

# Human Vision System and Its Application for Safety and Security

XL.ZHANG\*, Toru MURASE, Tsuyoshi KOBAYASHI, Shigenori TAKAGISHI and Masahiro MORIGUCHI

Drawing upon the human oculomotor neural system, we developed a vision sensor system with mechanisms to control stereo active cameras and mimic the human vision system such as smooth pursuit (target chasing), saccade (gaze shifting) and the binocular cooperative movements (conjugate and vergence movements). We established a trial machine and conducted a field experiment for simulated passage and crosswalk. In this paper, we describe the process of the system construction along with the results of the field evaluation work.

Keywords: bio-mimetics, active stereo camera, visual sensor

## 1. Introduction

Human vision developed in line with shifts in life settings and adaptation to new environments. Particularly, central and peripheral visions and saccade (described below), which developed to obtain foods or evade other animals, enable people to grasp instantaneously changes in situations and recognize distances from each other. In modern society, the above functions are still useful for securing safety and security in living spaces and in transit.

Drawing upon biomimetics, we developed an electronics- and mechanics-based vision sensor system equivalent to the human vision system, and are studying its application to monitoring and watching over pedestrian traffic in living spaces. We established a test model, successfully tested its principle and demonstrated its usefulness, as shown below.

## 2. Human Ocular Functions and Three-Dimensional (3D) Vision Sensor

### 2-1 Establishing the basic methodology for binocular 3D vision recognition

There are two types of 3D vision sensors: active and passive. The active vision sensor emits signals such as ultrasonic waves, electric waves or light, and measures the distance from a target using reflected signals. The principle is simple and false recognition is uncommon. Laser radar, which has recently shown remarkable developments, is used as a vision sensor for mobile robots. The active sensor, however, has some inescapable drawbacks; among those pointed out are mutual interference with active sensors of the same type, large energy consumption, and distance measurement accuracy that is dependent on target reflectivity.

The passive vision sensor, utilizing information from the environment, consumes less energy but poses difficulties in information processing. The most commonly-used passive 3D vision sensor has been a stereo camera with cameras fixed relative to each other (fixed stereo cam-

era). For this type of camera, however, it is difficult to increase the measurement accuracy because it uses lenses that basically do not require focusing. When telescopic lenses are used, in particular, it is not possible to make gaze shifts towards a different target that lies at a different distance. A several-decades-long study of 3D image processing using a fixed stereo camera has been approaching its limit. We expect the active stereo camera to be the next mainstay of 3D vision sensors, but only limited data have been available regarding how this camera moves and processes images.

Becoming aware of the need to reaffirm human ocular functions, we analyzed the principles of the ocular motor function and retinal information processing on the basis of physiological and anatomical knowledge. Outcomes of a series of studies, such as modeling of the brainstem neural system related to ocular motor control and retinal processing for visual information, led to our development of vision- and control-system-based basic methodology for binocular stereoscopic recognition, which mimics the ocular motor function, integrating conjugate movements for stereoscopic vision, saccade for gaze shifting, smooth pursuit for target chasing, and vestibulo-ocular reflex (VOR) as compensation for head movements.

### 2-2 The principles of binocular motor function

Since D. A. Robinson and associates proposed in the 1970s a monocular model linearly integrating vestibular and visual signals, various oculomotor neural system models have been presented mainly by physiologists, but almost none of them are practical in the engineering sense. Domestically, the Kawahito group proposed a monocular motor model and presented robot "eyes" capable of VOR and optokinetic eye movements. However, studies of the mutual relationship of binocular movements and, further, studies of motor control systems for artificial vision devices in the engineering field are mainly for the control of the "neck" of the head with fixed eyes. Research looking into the binocular motor system per se, like ours, is limited.

### 2-3 Analysis of binocular coordination of eye movements

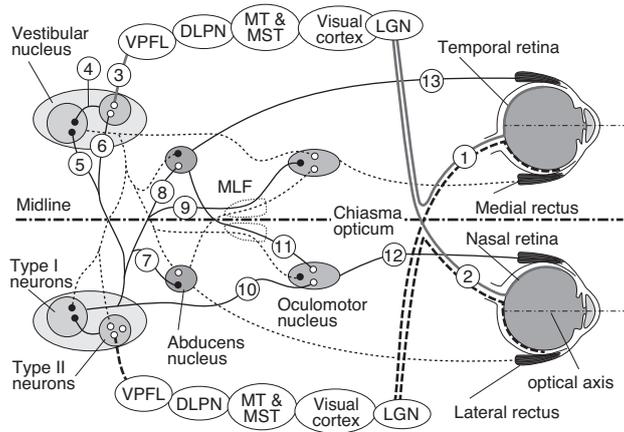
The most basic neural paths for oculomotor control are believed to reside in the brainstem, a primitive portion of the brain. A model of an oculomotor control nerv-

ous system therefore needs to be constructed solely for relevant neural paths. **Figure 1** shows neural paths for horizontal eye movements that have been presented in the physiological field. To construct a model of the paths in **Fig. 1**, a coordinate system representing the relationship between eyeballs, semicircular canals and head is established as in **Fig. 2**. From the data in **Figs. 1 and 2**, a mathematical model with each corresponding path of the neural system (**Fig. 3**) can be constructed<sup>(10)</sup>. Simplifying the model in **Fig. 3** results in the model in **Fig. 4**.

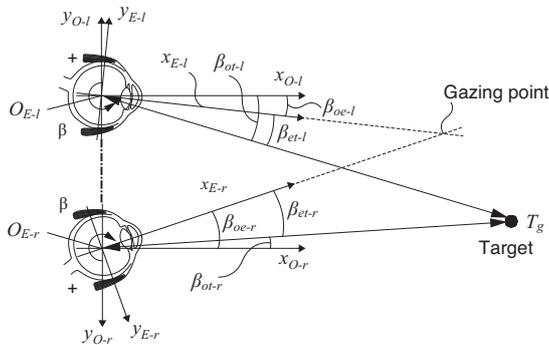
If it is assumed that  $\sigma_r, \eta_r, \rho_r = 0$  in **Fig. 4**, i.e., if all crossover paths in the system are shut, the binocular control loop is a common visual feedback control loop. What those crossover paths mean can be analyzed as follows.

$$\varphi_{oe-l}(s) = M(s) \left( \kappa + \frac{1}{s} \right) \left( (\rho(\sigma + \eta s) + \rho_r(\sigma_r + \eta_r s)) \varphi_{et-l}(s) - (\rho_r(\sigma + \eta s) + \rho(\sigma_r + \eta_r s)) \varphi_{et-r}(s) \right) \quad \dots(1)$$

$$\varphi_{oe-r}(s) = M(s) \left( \kappa + \frac{1}{s} \right) \left( (\rho(\sigma + \eta s) + \rho_r(\sigma_r + \eta_r s)) \varphi_{et-r}(s) - (\rho_r(\sigma + \eta s) + \rho(\sigma_r + \eta_r s)) \varphi_{et-l}(s) \right) \quad \dots(2)$$



**Fig. 1.** Neural paths related to horizontal binocular visual feedback control



**Fig. 2.** Coordinates of the oculomotor system

Eq. (1) + Eq. (2) and Eq. (1) - Eq. (2) results in the following:

$$\begin{aligned} & \varphi_{oe-l}(s) + \varphi_{oe-r}(s) \\ &= M(s) \left( \kappa + \frac{1}{s} \right) (\rho - \rho_r) \left( (\sigma - \sigma_r) + (\eta - \eta_r) s \right) (\varphi_{et-l}(s) + \varphi_{et-r}(s)) \quad \dots(3) \end{aligned}$$

$$\begin{aligned} & \varphi_{oe-l}(s) - \varphi_{oe-r}(s) \\ &= M(s) \left( \kappa + \frac{1}{s} \right) (\rho + \rho_r) \left( (\sigma + \sigma_r) + (\eta + \eta_r) s \right) (\varphi_{et-l}(s) - \varphi_{et-r}(s)) \quad \dots(4) \end{aligned}$$

From Eq. (3) and Eq. (4), the system in **Fig. 5** is obtained.

As shown in **Fig. 5**, vergence movements (eyeballs move in different directions) and conjugate movements (eyeballs move in the same direction) can be separated completely, and are controlled by different algorithms.

From **Fig. 2**,

$$\varphi_{et-r} = \varphi_{ot-r} - \varphi_{oe-r}, \quad \varphi_{et-l} = \varphi_{ot-l} - \varphi_{oe-l}$$

Eq. (3) and Eq. (4) can be further transformed into the following:

$$\begin{aligned} & \varphi_{oe-l}(s) + \varphi_{oe-r}(s) = \\ & \frac{M(s) \left( \kappa + \frac{1}{s} \right) (\rho - \rho_r) \left[ (\sigma - \sigma_r) + (\eta - \eta_r) s \right]}{1 + M(s) \left( \kappa + \frac{1}{s} \right) (\rho - \rho_r) \left[ (\sigma - \sigma_r) + (\eta - \eta_r) s \right]} (\varphi_{ot-l}(s) + \varphi_{ot-r}(s)) \quad \dots(5) \end{aligned}$$

$$\begin{aligned} & \varphi_{oe-l}(s) - \varphi_{oe-r}(s) = \\ & \frac{M(s) \left( \kappa + \frac{1}{s} \right) (\rho + \rho_r) \left( (\sigma + \sigma_r) + (\eta + \eta_r) s \right)}{1 + M(s) \left( \kappa + \frac{1}{s} \right) (\rho + \rho_r) \left( (\sigma + \sigma_r) + (\eta + \eta_r) s \right)} (\varphi_{ot-l}(s) - \varphi_{ot-r}(s)) \quad \dots(6) \end{aligned}$$

Vergence movements shown in Eq. (5) and conjugate movements in Eq. (6) are controlled by different transfer functions. When a control target is analyzed with first-order lag, the transfer function of vergence movements is found to have a smaller gain and a larger time constant compared with those of conjugate movements<sup>(10)-(12)</sup>.

Since 3D vision obtained with a stereo camera is based on the triangulation principle, target position measurement accuracy is highly dependent on depth and the

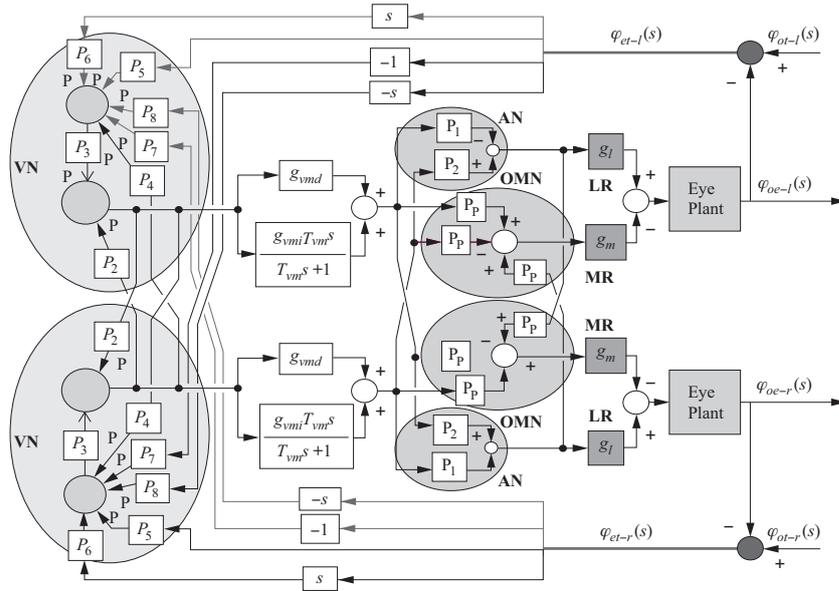


Fig. 3. Mathematical model of the neural paths shown in Fig. 1

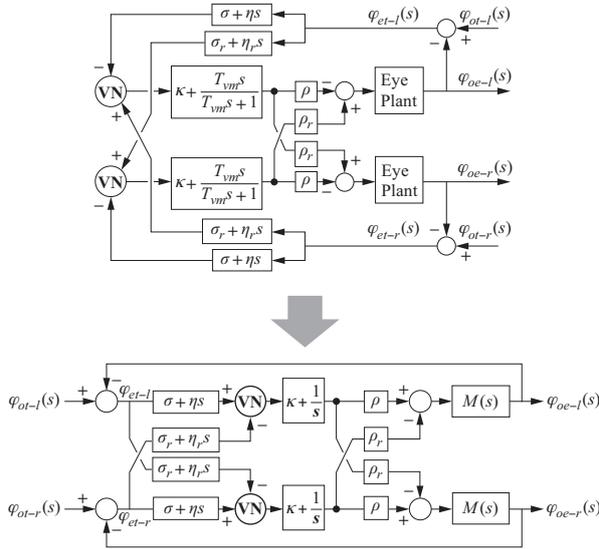


Fig. 4. Model obtained by simplifying the Fig. 3 model

direction of the tangent to the line of sight. Namely, while measurement accuracy for objects moving along the plane tangent to the line of sight is proportional to the distance, measurement accuracy for objects moving in depth is proportional to the square of the distance<sup>(3)</sup>. This means that in controlling camera motion, measurement in depth is more subject to the effect of disturbance. Therefore, for conjugate movements, which are made along the plane tangent to the line of sight, the control characteristic needs to be “light” to enable quick target tracking, while for vergence movements, which are depth tracking, the control characteristic needs to be “heavy” to stabilize tracking performance. Here, “light” and “heavy” can be interpreted as small and large time constants, re-

spectively, in terms of control transfer. Further details are described in references (10) to (12).

The control system in Fig. 5 shows that the mutually independently driven mechanism, like that of the eyeballs, can be transformed equivalently into separate mechanisms, one of which drives vergence movements and the other of which drives conjugate movements.

Based on the above, we designed a test model with a mechanism that controlled vergence and conjugate movements separately as in Fig. 5, with a relatively small target area such as a crosswalk at an intersection. For this mechanism, actuators and encoders can be selected that are most suitable for different vergence and conjugate commands.

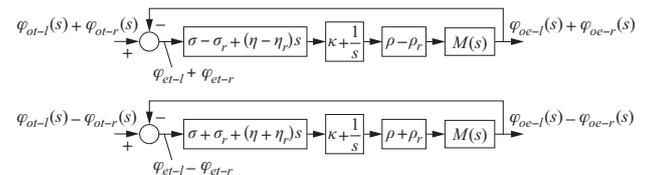


Fig. 5. Equivalent control system, dividing the Fig. 4 system into conjugate and relative movements

#### 2-4 Structures of peripheral and central visions

Figure 6 illustrates the ocular structure (top: cross section of the right eyeball). The retina of the human eye is divided into the part responsible for central vision and that responsible for peripheral vision. The central vision part is composed of cone cells, and the peripheral vision part, mainly of rod cells. Those two parts are not clearly demarcated and are distributed as shown in Fig. 7. Different visual cell structures of the central and the peripheral

vision parts of the retina mean different principles for visual information processing in those parts. In general, human eyes have the characteristic of automatically tracking an object being gazed at (smooth pursuit); retinal images in the central vision are basically stationary while those in the peripheral vision are mainly in motion.

We believe that central vision recognizes stationary images through differential processing spatially and integral processing temporally. Namely, retinal images are believed to undergo edge detection in tandem with the enhancement of the parts remaining unchanged over time. The peripheral vision, which mainly deals with moving image processing, is deemed to engage in integral processing spatially and differential processing temporally, in other words, low-pass filtering concurrently with the detection of parts changing with time.

With the aforementioned retinal characteristics, peripheral vision can be managed with a wide-angle camera, and central vision, with a telescopic camera. Furthermore, with certain camera control techniques, it is possible to search for an object in motion with a wide-angle lens, to shift a gaze at that object and to recognize it in central vision. Gaze shifting is called saccade in physiology; it involves very rapid eye movements (max. 800 degrees/second).

In the present attempt, we used our test model installed at a fixed point. The main role of peripheral vision was to detect objects in motion (e.g., automobiles, bicycles, pedestrians), with the peripheral vision system also

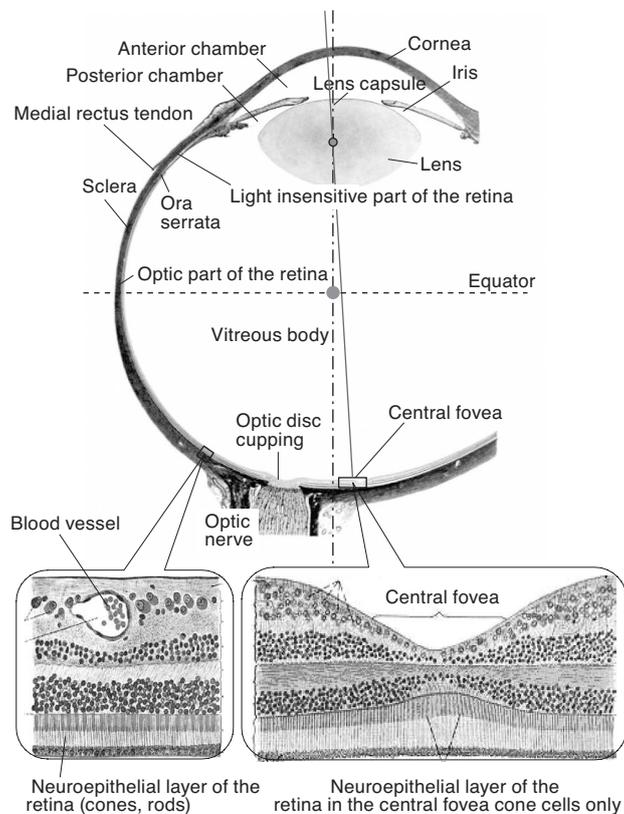


Fig. 6. Eyeball and its retinal structure

fixed, considering the relationship with objects for detection. A wide-angle camera was installed on the base, as shown in **Photo 1**.

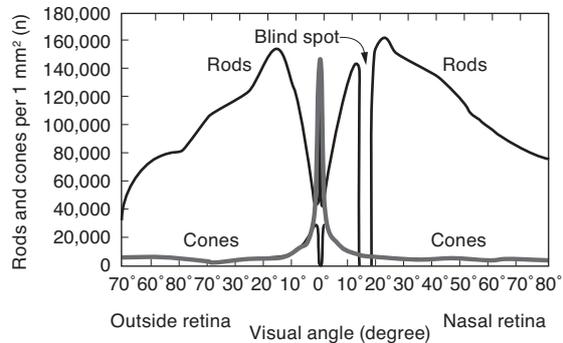


Fig. 18.5.1, p. 924, Handbook of Sensation and Perception

Fig. 7. Cell distribution in the retina

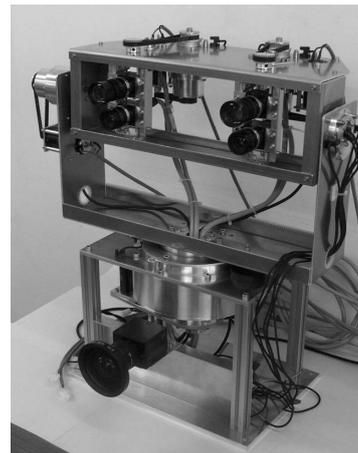


Photo 1. Test model of a binocular vision sensor for monitoring

### 3. Superiority of the Active Binocular Camera System

In the present study, the following were achieved by employing a two eyes' relative motion system, as in the human ocular system, instead of vision recognition and target tracking systems using a conventional monocular device or a binocular camera system with two lenses fixed relative to each other.

#### 3-1 High-accuracy 3D measurement

The vergence movement capability of our system enables highly accurate measurement of the target position at a larger distance. Both the right and left cameras, capable of gazing at a target, are able to zoom it in; the target position is measured based on camera rotation angles. This greatly increases the spatial resolution of the binocu-

lar camera system. For example, with horizontally fixed cameras without vergence movement capability, a theoretical error in the depth measurement of a target at a distance of 100 m stands at not less than 12 m, with the inter-cameras distance at 0.4 m, 1000 × 1000 pixels, and an angle of view of 30 degrees. In contrast, the error is not more than 0.2 m under the same conditions, if a binocular zoom camera system capable of vergence movements is used coupled with a highly accurate encoder. This feature enables high resolution photographing and creation of high accuracy 3D images for the monitoring camera.

### 3-2 Rapid gaze shifting and smooth tracking

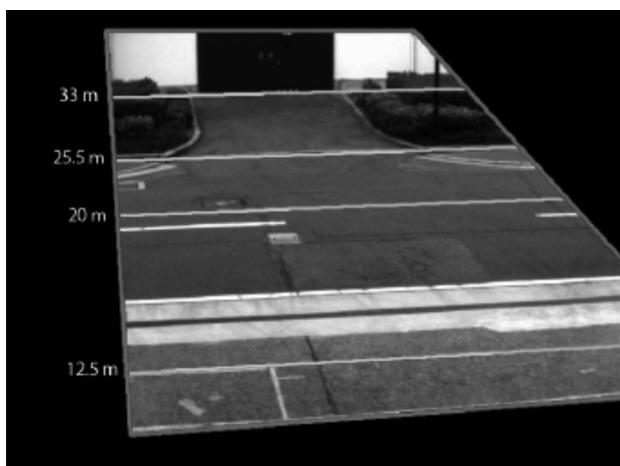
A telescopic lens and a wide-angle lens can be used to simulate the functions of the human foveal retina and of other parts of the retina, respectively. Use of these lenses achieves wide-angle monitoring together with high-resolution gaze. Integrating such a characteristic and binocular motor characteristics makes possible rapid gaze shifting (aiming) and smooth tracking of targets being gazed at, as human eyes do. Given the above, the following functions can be obtained: distinguishing separate individuals from among multiple people, creating a 3D map of the vicinity of a monitored intersection, measuring the size of obstacles in the monitored area, and keeping track of suspicious objects and individuals.

## 4. Outline of Experiment Using Our Test System

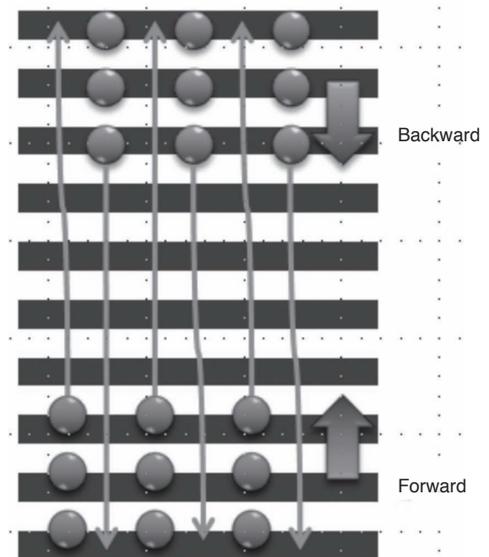
### 4-1 Field experiment

Using the test model shown in **Photo 1**, we conducted a field experiment for simulated passage and crosswalk. The field size is provided in **Photo 2**. The pedestrian traffic tested in the experiment is given in **Fig. 8**.

Under the above conditions, subjects were requested to walk at a constant pace, and groups of people moving in two opposite directions were tracked. In addition, assuming application of our system to traffic lights for pedestrians and to traffic light adaptive control in accor-



**Photo 2.** Field used in the experiment



**Fig. 8.** Pedestrian traffic in the experiment



**Photo 3.** Tracking pedestrian traffic in the experiment

dance with crossing pedestrians, the time when the last person walking in any given group of pedestrians (this person is replaced by another from moment to moment in the group) finished crossing was estimated, and the accuracy of the estimation was assessed. Furthermore, simulated actual traffic, including automobiles and bicycles, was created and evaluated. **Photo 3** is a scene from the simulated actual traffic testing.

### 4-2 Results of the field experiment

**Table 1** summarizes the result of the above experiment. In targeting pedestrians only, none was overlooked. Under daytime conditions, the tracking rate was nearly 100%. Although our test system mimicking human vision was found to have certain defects that required system corrections, depending on conditions such as viewing angles, the results we eventually obtained were mostly satisfactory in accomplishing the aim of detecting and tracking pedestrians.

**Table 1.** Pedestrian overlooking rates

Trial	Pedestrians (total) (n)	Overlooked (n)	Overlooking rate (%)
1	84	0	0.0
2	105	0	0.0
3	105	0	0.0
4	84	0	0.0
5	84	0	0.0
6	84	0	0.0
Total	546	0	0.0

In measuring the time spent by pedestrians to cross a street, their walking rates were set at 1.0 to 1.3 m/sec. Variability in the walking rates of individuals needs to be considered, but it was eventually found that the error remained within 2 sec. in the estimation of the time required for pedestrians to complete crossing of a 20-m wide path (**Table 2**). In our aging and declining birthrate society, diversely mixed traffic conditions are expected including slow pedestrian traffic, such as elderly people, and individuals in relatively fast-moving wheelchairs. The experimental results indicate that our system can be applied to monitoring individuals with various moving speeds.

**Table 2.** Estimated time required for pedestrians to complete crossing

Moving rate		Estimated time for complete crossing	
Mean	Standard deviation	Mean error	Standard deviation
1.03	0.17	4.59	1.91
1.03	0.20	2.61	1.06
1.18	0.24	2.13	1.42
1.43	0.30	1.94	2.60
1.17	0.27	1.04	0.64
1.29	0.25	1.49	1.09

Furthermore, bicycles crossing and passing, which have been hard to detect, as well as automobiles, must be taken into account as factors involved in passage and crosswalk monitoring, in addition to pedestrians. **Table 3** provides the results of experiments with these additional factors included. Since the number of trials was not sufficiently large in the present experiment, the findings may not be fully convincing, but they suggest that the system is potentially effective in tracking objects in motion in actual urban traffic environments.

Considering the future course of system development based on the above achievements and challenges faced in the present experiment, processing with a fixed wide-angle camera is expected to develop into a satisfactorily practical system if its performance is further improved in the current course of development. On the other hand, the part regard-

ing distance measurement, i.e., image processing and motor control of an active telescopic camera, was found to require improvements in both driving speed and accuracy.

**Table 3.** Overlooking rates in mixed pedestrian and vehicle traffic

Trial	Pedestrians (n)	Bicycles (n)	Automobiles (n)	Overlooked (n)	Overlooking rate (%)
1	5	0	4	0	0.0
2	10	0	2	0	0.0
3	6	1	0	0	0.0
4	10	3	0	0	0.0
5	19	2	1	1	4.5
6	7	0	1	0	0.0
Total	57	6	8	1	1.4

## 5. Summary and Conclusions

Based on a physiological analysis of the human ocular function, we developed a system equivalent to the human vision system using electronic and highly sophisticated drive unit parts. We then conducted a field experiment using simulated urban traffic environments, such as passages and crossroads, and tested the effectiveness of the system with successful results. Given this, the system was validated in principle in terms of capturing the position and motion of each individual in any given group of pedestrians. It is possible to apply this system to the control and understanding of the motion of people in response to traffic lights for pedestrians and at entrances and exits.

On the other hand, this test system was found to have certain aspects to be improved in some required unit parts. Improving the system's driving speed and accuracy within the boundary of the overall cost will lead to further development of a reasonably-priced and practical vision device.

## References

- (1) S. C. Cannon, D.A. Robinson, "Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey," *J. Neurophysiol.* 57:(5) (1987), 1383-1409.
- (2) S. C. Cannon, D.A. Robinson, "An improved neural-network model for the neural integrator of the oculomotor system: more realistic neuron behavior." *Biol. Cybern.* 53:(2) (1985),93-108.
- (3) YZ. GU, M. Sato, XL. Zhang: An Active Stereo Vision System Based on Neural Pathways of Human Binocular Motor System, *Journal of Bionic Engineering*, Vol.4 No.4, 185-192(2007/12)
- (4) E. L. Keller, D.A. Robinson, "Abducens unit behavior in the monkey during vergence movements," *Vision Research.* 12:(3) (1972), 369-82.
- (5) A. Komatsuzaki, Y. Shinoda and T. Maruo, *Neurology of Ocular Movement*, Igaku-shoin (1985)
- (6) D. A. Robinson, "Linear addition of optokinetic and vestibular signals in the vestibular nucleus," *Exp. Brain Res.*, Vol.30, (1977), 447-450.
- (7) D. A. Robinson, "The use of control systems analysis in the neurophysiology of eye movements." *Annu. Rev. Neurosci.* Vol. 4, (1981), 463-503.

- (8) H. L. Smith, H.L. Galiana, "The role of structural symmetry in linearizing ocular reflexes," Biol. Cybernetics, 1991.
- (9) XL. Zhang and H. Wakamatsu, An united adaptive oculomotor control model. International Journal of Adaptive Control and Signal Processing, Vol. 15, No. 7, pp. 697-713, 2001.
- (10) XL. Zhang, An object tracking system based on human neural pathways of binocular motor system. ICARCV2006, pp. 285-292, 2006.
- (11) XL. Zhang and Y. Sato, Cooperative movements of binocular motor system. IEEE Conference on Automation Science and Engineering (CASE 2008), pp375-380, 2008.
- (12) XL. Zhang, and H. Wakamatsu, Architecture of Mathematical Model of Binocular Movements and Control of Axis of Sights, The Robotics Society of Japan, Vol. 20, No. 1, pp. 89-97, 2002.
- (13) XL. Zhang, and H. Zhang, Axis of Sights and Its Control, Pat. Number 3862087, 2004.

~~~~~

**Contributors** (The lead author is indicated by an asterisk (\*)).

### XL. ZHANG\*

- Ph.D.

Associate professor, Information Processing and Recognition, Dept. of Advanced Information Processing, Precision and Intelligence Lab., Tokyo Institute of Technology.

He is engaged in the analytical research of eye movement, neuron and robotics vision.

Member of IEEE and the Robotics Society of Japan.



### T. MURASE

- Ph.D.

Chief Engineer, Materials and Process Technology R&D Unit,  
Chief Engineer, Information and Communications Technology R&D Unit.

Leader of Security, Safety and Ubiquitous NW WG.

He was a visiting fellow at Stanford University and Carnegie Mellon University.

Member of ISCIE and IEICE.



### T. KOBAYASHI

- Researcher, Tokyo Institute of Technology

### S. TAKAGISHI

- Dr. Eng

Senior Assistant General Manager, R&D General Planning Division

Security, Safety and Ubiquitous NW WG

### M. MORIGUCHI

- Assistant General Manager, R&D General Planning Division

Security, Safety and Ubiquitous NW WG