MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

13—22

Enhancement of 3-D Information Acquisition from 2-D Spatio-Temporal Image

Chunxiao Li, Katsushi Ikeuchi and Masao Sakauchi *
Institute of Industrial Science
University of Tokyo

Abstract

In order to cope with the enormous amount of 3-D Spatio-Temporal Image data we have proposed a method to transform 3-D Spatio-Temporal Image into 2-D Spatio-Temporal Images to reduce processing time for 3-D information acquisition. In this paper, we will extend the camera motion to enhance our previous method for building 3-D information acquisition from urban street images. The creation of 2-D Spatio-Temporal Image and the theoretical analysis of 3-D information acquisition will be given first in the paper, then based on the error analysis slit selection for 2-D Spatio-Temporal Image creation will be discussed in the paper, experiments will reveal the accuracy and the efficacy of the approach.

1 Introduction

The research to acquire 3-D information of objects in the camera moving environment from a sequence of images taken by a camera while moving has been doing for a long time. 3-D Spatio-Temporal Image processing as an effective method to cope with the matching problem in stereo vision is well known [1][2][3][4]. However, 3-D Spatio-Temporal Image includes too large an amount of data, its processing needs too much processing time. In order to decrease the amount of successive image data and processing time, the research to transform 3-D Spatio-Temporal Image into 2-D images and then to obtain the 3-D information from the 2-D Spatio-Temporal image has been being developed[5][6][9]. We have developed a method to transform 3-D Spatio-Temporal Image into 2-D Spatio-Temporal Image using a hyperbolic slit[7]. It has been shown effective for building 3-D information acquisition from street image sequence. In this paper, we will change the camera direction when taking picture to extend the camera motion to enhance our previous method. The remainder of this paper is organized as the following: A theoretical analysis will be given first. After the analysis, slit selection and creation of 2-D Spatio-Image for 3-D information acquisition will be discussed. Several experiments will show the accuracy of the new method discussed in this paper. At the end, a conclusion with discussion of future work will be given.

*Address: 7-22-1 Roppongi, Minato-ku, Tokyo 106 Japan. E-mail: lcx@sak.iis.u-tokyo.ac.jp

2 Theoretical Analysis

When a camera takes pictures while moving along a straight line, the objects beside the path which the camera moves along will appear in the picture in different sizes, positions and even shapes if the camera faces different directions. If a camera moves and faces forward when taking picture, a point $P(X、Y、Z)$ in the world space is, as we have previously discussed[7], projected to the point $p(x, y)$ in the image plane as described in Equation 2. Now we discuss the case that the camera turns to a side while taking picture at a certain yaw angle as shown in Figure 1.

When the camera moves along the positive direction of Z axis and faces forward while keeping a certain yaw angle α to the axis, the projective Equation 2 becomes Equation 3 from the coordinate conversion as shown in Equation 1.

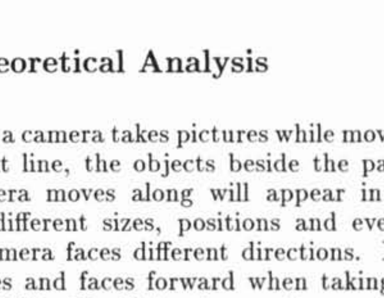

Figure 1: Camera Rotation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1)$$

$$\frac{Y}{y} = \frac{X}{x} = \frac{Z - vt}{f} \quad (2)$$

$$\frac{X\cos(\alpha) - (Z - vt)\sin(\alpha)}{x} = \frac{X\sin(\alpha) + (Z - vt)\cos(\alpha)}{f} \quad (3)$$

$$\frac{f}{x + f\tan(\alpha)} = \frac{\cos^2(\alpha)(Z - vt) + X\sin(\alpha)\cos(\alpha)}{X} \quad (4)$$

From Equation 3 we can get Equation 4. Here, we choose a hyperbola, as shown in Equation 5, as slit to create the 2-D Spatio-Temporal Image. Figure 8 shows a hyperbolic slit set on an image taken by a camera which is set with a certain angle to its moving path.

$$C = (x + ftan\alpha)(y + a) \quad (5)$$

Now we analyze the projection. From Equation 5 and 4 we get Equation 6.

$$y = -a + \frac{C\sin(2\alpha)}{2f} + \frac{ZC\cos(\alpha)^2}{Xf} - \frac{Cv\cos(\alpha)^2}{Xf}t \quad (6)$$

Here, when X, Z, v, f, c and α in the equation are constants, the Equation describes a straight line in a 2-D $y - t$ plane.

In the world space, a line which is with all the points on it having constant coordinates X and Z, and always has a crosspoint with the slit on the projective plane is a straight line parallel to the Y axis. According to Equation 6, as discussed previously, by extracting the straight lines from the the 2-D Spatio-Temporal Image we can calculate the X coordinate of the straight line in the world space from its inclination. Furthermore, the Z coordinate of the line can be derived from Equation 5 and 6 using this X coordinate and the y and t coordinates of some points on the line in the $y - t$ 2-D image.

Consequently, if we use the hyperbolic slit described in Equation 5 to transform the continuous images into a 2-D Spatio-Temporal Image, we can acquire the 3-D information about a straight line parallel to the Y axis in the moving environment by only extracting the projection of the straight line from the 2-D image.

In the same way, when camera takes pictures while moving along a straight line and facing forward with a certain angle to the moving direction in vertical direction, we can use a proper slit to transform the continuous images into an $x - t$ 2-D Spatio-Temporal Image, then to extracts straight lines from the 2-D image to acquire the 3-D information about straight lines parallel to the X axis in the world space.

3 Camera Rotation and Projective Distance

In the following, we discuss the advantage of camera rotation when taking pictures. For the convenience of explanation, we only investigate the case of creating $y - t$ 2-D Spatio-Temporal Image. The case of $x - t$ 2-D Spatio-Temporal Image basically is the same.

1. Extension of View Field

   When taking pictures, the camera is moving along a straight line in the horizontal plane as shown in Figure 2. Consider the case that the camera takes pictures while moving along the straight line from o_1 to LL and there are some objects in the moving environment. We can concentrate, for instance, on two objects box A and B. As shown in the figure, when the camera faces exactly forward as in the position o_1, due to the occlusion of the object A, the vertical edge line p of the object B would not be projected into the pictures. In contrast, when the camera is rotated with an angle α to the right of the moving path, as shown in position o_2, the straight-edge line p of the object B will appear in the pictures. Although it does not necessarily make the view field wide, it enables the object to clearly and completely appear in the pictures. This is what we need for acquiring the 3-D information of the objects in the moving environment from images. In fact, the objects are rarely on the way camera is moving along. They are mostly beside the way. When camera is turned to the objects, the objects will appear in the pictures, possibly larger and clearer than it is not turned.

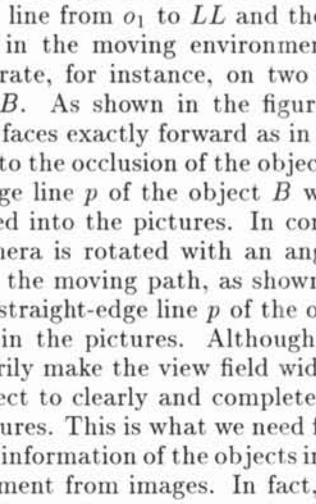

Figure 2: Extension of View Field

2. Reduction of Projective Distance

   Now we look at another factor of projection, Projective Distance. The farther an object is from the camera when a picture is taken, the smaller the projection of the object appears in the pictures, and the bigger the area of object surface which is projected into a pixel on the picture becomes. Figure 3 shows different projective distances.

Consider the case that a point P on the object AA is projected in the environment as shown in 3. The point P has a distance D to the moving line oy. Now we calculate the projective distance when the point P is projected on the vertical line which is dd far from the central vertical line in the picture. If the camera does not turn at an angle, as shown in the figure, the projective distance LL will be $D * csin(\alpha)$. When the camera turns to the right with an yaw angle β, the projection distance becomes $D * csin(\alpha + \beta)$. It is clear that the project distance became shorter.

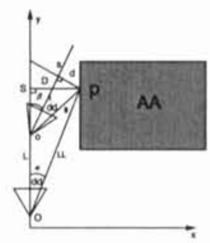

Figure 3: Reduction of Projective Distance

4 Error Analysis and Slit Selection

As we have discussed previously[7], there is error when the straight lines are extracted from 2-D Spatio-Temporal Image and the error has influence on the acquired 3-D information of the objects in the camera moving environment. In this section, we discuss the slit selection for the creation of 2-D Spatio-Temporal Image from continuous images taken by a camera which is set facing forward with a certain angle to the moving direction. Here we discuss as an example the case that the object on the right side of the path which camera moves along. The camera is set facing to the right at an certain angle α to its moving direction.

As discussed in Section 2, the X coordinate of a vertical line in the moving environment can be acquired by extracting its projection in the $y-t$ 2-D Spatio-Temporal Image. Now we consider the error of the 3-D information of the straight line which is caused by the error yielded when extracting the straight lines from the $y-t$ image.

$$\rho = xcos\theta + ysin\theta \qquad 0 \le \theta < \pi \tag{7}$$

From Equation 3 and 5, we have had Equation 6. And the correspondent relation between the inclination of the line 6 and the angle θ of the line 7, as shown in Figure 4, can be described with Equation 8. In the following we first calculate the error of 3-D coordinate X caused by the error of angle θ when extracting the straight line from 2-D image. Then we look at the selection of slit for creating 2-D Spatio-temporal Image.

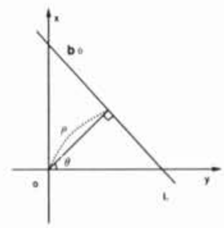

Figure 4:
A Line in a Plane

$$cot(\theta) = \frac{vc\cos(\alpha)^2}{Xf} \tag{8}$$

Now we calculate the error $\Delta\theta$ of the angle which occurred when the line is extracted from the 2-D Spatio-Temporal Image.

From Equation 8, we have:

$$X = \frac{vc\cos(\alpha)^2}{f} * tan(\theta) \tag{9}$$

$$\Delta X = \frac{vc\cos(\alpha)^2\Delta\theta}{f} * \frac{1}{cos^2(\theta)} \tag{10}$$

From the same Equation 8, we can also get:

$$\begin{aligned} &\frac{1}{cos^2(\theta)} \\ &= 1 + tan^2(\theta) \\ &= 1 + \frac{X^2 f^2}{v^2 c^2 \cos(\alpha)^4} \end{aligned} \tag{11}$$

Then from Equation 11 and 10, we get:

$$\begin{aligned} &\Delta X \\ &= \frac{vc\cos(\alpha)^2\Delta\theta}{f}\left(1 + \frac{X^2 f^2}{v^2 c^2 \cos(\alpha)^4}\right) \\ &= \frac{vc\cos(\alpha)^2\Delta\theta}{f} + \frac{\Delta\theta X^2 f}{vc\cos(\alpha)^2} \end{aligned} \tag{12}$$

The minimum value of ΔX is calculated as:

$$\frac{d(\Delta X)}{d(c)} = \Delta\theta\left(\frac{v\cos(\alpha)^2}{f} - \frac{X^2 f}{vc^2\cos(\alpha)^2}\right) = 0 \tag{13}$$

When

$$|c| = \left|\frac{fX}{v\cos(\alpha)^2}\right| \tag{14}$$

ΔX is minimum :

$$|\Delta X| = |2\Delta\theta X| \tag{15}$$

When ΔX is minimum

$$\theta = \frac{\pi}{4} \qquad (or \frac{3\pi}{4}) \tag{16}$$

From above equations we can see that when C equals $\frac{fX}{v\cos(\alpha)^2}$ the error of X caused by the error $\Delta\theta$ is minimum.

Because X is unknown we need to estimate it from the approximate position of objects when we select a hyperbolic slit. When parameter c is determined, we can adjust the slit's position using parameter a in the 2-D image to set the slit.

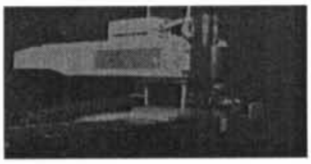

Figure 5: Device for Camera Operation

5 Experiments

This section gives several experiments to verify the accuracy of the presented method. In the experiments, the camera takes pictures of the street scene while moving on an urban road. Our experiments try to acquire the 3-D information about the buildings using the images taken by the camera. We have the camera run through the same road twice. The first time, the camera faces exactly forward while it is moving and taking pictures of buildings. The second time, we make the camera face forward while horizontally keeping a certain angle to the moving direction as we have discussed. From each sequence of continuous images taken on the road, we calculate the 3-D information about the buildings on the road, and then give a comparison of the results.

5.1 Experiment Device and Environment

In the experiment, the camera is set on a robot which has 5 moving freedoms. Figure 5 shows the robot. The camera is set horizontally and moves along a straight-line path. There are building models standing beside the robot's moving path. Pictures of the building models are taken while the camera is moving. The pictures are used as continuous images to create 2-D Spatio-Temporal Image to obtain the 3-D information about the building models.

5.2 3-D Information Acquisition of Buildings

1. Experiment 1

   In this experiment, the camera is set exactly forward while moving along a straight-line path. Figure6 is one of the pictures taken by the camera. There are 250 frames of such images taken. A hyperbolic slit is set on each of them. With

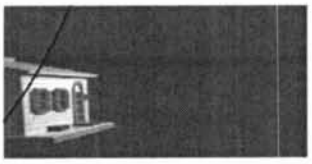

Figure 6: Hyperbolic Slit

the pixels on the slit of all frames we create the $y - t$ 2-D STI. Figure 7 shows the 2-D Spatio-Temporal Image created from the continuous images.

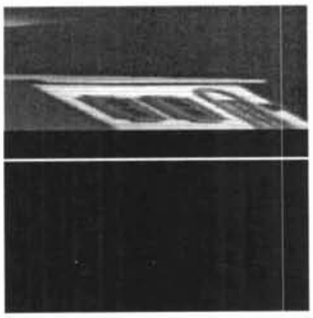

Figure 7: 2-D STI

After the creation of 2-D Spatio-Temporal Image as described previously we can calculate the 3-D information about the buildings on the road by extracting the projection of the straight-line edges of the buildings from the 2-D Spatio-Temporal Image. Figure 7 shows the results of extracting lines from the 2-D image. Using the parameters of the straight lines in the 2-D image we can obtain the 3-D information about the buildings. In Figure 10, the upper half of the figure shows the building surface recovered using the 3-D information of the building acquired from the 2-D Spatio-Temporal Image[8].

Figure 8: Hyperbolic Slit

2. Experiment2

   In the second experiment, we make the camera go through the same route as in the first experiment again. The only change is to set the camera obliquely, as shown in Figure 1. The angle α is 30 degree. 400 frames of the road scene images, as shown in Figure 8, are taken in the experiment. In order to create the 2-D Spatio-Temporal Image, we set a hyperbolic slit as described in Equation 5 on each frame. Then a $y-t$ 2-D Spatio-Temporal Image is created in the same way as done in the previous. Figure 9 shows the 2-D Spatio-Temporal Image created from the continuous images. The projection of straight-line edges of the building are then extracted. As discussed in Section 2, from the parameters of the straight lines we have been able to obtain the 3-D information of the buildings.

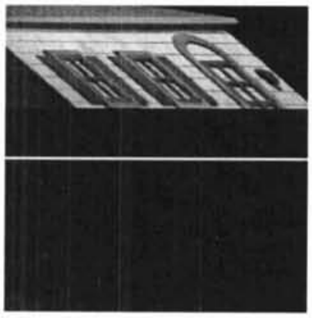

Figure 9: 2-D STI

Figure 10: Surface of Building

The lower part in Figure 10 shows the building surface recovered using the 3-D information of the building acquired from the 2-D Spatio-Temporal Image[8].

As shown in figure 10, the rotation of camera makes the projective distance short and enables us to get clear picture. In fact this makes it possible to acquire more precise 3-D information about the objects as seen in the experiments.

6 Summary

Building is a kind of man-made object existing typically and greatly in urban environment. In many vision application systems, building 3-D information acquisition and surface recovering from the environment scene are involved[10][11][12]. In order to improve our 2-D Spatio-Temporal Image processing method proposed previously the Extension of camera rotation in yaw was discussed in this paper. The rotation of camera spreads the view field of camera and decreases the possibility of taking the object into picture partially. The rotation of camera also reduces the projective distance. The shorter projective distance means that the picture can be taken clearer and we are able to obtain more precise 3-D information of the objects in the environment. The experiments shaw the efficacy. As future work, the accurate camera parameter acquisition and the application to real image can be considered.

References

[1] 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, Vol. 1, pp.7-55, 1987.

[2] H.H. Baker and R.C. Bolles , "Generalizing Epipolar-Plane Image Analysis on the Spatiotemporal Surface," International Journal of Computer Vision, Vol. 3, pp.33-49, 1989.

[3] T. Yasuno and T. Hamano, "3D Structure from motion using homocentric spherical spatiotemporal image analysis," Proceedings of IAPR Workshop on Machine Vision Applications," pp. 371-374, Tokyo, November 1990.

[4] T. Yasuno and S. Suzuki, "Surface Reconstruction Using Occlusion Analysis of Spatiotemporal Images," IJSP Trans., Vol. 34, No. 10, pp. 2174-2183, 1993.

[5] J.Y. Zheng and S. Tuji, "From Anorthoscope Perception to Dynamic Vision," Proc.IEEE Int. Conf. Robotics and Automation 1990.

[6] J.Y. Zheng, M. Asada and S. Tsuji, "Color-Based Panoramic Representation of Outdoor Environment for a Mobile Robot," Proc. 9th Inter. Conf. Patt. Recog., 801-803, 1988.

[7] C.X. Li, H.T. Zen, and M. Sakauchi, "3D Information Acquisition from Spatiotemporal Image Created by a Hyperbolic Slit", IAPR Workshop on Machine Vision Applications, pp.54-57, 1994.

[8] C.X Li, K. Ikeuchi and M. Sakauchi, "Building 3D Information Acquisition from 2D Spatio-Temporal-Image", International Workshop on Urban Multi-Media/3D Mapping 1998 (UM3'98).

[9] J.Y. Zheng, "Dynamic Projection, Panoramic Representation and Route Recognition," Ph.D Dissertation in Osaka University, Dec. 1989.

[10] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3-D Structures," IEEE Transections on Pattern Analysis and Machine Intelligence, Vol. 11, No. 11, 1989.

[11] P.L. Liu, W. Wu, K. Ikeuchi and M. Sakauchi, "Recognition of Urban Scene Using Silhouette of Building and City Map Database," Asian Conference on Computer Vision, 1998.

[12] C. Lin, A. Huertas and R. Nevatia, "Detection of Buildings Using Perceptual Grouping and Shadows," IEEE Proceedings of Computer Vision and Pattern Recognition, pp. 62-69, 1994.