MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

3D STRUCTURE FROM MOTION USING HOMOCENTRIC SPHERICAL SPATIOTEMPORAL IMAGE ANALYSIS

Takayuki YASUNO and Teruo HAMANO

NTT Human Interface Laboratories

Nippon Telegraph and Telephone Corporation

1-2356, Take, Yokosuka-Shi, Kanagawa 238-03, Japan

E-mail:yasuno%nttcvg.ntt.jp@relay.cs.net

ABSTRACT

A technique for building a three-dimensional description of a static scene from a spatiotemporal image is presented. This technique utilizes *Homocentric Spherical Spatiotemporal Image (HSSI)* analysis and makes allowances for camera rotation, which is a wider class of camera motion than that in ordinary spatiotemporal image analysis. Experiments with real images from a translating and rotating camera were preformed, and the three dimensional structures in a static scene were reconstructed. With this technique, it is possible to distinguish objects with a rotating camera at long distances and to measure them with much greater accuracy.

1 Introduction

Spatiotemporal image analysis is based on epipolar-plane image analysis, which was first presented by Bolles and Baker at. el. [1][2][3]. The spatiotemporal image is a dense sequence of images taken in such rapid succession that they form a single solid block of data in which the temporal continuity from image to image is approximately equal to the spatial continuity in an individual image. For straight-line camera motion, tracks of objects appear as linear structures on *epipolar-plane images (EPIs)* , which are slices of the spatiotemporal image containing epipolar lines. The distances from the camera to the objects are determined from the inclination of tracks. With EPI analysis, it is much easier to compute the three dimensional positions of object features. Any corresponding problems between images can be solved by detection of feature-lines from the EPI [4].

Baker and others found that camera motion was not limited to only the straight lines perpendicular to the optic axis in spatiotemporal image processing when using a cylindrical co-ordinate system with a camera path axis. Instead, they found multiple free moving axes. In such cases, any feature of a scene is restricted to a single epipolar plane. The tracks of the features draw random curves; therefore, it is only possible to distinguish them one after another [5].

In this article, an expansion of camera motion is presented for spatiotemporal image analysis using spherical projection transformation. Shapes in images that are projected on a spherical surface do not vary with rotation when the axis pierces the len's center. The spherical projection transformation makes it easier to revise image-changes caused by camera rotation, and it also becomes possible to separate the rotating and translating effects from image-changes due to camera motion. Object features in the images move along longitudinal lines of polar co-ordinates with the camera path axis. An HSSI is composed of images projected on homocentric spherical surfaces with radii in proportion to the distance of movement of the camera. An object feature is drawn as a curve on a plane containing the longitudinal lines of the HSSI. Therefore, for straight-line camera motion, feature tracks can be denoted as a function and extracted robustly with Hough transformation. In another method, Hough transformation has been utilized only for determining three dimensional lines [6]. But in our method, it is possible to ascertain the 3-D positions of feature points as well as the lines.

2 3D reconstruction using HSSI

2.1 3D reconstruction using a spatiotemporal image

A spatiotemporal image consisting of images obtained with a moving camera in a static scene at regular intervals is shown in Figure 1. When camera motion is linear and perpendicular to the camera axis at a constant speed, the object's features appear as lines on a horizontal slice of the spatiotemporal image, which is parallel to direction of movement to the image plane. This slice is called an epipolar plane image (EPI).

The global co-ordinates $O\text{-}XYZ$ are defined. Here, the view point moves on the X axis at a constant speed, with the optic axis being parallel to the Z axis. If the focus length is F, an object's image is projected on the plane; $Z = F$. If the projection plane is $o\text{-}xy$, the spatiotemporal image is an accumulation of $o\text{-}xy$. Point $P(X,Y,Z)$

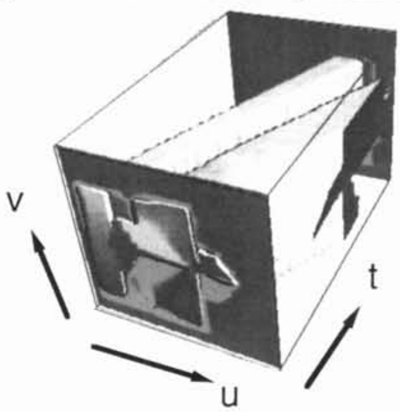

Figure1: A spatiotemporal image.

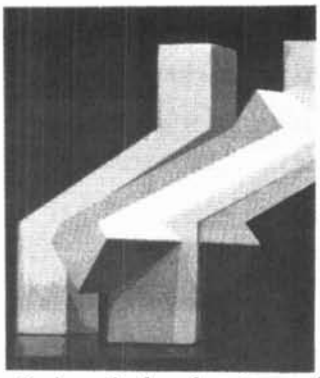

Figure2: An epipolar plane image (EPI).

in the global co-ordinates is projected as $p(x, y)$ on the projection plane. Point $p(x, y)$ can be expressed as

$$x = \frac{F}{Z}(X - vt) \tag{1}$$

$$y = \frac{F}{Z}Y, \tag{2}$$

where v is the velocity of view point and t is time.

The projection p is always on a slice of a spatiotempral image expressed as equation (2). This slice, which is the accumulation of epipolar lines, is called an epipolar plane image (EPI). When the velocity v is constant, p draws a line on the EPI. The depth of p is determined from the line's inclination $-Z/(Fv)$. Therefore the three dimensional positions of an object's features are established by line detection on the EPI using Hough transformation. Figure 2 is an EPI.

2.2 Spherical projection transformation

Changes in projected images are brought about by translation and rotation of the camera. Images are transformed into *spherical projection transformation*, i. e., projecting images on a sphere to separate into translation and rotation effects. It becomes easy to extract translation effects from image changes because images projected on a sphere are not changed by camera rotation.

With spherical projection transformation, a plane projection image I is projected on a sphere S with the center as the view point, as shown in Figure 3.

A spherical projection image is represented by the spherical co-ordinate. Rotation with an axis as the view point is expressed as composition rotation with x, y, z axes. Point $p(x, y)$ on plane projection image I is represented as $p'(\theta, \gamma)$ on the sphere. The $p'(\theta, \gamma)$ is a spherical co-ordinate expression with an axis as the optic axis. The following equations holds for p and p'.

$$\tan\theta = \frac{y}{x} \tag{3}$$

$$\tan\xi = \frac{x}{F} = \cos\theta \tan\gamma \tag{4}$$

thus

$$\theta = \arctan(\frac{y}{x}) \tag{5}$$

$$\gamma = \arctan(\frac{x}{F\cos\theta}). \tag{6}$$

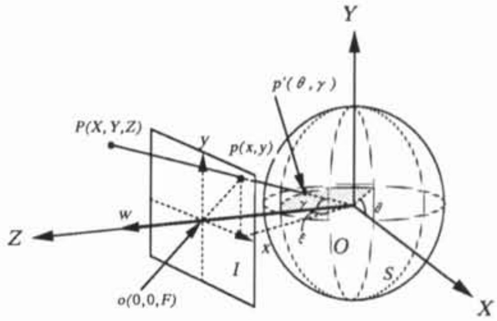

Figure3: Spherical projection transformation.

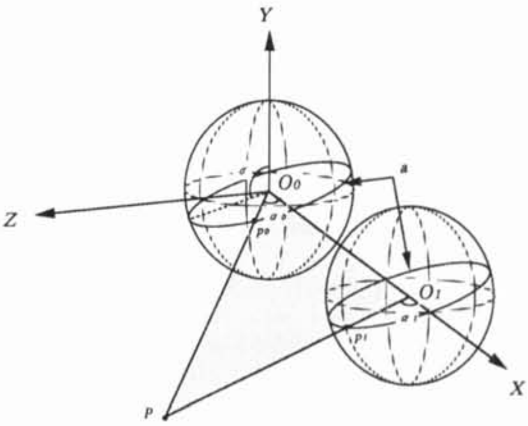

Figure4: Moving view point.

2.3 The Epipolar-arc

Shapes in images that are projected on a spherical surface are uninfluenced by rotation when the axis pierces the len's center. That is when the camera is rotated with the axis piercing the len's center, images that are projected on a spherical surface slide on the sphere but are neither distorted nor warped. When the camera is translated, object features in the images move along the longitudinal lines of polar co-ordinates with the camera path axis and a pole as the focus of expansion (FOE). As shown in Figure 4, the center O of the sphere moves from O_0 to O_1 along the X axis. p_0 and p_1 are images of 3D point P on the sphere O_0 and O_1, respectively. The image p moves along great circle a which is included in a plane determined by points PO_0O_1. In the spherical co-ordinate with an axis parallel with direction of movement, p_0, p_1 are expressed as $p_0(\sigma, \alpha_0)$, $p_1(\sigma, \alpha_1)$, respectively. Both points are on great circle a at longitude σ. The great circle is called an *Epipolar-arc*. An Epipolar-arc corresponds to an Epipolar-line in the plane projection. The speed at which an image point moves on an Epipolar-arc is determined according to the distance from the view point to 3D object. Therefore, when camera motion is known, the extrac-

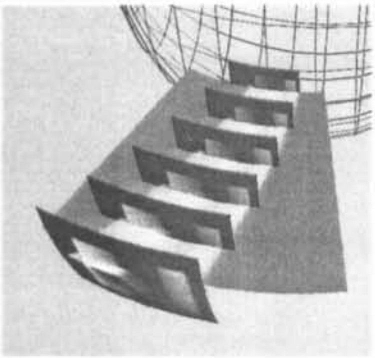

Figure5: Homocentric Spherical Spatiotemporal Image (HSSI).

tion of translation effects from image changes is possible with rotational revision for spherical projection images. Three dimensional depth of objects are estimated from velocities of images on Epipolar-arcs.

2.4 the Homocentric Spherical Spatiotemporal Image (HSSI)

A three dimensional image composed of images projected on homocentric spherical surfaces with radii in proportion to the distance of movement of the camera is defined as *the Homocentric Spherical Spatiotemporal Image (HSSI)*. An HSSI is shown in Figure 5. The distance from its center corresponds to time t. An object feature is drawn as a curve on a slice of the HSSI containing longitudinal lines. This slice corresponds to the EPI explained in 2.1. Therefore, for straight-line camera motion, feature tracks can be denoted as a function and extracted robustly with Hough transformation.

Steps of three-dimensional regeneration using HSSI analysis are as follows.

1. At first, a sequence of image is taken as a spatiotemporal image.
2. Then the image is transformed in spherical projection according to equations (5) and (6).
3. Rotational revision is done to make a HSSI in spherical co-ordinate having an axis which is parallel with the direction of movement.
4. Next, object features are tracked on slices containing epipolar-arcs.
5. Finally, three dimensional positions are calculated.

2.5 Feature detection in the EPI

Equation (7) holds for object P and its image p on an epipolar-arc, as shown in Figure 6.

$$x/d = 1/\tan\alpha, \tag{7}$$

where α is the angle of the direction of movement and OP. Thus tracks on the EPI (α, t) are on curves represented in the following equation:

$$t = \frac{d}{x'}\left(\frac{1}{\tan\alpha} - \frac{1}{\tan\alpha_0}\right), \tag{8}$$

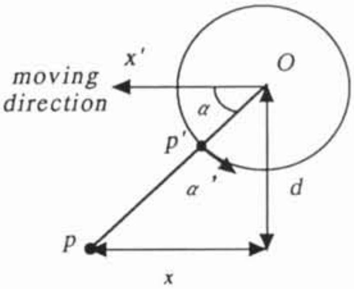

Figure6: Image on the Epipolar-arc.

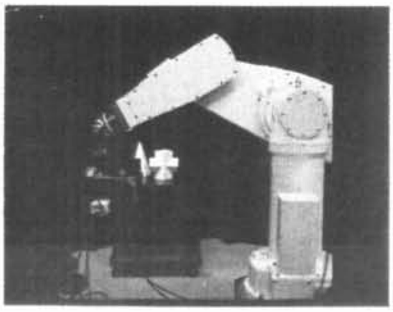

Figure7: Manipulator camera.

where α_0 is the initial value of α. With constant camera speed, equation (8) is an hyperbola, and it can establish the three dimensional position of P with Hough transformation expressed in terms of d and $\tan\alpha_0$. Feature points on EPIs are transformed with Hough transformation, and relative maximums in the parameter space are detected. Then, feature tracks can be approximated.

3 Experimentations with real images

Experiments with real images were conducted. An image input system for the experiments was constructed with a camera on a manipulator capable of controlling 6 axes, as shown in Figure 7. With the camera going straight ahead, an experiments with and without rotation were preformed and the results compared.

The camera, with its optic axis parallel to the Z-axis, is moved straight and parallel with the X-axis.

200 real images were obtained at 1 millimeter intervals. Two spatiotemporal images were obtained; one using one-parallel-translation and one-rotation at 0.025^o per step, and one using only one-parallel-translation. Steps of the results of the various spatiotemporal image processing are shown in Figure 8. In Figure 8, the abscissa is denoted as $1/\tan\alpha$, so edge tracks become linear and features can be detected with the root mean square error fitting in this plane. To remove errors in the Hough transformation for estimating parameters d, s, we use the root mean square error fitting method for real images.

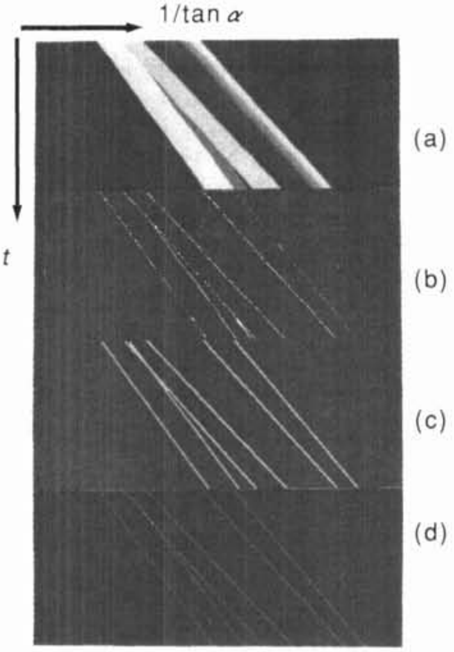

Figure8: Spatiotemporal image processing (a) EPI (b) Edge image (c) Root mean square error fitting (d)Detected lines.

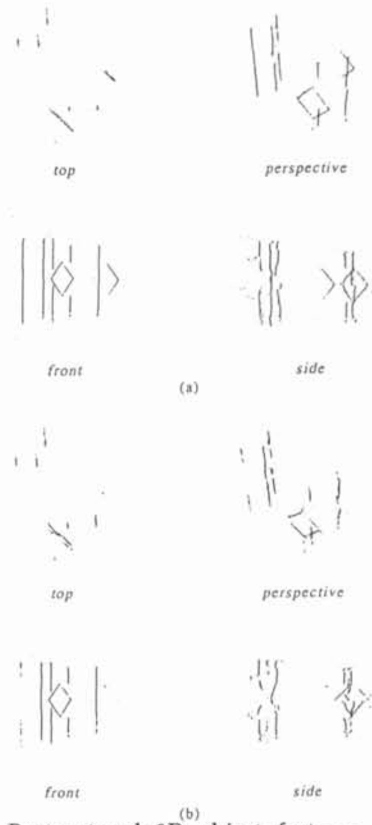

Figure9: Restructured 3D object features (a)Motion with 1 translation and 1 rotation (b)Motion with 1 translation.

Three dimensional structures of object features restructured are shown in Figure 9. Result (b) in Figure 9 has less edges in the right part and is more disarranged than result (a). This is because case (b) is that without rotation. Objects went out of the image plane earlier and accuracy was worse than in case (a). With camera rotation, it is possible to watch objects longer without missing any and to measure them with greater accuracy than without rotation.

4 Conclusion

This paper has proposed a technique for reconstructing a three-dimensional structure that makes allowances for wider range of camera motion with the Homocentric Spherical Spatiotemporal Image (HSSI) analysis. Shapes in images that are projected on a spherical surface do not vary with by rotation when the axis pierces the lens's focus. Using spherical projection transformation, it is possible to separate the rotating and translating effects from image-changes due to camera motion. Therefore, three dimensional structures can be reconstructed robustly with Hough transformation using the HSSI analysis.

Acknowledgments

The authors would like to thank Dr. Yukio Kobayashi, executive manager, and Kenichiro Ishii of NTT Human Interface Laboratories for their encouragement and helpful advice.

References

[1] R. C. Bolles and H. H Baker. *"Epipolar-Plane Image Analysis : A Technique for Analyzing motion Sequences,"* Proceedings of the Third Workshop on Computer Vision, pp.168–178, October 1985.

[2] D. H. Marimont. *"Projective Duality and Analysis of Image Sequences,"* Proceedings of the Workshop on Motion: Representation and Analysis, IEEE Computer Society, Kiawah Island, South Carolina, pp.7–14, May 1986

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

[4] M. Yamamoto. *"Determining 3-D structure of scene from image sequences obtained by horizontal and vertical moving camera,"* Lect Notes Comput Sci, Vol.301, pp.458–467, 1988.

[5] H. H Baker and R. C. Bolles. *"Generalizing Epipolar-Plane Image Analysis on The Spatiotemporal Surface,"* Proc. of CVPR, pp.2–9, 1988.

[6] T. Morita, Y. Yasukawa, Y. Inamoto, T. Uchiyama and S. Kawakami. *"Measurement in Three Dimensions by Motion Stereo and Spherical Mapping,"* Proc. of CVPR, pp.422–428, 1989.