Binocular Tracking Based on Virtual Horopters

†S. Rougeaux, N. Kita, Y. Kuniyoshi, S. Sakane, †F. Chavand

Autonomous Systems Section
Electrotechnical Laboratory
1-1-4 Umezono,Tsukuba,Ibaraki 305,JAPAN

†Laboratoire de Robotique d'Evry
Institut d'Informatique d'Entreprise
18 Allee Jean Rostand,91000 Evry,FRANCE

Abstract

This paper presents a stereo active vision system which performs tracking tasks on smoothly moving objects in complex backgrounds. Dynamic control of the vergence angle adapts the horopter geometry to the target position and allows to pick it up easily on the basis of stereoscopic disparity features. We introduce a novel vergence control strategy based on the computation of "virtual horopters" to track a target movement generating rapid changes of disparity. The control strategy is implemented on a binocular head, whose right and left pan angles are controlled independently. Experimental results of gaze holding on a smoothly moving target translating and rotating in a complex surrounding demonstrate the efficiency of the tracking system.

1 Introduction

The importance of eye movement to biological visual systems is obvious. In contrast, controlled camera movement have played a small role in computer vision research, but are becoming increasingly recognized as important capabilities in robotic visual perception [1, 2, 3, 4]. In fact, active motion of camera provide many advantages: Since tracking involves that the visual target remains near the center of the image, it allows the use of localized visual processing and stereo algorithms that accept only a limited range of disparity. Moreover, since the eyes follow the target, the target image tends to have slow motion across the retina. On contrary, surrounding distractors move rapidly across the retina and suffer from motion blur. As a result, the signal of the target is emphasized over the background.

In binocular systems, whose cameras have their optic axis in the same plane, *gaze control* is the process of adjusting pan angles so that both eyes are looking at the same world point. Gaze control may be broken into two different tasks: *gaze holding*, which involves maintaining a gaze point on a moving (or motionless) visual target from a moving (or motionless) gaze platform and *gaze shift*, which represents the ability of the vision module to transfer a gaze point from one visual target to another. We describe in this paper a method for holding gaze on a smoothly moving objects with a binocular head whose gaze point is controlled in real-time , and show experimental results.

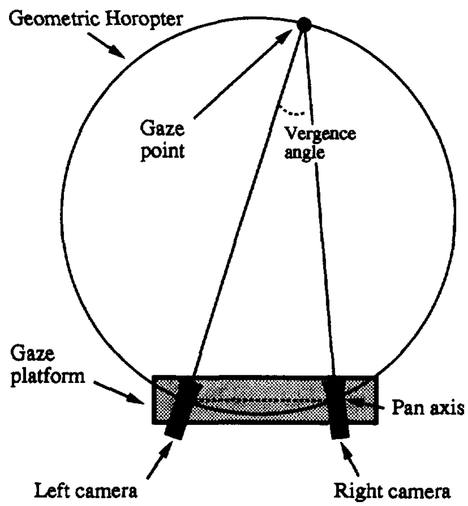

Figure 1: *View of binocular fixation.*

2 Tracking within a horopter

Our tracking algorithm is based on stereoscopic disparity features: we assume that the target is initially at a gaze point. Since during successful tracking the target is always near the gaze point, the projections of the target keep mostly zero disparity. By definition, any point of zero disparity projects onto left and right image points with identical coordinates. The set of such points (the geometric *horopter*, shown in figure 1) is located on a circle, also called the Vieth-Müller circle, passing through the two nodal points of the cameras and the gaze point. Thus, objects which stay in this *horopter* can be easily picked up from many other objects by suppressing features of non-zero disparity (fig. 2). This well known method has been in particular implemented for the Rochester robot head [5]. The principle of the zero disparity filter (ZDF) is simple: first vertical edges from both left and right images are extracted. Then, the stereo edge images are compared in corresponding pixel location. As a consequence, only edges of objects laying in the horopter may remain on the matching output. Considering that only one object is located on the common space of both cameras field of view and the horopter, the computation of the center of the gravity of the ZDF output is enough to obtain a rough measurement of the target location.

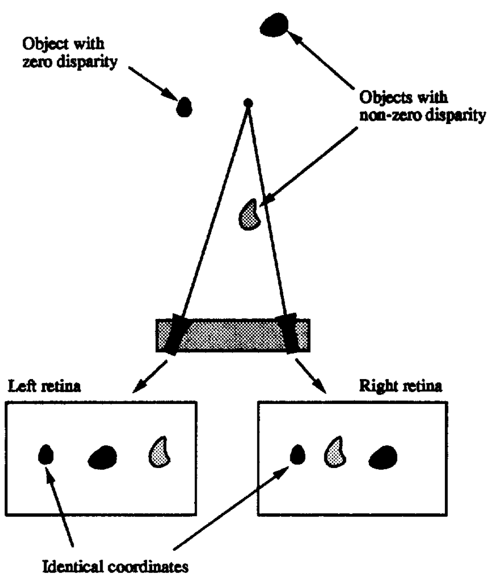

Figure 2: *Stereoscopic disparity.*

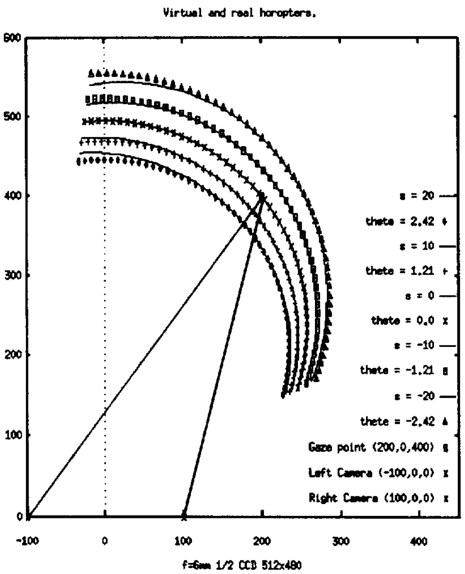

Figure 3: *Virtual horopters (solid lines) and the corresponding real horopters (dotted lines).*

3 The virtual horopter

Unfortunately, the strategy of zero disparity filtering does not work for object moving across the horopter. Recently, Coombs and Olson [6] developed a method based on *Cepstral filtering* [7] for measuring the disparity error, but experiments showed that this process requires the use of powerful hardware in order to keep a real-time feature. We propose a novel approach for the localization of object moving across the horopter. To cope with vergence error, ZDF is extended with a simple algorithm based on the computation of *virtual horopters*. This *virtual horopter* is the horopter generated by shifting horizontally the right image by a certain amount of pixel. As shown in figure 3, small shifts (s pixel) of the right image are almost equivalent to small *virtual* rotations ($\Delta\theta_r$) of the right camera. Here,

$$\Delta\theta_r = \tan^{-1}(s/f_x),$$

$$f_x = f \times h_{pixel}/h_{width},$$

f is the focal length, h_{pixel} is the pixel width of the images, and h_{width} is the horizontal length of the image planes.

Virtual horopters enable the quick test of the target position on the different horopters, because it doesn't need actual rotations of the cameras. The possible shift for the horizontal translation of the right image is estimated according to the ZDF output: firstly both left and right vertical edge images are blurred so as to have the approximately same edge width "w". This

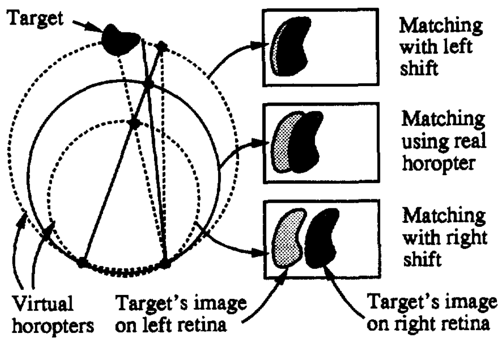

Figure 4: *Matching in the virtual horopters.*

results in making an overlap of the edges which have slight disparities, and the width of the overlapping is w for the edge on the horopter and decreases as the edge is getting far from the horopter. Thus, we can estimate the proper shift "s" from the overlapping width "w_o" of a vertical edge as follows.

$$s = w - w_o$$

The target, however, usually consists of many vertical edges, then we count the total number "N" of pixel in the ZDF output and convert this into a pixel shift "s" by the following equation.

$$s = w \times (1 - \frac{N}{N_{max}})$$

Here, N_{max} is the pixel number in the ideal case, that is when the target is on the horopter, and N_{max} is actually the pixel number on the ZDF output at the beginning of the tracking. "s" is used to translate the right vertical edge image in both horizontal direction. Doing so, three different matching, corresponding to one real and two virtual positions of the horopter (figure 4) are computed. As a result, in the horopter producing the best matching, target is located. The horizontal shift amount of the right image and the position of the gravity center are employed to generate proper control command for right and left pan angles. Thus the target motion comprising both motions across and along a horopter can be tracked by slight extension of ZDF, without such special hardware that has been used to execute Cepstral filtering in a real-time.

Besides realizing gaze holding, this method brings us several benefits. The first one is that usual motor control error problems are solved by the virtual rotation of the camera generating automatic and precise fixation on the ideal gaze point. Moreover, by taking the previous shift as a basis for the next frame computation, you can follow an object with small range motion without dynamic rotation of the cameras. Finally, switching from one horopter to another one instantaneously makes it possible to locate at the same time several objects laying in distinct horopters.

4 Gaze platform

Our gaze platform shown in figure 5 has two degrees of freedom corresponding to the two rotations of the cameras around the vertical axis. It is equipped with DC motors and potentiometers for pan angle range measurements. The stereo image acquisition is performed by two 1/2 inch CCD color cameras. However, due to the limitations of our image processing system, we only use gray scale information. This platform is a prototype designed to be easily manipulated and mounted on a robot's arm effector or on a small autonomous mobile robot. The table 1 shows its physical properties and dynamic performance of camera rotation.

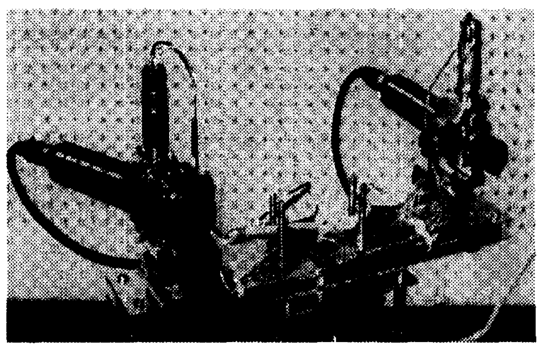

Figure 5: *The gaze platform.*

Table 1: Properties of the platform

Physical		Performance	
width	305 mm	range	±30 deg
height	150 mm	precision	±0.4 deg
depth	170 mm	max.speed	50 deg/s
weight	1.0 kg		
baseline	200 ∼ 250 mm		

5 Tracking system

The tracking system is managed by MaxTD, a workstation composed by a LynxOS kernel and Maxvideo 200, a real-time image processing hardware. Stereo image acquisition is performed by Digicolor using red and green components of the video signal. Figure 6 shows the flow of data in the system between the different devices. The workstation is also connected through a serial port to a controller board which forms a local servo loop for managing cameras rotation around pan axis. The servo loop is controlled by a classic PD controller and the feedback frequency is 200 Hz.

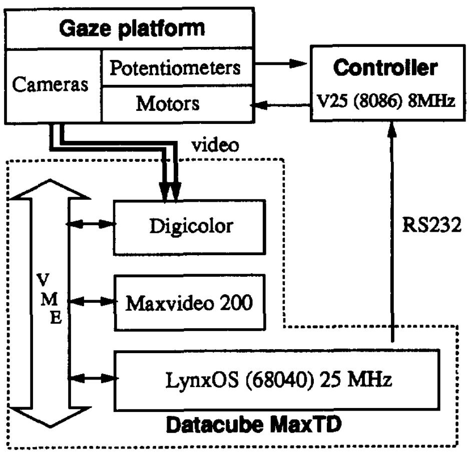

Figure 6: *Overview of the stereo tracking module.*

Central 260x180 portions of 512x480 gray scale images digitized on 8 bit per pixel are processed by Maxvideo 200. Frame acquisitions, edge extractions and three ZDF processing are performed within the continuous three frame time in the pipeline fashion. Then the global visual feedback to the PD controller is given with 30 Hz and the latency is about 90 msec.

6 Experimental results

In the following experiments, the interocular length was set to 200mm and 6mm focal length lenses were used. In the first experiment, a turn table was set in front of the gaze platform at the distance of 1 meter and a target was put on the turn table at 30 cm from the center. Figure 7 shows motor response while tracking the target when the turn table was rotating at 0.4 Hz. Error means the angle difference between the optimal (target) position and the real positions of the cameras. This shows the system track a target with the latency of 90 msec. After many trials with various rotating speeds of the turn table and distances between the target and the center of the table, the speed limits which the system can track were about $50^o/sec$ for the movements along the horopter and about $15^o/sec$ for the movements across the horopter, in terms of the angular speed of only right camera rotation. The reason of the former limitation is that the maximum speed of camera rotations are $50^o/sec$. The latter limitation corresponds to the 4 pixel shift in one frame time, which yields 4 pixel disparity between the continuous images. Since the average width of blurred edges during ZDF processing was set to 6, this demonstrates that the system could correctly estimate the vergence error smaller than about 70 % of the blurred edge width.

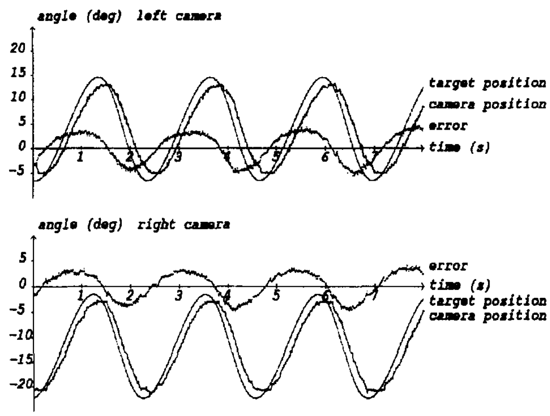

Figure 7: *Motor response while tracking a target on a turn table.*

Various shaped objects, not only objects consisting of vertical edges like a pencil put vertically, but objects like a miniature car or a doughnut, could be tracked with several distractors (figure 8). This demonstrates that the ZDF method based on vertical edges has fairly well generality about the target shape and capable to discriminate the target from backgrounds or distractors. Only when a lot of vertical edges besides ones of the target appeared in both fields of views of right and left cameras, the system sometimes mistook to track the target. This is caused by the accidental matches of edges in the ZDF output even if those object are not in the horopter. Though this happenings will be able to be avoided by using richer property of matching features, this will lead much increasing of computational load.

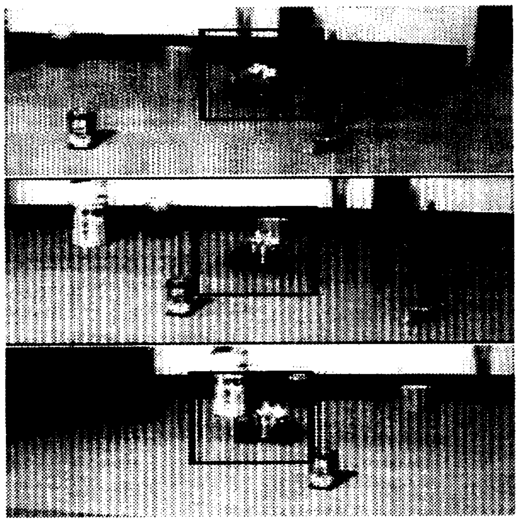

Figure 8: *Three different frames of the left camera during a tracking sequence.*

This system also roughly estimates the target position in the 3D world environment (± 6 cm in depth for a target moving at 1 meter from the baseline of the gaze platform)(figure 9). The inaccuracy is essentially caused by the noisy output of the potentiometers reporting cameras pan angle rotation: it has around $\pm 0.4^{o}$ error.

7 Conclusion

We proposed a novel method based on zero disparity filtering for estimating vergence error during tracking of moving object. By using the method, the estimation of the target position became able to be obtained simultaneously with the discrimination of the target from backgrounds or distractors, which also rely on zero disparity filtering. This simplified the tracking system very much in comparing with the conventional system, which demanded two different kinds of image processing modules for the discrimination and the vergence error estimation. Since this enabled much efficient visual feedback, tracking performance was eventually improved.

We implemented the proposed method on our active vision system. The tracking experiments demonstrated that the system had the ability to track various shaped objects in the fairly complicated backgrounds and with several moving distractors. Now the active head has been mounted on a small mobile robot, and the tracking capability has been used as visual active sensor for the several kinds of cooperative behaviors [8].

References

[1] R. Bajcsy: "Active Perception vs. Passive Perception" , *Proc. of IEEE Workshop on Computer Vision*, 55–62, 1985.

[2] J. Aloimonos, I. Weiss and A. Bandyopadhyay: "Active Vision" , *Proc. of First ICCV*, 552–573, 1987.

[3] J. K. Tsotsos: "Active vs. Passive Visual Search: Which is more Efficient?" , *Tech. Rep. on Univ. of Toronto*, RBCV-TR-90-34, 1990.

[4] Dana H. Ballard: "Animate Vision" , *Artificial Intelligence*, no.48, 57–86, 1991.

[5] P. von Kaenel, C. M. Brown and D. J. Coombs: "Detecting Regions of Zero Disparity in Binocular Images" , *Tech. Rep. of University of Rochester*, TR388, 1991.

[6] T. J. Olson and D. J. Coombs: "Real-Time Vergence Control for Binocular Robots" , *Tech. Rep. of University of Rochester*, TR348, 1990.

[7] Y. Yeshurun and E. Schwartz: "Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation" , *IEEE Trans. Pattern Anal. Machine Intelli.*, vol.PAMI-11, no.7, 759–767, 1989.

[8] Y. Kuniyoshi, S. Rougeaux, M. Ishii, N. Kita, S. Sakane and M. Kakikura: "Cooperation by observation –The framework and basic task patterns–" , *to appear in Int. Conf. on Robotics and Automation*, 1994.

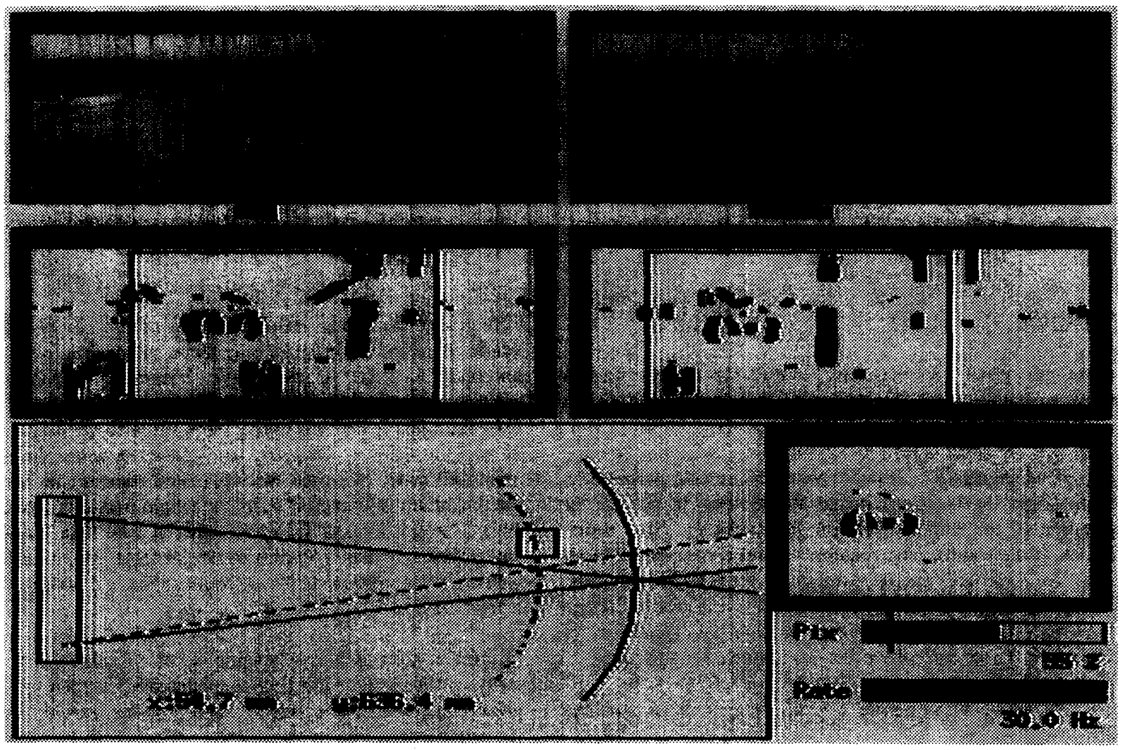

Figure 9: *View of the monitoring display. The rectangles drawn on the edges images show the input matching areas for the ZDF processing. You can notice that the right image has been shifted to the left in order to correct the vergence error. The lower left corner of the display shows the estimated target position and both* real *(continuous line) and* virtual *(dotted line) horopters. This figure clearly shows that the computed virtual horopter is well adapted to the target position.*