3-D Tracking of a Moving Object by an Active Stereo Vision System

Maki TANAKA, Noriaki MARU and Fumio MIYAZAKI

Faculty of Engineering Science, Osaka University
Machikaneyama 1-1, Toyonaka, Osaka 560, Japan
phone: +81-6-844-1151 ex.4507; fax: +81-6-843-9730
e-mail: maru@me.es.osaka-u.ac.jp

Abstract - We present a 3-D tracking method of an object which is moving in the complicated scene by an Active Stereo Vision System. The system uses binocular vision robot, which can simulate the human eye movements. Holding gaze on an target object with the controlled cameras keeps the target's stereo disparity small, and simplify the visual processing to locate the target for pursuit control. The novel point of our tracking method is the disparity-based segmentation method of the target object. The method utilizes sero disparity filter (ZDF) and correlation to separate the target object with small disparity from distracting background. Furthermore, using correlation method to estimate stereo disparity makes it possible to fixate on a surface of the target object. We show the experimental results with the complicated scene to demonstrate the effectiveness of the proposed method.

INTRODUCTION

The controlled camera movements are becoming increasingly recognized as important capabilities in robotic visual perception (see [1][2]). In binocular systems, whose cameras have their optic axis in the same plane, binocular gaze holding is the process of adjusting pan angles so that both eyes are looking at the same world point. Binocular gaze holding of a moving object provides a number of benefits. For example, not only does it allow greater viewing time of a part of a scene or object, but also it allows the use of localized visual processing and makes a number of vision problems simpler and/or better posed[2]. Furthermore, the signals of the target object is emphasized over the background, because retinal images of a target object are stabilized, reducing motion blur, while surrounding distractors move rapidly across the retina and suffer from motion blur. In addition to it, active vergence control facilitates stereo fusion. The fixation point has a stereoscopic disparity of zero, and points nearby tend to have small disparities. This makes it possible to use stereo algorithms that accept only a limited range of disparities.

Coombs[4] proposed a method which isolates the target object from distracting background by maintaining vergence. Holding vergence on the target enables the object to be isolated by using simple ZDF(Zero Disparity Filter). ZDF is the logical AND of stereo vertical-edge images and detects objects at the fixation distance. On the other hand, Rougeaux[6] proposed a method which obtain the location of the target moving across the horopter by using ZDF and virtual horopter. Horopter is the set of points located on a circle passing through the two nodal points of the cameras and the gaze point (see **Fig.1**).

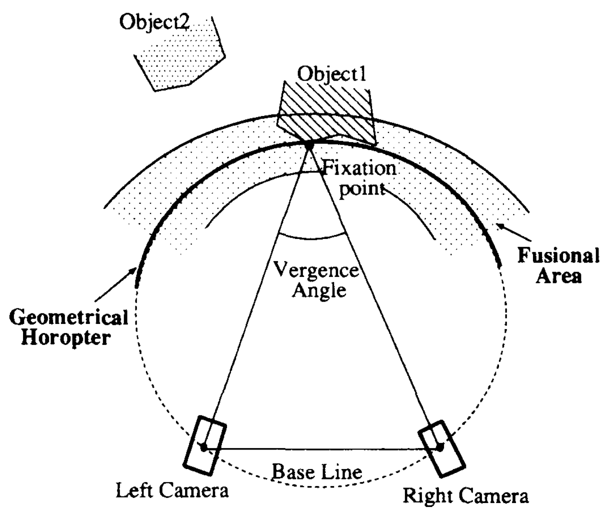


Fig.1 Geometrical Horopter.

The virtual horopter is the horopter generated by shifting horizontally the right image by a certain amount of pixel. Low shift values are equivalent to small virtual rotations of the right camera. Therefore, the location of the target moving across the horopter is obtained by comparing the outputs of ZDF on three virtual horopters. Although ZDF is an effective method which can obtain the location of the target in a simple scene, it tend to output many false zero disparities in the complicated scene, because it is just the logical AND of stereo vertical-edge images. Therefore, both Coombs method and Rougeaux method will fail in obtaining the location of the object in the complicated scene.

We present a 3-D tracking method of an object which is moving through the complicated scene. We propose a new disparity-based segmentation method of the target object to separate the target object from distracting background. The method utilizes ZDF and correlation to separate the target object with small disparity from distracting background. Furthermore, using correlation method to estimate stereo disparity makes it possible to fixate on the surface of the target object.

We have developed active stereo vision system to demonstrate the effectiveness of our tracking method. The system consists of vision robot which has two cameras mounted on a head. The cameras tilt together and pan independently on a common platform very precisely and the head pan and slide independently to demonstrate the human eye movements. We show the experimental results of the proposed 3-D tracking method with the active stereo vision system.

3D TRACKING OF A MOVING OBJECT

Estimation of the target location

Now let us consider that the stereo camera fixate on a target object as shown in **Fig.1**. The gaze-holding problem is to maintain fixation on a moving target. To do this, the errors in camera orientation must be determined, so the location of the target's image on the retina must be found. Our method exploits binocular cues and the fact that the cameras are actively following the target to avoid requiring object recognition. We assume that the target is initially at the fixation point. Any point of zero disparity projects onto the left and right image points with identical coordinates. The set of such points, which is called *horopter*, is located on a circle passing through the two nodal points of the cameras and the gaze point. Objects which stay in this horopter can be easily picked up from many other objects by suppressing features of non-zero disparity. Since during successful tracking the target is always near the gaze point, its images keep mostly zero disparity. We propose a new disparity-based segmentation method of the target object by combining the ZDF and correlation of the stereo images.

Let the images obtained from the left and right cameras be $L(x, y)$ and $R(x, y)$ respectively. We divide $L(x, y)$ and $R(x, y)$ into the partial images $L_{ij}(m, n)$ and $R_{ij}(m, n)$ whose size is $w \times h$ respectively as shown in **Fig.2** and let a pair of these images be Seg_{ij}. We define the correlation value C_{ij} of $L_{ij}(m, n)$

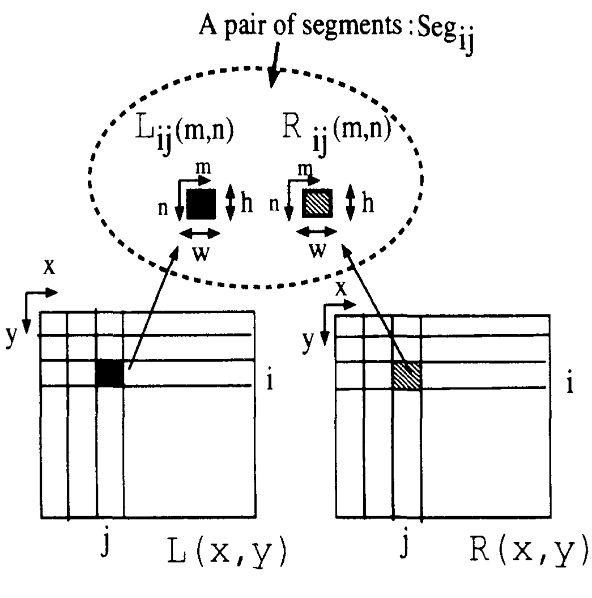


Fig.2 Segmentation of the Stereo images.

and $R_{ij}(m, n)$ as

$$C_{ij} = \sum_{w} \sum_{h} |R_{ij}(m, n) - L_{ij}(m, n)|. \tag{1}$$

If C_{ij} is greater than a threshold T_S, we can estimate that $L_{ij}(m, n)$ is not similar to $R_{ij}(m, n)$. Therefore, we can acquire the reliable zero disparity by inhibiting the corresponding output of ZDF. We assume that only one object is located on the common space of both cameras field of view and the horopter. We select the center of the gravity mass of the output of this filter which is a rough measurement of the target location as the fixation point (x_g, y_g).

Estimation of Binocular Disparity

We estimate the stereo disparity of the fixation point by using the correlation method to fixate on a surface of the target object. This is important to facilitate stereo fusion by limiting the range of binocular disparity. We take the reference image R_f whose size is $w' \times h'$ around the fixation point (x_g, y_g) in $R(x, y)$ and also take the search area S whose size is $w_s \times h_s$ around (x_g, y_g) in $L(x, y)$. We compute the correlation value between the reference image R_f and the all of the partial images L_s in S and find the most similar partial image L_f. Then we estimate the binocular disparity as the horizontal distance between L_f and R_f (see **Fig.3**).

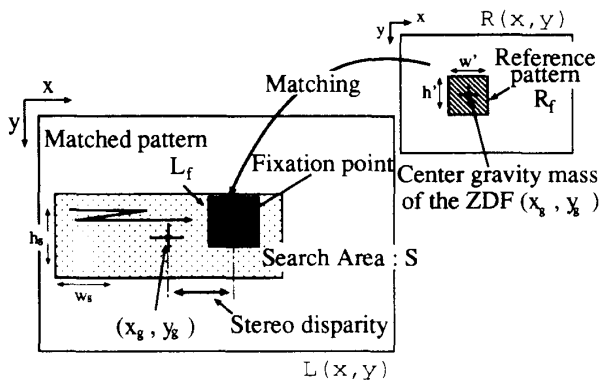


Fig.3 Estimation of Binocular Disparity by Correlation.

3-D Tracking procedure

By combining the methods mentioned 2.1 and 2.2, we can track a target object which is moving through a complicated environment. First of all, we teach the target object by fixating the object manually. Then 3-D tracking is realized by iterating the following procedure.

1. Divide the original stereo images into partial image pair: Seg_{ij}.
2. Segment the stereo images by computing the correlation values for each pair of partial images.
3. Compute ZDF with Seg_{ij} which does not satisfies $C_{ij} > T_s$.
4. Estimate target position in $R(x, y)$ by computing the center gravity mass of the ZDF output.

5. Estimate binocular disparity by correlation.
6. Decide the desired velocities of each motors.

The desired velocities of each motors are realized with PD control by each motor drivers. We show the flowchart of this procedure in **Fig.4**.

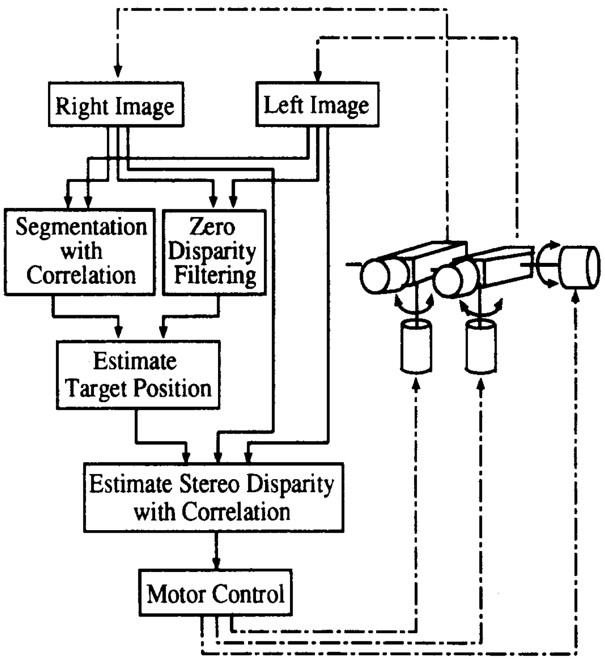


Fig.4 3-D tracking System.

ACTIVE STEREO VISION SYSTEM

We have developed an active stereo vision system to demonstrate the effectiveness of our tracking method. The active stereo vision system consists of the *vision robot*, motor control system and image processing system (see **Fig. 5**). We show the portrait of the vision robot in **Fig.6**. The cameras tilt up and down together on a common platform, and pan independently from side to side, driven by three motors, one for the tilt platform and twin pan motors. Besides the cameras are mounted on a head which pan and slide independently, driven by two motors, one for the pan of the head and one slide motor (see **Fig.7**).

The stereo images from the two CCD cameras(PULNiX TM540) are digitized into the frame memory by the image board (Concurrent Systems TRP-IMG) and transported to the DSP board (TRP-860) by the two Transputers(IMOS T805). Image processing (ZDF, segmentation, stereo disparity computation) are performed by the DSP (Intel i860) and send the desired velocity of each cameras to the Transputer(T805) for motion control. The Transputer output the motor command to the driver of the motor and both rapid and smooth velocity control is realized. The performance specifications are summarized in **Table 1**. This table tell us that this system has enough performance to simulate some movements of human eye. A mechanical advantage of our head-eye system is its simplicity: the

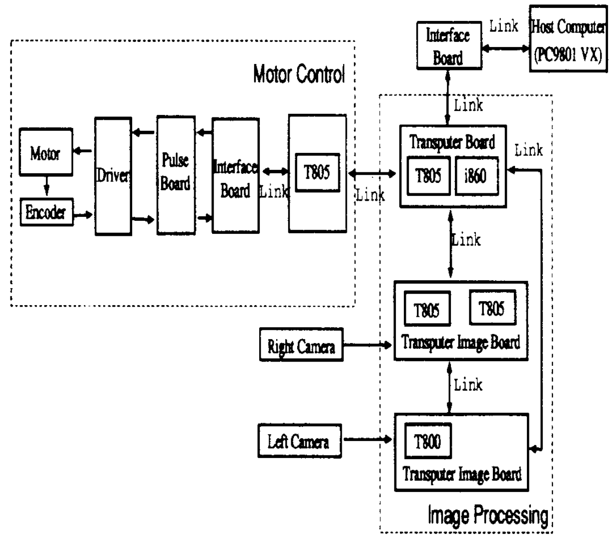


Fig.5 Configuration of the Active Stereo Vision System.

compact mechanism and fairly direct linkages facilitate both rapid saccade and smooth pursuit.

EXPERIMENT

We show the effectiveness of our tracking method by using the active stereo vision system. The focal length of the camera is $16mm$ and the baseline length is $120mm$. Central 176×192 portions of 256×240 gray scale images digitized on 8 bit per pixel are processed. The size of region for segmentation is 8×8(pixel) and the number is 22×24. The size of the reference image and search are in detecting binocular disparity is 16×16(pixel) and 40×20(pixel) respectively.

Fig.8 shows the original stereo images taken by fixating on a surface of the target object (doll). There are many vertical edges in the experimental scene. The target object is placed at the distance of about $1m$ from the vision robot. **Fig.9** show the output of ZDF. In this figure, the left and right figure denotes the case with and without segmentation respectively. In these figures, the black portion denotes the region which is removed from computing ZDF and white portion denotes the edge with zero disparity. ' + ' mark denotes the center of gravity mass of the ZDF output.

Fig.10 and **Fig.11** show the motor response while tracking a target object by moving a slide motor in a sine wave with an amplitude of $110mm$ instead of moving a target. **Fig.10** and **Fig.11** corresponds to the case with and without segmentation respectively. Error means the angle difference between the target position and the real positions of the cameras. In these figures, the bold line denotes the actual angle of the camera and the break line denotes the desired angle which is calculated by using the actual head motions. From these figures, we can see that the conventional method fail in tracking, but the proposed method success. This is because ZDF tend to output many false zero disparities when many objects exist behind a targe object from **Fig.9**.

We show the time-chart of this experiment in **Fig.12**. In this experiment, it takes about $167msec$ for the whole process.

Table 1 Specifications of Vision Robot

Motor	yaw	tilt	pan	slide
Maximum speed	600 deg/s	204 deg/s	270 deg/s	33.3 mm/s
Axis position sensing resolution	0.0072 deg	0.00409 deg	0.0327 deg	0.0008 mm
Ranges of motion	38 deg(in) 31 deg(out)	48 deg(down) 54 deg(up)	360 deg	550 mm

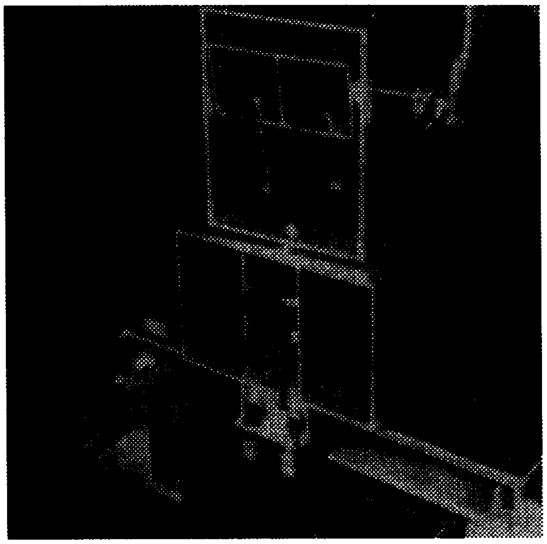


Fig.6 Portrait of the Vision Robot.

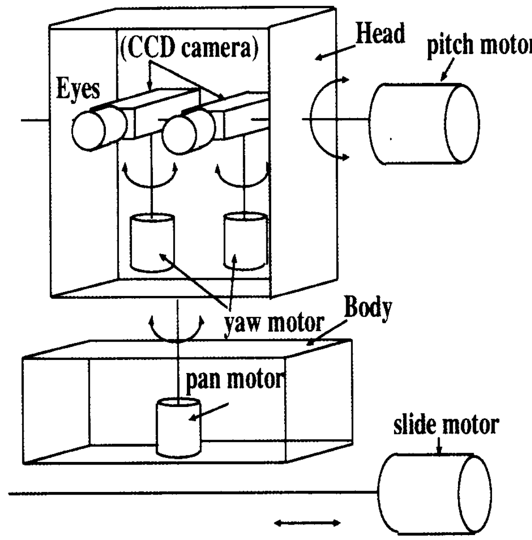


Fig.7 Hardware Configuration of the Vision Robot.

Details are $53msec$ for image transportation and $20msec$ for segmentation, $85msec$ for ZDF, $4msec$ for the computation of binocular disparity.

CONCLUSION

We have presented a 3-D tracking method of an object which is moving through the complicated scene with the active stereo vision system. The system uses the binocular vision robot, which can simulate the human eye movements. Holding gaze on an target object with the controlled cameras keeps the target's stereo disparity small, and simplify the visual processing to locate the target for pursuit control. The novel point of our tracking method is the disparity-based segmentation method of the target object. The method utilizes zero disparity filter (ZDF) and correlation to separate the target object with small disparity from distracting background. Furthermore, using correlation method to estimate stereo disparity makes it possible to fixate on the surface of the target object. We have shown the experimental results of the 3-D tracking with the complicated scene to demonstrate the effectiveness of the proposed method.

In this work, we have just selected the center of mass gravity of the ZDF output for the fixation point and found the binocular disparity of its point. However, there may be no enough features near the point for the correlation matching. We should select the fixation point so that there is enough features for the correlation matching.

References

[1] D.H.Ballard and C.M. Brown, "Principles of animate vision", *CVGIP: Image Understanding*, 1992.

[2] J. Aloimonos, A. Bandopadhay, and I. Weiss, "Active Vision", *Proc. of 1st Int'l Conf. on Computer Vision*, 35-54, 1987

[3] D. Coombs and C. Brown, "Real-Time Binocular Smooth Pursuit", *Int'l Journal of Computer Vision*, Vol.11 No.2, 147-164, 1993.

[4] D.J.Coombs and C.M.Brown, "Cooperative Gaze Holding in Binocular Vision", *IEEE Control Systems*, pp.24-33, June 1991.

[5] T.J.Olson and D.J.Coombs, "Real-Time Vergence Control for Binocular Robots", *Int. Journal of Computer Vision*, Vol.7, No.1, pp.67-89, 1991.

[6] S.Rougeaux, N.Kita, Y.Kuniyoshi and S.Sakane, "Tracking a Moving Object with a Stereo Camera Head", Preprints of 11th annual conference of Robotic Society of Japan , pp.473-476, November 1993.

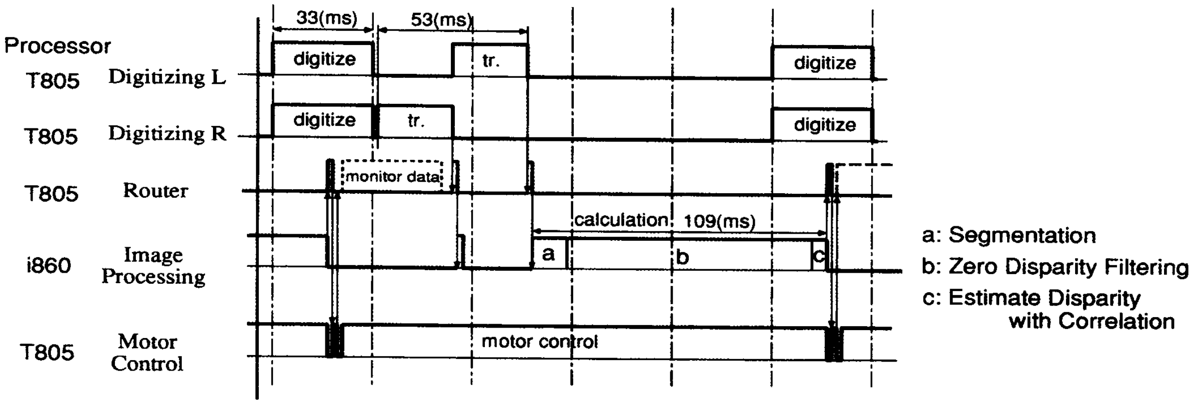


Fig.12 Timing Diagram of Tracking.

left image

right image

Fig.8 Original Stereo Images.

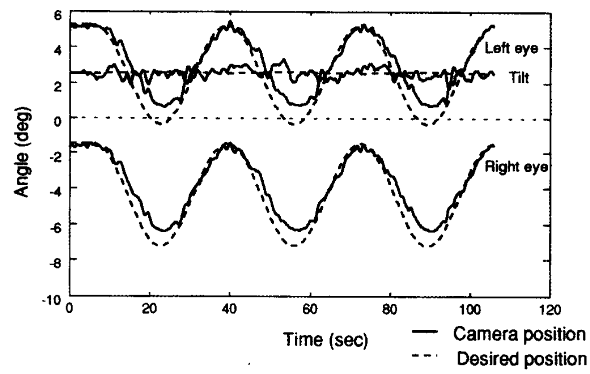
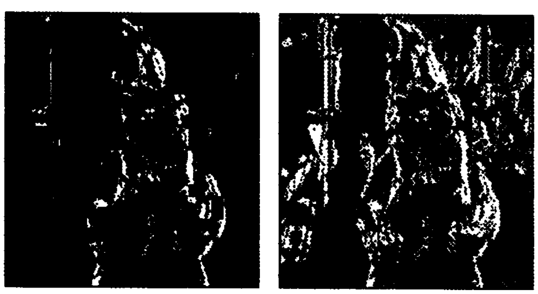


Fig.10 Camera Motions while Tracking a Moving Object with Segmentation.



with segmentation

without segmentation

Fig.9 Zero-disparity Images produced by ZDF

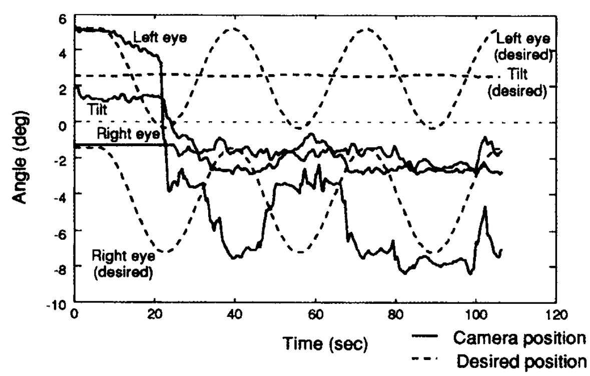


Fig.11 Camera Motions while Tracking a Moving Object without Segmentation.