

An Active Stereo Vision System Based on Neural Pathways of Human Binocular Motor System

Yu-zhang Gu¹, Makoto Sato², Xiao-lin Zhang²

1. *Department of Advanced Applied Electronics, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama, 226-8503 Japan*

2. *Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama, 226-8503 Japan*

Abstract

An active stereo vision system based on a model of neural pathways of human binocular motor system is proposed. With this model, it is guaranteed that the two cameras of the active stereo vision system can keep their lines of sight fixed on the same target object during smooth pursuit. This feature is very important for active stereo vision systems, since not only 3D reconstruction needs the two cameras have an overlapping field of vision, but also it can facilitate the 3D reconstruction algorithm. To evaluate the effectiveness of the proposed method, some software simulations are done to demonstrate the same target tracking characteristic in a virtual environment apt to mistracking easily. Here, mistracking means two eyes track two different objects separately. Then the proposed method is implemented in our active stereo vision system to perform real tracking task in a laboratory scene where several persons walk self-determining. Before the proposed model is implemented in the system, mistracking occurred frequently. After it is enabled, mistracking never occurred. The result shows that the vision system based on neural pathways of human binocular motor system can reliably avoid mistracking.

Keywords: binocular motor system, neural pathway, gaze, mistracking

Copyright © 2007, Jilin University. Published by Elsevier Limited and Science Press. All rights reserved.

1 Introduction

For an active stereo vision system, the cooperative camera movements are very important. Generally, two cameras gazing at the same spatial point can provide an overlapping field of vision and small stereo disparities around the gaze point, which makes stereo fusion and 3D reconstruction easier. Conventionally, image processing approaches are applied to make sure that the two cameras are tracking the same target object. There are many methods proposed to solve this control problem. Coombs proposed a method called Zero Disparity Filter (ZDF) which can isolate the target object from a distracting background^[1]. It is an effective method to obtain the target position in a simple scene, but it tends to output a lot of false zero disparities in a complicated

scene, since ZDF is just the logic AND of stereo vertical edge images. Rougeaux proposed a method which can obtain the location of the target moving across the horopter by using ZDF and virtual horopter^[2]. Because it relies on ZDF, it is apt to fail in complicated scene also. Tanaka proposed a disparity based segmentation method to reduce the false output of ZDF^[3], but it cannot resolve the problem that there are many objects on the horopter. On the whole, some of them can make good performance only when the scene is simple, others are designed to adapt certain special scene. Evidently, it is insufficient to solve the problem only with image processing approaches.

It is well known that the ocular motor control system is an effective device for capturing an object in the central pits of the retinas. Human eyes always orientate

Corresponding author: Yu-zhang Gu

E-mail: gu.y.ab@m.titech.ac.jp

their lines of sight to keep the image of the object from leaving their central pit, i.e. eyeballs never let an object image moving freely on their retina. This is the motivation that we develop an active stereo vision system based on human binocular motor system model.

In this paper, first we briefly introduce our mathematical model derived from the neural pathways of human binocular motor system proposed in our previous work. Second, we do some software simulations to optimize the parameters of our control model, as well as exhibit the unique feature. Then, we introduce our active stereo vision system which is made of two PTZ cameras, one omnidirectional camera, and one personal computer. Finally, we launch it in our laboratory where several persons walk self-determining to test the robustness of the system in a relatively complicated environment.

2 Mathematical model

Eye movement includes horizontal rotation and vertical rotation. Since two eyes are arranged horizontally, to keep gazing on a moving object the two eyes have to do horizontal conjugate movement or vergence movement from time to time. Therefore the horizontal binocular eye movement control is more complex than that of vertical binocular eye movement. Here we propose the horizontal mathematical model derived from our previous work^[4,5]. For vertical binocular eye movement control, just a few modifications of the horizontal model are necessary.

Because of the complexity of human being's ocular motor system, some assumptions are introduced to simplify the model. First, all the transfer functions in the ocular motor system model are considered linear^[6-8]. Second, delays caused by image processing in the visual cortex and signal transmission in the nerves are disregarded. Third, cognition in the cerebrum, including synthetic inference, prediction, and the influence of attention^[9] is neglected.

The proposed mathematical model is shown in Fig. 1. φ_l and φ_r are respectively desired values to fix the lines of sight of left and right eyes onto the visual target. E_l and E_r are respectively the controlled rotation angles of left and right eyes. σ , v , κ , η , ρ_1 and ρ_2 are gains modeled from neural pathways of human binocular motor system,

which are described in our previous work in detail^[5].

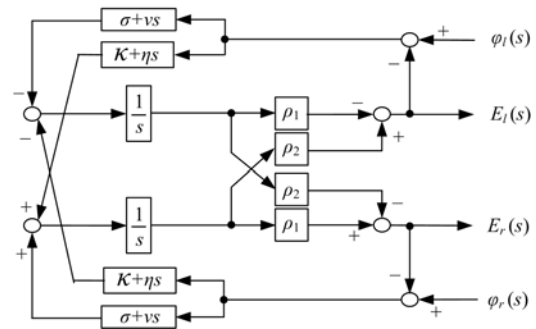


Fig. 1 Horizontal binocular motor system model.

The model proposed in this paper is a simplified version of the previous one. The block of vestibular nucleus is erased. It means that the model is appropriate for head-fixed binocular motor system, rather than that mounted on a vehicle or ego-motion robot. The simplification facilitates us to construct a real system, and it does not affect the essential binocular cooperative movement character of the original model.

According to Fig. 1, Eq. (1) and Eq. (2) can be obtained,

$$E_l(s) = \frac{\rho_1}{s}(\sigma + vs)(\varphi_l(s) - E_l(s)) + \frac{\rho_2}{s}(\kappa + \eta s)(\varphi_l(s) - E_l(s)) + \frac{\rho_2}{s}(\sigma + vs)(\varphi_r(s) - E_r(s)) + \frac{\rho_1}{s}(\kappa + \eta s)(\varphi_r(s) - E_r(s)), \tag{1}$$

$$E_r(s) = \frac{\rho_1}{s}(\kappa + \eta s)(\varphi_l(s) - E_l(s)) + \frac{\rho_2}{s}(\sigma + vs)(\varphi_l(s) - E_l(s)) + \frac{\rho_2}{s}(\kappa + \eta s)(\varphi_r(s) - E_r(s)) + \frac{\rho_1}{s}(\sigma + vs)(\varphi_r(s) - E_r(s)). \tag{2}$$

By summing up Eq. (1) and Eq. (2), Eq. (3) can be obtained,

$$E_l(s) + E_r(s) = \frac{(\sigma + vs + \kappa + \eta s)(\varphi_l(s) + \varphi_r(s))}{\sigma + \kappa + (\frac{1}{\rho_1 + \rho_2} + v + \eta)s}. \tag{3}$$

By subtracting Eq. (2) from Eq. (1), Eq. (4) can be obtained,

$$E_l(s) - E_r(s) = \frac{(\sigma + \nu s - \kappa - \eta s)(\varphi_l(s) - \varphi_r(s))}{\sigma - \kappa + \left(\frac{1}{\rho_1 - \rho_2} + \nu - \eta\right)s}. \quad (4)$$

From Eq. (3) and Eq. (4), Eq. (5) can be obtained,

$$\begin{aligned} \begin{bmatrix} E_l(s) \\ E_r(s) \end{bmatrix} &= \frac{1}{2} \frac{\sigma + \kappa + (\nu + \eta)s}{\sigma + \kappa + \left(\frac{1}{\rho_1 + \rho_2} + \nu + \eta\right)s} \begin{bmatrix} (\varphi_l(s) + \varphi_r(s)) \\ (\varphi_l(s) + \varphi_r(s)) \end{bmatrix} \\ &+ \frac{1}{2} \frac{\sigma - \kappa + (\nu - \eta)s}{\sigma - \kappa + \left(\frac{1}{\rho_1 - \rho_2} + \nu - \eta\right)s} \begin{bmatrix} (\varphi_l(s) - \varphi_r(s)) \\ (\varphi_l(s) - \varphi_r(s)) \end{bmatrix}. \end{aligned} \quad (5)$$

Obviously, the first part of the right side of Eq. (5) is the transfer function of conjugate eye movement, and the second part is the transfer function of vergence eye movement. The time constant of the conjugate transfer function T_c and that of the vergence transfer function T_v can be obtained, as shown in Eq. (6) and Eq. (7).

$$T_c = \frac{1 + (\rho_1 + \rho_2)(\nu + \eta)}{(\rho_1 + \rho_2)(\sigma + \kappa)}, \quad (6)$$

$$T_v = \frac{1 + (\rho_1 - \rho_2)(\nu - \eta)}{(\rho_1 - \rho_2)(\sigma - \kappa)}. \quad (7)$$

To make sure that the system is stable, condition Eq. (8) or Eq. (9) should be met^[5]

$$\sigma > \kappa, \nu > \eta, \rho_1 > \rho_2, \quad (8)$$

$$\sigma < \kappa, \nu < \eta, \rho_1 < \rho_2. \quad (9)$$

Either Eq. (8) or Eq. (9) makes $T_c > T_v$. In other words, in a stable system the time constant of vergence eye movement is longer than that of conjugate eye movement.

This conclusion accords with normal physical phenomenon. When the visual target moves in a plane perpendicular to the line of sight approximately, its image in the retina moves quickly. Short conjugate eye movement time constant enables eye rotation to respond quickly to keep smooth pursuit. In case of the visual target moves along the line of sight, its image in the retina moves slowly. Even though the time constant of vergence eye movement is longer, the two eyes still can follow the visual target.

The different time constants between conjugate and vergence eye movements can efficiently prevent binocular motor system from mistracking. In this paper, mistracking means two eyes track two different objects separately.

3 Mistracking problem

Usually mistracking occurs when an object similar to the visual target object being tracked crosses the target object.

The reason of mistracking is that the system cannot distinguish two similar objects correctly. Of course, skillful image processing algorithm can improve the ability of distinguish similar objects to a certain extent. For example, we can calculate the velocity vector of target object. But in a real complicated scene, this kind of processing may become very unreliable. It is not a practicable approach to solve the mistracking problem through complicated image processing algorithm only.

The character of the proposed binocular motor system model can be used to solve the mistracking problem. As described in section 2, the time constant of vergence eye movement is longer than that of conjugate eye movement. This character provides us a way to prevent the system from mistracking.

In Fig. 2, there is an object B similar to the tracked target object A. At first the two cameras track object A. When object A and object B cross each other, the images of them are too similar to be distinguished. Suppose that the left camera keeps tracking object A. But the right camera is confused by similar objects A and B, and it tries to track object B. At this time mistracking occurs, and the system switches from the original conjugate eye movement to vergence eye movement. Because of the longer time constant of vergence eye movement, neither

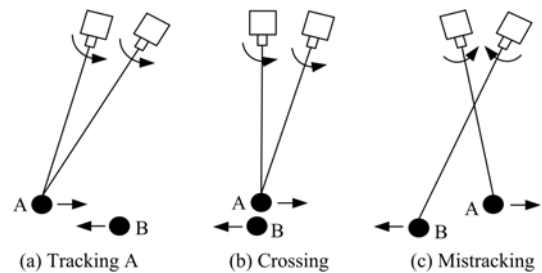


Fig. 2 Mistracking problem.

of the eyes can follow its own visual target object. Accordingly, the image of the visual target object will depart from the center of camera sensor quickly. In other words, the abrupt departure indicates that mistracking may occur. At this time both objects A and B are still near the gazing point. Now if the system searches all possible visual target objects nearby the center of camera sensor immediately, considering that the binocular disparity of object near the gazing point is small enough, the system can review whether the two cameras are attempting to track different objects, and return to normal tracking condition.

In next section, some software simulation experiments are done to verify the mistracking prevention effect of the proposed binocular motor system model.

4 Software simulation

Theoretically, to track a target as fast as possible, both T_c and T_v should be kept as shorter as possible. But in nature, as described in section 2, fast vergence eye movement is not necessary. According to the stable condition requirement described in Eq. (8), we propose a set of effective parameters as $\sigma = 3$, $\nu = 0.5$, $\kappa = 2.5$, $\eta = 0.2$, $\rho_1 = 1.5$ and $\rho_2 = 0.5$. According to Eq. (6) and Eq. (7), the conjugate time constant is $T_c = 0.191$ s, and the vergence time constant is $T_v = 2.6$ s. Fig. 3 and Fig. 4 show the step response of conjugate eye movement and vergence eye movement of this model.

Suppose there are two objects A and B move horizontally in front of the binocular motor system. The distance from the objects to the base point of the binocular motor system is far longer than that between the two cameras. At first the two cameras of the system track

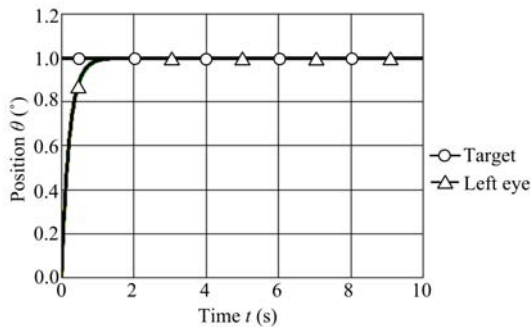


Fig. 3 Conjugate step response.

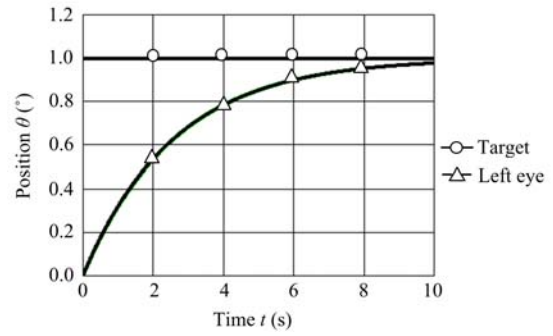


Fig. 4 Vergence step response.

object A. When objects A and B cross each other, the left camera keeps track object A, but the right camera begins to track the other object B (as shown in Fig. 2). Fig. 5 shows the simulation result, in which both object A and object B move toward each other at speed of $5 \text{ deg}\cdot\text{s}^{-1}$ (relative speed is $10 \text{ deg}\cdot\text{s}^{-1}$). At time $t = 3$ s, A and B cross each other, and mistracking occurs. For the reason explained in section 3, neither of the cameras can follow its own visual target object.

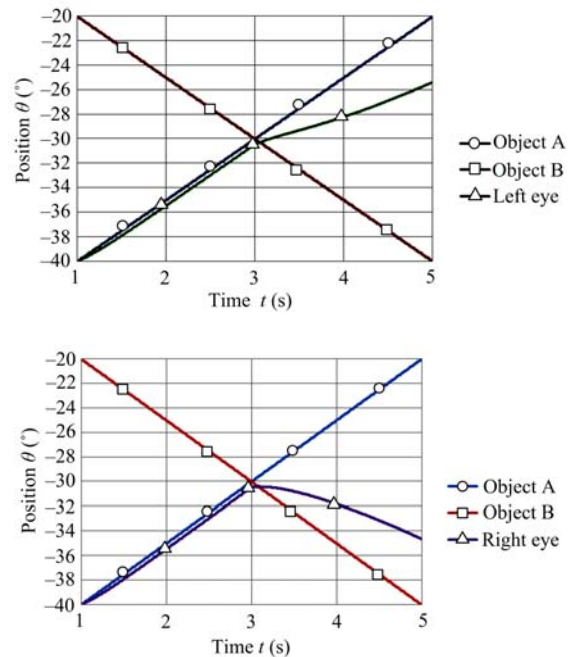


Fig. 5 Mistracking simulation result.

Let D_t denote the binocular disparity of target, D_a denote binocular disparity of object A, and D_b denote binocular disparity of object B. Eq. (10), Eq. (11) and Eq. (12) can be given,

$$D_t = (\theta_{\text{left_target}} - \theta_{\text{left_eye}}) - (\theta_{\text{right_target}} - \theta_{\text{right_eye}}), \quad (10)$$

$$D_a = (\theta_{\text{object_A}} - \theta_{\text{left_eye}}) - (\theta_{\text{object_A}} - \theta_{\text{right_eye}}) = \theta_{\text{right_eye}} - \theta_{\text{left_eye}} \quad (11)$$

$$D_b = (\theta_{\text{object_B}} - \theta_{\text{left_eye}}) - (\theta_{\text{object_B}} - \theta_{\text{right_eye}}) = \theta_{\text{right_eye}} - \theta_{\text{left_eye}} \quad (12)$$

Fig. 6 shows the changes of D_t and D_a (D_b) during the tracking process of Fig 5. When mistracking occurs at time $t = 3$ s, the binocular disparity of the visual target increases dramatically. The increase in D_t can be treated as a signal which indicates that mistracking has occurred.

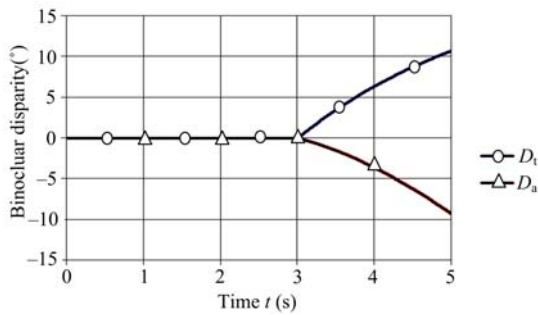


Fig. 6 Binocular disparity.

Table 1 shows the relation between D_t and D_a (D_b) when mistracking occurs.

Table 1 D_t and D_a in Fig. 6

t (s)	D_t (°)	D_a (°)
2.98	0	0
2.99	0	0
3.00	0	0
3.01	0.077	-0.023
3.02	0.154	-0.046
3.03	0.230	-0.070
3.04	0.306	-0.094
3.05	0.382	-0.118
3.06	0.458	-0.142
3.07	0.533	-0.167
3.08	0.607	-0.193
3.09	0.681	-0.219
3.10	0.756	-0.244

Obviously, the absolute value of D_t is greater than that of D_a (D_b) after mistracking. That means if the system compares current binocular disparity of the target with all possible binocular disparities of other objects near the gazing point when mistracking occurs, it can distinguish different objects and track only one of them

although it may not be the correct one.

This means that, because D_a equals D_b theoretically, the model cannot ensure that the system still track object A, even though it can prevent system from mistracking.

To illustrate the unique effect of our model, the simulation result of a binocular system with two independent control blocks for its two cameras is given in Fig. 7. Suppose the independent control block is a first order lag transfer function with gain = 1 and time constant = 0.05 s.

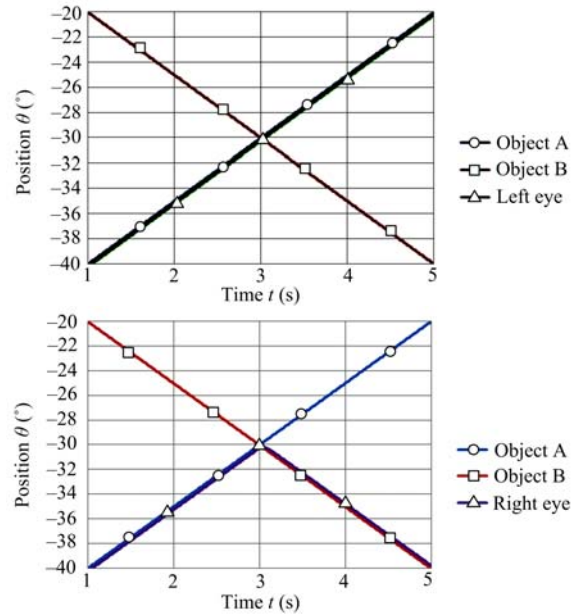


Fig. 7 Mistracking simulation result (without proposed model).

Obviously, the two cameras can keep tracking their own visual targets even if the right camera confused object B with object A after time $t = 3$ s.

Fig. 8 shows the changes of D_t and D_a (D_b) during tracking process of Fig. 7.

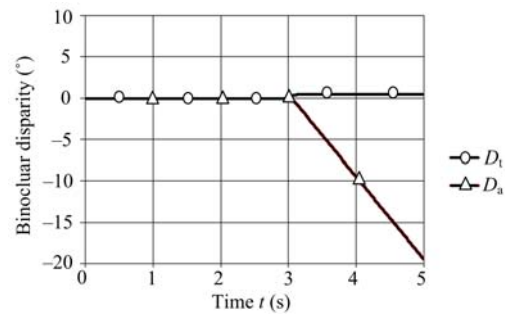


Fig. 8 Mistracking simulation result (without proposed model).

Similarly, Table 2 shows the relation between D_t and D_a (D_b) when mistracking occurs in Fig. 8.

Table 2 D_t and D_a in Fig. 8

t (s)	D_t (°)	D_a (°)
3.00	0	0
3.01	0.091	-0.009
3.02	0.165	-0.035
3.03	0.226	-0.074
3.04	0.275	-0.125
3.05	0.316	-0.184
3.06	0.349	-0.251
3.07	0.377	-0.323
3.08	0.399	-0.401
3.09	0.417	-0.483
3.10	0.432	-0.568
⋮	⋮	⋮
3.40	0.500	-3.500
3.41	0.500	-3.600

Unlike that shown in Fig. 6, in D_t Fig. 7 only increases a little soon after mistracking. Then D_t keeps at a low value in the rest of the tracking process. Theoretically, a little change of D_t at time $t = 3$ s can be treated as a signal indicating mistracking. But in practice, such a low level signal is too weak to be picked up reliably from noise. Mistracking is unavoidable once it started in such kind of systems.

The simulation mentioned previously is an example to explain the mechanism of how the proposed model can prevent mistracking from occurring. In practice, objects A and B can cross with each other at different relative speed. Fig. 9, Fig. 10 and Fig. 11 show the relation between D_t and D_a (D_b) when A and B cross each other at different relative speed v .

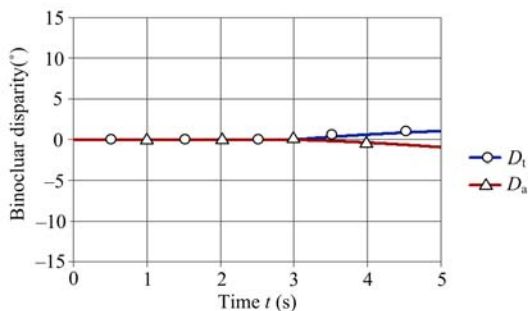


Fig. 9 Binocular disparity ($v = 1 \text{ deg}\cdot\text{s}^{-1}$).

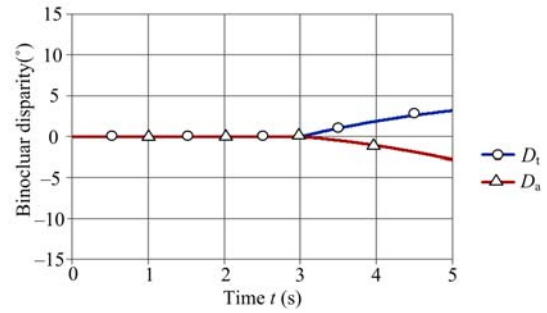


Fig. 10 Binocular disparity ($v = 3 \text{ deg}\cdot\text{s}^{-1}$).

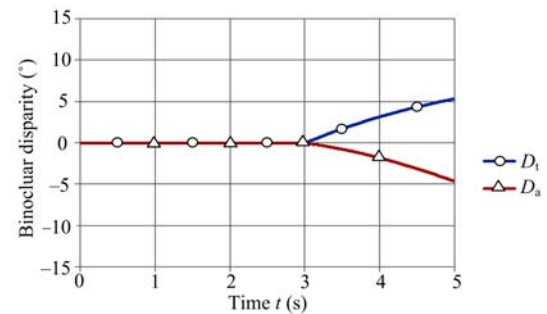


Fig. 11 Binocular disparity ($v = 5 \text{ deg}\cdot\text{s}^{-1}$).

Obviously, at different relative speeds v , the increase style of D_t and D_a is the same, and the increase speed of D_t and D_a is in direct proportion to the relative speed of objects A and B.

There are two things that are independent to the relative speed v . First, D_t will keep increase after mistracking. Second, the absolute value of D_t is greater than that of D_a all the time after mistracking. Therefore, in real system a relatively low threshold, which is slightly greater than its noise level, can be set to detect mistracking. When D_t gets greater than this threshold, mistracking is detected and the system will get a chance to review its visual target. Because $|D_a| < |D_t|$, at least one of objects A and B is close to current gazing point, so the system is able to find out the correct visual target easily and return to track it immediately.

5 Experiment system

An active stereo vision system is developed to implement the proposed binocular motor system model. Fig. 12 shows the active stereo vision system developed. The main parts of the system are two Canon VC-C50i PTZ cameras. VC-C50i outputs NTSC model video

signal, and its 1/4 inch CCD sensor has 340,000 effective pixels. The focal length is from 3.5 to 91.0 mm, providing 26 \times power zoom. The pan rotation angle range is left 100 $^\circ$ to right 100 $^\circ$, and rotation speed is 1 $^\circ$ to 90 $^\circ$ per second. The tilt rotation angle range is up 90 $^\circ$ to down 30 $^\circ$, and rotation speed is 1 $^\circ$ to 70 $^\circ$ per second. There is an omnidirectional camera between the two PTZ cameras. This camera is utilized to monitor a wider scene, since the VC-C50i can only provide maximal 40 $^\circ$ angle of view. To real-time process images from three cameras, we use a personal computer with a 3.2 GHz Pentium D CPU.



Fig. 12 Active stereo vision system.

Images from the omnidirectional camera are used to detect moving object in the monitored scene. A simple background subtraction algorithm is implemented to do moving object detection. When a moving object is detected, the two PTZ cameras are rotated to make the image of the moving object shift to the sensor's center. Then the system detects the face from the images of the two PTZ cameras. We implement boosted cascade of simple features object detection algorithm to detect face^[10,11]. After the face was detected, the two PTZ cameras are rotated to make the image of the detected face shift to the sensor's center. Then the system begins to smoothly pursuit the moving face. At first we tried to use dynamic updating template matching method to track the visual target, but we found it is not reliable enough. Finally we adopted Continuously Adaptive Mean Shift (CAMSHIFT) method^[12], which is more effective and robust to noise.

We do an experiment to track person walking in our laboratory. The background is relatively complicated

and there are several persons to confuse the system. Because skin color is the only feature used for object tracking, the system is easily confused when two faces cross each other in front of it. Before the proposed model is implemented in the system, mistracking occurred frequently. After we enabled the proposed model, mistracking never occurred. The result shows that the vision system based on neural pathways of human binocular motor system can reliably avoid mistracking problem. The video clips can be found in our web site <http://www.zhang.pi.titech.ac.jp>.

6 Conclusions

In this paper, an active stereo vision system based on a model of neural pathways of human binocular motor system is proposed. The model includes conjugate eye movement and vergence eye movement. The vergence time constant is far greater than the conjugate time constant. By simulation, we found this feature can guarantee that the two cameras of the active stereo vision system can keep their lines of sight fixed on the same target object during tracking. This feature is very important for active stereo vision systems, since not only 3D reconstruction needs the two cameras have an overlapping field of vision, but also it can facilitate the 3D reconstruction algorithm.

References

- [1] Coombs D J, Brown C M. Cooperative gaze holding in binocular vision. *IEEE Control Systems Magazine*, 1991, **11**, 24–33.
- [2] Rougeaux S, Kita N, Kuniyoshi Y, Sakane S. Tracking a moving object with a stereo camera head. *Proceedings of the 11th Annual Conference of Robotics Society of Japan*, Japan, 1993, 473–476.
- [3] Tanaka M, Maru N, Miyazaki F. Binocular gaze holding of a moving object with the active stereo vision system. *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, USA, 1994, 250–255.
- [4] Zhang X, Wakamatsu H. A unified adaptive oculomotor control model. *International Journal of Adaptive Control and Signal Processing*, 2001, **15**, 697–713.
- [5] Zhang X, Wakamatsu H. Mathematical model for binocular movements mechanism and construction of eye axes control system. *Journal of the Robotics Society of Japan*, 2002, **20**,

- 89–97.
- [6] Robinson D A. The use of control systems analysis in the neurophysiology of eye movements. *Annual Review of Neuroscience*, 1981, **4**, 463–503.
- [7] Galiana H L, Outerbridge J S. A bilateral model for central neural pathways in vestibuloocular reflex. *Journal of Neurophysiology*, 1984, **51**, 210–241.
- [8] Raphan T, Wearne S, Cohen B. Modeling the organization of the linear and angular vestibulo-ocular reflexes. *Annals of the New York Academy of Sciences*, 1996, **781**, 348–363.
- [9] Uchino Y, Sasaki H, Sato H, Imagawa M, Suwa H, Isu N. Utriculoocular reflex arc of the cat. *Journal of Neurophysiology*, 1996, **76**, 1896–1903.
- [10] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, 2001, 511–518.
- [11] Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection. *IEEE International Conference on Image Processing*, Rochester, New York, 2002, 900–903.
- [12] Bradski G R. Computer vision face tracking for use in a perceptual user interface. *Intelligence Technology Journal*, 1998, **2**, 1–15.