Extracting 3D Information Using Spatio-Temporal Analysis of Aerial Image Sequences

Guoqing Zhou, Joerg Albertz, and Klaus Gwinner

Abstract

To overcome depth discontinuities and occlusion problems in three-dimensional (3D) surface information extraction using traditional stereophotogrammetric matching, a new approach called spatio-temporal analysis *of aerial image sequences is proposed. In the proposed method, a set of spatio-temporal solid data is first formed from a sufficiently dense sequence of images taken by a camera moving along a straight-line path. Second, the set of spatio-temporal solid of data is sliced along a temporal dimension into epipolar-plane images (EPIs), and features in these slices are extracted and described. Finally, three-dimensional coordinates in a ground coordinate system are computed for the features in the EPIs. This method is fairly radically different from traditional two-view stereophotogrammetric matching; therefore, we discuss in detail the estimation accuracy, error resources, and sensitivities to occlusion and depth discontinuities. The experimental results from three test fields in Berlin, Germany show that the method is a useful tool for solving the problems of depth discontinuities and occlusion with which photogrammetrists have been wrestling for a decade.*

Introduction

Extracting three-dimensional (3D) surface information from large-scale aerial images is an important research field in photogrammetry, computer vision, and robot vision navigation. Over the last decade of photogrammetric development, a large number of efforts have been pursued. However, because the scenes may be very complex and diverse, many existing algorithms, such as traditional stereo matching, have shown an inability to cope with such scenes. Therefore, when two-view stereophotogrammetric matching is used for 3D information extraction, the following problems may occur:

- *Large gray-value differences between corresponding points.* Because a flight would have captured aerial images from varying camera positions (a large baseline length), the gray values between corresponding points show large differences. Moreover, it is very difficult to model the gray differences. If traditional stereo matching schemes are employed, it would be difficult to get accurate depth information from the various scenes.
- *Poor texture appearance.* In some homogeneous gray areas, such as water, house tops, grassy, etc., extracted 3D information is unreliable when using points as the matching elements.

- *Repeated texture appearance.* Some patterns, such as walls or house groups in residential regions, appear repeatedly. Thus, it is difficult to find the matching corresponding points when using traditional stereo matching schemes.
- *Depth discontinuities.* Almost all buildings stand above the surrounding terrain, which gives rise to depth discontinuities. Such features cause three problems in 3D information extraction. First, it is difficult to use an additional smoothing constraint in the stereo matching process. Second, occlusions occur in the image planes. If the occluded regions cannot be recognized correctly, the gray information will disrupt normal matching, or even mislead matching. Third, the depth discontinues often cause an abrupt change in the apparent surface brightness due to shadow.

Obviously, none of the traditional approaches has so far been ablt to solve the above mentioned problems successfully. For example: (1) the correlation coefficient approach may overcome large gray differences, but it cannot resolve non-linear gray-value components, and it is sensitive to geometric distortion caused by depth discontinuities; (2) multiple-point least-squares matching (LSM) can theoretically deal with repeated textures and poor textures; however, because the algorithm depends on the constraint of surface continuity, many difficulties are encountered when depth discontinuities exist; and (3) feature-based matching algorithms can achieve a satisfactory effect relative to area-based matching in handling large gray differences, but their applications are restricted by the effectiveness of the feature detectors and by 3D data interpolation error (for example, only some sparse heights are produced).

In this paper, we propose a *spatio-temporal analysis technique* to extract 3D surface information from aerial image sequences. The method was originally presented by Bolles and Baker (1985a), Bolles and Baker (1985b), Baker *et al.* (1986), and Baker *et al.* (1987) for the analysis of motion image sequences. Their initial experimental environment fully met the four constraint conditions listed in the next section. Bolles *et al.* (1987) extended the analysis to arbitrary motions using projective duality in space. Their generalizations are on (1) the generalization of their technique for varying viewing directions, including variation over time; (2) the provision of three-dimensional connectivity information for building coherent spatial descriptions of observed objects; and (3) a sequential implementation, allowing initiation and refinement of scene feature estimates while the sensor is in motion (Baker and Bolles, 1988; Baker and Bolles, 1989). This paper

---

G. Zhou is with the Department of Civil & Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275 (zhou1@hcgl1.eng.ohio-state.edu).

J. Albertz and K. Gwinner are with the Department of Photogrammetry and Cartography, Technical University of Berlin, Strasse des 17 Juni 135, Sekr. EB9, D-10623 Berlin, Germany.

Figure 1. Construction of a spatio-temporal data set.
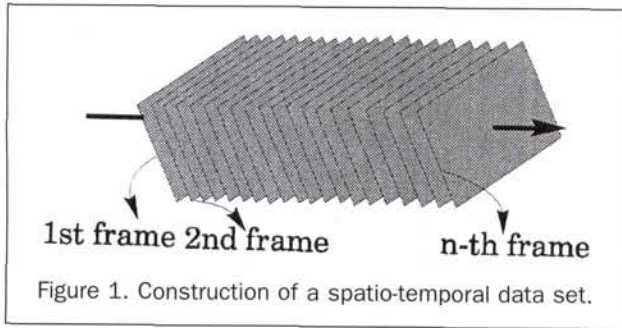
describes the acuracy to be reached, and the ability to accomodate occlusions and depth discounities when the technique is applied in aerial photogrammetry. Three test fields were established built in Berlin, Germany. The experimental results show that the technique is able to solve some of the problems which conventional photogrammetry is not able to solve.

Principle of Temporal and Spatial Analysis

The original method was based on the following four conditions:

- The camera's movement is restricted to a straight linear path,
- Image capture is rapid enough with respect to camera's movement and scene scale to ensure that the data are temporally continuous,
- The velocity of the camera's movement is a constant, and
- The camera's position and attitude at each imaging site are known.

Theoretically, if a camera mounted on an airborne platform flies along a straight line path (flight course), and the camera's optical axis is orthogonal to the direction of motion, this operational condition meets the above constraint conditions. The impact of irregular real flight movements will be addressed in the Error Analysis section.

The meaning of the second condition is that the image sequences captured by the camera are so close that none of

Figure 3. Construction of epipolar plane images (EPIS).
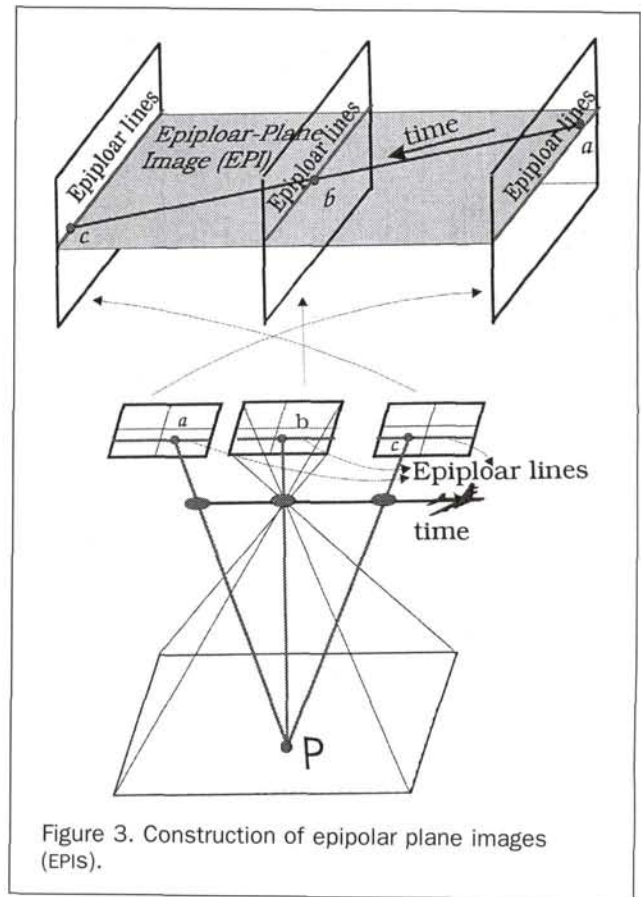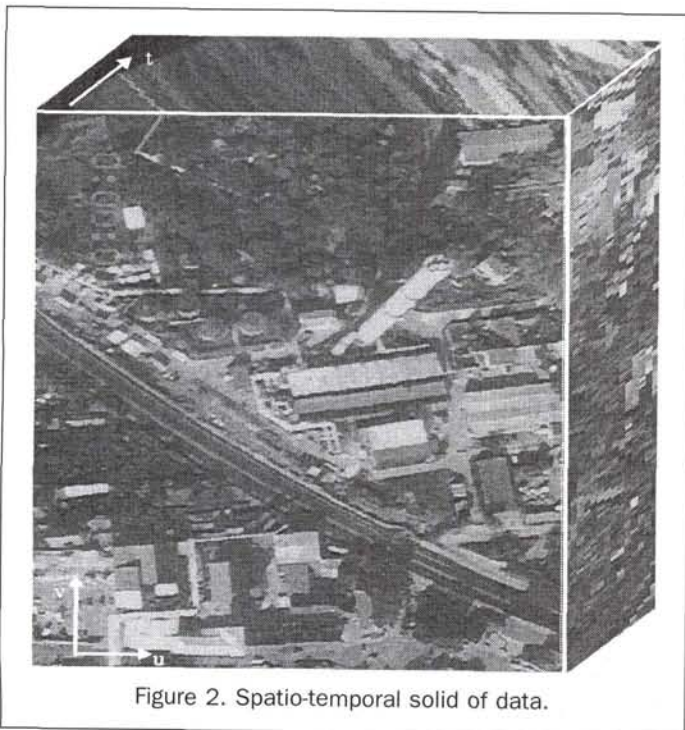
the image features moves more than a pixel or so. This sampling frequency guarantees continuity in the temporal domain that is similar to continuity in the spatial domain (Figure 1). Thus, an edge of an object in one image appears temporally adjacent to its occurrence in the following image. This temporal continuity makes it possible to construct a cube of data in which time is the third dimension and continuity is maintained over all three dimensions. This solid of data is referred to as *spatio-temporal data* (Figure 2). Figure 3 shows three individual images used to form the solid of data. Typically, 100 or more images are used, making the trajectory of a point, such as *P*, follow a continuous path.

With the first condition, suppose that there is a simple motion in which a camera moves from left to right, with its optical axis orthogonal to its direction of motion. For this type of motion, the epipolar plane for a point *P* is the same for all camera positions. We call this plane an *epipolar plane* of the point *P* for the whole motion. If the velocity of the camera is a constant (the fourth condition), the trajectories in the epipolar planes images (EPIs) are straight lines. Figure 3 illustrates the construction of the epipolar-plane images.

The projection of *P* onto the epipolar lines moves from the right to the left as the camera moves from the left to the right. The velocity of this movement along the epipolar line is a function of *P*'s distance from the line through the lens centers. The closer it is, the faster it moves. Therefore, a vertical slice of the spatio-temporal solid of data contains all the epipolar lines associated with one epipolar plane. If we slice the spatio-temporal data along the temporal dimension, a new "image plane," which is called the *epipolar-plane image* (EPI) is formed (Figure 3).

(*x, y, z*) coordinates can be computed from the fourth condition. We first derive an equation for the trajectory of a

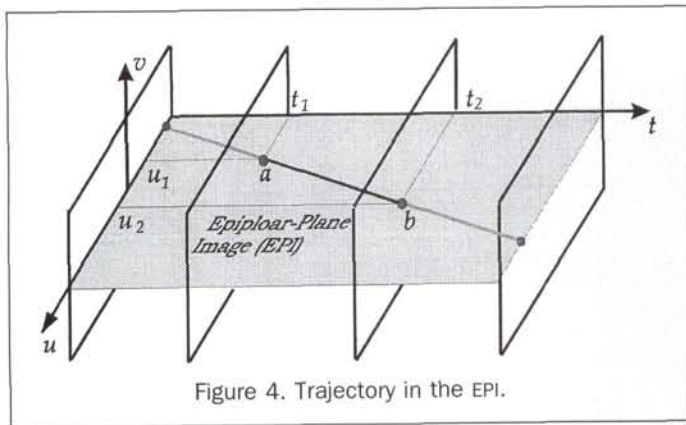Figure 2. Spatio-temporal solid of data.

Figure 4. Trajectory in the EPI.

scene point in the EPI constructed from a motion, and then explain how to compute the (x, y, z) coordinates of such a point. Figure 4 is a diagram of a trajectory in an EPI derived from the left-to-right motion illustrated in Figure 5. The image row at t_1 in Figure 4 corresponds to the epipolar line l_1 in Figure 5. Similarly, the image row at t_2 corresponds to the epipolar line l_2. (Remember that the EPI is constructed by extracting one line from each image taken by camera as it moves along the straight line from c_1 to c_2. Because the images are taken very close together, there would be several images taken between c_1 and c_2. However, to simplify the diagram, none of them is shown.) The point (u_1, t_1) in the EPI corresponds to the point (u_1, v_1) in the image taken by the camera at time t_1 and position c_1. Thus, as the camera moves from c_1 to c_2 in the time interval t_1 to t_2, the scene point moves in the EPI from (u_1, t_1) to (u_2, t_2).

The intent of above section is to characterize the shape of this trajectory. The following describes how to compute the three-dimensional coordinates when given the focal length of the camera and the camera's velocity.

In our analysis we define a right-hand coordinate system that is centered on the initial position of the camera.

Given the speed of the camera, s, which is assumed to be a constant, the distance B from c_1 to c_2 can be computed as (see Figure 5)

$$B = s \cdot \Delta t \qquad (1)$$

where $\Delta t = (t_2 - t_1)$. From similar triangles (Figure 5), we have

$$\frac{u_1}{h} = \frac{x_p}{D} \qquad (2)$$

$$\frac{u_2}{h} = \frac{B + x_p}{D} \qquad (3)$$

where u_1 and u_2 have been converted from pixel values into distances in the image plane, h is the distance from the projected lens center to the epipolar line in the image plane, x_p is the x-coordinate of P in the scene coordinate system, and D is the distance from P to the line through the lens centers. Because h is the hypotenuse of a right triangle, it can be computed as follows:

$$h = \sqrt{f^2 + v_1^2} \qquad (4)$$

where f is the focal length of the camera. From Equations 2 and 3, we get

$$\Delta u = (u_2 - u_1) = \frac{h(B + x_p)}{D} - \frac{h x_p}{D} = \frac{h}{D} B. \qquad (5)$$

Equation 5 shows that Δu is a linear function of B, while B is also a linear function of Δt. Thus, Δt is linearly related to Δt. This means that all trajectories in the EPIs are straight lines in the constrained straight-line motion.

The (x, y, z) coordinates of P can be computed from u_1, u_2, and f. From Equation 5, we define

$$m = \frac{D}{h} = \frac{B}{\Delta t} \qquad (6)$$

which represents the slope of the trajectory computed in terms of the distance traveled by the camera (B as opposed to Δt) and the distance which the point moves along the epipolar line. From similar triangles, we have

$$(x, y, z) = (\frac{D}{h}u_1, \frac{D}{h}v_1, \frac{D}{h}f). \qquad (7)$$

Equation 7 may be rewritten as

$$(x, y, z) = (mu_1, mv_1, mf). \qquad (8)$$

If the first camera position c_1, on an observed trajectory, is different from the camera position c_0, defining a global camera coordinate system, the x coordinate has to be adjusted by the distance traveled from c_0 to c_1. Thus,

$$(x, y, z) = ((t_1 - t_0)s + mu_1, mv_1, mf) \qquad (9)$$

where t_0 is the time of the first image and s is the camera's speed. This correction is equivalent to computing the intercept x of the trajectory, which is the first camera's position. Therefore, the (x, y, z) coordinates of the points can be easily computed from the slopes and intercepts of the trajectories.

Experiments of Spatio-Temporal Analysis

Three test fields in the region of Berlin (Berlin city, Schönefeld, and Werder) were used to test this technical application in aerial photogrammetry. In October 1995, the image sequences for the three experimental fields were captured using a video camera mounted on a Cessna 207 T flying platform. Details of the imaging parameters are listed in Table 1.

The original data were recorded on a video cassette, and the digital image sequences were obtained by resampling at a frequency of 10 frames/second. For further analysis, we developed software to implement the programs, which include the following steps:
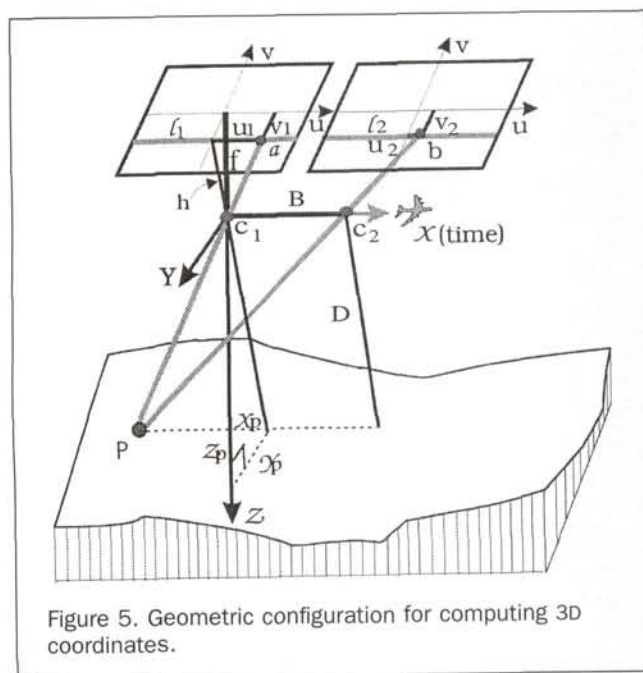

Figure 5. Geometric configuration for computing 3D coordinates.

TABLE 1. IMAGING FLIGHT PARAMETERS.

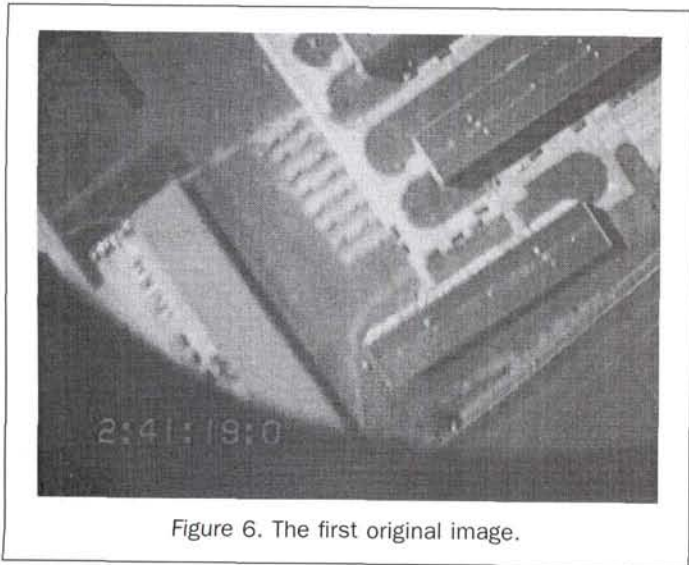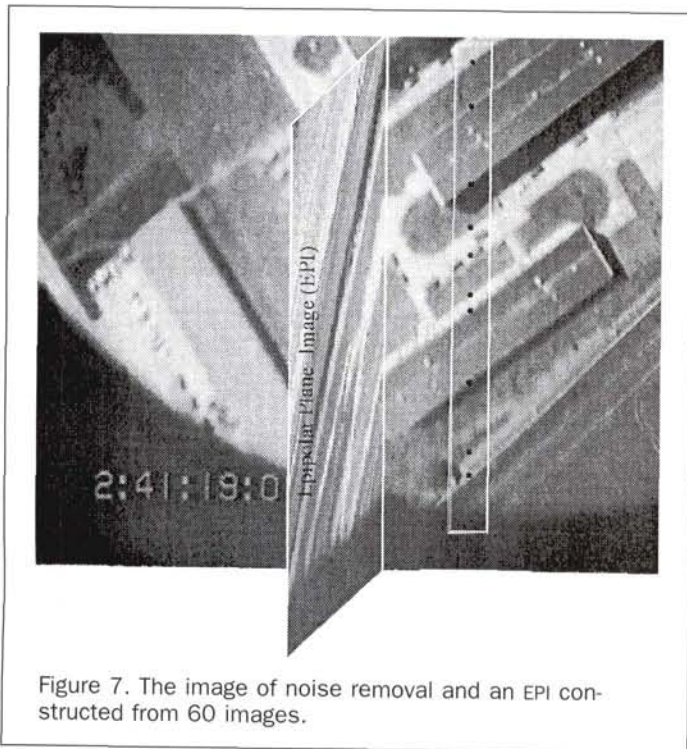| Name	Parameters
Platform	Cessna 207T
Flight height	about 800 m
Focal Length	35 mm
Flight velocity	100 Knots
Camera type	S-VHS, Panasonic Videorecorder
Scale	1:2500

Figure 6. The first original image.

Figure 7. The image of noise removal and an EPI constructed from 60 images.

- The first step consists of noise removal algorithm. The original image is depicted in Figure 6, while Figure 7 shows the pre-processing image after noise removal.
- The second step generates EPIs from the spatio-temporal solid of data. Figure 8 illustrates the 500th EPI constructed from a sequence of 60 images.
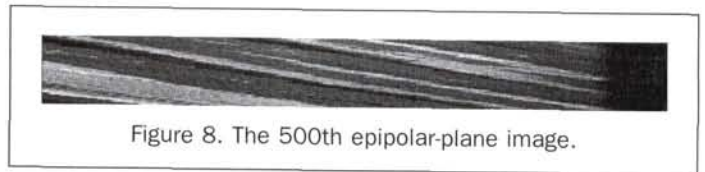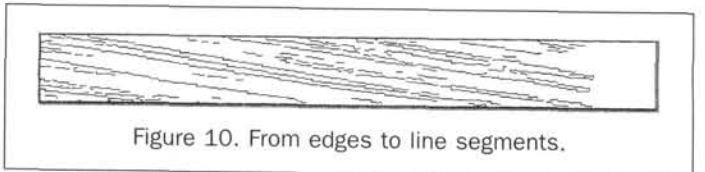
- The third step applies an edge detector to detect "edge-like" features in the EPIs. We use the zero-crossing operator to detect the edges (see Figure 9), whose original image is shown in Figure 8.
- The fourth step fits linear segments to the edges. It is divided into two passes. The first pass partitions the edges at sharp corners by analyzing the curvatures of the edges. The second pass applies a regression algorithm to recursively partition the smooth segments into continuous straight line segments. Figure 10 shows the linear segments derived from the edges in Figure 9.
- The fifth step builds a description of the line segments that aligns together those that are collinear. The line segments are used to identify sets of lines that belong to the same feature. Figure 11 shows the edge feature descriptions for Figure 10 that are linked by bridging gaps caused by occlusions or other effects (for example, lines broken by the edge detection operator).
- The sixth step computes the (x, y, z) coordinates of the world features corresponding to the EPI features. The world coordinates are uniquely determined by the slope and intercept of the trajectories in EPI. Figure 12a shows the (x, y, z) coordinates for world features marked with black points inside the white rectangle in Figure 7.
- The seventh step links x-y-z points between EPI to obtain structure information. Figure 12b illustrates the structure information for world features marked with the black points inside the white rectangle in Figure 7.

For the first experimental field, we concentrated on estimating the accuracy of the (x, y, z) coordinates for the points on a profile. The eleven profile points are marked with black points inside the white rectangle in Figure 7. The profile corresponding the 500th EPI (Figure 8) includes houses, ground, and shadows (see Figure 7). The (x, y, z) coordinates of the profile points are illustrated in Figure 12. The experimental result shows that the heights of two houses are clearly distinguished from the ground surface.

Figure 8. The 500th epipolar-plane image.
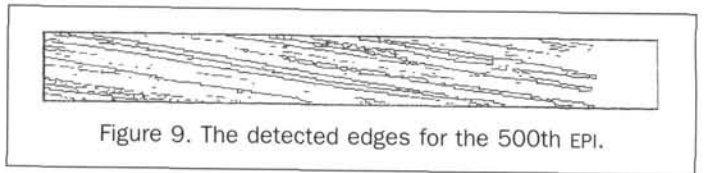

Figure 9. The detected edges for the 500th EPI.

Figure 10. From edges to line segments.
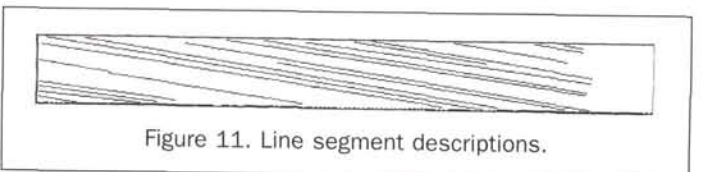

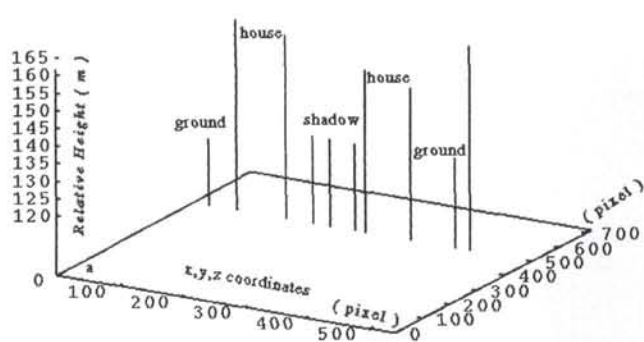
Figure 11. Line segment descriptions.

Figure 12. The of (x, y, z) coordinates of world features corresponding to the black points in Figure 7.
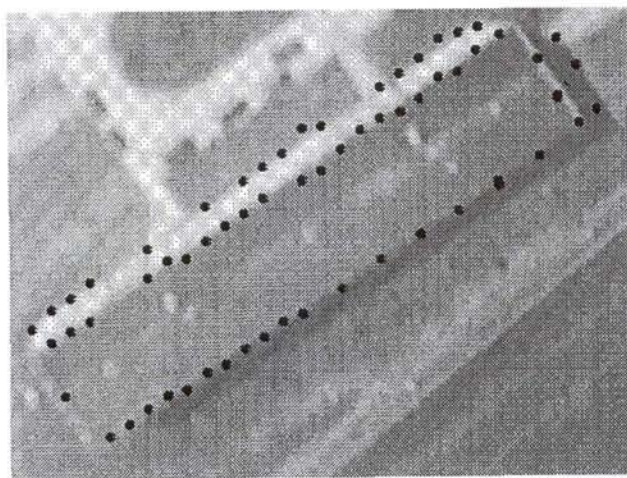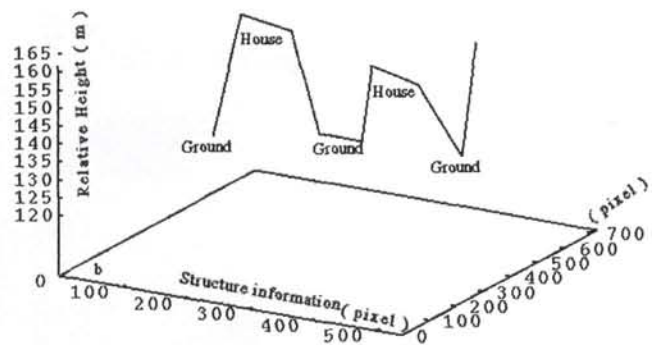

Figure 13. Subsection of Figure 7. The black points indicate the extracted information using EPI analysis at a 5-pixel EPI interval.
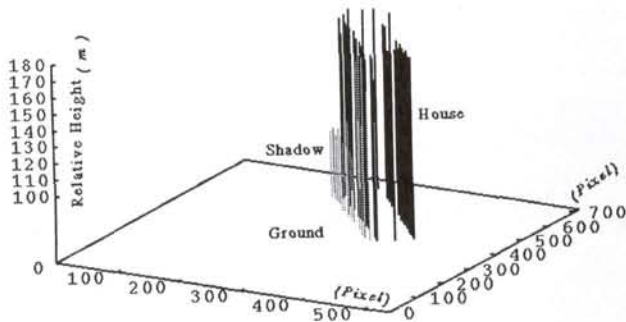

Figure 14. (x, y, z) coordinates of the house and the shadow.

Figure 13 is a subsection of Figure 7, in which a house is contained. The black points indicate the information extracted at a 5-pixel EPI interval. The (x, y, z) coordinates for the shadows and the houses are shown in Figure 14, in which the dotted lines are (x, y, z) coordinates of a shadow. By examining Figure 14, we find that the shadow and ground have different heights.

In order to have an evaluation of the accuracy of the entire 3D surface, the DEM is generated by a simple nearest-height interpolation algorithm from sparse (x, y, z) coordinates extracted using the spatio-temporal analysis (see Figure 15). By examining Figure 15, we believe that the DEM data shows the height of the scene correctly.

For the second experimental field, the imaging parameters are the same as for the first test field. In the scene an obvious landmark — a chimney — is chosen to test the sensitivity to occlusion and depth discontinuities. The black points depicted inside the white rectangle of Figure 16 are the extracted profile features, whose corresponding EPI is shown in Figure 18 [the 220th EPI]. The (x, y, z) coordinates for these world features and their structure information are illustrated in Figure 19. A magnified view of the chimney top shows the locating accuracy of the chimney, as illustrated in Figure 17. The DEM data generated from sparse coordinates (x, y, z) are illustrated in Figure 20. By examining Figure 20 and the 220th EPI (Figure 18), as well as the (x, y, z) coordinates (Figure 19), we find that (1) the estimation precision for (x, y) is apparently better than that for z; (2) the technique can effectively deal with occlusions because the trajectories of the occluded objects are broken temporally (see the 220th EPI in Figure 18); (3) the technique easily handles depth discontinuities because it does not need to determine the magnitude of horizontal parallaxes, and 3D coordinates can be obtained from the trajectories of the features in EPIs directly; and (4) the DEM data are relatively rough in the scene, but the higher buildings can be reflected correctly.
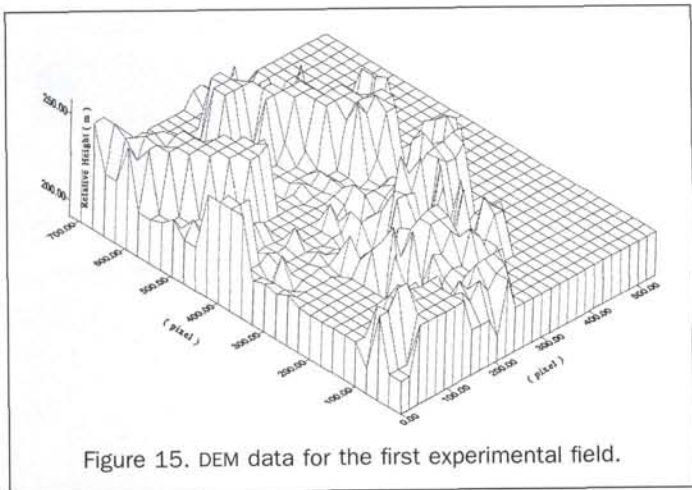

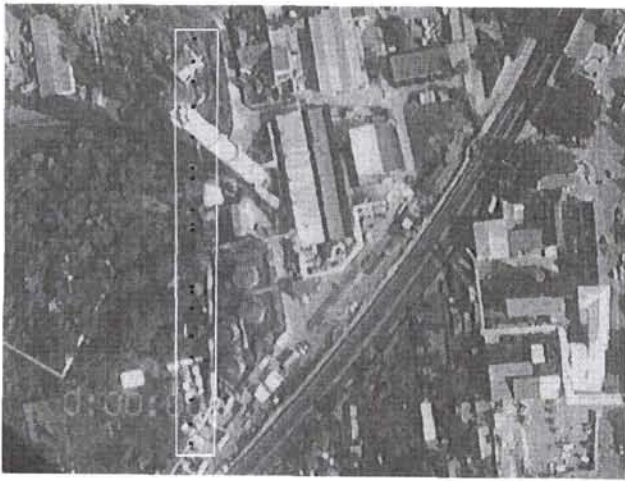Figure 15. DEM data for the first experimental field.

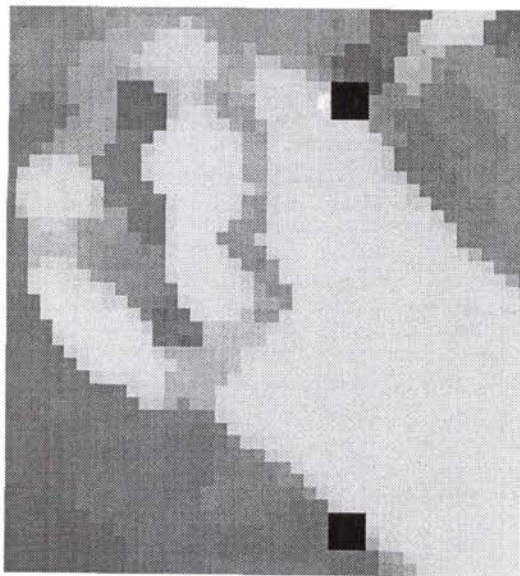Figure 16. The first image in the second experimental field.

Figure 17. Magnified image of chimney with located feature points.

Figure 18. The 220th epipolar-plane image.

The third experimental field is a more complex scene, involving trees, bushes, forest, houses, and roads. Many houses are not easily identified because of the lower image resolution. A road is chosen to evaluate the accuracy of the spatio-temporal analysis. The black points inside a white rectangle in Figure 21 are the world features extracted by using the spatio-temporal analysis at a 5-pixel EPI interval. The coordinates (x, y, z) of the road are illustrated in Figure 22. Similarly, the DEM, visualized in Figure 23, is generated by a

simple nearest-height interpolation algorithm from sparse (x, y, z) coordinates. By examining the experimental results (Figures 22 and 23), we find that the z coordinates of the road still show a relatively large variance. Moreover, the DEM data are relatively wild in the scene.
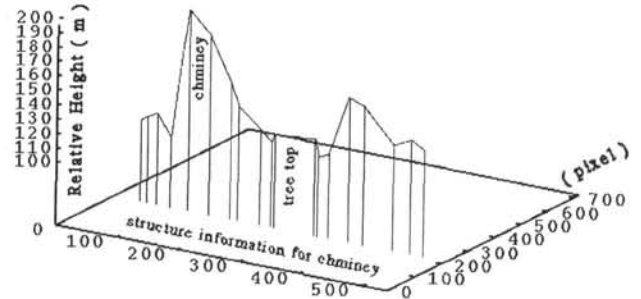

Figure 19. The (x, y, z) coordinates and the structure information corresponding to the profile marked by black points in Figure 16.
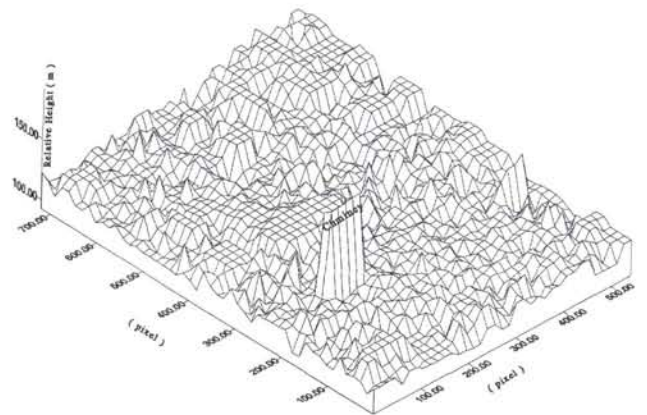

Figure 20. DEM data for the second experimental field.

Figure 21. Third experimental field and the extracted road's feature points using EPIS at a 5-pixel EPI interval.
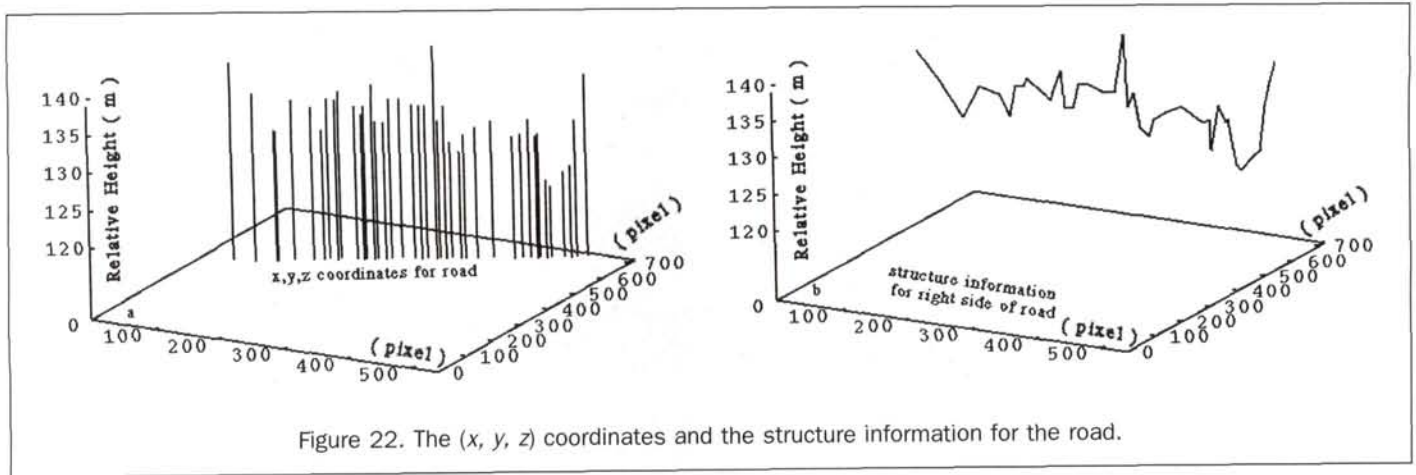
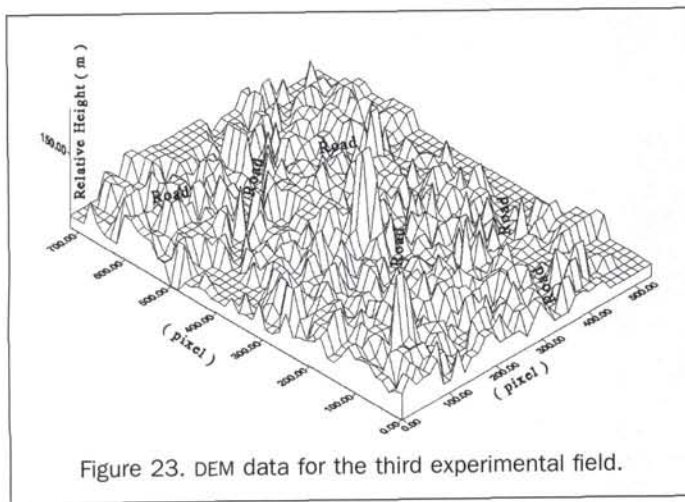Figure 22. The (x, y, z) coordinates and the structure information for the road.

Figure 23. DEM data for the third experimental field.

Error Analysis

It should be noted that the operational conditions of our airborne experiment are much more complicated than we had supposed. The real operation of the camera shows considerable deviations from an ideal linear movement. Atmospheric turbulence, flight vibration, and other influences result in a rather irregular flight course. This is why the trajectories of the features in EPIs are not straight lines. On the other hand, the camera's viewing direction randomly varies due to changes in the flight attitude parameters, including the roll, pitch, and yaw angles. Thus, it is not possible to guarantee that the aerial image sequences exactly meet the four constraints. In addition, the resolution of the images captured by the video CCD camera is not very high. In the second resampling, some information may be lost again. Therefore, these various error sources cause rather large errors in the world coordinates. Here we analyze several kinds of error paradigms.
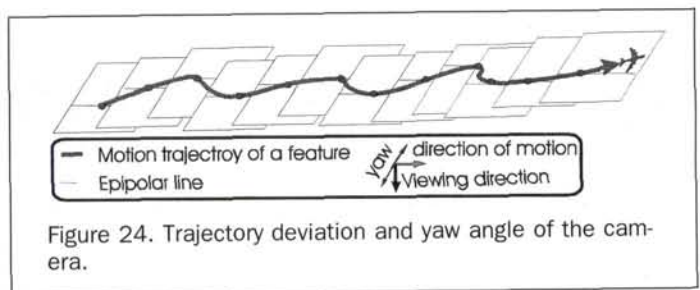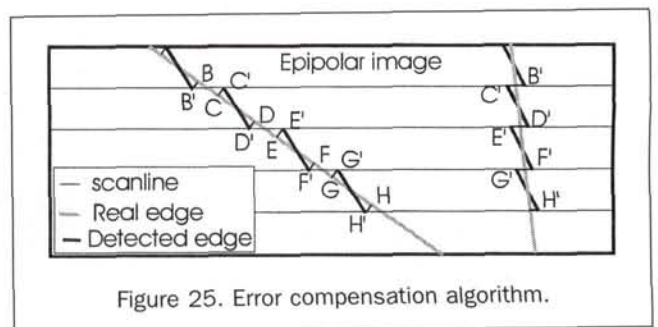
Deviation from the ideal camera path in the horizontal plane
When the camera deviates from the ideal path in the horizontal plane, that path is distorted (Figure 24) and the trajectories for the features in the EPIs show a "zigzag" shape. Figure 26 is an EPI from the first experimental field. The edges that appear like narrow white bars in Figure 26 show an obvious "zigzag" shape. If the deviation is restricted within a finite range and follow a stochastic process, the errors can be compensated by our program automatically because the trajectories in the EPIs are generated by more than

a hundred images, and line segments are described by applying a regression algorithm associated with gross error eliminating technique. Figure 25 illustrates our compensation principle employed in our program.

Deviation from the ideal camera path in the vertical plane
When the optical axis of the camera (the camera viewing direction) is not orthogonal to the direction of motion (Figure 27), the error can be modeled as follows (Wang, 1986):

$$\delta_r = - \frac{\frac{r^2}{f} \sin \varphi \sin \theta}{1 - \frac{r}{f} \sin \varphi \sin \theta} \tag{10}$$

where $\delta_r = r - r_0$ is a radius difference (see Figure 27). When the angle θ is very small, Equation 10 can be approximated as

$$\delta_r \approx \frac{r^2}{f} \sin \varphi \sin \theta. \tag{11}$$

Figure 24. Trajectory deviation and yaw angle of the camera.

Figure 25. Error compensation algorithm.

Figure 26. The trajectories of features and deviations of the camera path in the horizontal plane.
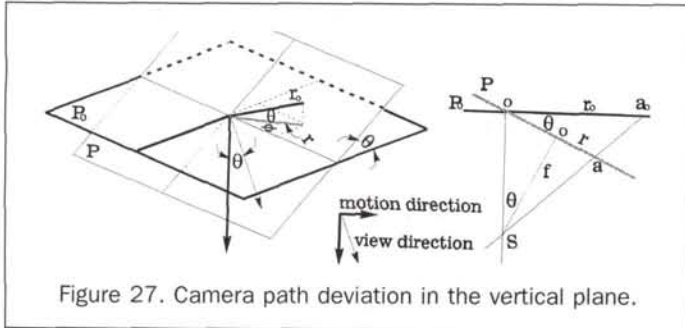

Figure 27. Camera path deviation in the vertical plane.

In Equation 11, when ϕ equals 0° or 180°, $\delta_r = r - r_o$ reaches a maximum value, i.e., the deviation is greatest; when ϕ equals 90° or 270°, the deviation is smallest. Moreover, the positive and negative signs change randomly with angles varying from 0° to 360°. Thus, the deviation can still be considered as a stochastic process. This kind of error can also be compensated automatically by our program with the same compensation principal.

However, if the camera's viewing direction deviates at a fixed angle relative to the direction of motion, which means that the deviation is not a stochastic process, the trajectories of the features in EPIs are simple hyperbolas (Baker and Bolles, 1988; Baker and Bolles, 1989). If the angle of the camera's viewing direction relative to the direction of motion shows systemic errors, it is impossible to partition the scene into a fixed set of planes, which in turn means that it is not possible to construct EPIs for such a motion. In these two cases, our program is not able to compensate the errors.

The velocity of the camera's movement is not a constant
We can derive the trajectories of the features in EPIs. The theoretical analysis (see Appendix A) shows that the trajectories of features in EPIs are composed of several line segments with various slopes rather than being simple straight lines. If the velocity of the camera's movement varies only occasionally during the entire movement process, our program can automatically compensate for such an error by employing a blunder elimination algorithm; however, if the velocity during the entire movement process varies continuously, our program loses this kind of compensation capability. On the other hand, if the error cannot be controlled effectively, it will reach rather large (x, y, z) deviations because it can change the slope of the trajectory (see Appendix A). Figure 28 illustrates the true trajectories of the features in the EPI from the third experimental field.

The sampling frequency cannot guarantee temporal continuity
This means that an edge of an object in an image does not appear temporally adjacent to its following images. Thus, the trajectory of a feature consists of several line segments (broken lines) rather than an entire straight line. Our program can bridge the gap automatically. Figure 28 illustrates the fact that the low sampling frequency cannot guarantee continuity in the spatial domain (several broken line segments).

Edge location error
We can determine the (x, y, z) error values caused by the edge location error in terms of error propagation laws. The mean error for (x, y, z) can be given by (from Equation 9)

$$\Phi_x = m\Phi_{u_1} = \frac{\Delta x}{\Delta u}\,\Phi_{u_1} = s \cdot slope \cdot \Phi_{u_1} \text{ (when } t = t_0)$$

$$\Phi_y = s \cdot slope \cdot \Phi_{v_1} \tag{12}$$

$$\Phi_z = s \cdot slope \cdot f$$

Equation 12 shows that the mean errors of the coordinates (x, y, z) are a linear function of u and v, and the scale factor is the velocity of the camera's movement.

If $s = 170$ (m/second), and $\Phi_{u_1} = \Phi_{v_1} = $ one pixel $= 0.001$ m, i.e., the trajectory of a feature shifts one pixel along the u and v directions in the EPI, respectively, and the slope $= 1$ (45°) (Figure 29). When the *slope* value of the trajectory remains constant, we have

$$\Phi_x = 170 \cdot slope = 0.17 \text{(m)} \tag{13a}$$

$$\Phi_y = 170 \cdot slope = 0.17 \text{(m)} \tag{13b}$$

$$\Phi_z = 170 \cdot slope \cdot f \tag{13c}$$

Equations 13 show that, if the trajectory of a feature shifts one pixel along the u and v axes in the EPI, the x and y coordinates in the world can reach a 0.17-meter deviation, whereas the z-coordinate does not change.

In the same way, if the slope of the trajectory of a feature deviates one pixel (this is in fact quite possible), the error in the (x, y, z) coordinates can reach a rather large deviation. Therefore, high accuracy edge location is significant.

Further Work

Even though this technique shows that the (x, y, z) coordinates have errors, the precision of the (x, y, z) coordinates in the first experiment is still acceptable; the second experimental results show that the technique is a very useful tool in overcoming depth discontinuities and occlusions. In order to improve the estimation precision of (x, y, z) coordinates and develop this technique to a practical tool in photogrammetry, the above-listed error sources must be reduced or eliminated by the software automatically. The following efforts are necessary in the future:

- From Equation 9, the (x, y)-coordinates of object points depend on the intercept of the trajectories of the features in EPIs, and the z-coordinate of object points depends on the slopes of the trajectories of the features in EPIs. If the coordi-

Figure 28. Trajectories and varying movement velocity, low frequency resampling.

Figure 29. Detected and real edge. (a) The trajectory deviates from ideal direction. (b) The trajectory shifts along the u and v axes.
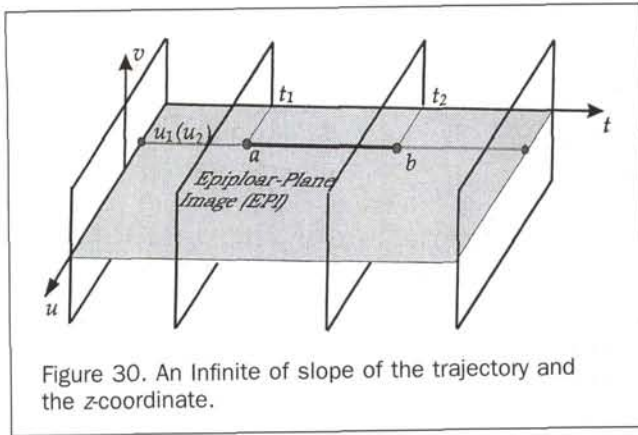
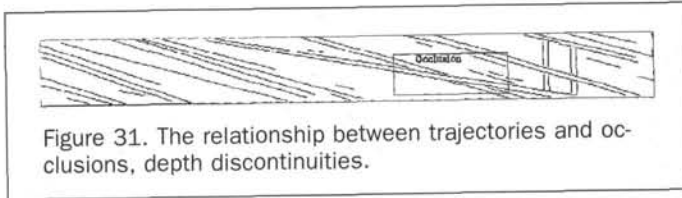Figure 30. An Infinite of slope of the trajectory and the z-coordinate.

Figure 31. The relationship between trajectories and occlusions, depth discontinuities.

nates at the points (u_1, t_1) and (u_2, t_2) in the EPI, which correspond to the points (u_1, v_1), (u_2, v_2) in the image taken by the camera at time t_1, t_2 and positions c_1, c_2 are equal, the slope of the trajectory is infinity (Figure 30). This means that the (x, y, z) coordinates of world features corresponding to the trajectories in the EPI reach infinity. In this case, we must analyze them carefully. For example, if the directions of a road and a house are approximately parallel to the direction of motion, the trajectories for the road and the house in the EPIs are identical, but their z-coordinates are in fact different.

- Geometric rectification of image sequences in advance, without (or with) known orientation parameters of the cameras, is imperative. This has been demonstrated by the analysis of irregular camera movement.
- We have demonstrated that, when the slope of a trajectory in an EPI deviates by one pixel, the (x, y, z) coordinates in the world coordinate system will deviate up to several hundred times. In the same way, when a trajectory in an EPI deviates by one pixel along the u and v axes, the x and y coordinates in the world coordinate system can also deviate up to several hundred times. Thus, high accuracy edge detection is very significant.
- We must assume that the computed (x, y, z) coordinates are relative coordinates instead of geodetic ones. This is because we really do not have information on (1) the camera's position and attitude at each imaging site, (2) the exact velocity of the camera's movement, and (3) the initial camera position. The bundle adjustment combining GPS and image data (Kruck, 1996; Schade, 1996) makes the technology applicable to aerial photogrammetry.
- The operational conditions of the camera, the resampling frequency, and the camera's position and attitude at each imaging site are more easily controlled in close-range measurement. Moreover, the imaging resolution is higher than that of aerial photography. Thus, the approach can lead to practical applications in close-range photogrammetry (e.g., in architectural photogrammetry).

Conclusions

Compared to traditional two-view stereo photogrammetric matching, this approach has some apparent advantages. First, because the rapid image sampling results in a minimal change from frame to frame, thus avoiding disparate views (a large baseline length), the correspondence problem (stereo matching) is eliminated. Second, because all trajectories in

the EPIs are straight lines, only very similar data are processed. Thus, the approach is much simpler and more robust. Third, this technique involves the processing of a very large number of images acquired by a moving camera, i.e., a great deal of redundant information is produced, which is not used in the conventional stereophotogrammetric approach. Fourth, the approach is feature-based, but is not restricted to point features. Linear features that are perpendicular to the direction of motion can also be used. Fifth, spatial structures in the EPIs are much simpler than in the original, which means that they are easy to be interpreted and analyzed.

By examining the described EPI in Figure 31, which corresponds to the original EPI (Figure 18), we find that the trajectories of the occluded objects are temporally broken by other trajectories, which correspond to the high objects. This means that occlusions are quite apparent in the EPIs. Thus, there are many chances to detect them. We can theoretically prove (see Appendix B) that, if an object temporally occludes another object, the trajectories corresponding to the occluding and occluded feature intersect in the EPI. Moreover, for each intersection, the feature with the smaller slope occludes the one with the steeper slope. This means that the higher building occludes the lower one. Additionally, the trajectories of higher buildings, which cause depth discontinuities, can obviously be reflected in the EPIs because of their steeper slopes. As mentioned above, the 3D coordinates are directly obtained from the trajectories of the high buildings in the EPIs without determining the values of horizontal parallaxes. Therefore, the approach is more robust in determining the occlusions and depth continuities than is the traditional multiple (two)-view matching.

Nevertheless, because this is just the first practical application of this technique to aerial photogrammetry, more testing is needed. Particularly, because the operational conditions are very complex, the real camera's path deviates considerably from an ideal one. Rectifying the distorted image sequences is imperative. Additionally, the bundle adjustment combining GPS and image data to get the orientation and attitude parameters of the camera is fundamental to practical applications.

In a word, we have had reasons to believe that the approach can become a useful tool in solving occlusion and depth discontinuities, with which stereo photogrammetrists have been wrestling for a decade.

References

Baker, H.H., 1987. Multiple-Image Computer Vision, *41st Photogrammetric Week*, Stuttgart Germany, September, pp. 7–13.

Baker, H.H., and Robert C. Bolles, 1988. Generalizing Epipolar-Plane Image Analysis on the Spatiotemporal Surface, *IEEE Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, June, pp. 2–9.

———, 1989. Generalizing Epipolar-Plane Image Analysis on the Spatiotemporal Surface, *International Journal of Computer Vision*, 3(1) pp. 57–91.

Baker, H.H., R.C. Bolles, and D.H. Marimont, 1996. A New Technique for Obtaining Depth Information from a Moving Sensor, *Proceeding of the ISPRS Comm. II Symposium on Photogrammetric and Remote Sensing System for Data Processing and Analysis*, Baltimore, Maryland, May 1986, pp. 120–129.

Bolles, Robert C., and H.H. Baker, 1985a. Epipolar-Plane Image Anal-

ysis: A Technique for Analyzing Motion Sequence, *Int. Symposium on Robotics Research*, Gouvieux, France, pp. 41–48.

———, 1985b. Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequence, *IEEE Third Workshop on Computer Vision: Representation and Control*, Bellaire, Michigan, pp. 168–176.

Bolles, R.C., H.H. Baker, and D.H. Marimont, 1987. *Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion*, International Journal of Computer Vision, 1(1):7–55.

Förstner, W., 1986. A Feature Based Correspondence Algorithm for Image Matching, *Int. Arch. of Photogrammetry*, 26–III(3):150–166.

Larry, M., and Takeo Kanade, 1989. Kalman Filter-based Algorithms for Estimating Depth from Image Sequences, *International Journal of Computer Vision*, 3(1):209–236.

Kruck, Erwin, 1996. Advanced Combined Bundle Block Adjustment with Kinematics GPS Data, *Int. Archives of Photogrammetry and Remote Sensing*, 31(Part B3):294–398.

Schade, Holger, 1996. On the Use of Modern GPS Receiver and Software Technology for Photogrammetric Applications, *Int. Archives of Photogrammetry and Remote Sensing*, 31(Part B3):729–734.

Wang, Zizuo, 1991. *Photogrammetric Principle*, House of Publishing, Surveying and Mapping, Beijing, China, 43 p.

Appendix A

We use the counterevidence method to prove that, if the speed of the camera is not constant, the trajectory of a feature in an EPI is not a straight line (instead of consisting of many line segments with various slopes).

Without loss of generality, assume that the camera's speed of movement falls between s_1 and s_2 and that s_1 is not equal to s_2: that is,

$$s_1 \neq s_2, \tag{A1}$$

From Equation 9, we have

$$z_1 = m_1 \cdot f_1 = s_1 \cdot slope_1 \cdot f_1 \tag{A2}$$

$$z_2 = m_2 \cdot f_2 = s_2 \cdot slope_2 \cdot f_2 \tag{A3}$$

where $slope_1 = \Delta t_1/\Delta u_1$ and $slope_2 = \Delta t_2/\Delta u_2$. Suppose that the slopes corresponding to the speeds s_1 and s_2 are $slope_1$ and $slope_2$, respectively. Moreover, assume that the slopes are equal, i.e., $slope_1 = slope_2$. We then have

$$z_1 = s_1 \cdot slope \cdot f \tag{A4}$$

$$z_2 = s_2 \cdot slope \cdot f \tag{A5}$$

For a feature in the scene, the z coordinate does not change with the camera's speed s_1, s_2, i.e., $z_1 = z_2$. From Equations A4 and A5, the following equation can then be obtained:

$$s_1 = s_2. \tag{A6}$$

Equation A6 shows that the speeds s_1 and s_2 should be equal. The conclusion contradicts the original proposition (Equation A1).

Therefore, when the speed of the camera's movement is not constant, the trajectory of a feature is not a straight line. If the camera's movement accelerates or decelerates smoothly, the trajectory of a feature is a smooth curve; if the camera's movement accelerates or decelerates suddenly, the trajectory of a feature is a discontinuous curve.

Appendix B

Suppose that there are two feature points P_1 and P_2 in the scene, which correspond to the heights z_1 and z_2 and suppose that $z_1 > z_2$. If the flight speed is a constant, s, then from Equation 9, we have

$$z_1 = m_1 \cdot f_1 = s \cdot slope_1 \cdot f_1 \tag{B1}$$

$$z_2 = m_2 \cdot f_2 = s \cdot slope_2 \cdot f_2 \tag{B2}$$

where $slope_1 = \Delta t_1/\Delta u_1$ and $slope_2 = \Delta t_2/\Delta u_2$. Because the focal length of the camera is a constant,

$$f_1 = f_2. \tag{B3}$$

Substituting Equation B3 into Equation B1 and B2 and considering that $z_1 > z_2$, we get

$$slope_1 > slope_2. \tag{B4}$$

Therefore, we can conclude that, if the features in the scene have various heights (z coordinates), the trajectories of these features will intersect in the EPI or at infinity. If a feature is too much higher than another feature in the scene, i.e., if $z_1 >> z_2$, then one feature is occluded by the other, and the trajectories of the two features should intersect in the EPI.