorporate the localization error of that ray as well. Symmetrically, the localization of the point in the right mosaic has the same amount of error, but the result of linear depth error characterization will remain the same.

How good is this linear error characterization in the stereo mosaics, then? The analysis in [31] also shows that, even though the depth estimate from two successive views O_1 and O_2 cannot give us good 3D information, as shown by the large diamond error region in Fig. 9 (denoted by $\delta Z_{interframe}$), the localization error of the interpolated point (i.e., the left-viewing ray from viewpoint O_i) is much smaller, leading to significantly smaller depth estimation error (δZ_{mosaic}). The key factor is that the PRISM approach only needs interframe matches (which are much easier to obtain than in the large baseline case), but not the explicit depth information from interframe matches (which is subject to large errors). Quantitatively, it turns out that *the depth error of the real stereo mosaics introduced by the ray interpolation step 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 real stereo mosaics*. Obviously, the

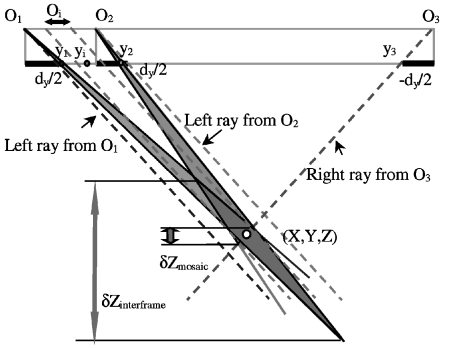


Fig. 9. Error analysis of ray interpolation. While depth estimation for two consecutive frames is subject to large error ($\delta Z_{interframe}$), the localization error of the interpolated ray for stereo mosaics turns out to be very small and so does the depth error of the real stereo mosaics (δZ_{mosaic}).

"real" stereo mosaic approach provides a systematic way to achieve such "optimal" configurations.

From the derivation of the localization error of the interpolated point for ray interpolation, we have found that this error is proportional to Z/F, where F is the focal length, but *is independent of the interframe translational magnitude* [31]. 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 and are not subject to the occlusion problems (as discussed in Section 4.3). An intuitive explanation is that, even though the depth errors via interframe stereo are inversely proportional to the magnitudes of the interframe motion, the projections of those error regions with different motion magnitudes to the parallel ray direction remain the same.

6 Concluding Remarks and Future Work

We have studied the representation and generation of a parallel-perspective stereoscopic mosaic pair from an image sequence captured by a camera with constrained 3D rotation and 3D translation. The 3D mechanism for the stereo mosaics includes two aspects: 1) a 3D mosaicing process consists of a global image rectification that eliminates rotation effects, followed by a fine local transformation and ray interpolation that accounts for the interframe motion parallax due to the 3D structure of a scene; and 2) the final mosaics are a stereo pair that embodies 3D information of the scene derived from optimal baselines. The core idea of ray interpolation in the PRISM algorithm, which leads to better and more accurate mosaics, can be generalized to mosaics with other types of projections.

Since parallel-perspective stereo mosaics provide adaptive baselines and large constant disparity, better depth resolution is achieved than in perspective stereo and the recently developed multiperspective stereo with circular projection. We have arrived at several important conclusions. Ray interpolation between two successive views is actually very similar to image rectification, thus the accuracy of a three-stage matching mechanism (i.e., matching for poses, mosaicing, and correspondences) for 3D recovery from stereo mosaics is comparable to that of perspective stereo with the same adaptive/optimal baseline configurations. Apparently, the stereo mosaic mechanism provides a nice way to achieve such "optimal" configurations. We have proven that the depth error of stereo mosaics from real video is a linear function of the absolute depth, which extends our understanding of the parallel-perspective stereo from the previous observation of constant depth resolution. 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.

Given the nice stereo geometry of the parallel-perspective stereo mosaics, there are several open issues for future research. The first important issue is camera orientation estimation for accurate geo-registered stereo mosaics. Bundle adjustment is an obvious approach, but, in order to apply the techniques automatically and efficiently (without or with little human intervention) to very long image sequences (usually with more than a thousand images), the robustness, convergence, and computational efficiency problems need to be studied.

The second important issue is the interframe matching and triangulation for ray interpolation in generating stereo mosaics. In our current implementation, a simple correlation approach may be sufficient for the forest scenes with strong textures and with quite dense image sequences. But, for a cultural scene with many textureless areas but obvious depth boundaries, an accurate and robust feature selection and matching method is required to build the correspondences between the two slices in the successive frames for ray interpolation. Since ray interpolation actually deals with 3D information (even though not explicit 3D recovery), physical constraints such as homogeneous texture region constraints and boundary detection could be used to define the matching primitives and to provide better triangulation for ray interpolation.

The third important issue is the stereo correspondence problem between a pair of stereo mosaics. The advantage of stereo mosaics for 3D reconstruction is the strong stereo effect from two widely separated viewing directions, creating large constant disparity, and adaptive baselines. However, large and adaptive baselines bring in difficulties in stereo matching. As one of the possible solutions, for example, we can extract multiple (i.e., more than 2) pairs of stereo mosaics with small viewing angle differences (i.e., the disparity d_y) between each pair of nearby mosaics—thus, constructing a "multidisparity" stereo mosaic system [28] , analogous to a multibaseline stereo system [11]. Multidisparity stereo mosaics could be a natural solution for the problem of matching across large oblique viewing angles.

Acknowledgments

This work is partially supported by US National Science Foundation Challenges CISE (Grant Number EIA-9726401), CNPq EIA9970046, SGER EIA-0105272, and Army Research Office (DURIP) DAAD19-99-1-0016. The authors are grateful to anonymous reviewers for their insightful comments and suggestions that have greatly improved the presentation of the paper.

References

[1] J. Chai and H.-Y. Shum, "Parallel Projections for Stereo Reconstruction," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '00),* pp. 493-500, 2000.

[2] R. Gupta and R. Hartley, "Linear Pushbroom Cameras," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 9, pp. 963-975, 1997.

[3] H.-C. Huang and Y.-P. Hung, "Panoramic Stereo Imaging System with Automatic Disparity Warping and Seaming," *Graphical Models and Image Processing,* vol. 60, no. 3, pp. 196-208, 1998.

[4] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Efficient Representations of Video Sequences and Their Applications," *Signal Processing: Image Comm.,* vol. 8, no. 4, pp. 327-351, May 1996.

[5] M. Irani and P. Anandan, "Video Indexing Based on Mosaic Representations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 86, no. 5, pp. 905-921, 1998.

[6] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omnidirectional Stereo for Making Global Map," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '90),* pp. 540-547, 1990.

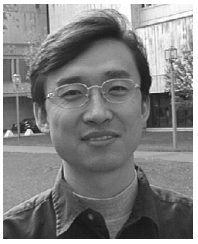
[7] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna, "Representation of Scenes from Collections of Images," *IEEE Workshop Representation of Visual Scenes,* pp. 10-17, 1995.

[8] R. Kumar, H. Sawhney, J. Asmuth, J. Pope, and S. Hsu, "Registration of Video to Geo-Registered Imagery," *Proc. IAPR Int'l Conf. Pattern Recognition (ICPR '98),* vol. 2, pp. 1393-1400, 1998.

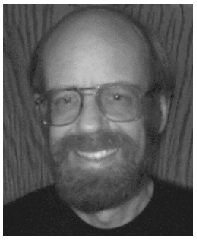
[9] D.L. Milgram, "Adaptive Techniques in Photo Mosaicing," *IEEE Trans. Computers,* vol. 26, pp. 1175-1180, 1977.

[10] D.D. Morris and T. Kanade, "Image-Consistent Surface Triangulation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '00),* pp. 332-338, June 2000.

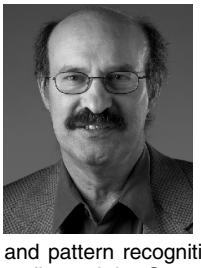
[11] M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 4, pp. 353-363, 1993.
[12] S. Peleg and J. Herman, "Panoramic Mosaics by Manifold Projection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '97),* pp. 338-343, 1997.
[13] S. Peleg and M. Ben-Ezra, "Stereo Panorama with a Single Camera," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '99),* pp. 395-401, 1999.
[14] S. Peleg, M. Ben-Ezra, and Y. Pritch, "OmniStereo: Panoramic Stereo Imaging," *IEEE Trans. Pattern Analysis and Machine Intelligence,* pp. 279-290, Mar. 2001.
[15] B. Rousso, S. Peleg, I. Finci, and A. Rav-Acha, "Universal Mosaicing Using Pipe Projection," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '98),* pp. 945-952, 1998.
[16] H.S. Sawhney, "Simplifying Motion and Structure Analysis Using Planar Parallax and Image Warping," *Proc. IAPR Int'l Conf. Pattern Recognition (ICPR '94),* pp. 403-408, 1994.
[17] H.S. Sawhney, R. Kumar, G. Gendel, J. Bergen, D. Dixon, and V. Paragano, "VideoBrush: Experiences Consumer Video Mosaicing," *Proc. IEEE Workshop Applications of Computer Vision (WACV '98),* pp. 56-62, Oct. 1998.
[18] H. Schultz, "Terrain Reconstruction from Widely Separated Images," *Proc. SPIE,* vol. 2486, pp. 113-123, Apr. 1995.
[19] H.-Y. Shum and L.-W. He, "Rendering with Concentric Mosaics," *Proc. SIGGRAPH '99,* pp. 299-306, Aug. 1999.
[20] H.-Y. Shum and R. Szeliski, "Stereo Reconstruction from Multiperspective Panoramas," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '99),* pp. 14-21, 1999.
[21] H.-Y. Shum and R. Szeliski, "Construction of Panoramic Image Mosaics with Global and Local Alignment," *Int'l J. Computer Vision,* vol. 36, no. 2, pp. 101-130, 2000.
[22] *Manual of Photogrammetry,* fourth ed. C.C. Slama ed., Am. Soc. of Photogrammetry, 1980.
[23] D. Slaymaker, H. Schultz, A. Hanson, E. Riseman, C. Holmes, M. Powell, and M. Delaney, "Calculating Forest Biomass with Small Format Aerial Photography, Videography and a Profiling Laser," *Proc. 17th Biennial Workshop Color Photography and Videography in Resource Assessment,* 1999.
[24] R. Szeliski and S.B. Kang, "Direct Methods for Visual Scene Reconstruction," *Proc. IEEE Workshop Representation of Visual Scenes,* pp. 26-33, 1995.
[25] J.Y. Zheng and S. Tsuji, "Panoramic Representation for Route Recognition by a Mobile Robot," *Int'l J. Computer Vision,* vol. 9, no. 1, pp. 55-76, 1992.
[26] Z. Zhu, A.R. Hanson, H. Schultz, F. Stolle, and E.M. Riseman, "Stereo Mosaics from a Moving Video Camera for Environmental Monitoring," *Proc. First Int'l Workshop Digital and Computational Video,* pp. 45-54, 1999.
[27] Z. Zhu, G. Xu, and X. Lin, "Panoramic EPI Generation and Analysis of Video from a Moving Platform with Vibration," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '99),* pp. 531-537, 1999.
[28] Z. Zhu, "PRISM: Parallel Ray Interpolation for Stereo Mosaics," http://www-cs.engr.ccny.cuny.edu/~zhu/StereoMosaic.html or http://www.cs.umass.edu/~zhu/StereoMosaic. html, 2000.
[29] Z. Zhu, E.M. Riseman, and A.R. Hanson, "Theory and Practice in Making Seamless Stereo Mosaics from Airborne Video," Technical Report #01-01, Computer Science Dept., Univ. of Massachusettes-Amherst, Jan. 2001, http://www.cs.umass.edu/zhu/UM-CS-2001-001.pdf.
[30] Z. Zhu, E.M. Riseman, and A.R. Hanson, "Parallel-Perspective Stereo Mosaics," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '01),* vol. I, pp. 345-352, July 2001.
[31] Z. Zhu, A.R. Hanson, H. Schultz, and E.M. Riseman, "Generation and Error Characteristics of Parallel-Perspective Stereo Mosaics from Real Video," *Video Registration,* M. Shah and R. Kumar, eds., Video Computing Series, Kluwer Academic, pp. 72-105, May 2003.
[32] A. Zomet, S. Peleg, and C. Arora, "Rectified Mosaicing: Mosaics Without the Curl," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '00),* pp. 459-465, June 2000.



Zhigang Zhu received the BE, ME, and PhD degrees, all in computer science from Tsinghua University, Beijing, China, in 1988, 1991, and 1997, respectively. He is currently an associate professor in the Department of Computer Science, the City College of the City University of New York. Previously, he was an associate professor at Tsinghua University, and a senior research fellow at the University of Massachusetts, Amherst. From 1997 to 1999, he was the director of the Information Processing and Application Division in the Computer Science Department at Tsinghua University. His research interests include 3D computer vision, Human-Computer Interaction (HCI), virtual/augmented reality, video representation, and various applications in education, environment, robotics, surveillance, and transportation. He has published more than 80 technical papers in the related fields. Dr. Zhu received the Science and Technology Achievement Award (second-prize winner) from the Ministry of Electronic Industry, China, in 1996 and the C.C. Lin Applied Mathematics Award (first prize winner) from Tsinghua University in 1997. His PhD thesis "On Environment Modeling for Visual Navigation" was selected in 1999 as a special award in the top 100 dissertations in China over the last three years, and a book based on his PhD thesis was published by China Higher Education Press in December, 2001. He is a member of the IEEE and a member of the ACM.



Allen R. Hanson received the BS degree from Clarkson College of Technology in 1964 and the MS and PhD degrees in electrical engineering from Cornell University in 1966 and 1969, respectively. He joined the Computer Science Department as an associate professor in 1981, has been a full professor since 1989. He has conducted research in computer vision, artificial intelligence, learning, and pattern recognition, and has more than 150 publications. He is the codirector of the Computer Vision Laboratory with a diverse range of recent research including aerial digital video analysis for environmental science, three-dimensional terrain reconstruction, distributed sensor networks, motion analysis and tracking, mobile robot navigation, under-vehicle inspection for security applications, object recognition, color analysis, and image information retrieval. He has been on the editorial boards of the following journals: *Computer Vision, Graphics and Image Processing* (1983-1990), *Computer Vision, Graphics, and Image Processing—Image Understanding* (1991-1994), and *Computer Vision and Image Understanding* (1995-present). He is a member of the IEEE.



Edward M. Riseman received the BS degree from Clarkson College of Technology in 1964 and the MS and PhD degrees in electrical engineering from Cornell University in 1966 and 1969, respectively. He joined the Computer Science Department as an assistant professor in 1969, has been a full professor since 1978, and served as chairman of the department from 1981-1985. He has conducted research in computer vision, artificial intelligence, learning, and pattern recognition, and has more than 200 publications. He has codirected the Computer Vision Laboratory since its inception in 1975, with a diverse range of recent research, including aerial digital video analysis for environmental science, three-dimensional terrain reconstruction, distributed sensor networks, motion analysis and tracking, mobile robot navigation, biomedical image analysis, under-vehicle inspection for security applications, object recognition, color analysis, and image information retrieval. He has served on the editorial board of *Computer Vision and Image Understanding* (CVIU) from 1992-1997, and the editorial board for the *International Journal of Computer Vision* (IJCV) from 1987 to the present, is a senior member of IEEE, and a fellow of the American Association of Artificial Intelligence (AAAI).