Binocular Tracking using Log Polar Mapping

Naoki Oshiro*, Noriaki Maru**, Atsushi Nishikawa* and Fumio Miyazaki*

*Faculty of Engineering Science, Osaka University
1-3 Machikaneyama-cho, Toyonaka city, Osaka 560, JAPAN

**Faculty of Systems Engineering, Wakayama University
930 Sakaedani, Wakayama 640, JAPAN

E-mail: oshiro@robotics.me.es.osaka-u.ac.jp

Abstract

This paper describes a new binocular tracking method using Log Polar Mapping (LPM) which approximately represents the mapping of the retina into the visual cortex in primate vision. Using LPM makes it possible not only to obtain both a high central resolution and a wide field of view, but also to significantly reduce processing image data. In this paper, LPM is performed in software by lookup table method. Our tracking method utilizes zero disparity filter (ZDF) for extracting the target object and virtual horopter method for estimating binocular disparities, respectively. The performance of both target extraction and disparity estimation is improved in comparison with the conventional methods, by using LPM. Some experimental results are also shown to demonstrate the effectiveness of the proposed method.

1 Introduction

In primate visual systems, the retina has a space variant structure; the resolution is high at the central region and becomes decreasingly low toward the periphery. Meanwhile, the visual cortex to which the retinal structure is mapped has a relatively space invariant structure. It is well-known that the mapping of the retina into the visual cortex can be approximately modeled by "*Log Polar Mapping*" (henceforth LPM) [1]. LPM can significantly reduce processing data in comparison with uniform sampling, because of the logarithmic radial compression. As a result, LPM makes it possible to realize a real-time robot vision system with both a high central resolution and a wide field of view.

In the previous works, we have developed an active vision system which can simulate the human eye movements [2] and proposed a method for using LPM to extract a gaze object by the vision system [3]. In this paper, we apply LPM to binocular tracking of a gaze object. LPM is performed in software by lookup table method. We use zero disparity filter (ZDF) [4] for extracting a gaze object, and virtual horopter method [5] for estimating binocular disparities, respectively.

Our tracking method based on LPM has several merits in comparison with the conventional methods:

- The number of stereo correspondence errors by ZDF is relatively small even in the case that there are many vertical edges in the background because peripheral information such as the background is globally dealt with.

- The detectable range of disparities of the target object is relatively large in the periphery because the zero disparity area of LPM images spreads from the gaze point toward the periphery.

In order to use ZDF, vertical edges must be detected from stereo images. The conventional vertical edge detectors such as vertical Sobel operator, however, are not applicable for LPM images because of their nonuniform pixel size, shape and connectivity. To overcome this problem,

- We introduce a new vertical Sobel operator for LPM images based on the connectivity graph (CG) [6] which can represent the connectedness relations in the LPM image.

All the processing for LPM images is performed in software. Some experimental results are also shown to demonstrate the effectiveness of the proposed method.

Proc. IROS 96
0-7803-3213-X/96/ $ 5.00 © 1996 IEEE

2 Zero Disparity Filter

The zero disparity filter (ZDF) is a well-known method of using a binocular visual system to extract objects on the zero disparity area (=horopter). The horopter is a circle which passes through the gaze point and each center of both eyes. When the target object exists on the horopter, both projected coordinates to the stereo images are the same. Therefore, we can easily extract and track the object on the horopter as follows: (i) vertical edges are detected from stereo images, (ii) the stereo edge images are matched by AND operation, (iii) as a result, edges on the horopter remain in the matching output. (iv) In order to track the object, eyes are moved so that the gaze point may be consistent with the matching point (= the 3D position of the centroid of the matching output).

This method, however, cannot extract objects moving from inside to outside of the horopter — in this case, tracking would fail. It is only applicable to the objects moving within the horopter †. In order to solve this problem, Rougeaux et al. [5] have developed the *extended* zero disparity filter (EZDF) based on the computation of *virtual* horopters. When the tracking fails, the target object are projected to stereo images with some disparities. In general, the disparity can be canceled by using one of the following two ways: ① *real* eye movements (convergence or divergence), ② *virtual* eye movements (positive or negative horizontal image shifts). The latter method, "virtual horopter method", is more efficient because it needs no *real* instrumental movements and reduces the cost for "vertical edge detection" to a minimum. In this method, after shifting the image horizontally, the ZDF method is applied to the resulting images in which disparities are virtually generated. The shift value corresponding to the maximal ZDF output is selected as the disparity of the target. It is suitable for a real-time robot vision system because its major process is only image translations whose computational complexity is very simple.

We construct our tracking algorithm by applying EZDF to LPM stereo images instead of original stereo images. Owing to the LPM horopter characteristics, our method can perform more appropriate stereo matching. This will be described below (see Section 5.2). Another difference between Rougeaux's method and ours is that our method uses multiple shift images at the same time to estimate the disparities. As a result, this produces "robust" disparity estimation.

†Therefore, the target object must be gazed (or held within the horopter) initially.

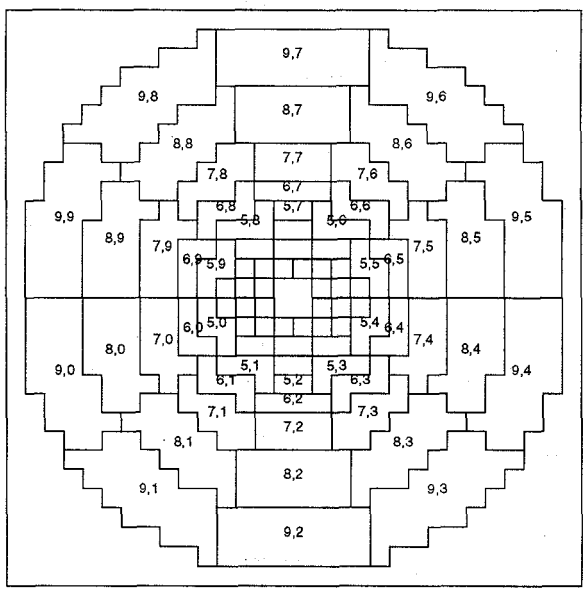


Figure 1: Lookup Table Representation of LPM ($30 \times 30 \rightarrow 10 \times 10$)

3 Log Polar Mapping

Log Polar Mapping (LPM) is formulated as follows: Let $I(x, y)$ be the original image and $L(u, v)$ be the LPM image to which $I(x, y)$ is transformed by LPM. Treating each pixel coordinate (x, y) and (u, v) as a complex number $z = x+iy$ and $w = u+iv$ respectively, LPM is

$$w = \ln(z + \alpha), \tag{1}$$

where α is the shift parameter to avoid the singularity at the origin ($\alpha = 0$ in this paper).

In practice, we represent LPM in a lookup table form [6]:

$$(u, v) = \mathrm{LUT}(x, y), \tag{2}$$

where LUT means the lookup table for an original image $I(x, y)$.

Figure 1 illustrates the lookup table that relates 10×10 LPM image with 30×30 original image.

4 LPM Image Operators based on Graphical Representation

Wallace et al. [6] have proposed some graphical representations for space variant image processing: They developed the *connectivity graph* (CG) which represents the connectedness relations in arbitrary space variant images including LPM images, and also introduced *transformation graph* which represents the ef-

fects of image transformations such as translations and rotations. In this section, we introduce two LPM image operators: *vertical edge detector* and *image translator* by using and extending these graphical representations.

4.1 Construction of Graph Structure

First of all, we explain how to construct a graph $G = (V, E)$ whose vertices V are associated with each mapping area and whose arcs E † present the adjacency relations between vertices. Each vertex $p = V(u, v)$ corresponds to each mapping coordinate (u, v) and has the following attributes: ① mean intensity $L(p)$, ② centroid $\mu(p)$, ③ area $a(p)$. Using these notation with respect to p, LPM forward and inverse map are respectively described as follows:

Forward map:

$$L(p) = \frac{1}{a(p)} \sum_{i,j} I(i,j) \quad | p = V(\mathrm{LUT}(i,j)), \qquad (3)$$

Inverse map:

$$I(i,j) = L(p) \quad | p = V(\mathrm{LUT}(i,j)). \qquad (4)$$

Note that all of the image processing are done in the LPM image, although we show inverse LPM images as examples or results in this paper. Now we denote the set of vertices associated with vertex p by $\{q | \mathcal{N}(p)\}$, and the arc that represents the adjacency relations between vertices p and q by $E(p, q)$. Notice that $\{q\}$ includes p itself.

Then, for convenience, we introduce the *directed arc* having weight (see Fig. 3). Denoting the LPM images before and after operation such as edge detection or image translation by L and L', respectively, and the weight for directed arc $E(p, q)$ whose direction is from p to q by w_q, we have

$$L'(p) = \sum_q w_q L(q) \quad | (p, q) \in E. \qquad (5)$$

4.2 Vertical Edge Detector

In order to detect vertical edges in LPM images, we developed a new vertical Sobel operator based on the connectivity graph described above.

The method for constructing this operator is summarized as follows [3]:

1. Make the connectivity graph $G_c = (V_c, E_c)$. Figure 2 shows the CG for vertex $V(7, 6)$ in Fig. 1.

†Although Wallace et al. [6] called this element as '*edge*', we call it as '*arc*' for avoiding the confusion with vertical *edges*.

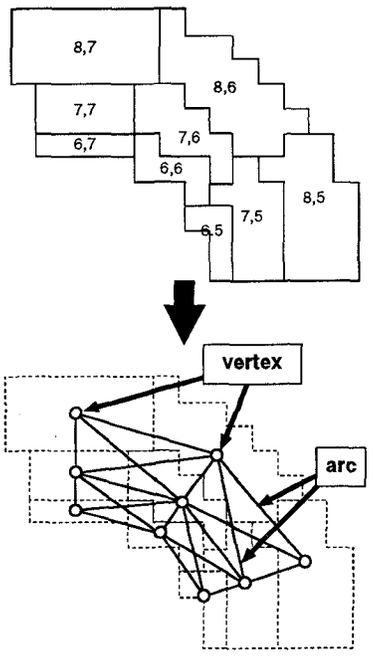


Figure 2: Construction of Connectivity Graph (CG)

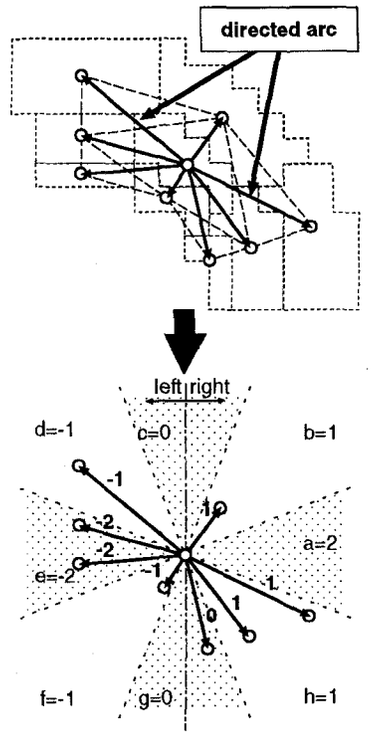


Figure 3: Construction of Vertical Sobel operator for LPM image

2. Based on the G_c, construct the graph G_{VSobel} whose weights are consistent with those of the conventional 3×3 vertical Sobel operator for rectangular coordinate images (see Fig. 3). As a result, weight w_q for each directed arc $E_c(p,q)$ is decided.

3. Normalize each weight w_q as follows:

   (i) Classify $\{q\}$ into $\{q_r\}$ (if q locates right with respect to p) or $\{q_l\}$ (otherwise).

   (ii) Calculate W_{q_r} and W_{q_l}, respectively:

$$W_{q_r} = \sum_{q\in\{q_r\}} w_q, \quad W_{q_l} = \sum_{q\in\{q_l\}} w_q. \tag{6}$$

   (iii) Normalize each weight w_q by using either W_{q_r} (if $q \in \{q_r\}$) or W_{q_l} (otherwise). This is a kind of compensation for unbalanced weights due to the space variant structure of LPM.

4.3 Image Translator

We use the virtual horopter based on the horizontal image translation to estimate disparities. Using the translation graph [6], image translation operation in LPM images $(i-\Delta i, j) \to (i,j)$ is given by

$$\begin{aligned} L'(q) &= \frac{\sum \kappa_T(p,q)L(p)}{\sum \kappa_T(p,q)}, \quad \kappa_T(p,q) = \sum_{i,j} 1 \\ &\mid (p,q) \in E_T, \\ &\quad p = \mathrm{LUT}(i-\Delta i, j),\ q = \mathrm{LUT}(i,j). \end{aligned} \tag{7}$$

Therefore, the weight w_q of arc $E_T(p,q)$ is

$$w_q = \frac{\kappa_T(p,q)}{\sum \kappa_T(p,q)} \quad \mid (p,q) \in E_T. \tag{8}$$

5 Binocular Tracking

5.1 Outline of the Binocular Tracking Process

We describe the outline of the binocular tracking process (see Fig. 4). Firstly, as the preprocessing step, lookup table (LUT), edge detector, and image translator are constructed, respectively (①~③ in Fig. 4). Now we suppose that the target object is initially at a gaze point †. Online binocular tracking process consists of the following steps (④~⑨ in Fig. 4): Firstly, stereo images are taken by the stereo camera (④) and then transformed into stereo LPM images by LPM (⑤). Next, vertical edges are detected from these LPM images by the edge detector (⑥). The ZDF method is applied to the resulting stereo edge images. Meanwhile, one side of these images is shifted horizontally not only to the right (minus virtual horopter) but also to the left (plus virtual horopter) by the image translator so that several stereo images with virtual disparities are generated (⑦). The ZDF method is also applied to these stereo images, respectively. As a result, we obtain multiple ZDF images. Furthermore, in order to improve the accuracy of matching, the resulting edge images are compared with intensity LPM images. Then, maximal matching output is selected from the ZDF images (⑧) and its centroid is calculated††. As a result, the target location is easily estimated (⑨). Finally, camera angles for gaze holding are calculated based on the estimated centroid and the shift width of virtual horopters. The vision system goes on tracking the target object by repeating all of these processes. Figure 5 shows an extraction result of a gaze object by ZDF. In this example, 366×416 original images were transformed into 50×50 LPM images.

5.2 Comparison with the Conventional Methods

Our tracking method based on LPM has several merits in comparison with the conventional methods using rectangular images:

1. In the conventional methods using rectangular images, ZDF may produce a lot of stereo correspondence errors when there are many vertical edges in the background. On the other hand, in our method based on LPM images, the number of correspondence errors by ZDF tends to be relatively small because peripheral information corresponding to the background is globally dealt with. Figure 6 shows the comparison with the conventional ZDF method. These are the ZDF output images whose input is Fig. 5(a). Notice that several vertical edges exist periodically in the upper right-hand part of the original images. We can see that the conventional ZDF (Fig. 6(a)) outputs

†This assumption means that our method is applicable after succeeding in detecting and gazing the target object by utilizing other methods. On realizing these processes, LPM may be also useful, because of its wide field of view and low data size. (this solution is beyond the scope of this paper).

††Denoting the matched vertices in stereo images by p_m, the centroid μ_{ZDF} is given by

$$\mu_{\mathrm{ZDF}} = \frac{\sum \mu(p_m)a(p_m)}{\sum a(p_m)}.$$

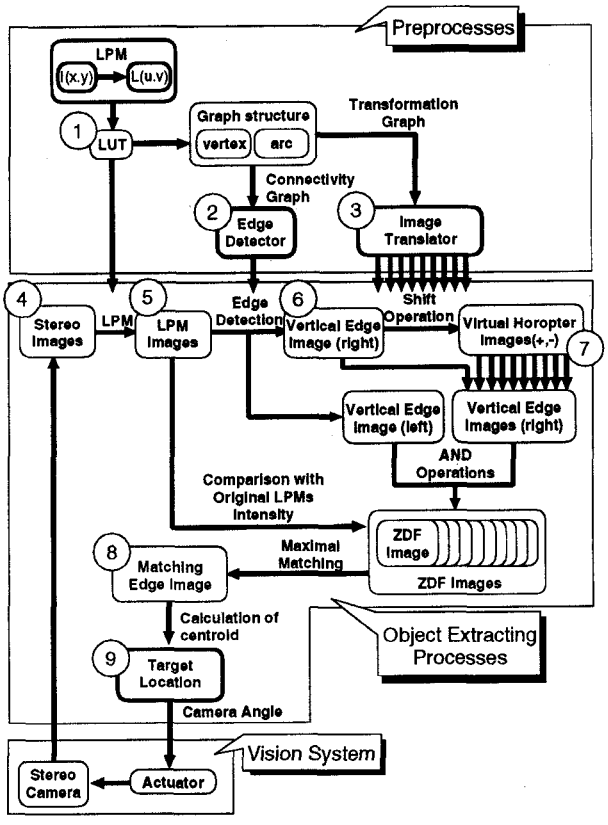


Figure 4: Diagram of the Binocular Tracking Process

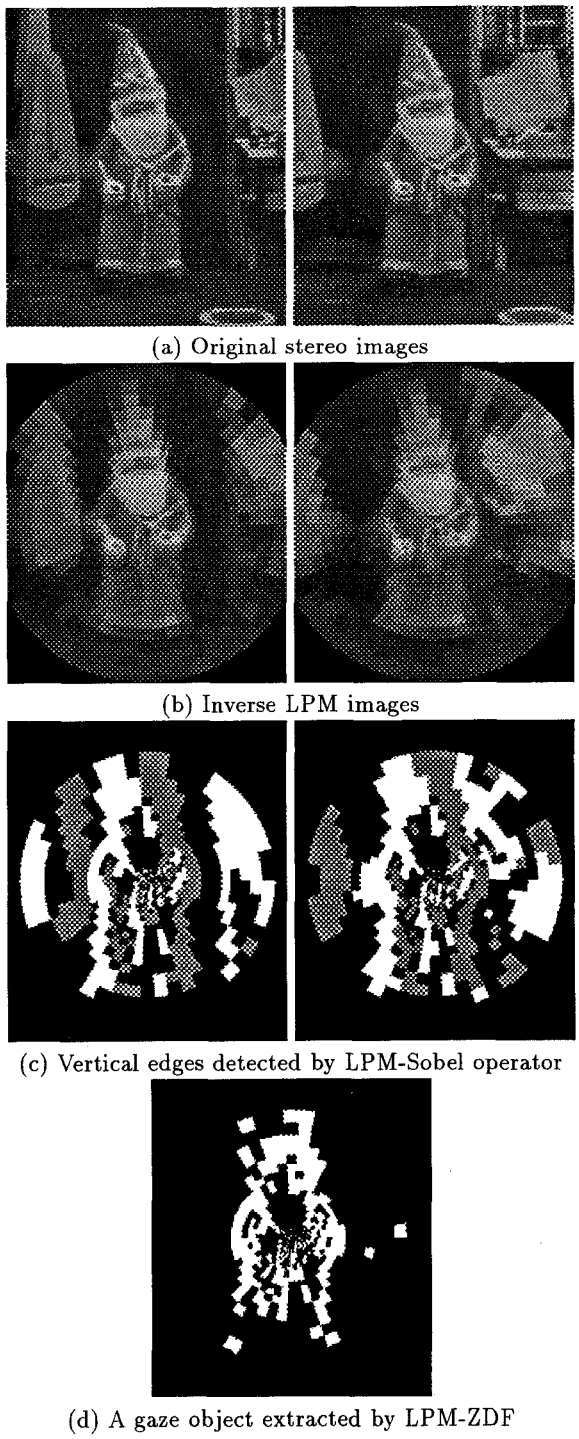
many false zero disparities (error rate ††† 10.1%), on the other hand, our method (Fig. 6(b)) hardly produces such false matches (error rate 0.4%).

2. Zero disparity area of LPM images spreads from the gaze point toward the periphery as shown in Fig. 7 (This figure illustrates Fig. 5's case). Consequently, ZDF can produce more stereo correspondence in comparison with applying to rectangular images, because the detectable range of disparities of the target object is relatively large in the periphery. This property is effective for detecting tracking objects and suitable for utilizing peripheral information.

6 Experiment

Now we show some experimental results of binocular tracking (target extraction and disparity estimation) to demonstrate the effectiveness of the proposed method.

†††Error rate is given by dividing the number of false matches by the total number of matched points.



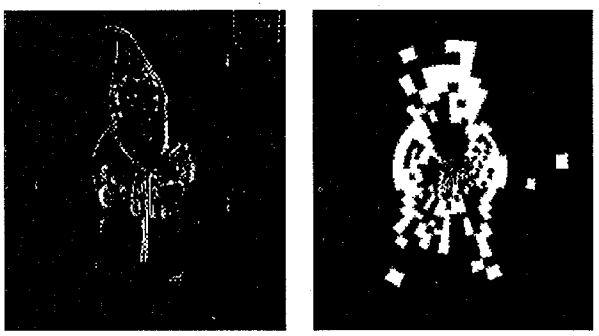
(a) Original stereo images

(b) Inverse LPM images

(c) Vertical edges detected by LPM-Sobel operator

(d) A gaze object extracted by LPM-ZDF

Figure 5: Extraction Results of a Gaze Object by LPM-ZDF ($366 \times 416 \rightarrow 50 \times 50$): Bright and dark pixels indicate vertical edges whose values are positive and negative respectively, in (c).



(a) ZDF (c) LPM–ZDF
correct matches (pixel):2916 correct matches (pixel):402
false matches (pixel) :329 false matches (pixel) :2

Figure 6: Comparison with the Conventional ZDF method

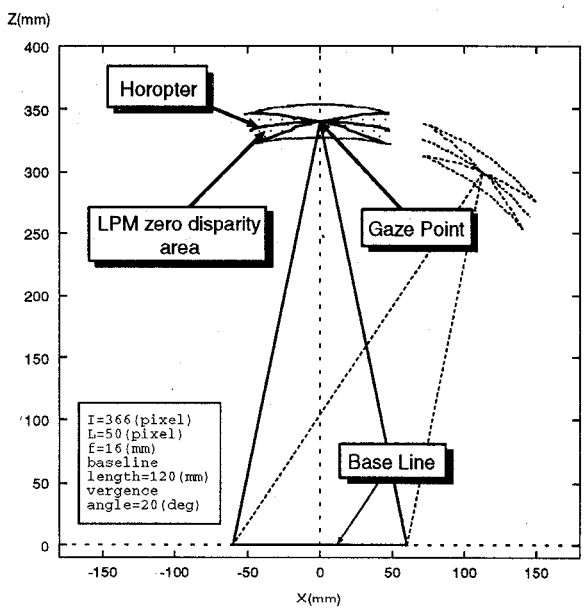


Figure 7: Shape change of the Horopter with LPM

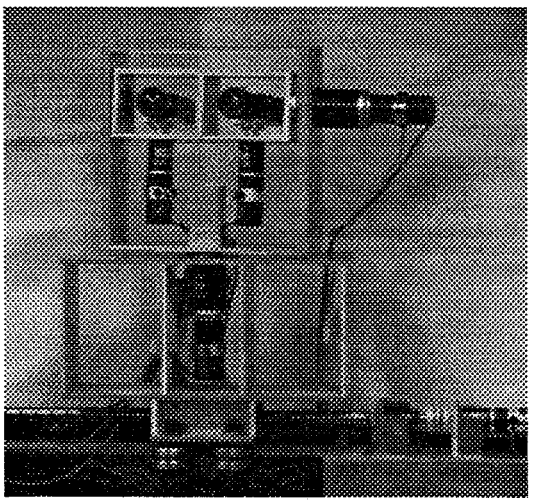


Figure 8: Overview of the Vision Robot

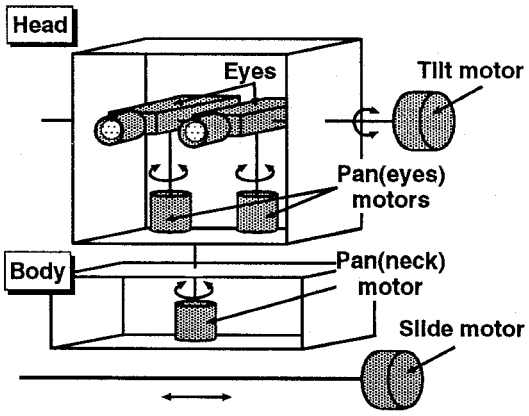


Figure 9: Kinematic Configuration of the Vision Robot

6.1 System Configurations

The experimental system (Vision Robot) [2] consists of two parts: image processing system and motor control system. Figure 8 shows the overview of the Vision Robot. 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. 9). In the experiment, we only use and control the twin pan (eyes) motors.

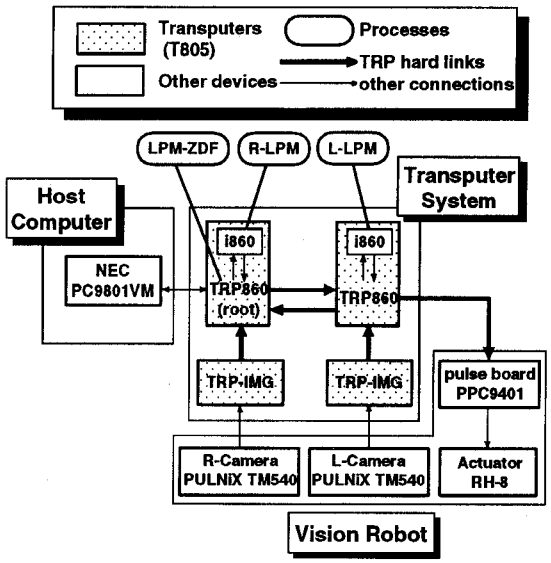


Figure 10: Configuration of Vision System

Stereo images from the two CCD cameras (PULNiX TM540) are digitized into the frame memory on the image board (Concurrent Systems TRP860) by the two Transputers (INMOS T805) as shown in Fig. 10. The baseline length is 120mm and focal length is 4.8mm respectively. LPM image processing is performed in software by the DSP (Intel i860) on the TRP-860. In the experiment, by LPM, 150×150 original images are transformed into 30×30 LPM images. The field of view is 60deg (1pixel~0.147deg for original images).

Motor control system consists of the pulse board (Nihon pulse motor PPC9401) and harmonic drive motors (see Fig. 10). Motor control is rapidly realized based on the pulse signal from the PPC9401. Dynamic performance of camera rotation (pan control only) is as follows: maximum speed 83deg/s, movable range ±30deg †.

6.2 Experimental Setup

At first, on the assumption that a target object initially at a gaze point has moved to another position across the horopter, the initial gaze point and the target object (a doll in this experiment) were set right in front of the Vision Robot at a distance of 1800mm, 2000mm respectively. In this experiment, nine disparity detectors $(0, \pm1, \pm2, \pm4, \pm8)$ and the corresponding nine image translators were used.

The ZDF method are applied to each image shifted by these image translators and the disparity detector value corresponding to the maximal ZDF output are selected as the 'estimated disparity' d.

The master eye (left camera) are controlled so that the centroid of the selected ZDF output may be consistent with the image center, and the slave eye (right camera) are simultaneously controlled so that the vergence angle may keep constant. Then, the rotation angle of the slave eye are compensated based on the estimated disparity.

†It may be too small to track moving objects. But the range can be easily spread, because our system can pan its head on which the cameras are mounted, by using the neck motor.

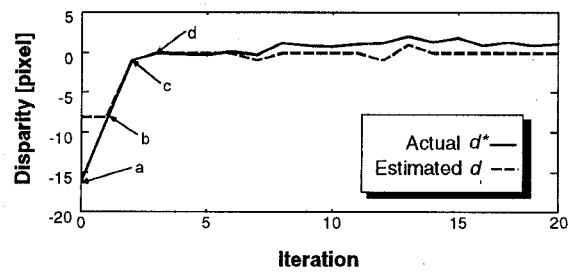


Figure 11: Results of Disparity Estimation

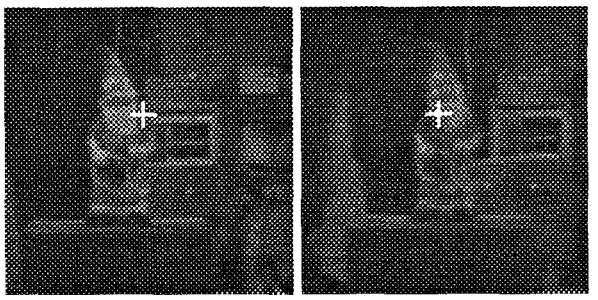
6.3 Experimental Results

Figure 11 shows the results of disparity estimation. In this figure, the bold line indicates the actual disparity d^*†† and the break line indicates the estimated disparity d. Figure 12 shows the stereo image sequences taken at the iteration step a–d in Fig. 11. In these figures, '+' mark denotes the centroid calculated by the ZDF method. We can see that the maximal disparity detector value '−8' is selected at the first step 'a', although the actual disparity is '−16.01' at the same step (see Fig. 11). This is because zero disparity area of LPM images spreads from the gaze point toward the periphery as described in Section 5.2 (see Fig. 7). This result substantiates that using LPM is effective for finding a target object in the periphery. From Fig. 11 and Fig. 12, we also see that disparities are appropriately estimated so that the proposed method successes in fixating the target object.

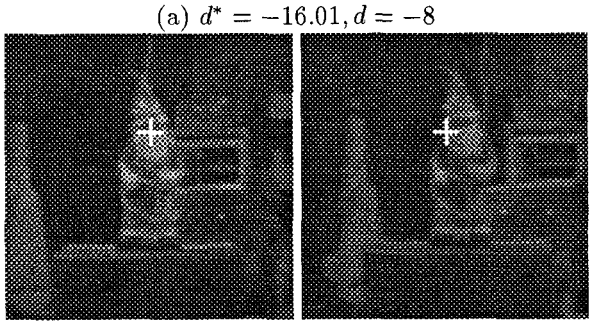
7 Conclusion

In this paper, we have proposed a new binocular tracking method of a gaze object using Log Po-

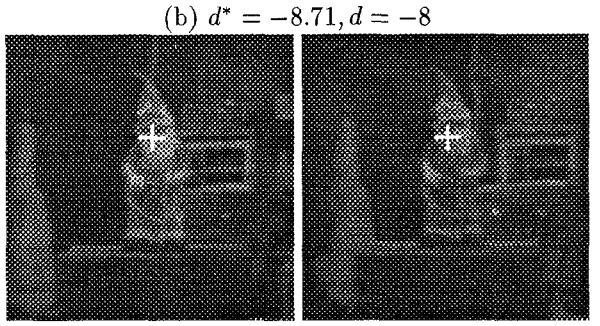
††We calculated *actual* disparities by segmenting the target object area in the stereo images by hand. Denoting the centroid of the target object in the left and right images by μ_l^* and μ_r^* respectively, actual disparity $d^* = \mu_l^* - \mu_r^*$



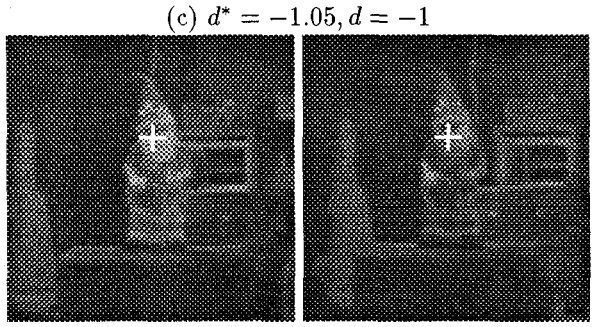
(a) $d^* = -16.01, d = -8$



(b) $d^* = -8.71, d = -8$



(c) $d^* = -1.05, d = -1$



(d) $d^* = -0.04, d = 0$

Figure 12: A part of Stereo Image Sequences in the Gaze Process using LPM-ZDF

lar Mapping (LPM). In the proposed method, extraction of the gaze object is performed by the ZDF method, and estimation of disparities of the object is done by the virtual horopter method. We have shown that the proposed method tends to decrease the number of stereo correspondence errors and increase the detectable range of disparities toward the periphery. These properties are due to the *peripheral vision function* of LPM. Therefore, using LPM is effective especially for binocular tracking of a moving object. We have also presented some experimental results of binocular tracking to demonstrate the effectiveness of our method.

We can more widely spread the detectable range of disparities of the target object only by increasing the number of disparity detectors. Even in this case, our algorithm can be speeded up to work in real-time, because the major process of each disparity detector is the ZDF operation whose computational complexity is very simple and each process is inherently parallelisable.

In this work, we have dealt with the situation that only 'one' object exists in the horopter. If there are many objects in the horopter, the proposed method may fail in tracking because of the wrong output of the target location. To overcome this problem, we are going to develop how to select the disparity detector corresponding to the target object by using the temporal information such as the location of the gaze point. Furthermore, we are going to speed up our tracking method by introducing the pipe-line process or decreasing the computational complexity of the LPM image transformation and operations.

References

[1] E. L. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex: a Spatial Mapping Approach to Perceptual Coding", *Vision Research*, Vol. 20, pp.645–669, 1980.

[2] M. Tanaka, N. Maru and F. Miyazaki, "3-D Tracking of a Moving Object by an Active Stereo Vision System", *IECON'94*, Bologna, Italy, pp. 816–820, 1994.

[3] N. Oshiro, N. Maru and F. Miyazaki, "Extraction of the Target Object in a Forward Log Polar Mapping Stereo Images", *RSJ'95*, No. 2, pp. 611–612, 1995. (in Japanese)

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

[5] S. Rougeaux, N. Kita, Y. Kuniyoshi, S. Sakane and F. Chavand, "Binocular Tracking Based on Virtual Horopters", *IROS'94*, Vol. 3, pp. 2052–2057, 1994.

[6] R. S. Wallace, P. Ong, B. B. Bederson and E. L. Schwartz, "Space Variant Image Processing", *Int. J. Computer Vision*, Vol. 13, No. 1, pp. 71–90, 1994.