High-speed Gaze Controller for Millisecond-order Pan/tilt Camera

Kohei Okumura, Hiromasa Oku and Masatoshi Ishikawa

Abstract—We developed an optical high-speed gaze controller, called the "Saccade Mirror", and used it to realize a high-speed pan/tilt camera with a high-speed image processor. Generally, in a pan/tilt camera, the gaze is controlled mechanically by rotational actuators. However, because pan/tilt cameras will be expected to use high-speed image processing (\geq 1000 fps), sufficiently high-speed performance of the gaze controller, comparable to the high frame rate, cannot be obtained with the usual method. In our system, the camera itself was fixed, and an external Saccade Mirror subsystem was used for optical gaze control. An arbitrary gaze direction within \pm 30 deg could be achieved in less than 3.5 ms.

I. INTRODUCTION

The purpose of our study was to realize a high-speed pan/tilt camera that has the ability to change its gaze direction extremely quickly, at speeds comparable to the processing speed of a high-speed image processor, while still providing a sufficient angle of view. The term "gaze" used here means the direction in which images are captured by the camera, and "high-speed image processor" means a kind of camera (a vision sensor) that can realize both capturing and processing of more than 1000 images per second. The background of this research follows.

It is said that more than 80 % of the sensory information that humans obtain is from the visual organs. For robots as well, visual servo systems have been widely used to grasp changing external environments. Vision sensors with a frame rate of 30 fps, such as CCDs, are mainly used in visual information processing. However, this approach cannot realize high-speed visual control because the image sampling rate in visual information processing is restricted to at most 30 Hz. In such cases, a high-speed image processor has been demonstrated to be quite useful [1].

On the other hand, in some applications, it is necessary to acquire images over a range wider than the camera's original angle of view. Hence, cameras whose gaze can be controlled, called pan/tilt cameras (or active vision systems), have been developed mainly for the purpose of monitoring, inspection, and so on. In a general pan/tilt camera, the camera is mounted on a two-axis rotational platform including actuators (Fig. 1(a)) [2].

If a high-speed image processor can control its gaze at a speed sufficiently high compared with its frame rate, various advanced applications are expected. Some examples are:

• Observation of high-speed dynamic objects without motion blur.

- Quasi-extension of the angle of view by combining multiple images taken from different gaze directions in real time.
- Displaying multiple images from different gaze directions in real time using only one vision sensor.

This kind of system is expected to be valuable in many vision applications, including robot vision, medical services, and broadcasting.

However, sufficient high-speed performance cannot be obtained even if the gaze direction of the high-speed image processor is controlled mechanically by rotational actuators. The time required to reach the desired gaze direction is usually more than 20 ms [3], whereas the imaging cycle of the high-speed image processor is less than 1 ms. This is a critical bottleneck to realize a high-speed pan/tilt camera, particularly for a visual servo system.

To solve this issue, we developed an external high-speed gaze controller, called the *Saccade Mirror*. This name is based on the rapid movement of the human eye known as a saccade. Two-axis rotational mirrors, the critical parts of the system, respond in millisecond-order to rapidly control the gaze direction optically (Fig.1(b)).



Fig. 1. (a) Illustration of a general pan/tilt camera, and (b) our method involving gaze control using two-axis rotational mirrors.

II. COMPONENTS OF SACCADE MIRROR

A. Two-axis Rotational Mirrors for High-speed Gaze Control

There have been several studies aimed at realizing a highspeed pan/tilt camera even with a conventional structure. However, the required level of performance has not been obtained. For example, Y. Nakabo has developed a 1 ms target tracking system, which is a good example of a pan/tilt camera with a high-speed image processor [3]. In this system, a high-speed image processor called Column Parallel Vision 1 (CPV1) is mounted on a two-axis rotational platform including high-speed actuators. Nevertheless, its response time for switching the gaze direction is more than 20 ms, much longer than the control cycle period of 1 ms. Also, the cutoff frequency of the gaze control in the visual feedback

K. Okumura, H. Oku and M. Ishikawa are with the Dept. of Information Physics and Computing, Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. kohei_okumura@ipc.i.u-tokyo.ac.jp

system is approximately 10 Hz, whereas the regulation, or imaging, rate is 1 kHz. For the millisecond-order control that we aim for, the inertia of the rotating parts must be reduced as much as possible.

We focused attention on two-axis rotational mirrors for gaze control, as described in the introduction section. In this case, because the only rotating parts are the two mirrors, the inertia of that subsystem can be considerably reduced. However, two-axis rotational mirrors have not been adopted often for pan/tilt cameras. They are mainly used for scanning non-diverging bundles of rays, such as laser beams. For example, J. Taylor et al. and H. Kim et al. both developed depth measurement systems (a kind of laser range finder) using a camera, a laser, and two-axis rotational mirrors [4], [5].

If we use rotational mirrors for the pan/tilt camera, one critical issue is that the obtained angle of view is significantly restricted because the mirrors are small and are constructed to rotate on two axes.

A model of the system geometry now follows. A pinhole camera model is shown in Fig. 2 [6]. The angle of view θ in Fig. 2 is defined as the angular extent of a given scene in the camera. Rays from the scene are concentrated at the pinhole. The point where the bundle of rays is concentrated is called the pupil. The relationship between θ and $S(\theta)$ in Fig. 2 is presented here. $S(\theta)$ is the area of a certain image plane a distance d away from the center of the pupil. If θ becomes larger, $S(\theta)$ will also obviously become larger.

For gaze control with mirrors, if a wide angle of view is required, large mirrors will also be required. Thus, it is not easy to realize both a wide angle of view and highspeed performance at the same time because large mirrors are difficult to drive quickly. This will be even more difficult particularly when using independent two-axis mirrors. We call the mirror near the pupil the "pan-mirror" and the other one the "tilt-mirror." If a certain wide controllable gaze range is needed, the required size of the tilt-mirror will be larger than that of the pan-mirror because the area through which rays pass on the tilt-mirror varies depending on the driven angle of the pan-mirror.



Fig. 2. Angle of view in the model of a pinhole camera.

B. Pupil Shift Lenses for Achieving Practical Angle of View

We solved the issue described above by using positive lenses called "pupil shift lenses" [7]. The basic theory of the function of a pupil shift lens with a pinhole camera model is shown in Fig. 3. A bundle of rays can be refracted by the positive lens. Some of the bundle of rays that would otherwise be concentrated at B are also concentrated at the desired point A using the positive lens. Thinking of A as the pupil of the camera, this vision system including the positive lens has two pupils, the original one A and the shifted one B.

This theory can be applied to our Saccade Mirror, enabling both a wide angle of view and high-speed performance to be realized at the same time. That is, if we need gaze control at a sufficiently high speed compared with the frame rate of a high-speed image processor, this combination should give a practical angle of view, or if we need a practical angle of view with a certain vision system, this combination give sufficiently high speed.

For example, to obtain a 30 deg angle of view without pupil shift lenses, the inertia of the tilt-mirror is estimated to be more than one-hundred times larger compared with the case where pupil shift lenses are used.



Fig. 3. Shift of the pupil using a positive lens.

III. DETAILED OPTICAL THEORY

A. Gaze and Visibility Model

Here, the gaze direction and appearance of an image through two mirrors are mathematically considered. The two mirrors are mounted as shown in Fig.4.



Fig. 4. (a) Mirror alignment, and (b) appearance of an image through the mirrors.

The pan-mirror surface $(z = (\tan \alpha)x)$ includes the origin, and its rotational axis is the y-axis. The tilt-mirror surface $(z = (\tan \beta)y + l_m)$ includes $\mathbf{O}'(0, 0, l_m)$, and its rotational axis is parallel to the x-axis. The original gaze direction lies in the +x-axis direction ($\mathbf{v} =^T (1, 0, 0)$).

A matrix S_p that gives a line-symmetric transformation with respect to the pan-mirror surface and a matrix S_t that gives a line-symmetric transformation with respect to the tiltmirror surface are expressed as:

$$\begin{cases} \mathbf{S}_{\mathbf{p}} = \begin{bmatrix} \cos(2\alpha) & 0 & \sin(2\alpha) \\ 0 & 1 & 0 \\ -\sin(2\alpha) & 0 & \cos(2\alpha) \end{bmatrix} \\ \mathbf{S}_{\mathbf{t}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(2\alpha) & -\sin(2\alpha) \\ 0 & \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix} \end{cases}$$
(1)

Position vectors, $\boldsymbol{a} =^{T} (a_x, a_y, a_z)$ and $\boldsymbol{b} =^{T} (b_x, b_y, b_z)$ that satisfy x < 0 near the x-axis are considered. \boldsymbol{a}^* , the line-symmetric displacement mapping of \boldsymbol{a} with respect to the two mirrors, is calculated as:

$$\boldsymbol{a^*} = \mathbf{S_t} \left(\mathbf{S_p} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ l_m \end{bmatrix} \right) + \begin{bmatrix} 0 \\ 0 \\ l_m \end{bmatrix} \quad (2)$$

When $a_1 = b_1 - 1$, $a_2 = b_2$, and $a_3 = b_3$ are assumed, b - a = v. Thus, the gaze direction through the mirrors, $b^* - a^* = v^*$, is calculated using (2):

$$\boldsymbol{v^*} = \begin{bmatrix} -\sin\phi_p \\ \cos\phi_p\cos\phi_t \\ \cos\phi_p\sin\phi_t \end{bmatrix}$$
(3)

where $2\alpha = \frac{\pi}{2} + \phi_p$, $2\beta = \frac{\pi}{2} + \phi_t$. In addition, a general pan/tilt camera has a gaze vector v_g :

$$\boldsymbol{v_g} = \begin{bmatrix} -\sin\phi_p\cos\phi_t\\\cos\phi_p\\\sin\phi_t \end{bmatrix}$$
(4)

On the other hand, when $a_1 = b_1$, $b_2 - a_2 = \cos \theta$, and $b_3 - a_3 = \sin \theta$ are assumed, b - a is a vector $u =^T (0, \cos \theta, \sin \theta)$ on the camera image plane. Then, its mapping with respect to the mirrors is as follows:

$$\boldsymbol{u^*} = \begin{bmatrix} \sin\theta\cos\phi_p \\ -\cos\theta\sin\phi_t + \sin\theta\sin\phi_t\cos\phi_t \\ \cos\theta\cos\phi_t + \sin\theta\sin\phi_p\sin\phi_t \end{bmatrix}$$
(5)

To compare u and u^* , u^* is rotated onto the x-axis using a matrix:

$$\mathbf{R} = \begin{bmatrix} -\sin\phi_p & \cos\phi_p\cos\phi_t & \cos\phi_p\sin\phi_t \\ -\cos\phi_p & -\sin\phi_p\cos\phi_t & -\sin\phi_p\sin\phi_t \\ 0 & -\sin\phi_t & \cos\phi_t \end{bmatrix}$$
(6)

Then,

$$\boldsymbol{u'} = \mathbf{R}\boldsymbol{u^*} = \begin{bmatrix} 0\\ -\sin\theta\\ \cos\theta \end{bmatrix} = \begin{bmatrix} 0\\ \cos(\theta + \frac{\pi}{2})\\ \sin(\theta + \frac{\pi}{2}) \end{bmatrix}$$
(7)

is obtained. That is, the image through the mirrors is inclined at 90 deg, as shown in Fig. 4(b), and the camera should be tilted.

B. Design of Pupil Shift Lenses

An object greater than 2 m away is assumed to be at infinity because the camera pupil is comparatively small (5–10 mm). Therefore, the pupil shift lenses are designed for an object at infinity.

First, two positive lenses are mounted, separated by a distance that is the sum of their two focal lengths in order to keep both the input and output bundles of rays as parallel light. The lens on the input side is called the "object lens", and that on the output side is called the "collimator." One more lens called a "field lens" is placed at the focal positions of these two lenses in order to prevent vignetting. The effect of the field lens is shown in Fig. 5 (a) and (b).



Fig. 5. Behavior of a bundle of rays through the lenses: (a) without a field lens, and (b) with a field lens.

C. Overall Optical Design

We now describe how each optical device is designed. The angle of view centered on the original camera pupil, α , and that centered on the shifted camera pupil, β , are generally different. Here, β means the real angle of view while the Saccade Mirror is used. The angles of view are related to the parameters of the pupil shift lenses as follows:

$$f_o: f_c = \tan\left(\frac{\alpha}{2}\right): \tan\left(\frac{\beta}{2}\right) = d_\beta: d_\alpha$$
 (8)

where f_o is the focal length of the object lens, f_c is that of the collimator, d_{α} is the diameter of the original camera pupil, and d_{β} is that of the shifted pupil [8]. These are shown in Fig. 6.



Fig. 6. Focal length and angle of view.

Here, each device is assumed to be selected from commercially available devices. However, the process cannot be uniquely determined. It varies on the intended purpose, use, situation (the limits of the obtainable devices etc.), and so on.

In our approach, we first determine the maximum angle of view, β , and that of the controllable gaze angle, ϕ . Rotational mirrors should be used at this point because ϕ is based on only rotational mirrors. Next, the positions of the two-axis mirrors are determined in consideration of the mirror size, β , and the shifted pupil position. Then, the position and diameter of the object lens are determined. An unfolded diagram of the optical system is shown in Fig. 7.



Fig. 7. Determining the mirror positions. We should avoid vignetting caused by the angles of the mirrors. This figure shows the case where both mirrors are driven by 15 degrees in both rotational directions.

With the paths of the bundles of rays in mind, the diameters of the collimator and the field lens, D_c and D_f , are determined. On the other hand, α can be calculated from D_c and the distance between the collimator and the camera pupil, l_{c-p} . Now both α and β are determined, and the focal lengths of the lenses, f_o , f_c and f_f , can be calculated by (8). This process is shown in Fig. 8.



Fig. 8. Simulation for determining the positions and the sizes of the lenses.

IV. PROTOTYPE SACCADE MIRROR

In this section, we describe a prototype Saccade Mirror developed with some commercially available devices.

A. Optical Setup

For high-speed gaze control, we used rotational mirrors originally designed for laser scanning (galvanometer mirrors M2 manufactured by GSI Group), and for the pupil shift lenses, we used three achromatic lenses (Edmund Optics). The setup is shown in Fig. 9, and the optical specifications were as follows:

- Gazing angle range, ϕ : $\pm 30 \text{ deg}$
- Maximum beam aperture: 30 mm
- Maximum angle of view, β: 38.6 deg (measured), 40 deg (designed)



Fig. 9. The optical setup of the prototype.

B. Response Performance

The response time for switching the gaze from a certain direction to any desired direction was measured using a highspeed camera. For the sake of simplicity, only the pan-mirror was driven, and the gazing angle was set to the maximum angle possible (60 deg).

For the desired gazing angle ϕ_{in} [deg], a ramp input that is a function of t [ms] was supplied to the prototype:

$$\phi_{in} = \begin{cases} -30 & (t \le 0) \\ -30 + 30t & (0 < t < 2) \\ 30 & (2 \le t) \end{cases}$$
(9)

The results are shown in Fig. 10. Even when the mirrors were scanned quite rapidly, the desired image captured by the high-speed camera appeared stable. The response time was only 3.5 ms. Compared with general mechanical pan/tilt cameras (response time ≥ 20 ms), we successfully achieved the expected millisecond-order response.



Fig. 10. Response time for switching gaze; image sequences captured by a high-speed camera.

V. MILLISECOND-ORDER HIGH-SPEED PAN/TILT CAMERA

A. Setup for Target Tracking

We developed a millisecond-order high-speed pan/tilt camera using the prototype Saccade Mirror described above and a high-speed image processor. We used the camera to implement a target tracking application (Fig. 11).

We used IDP-512e as the high-speed image processor [9] and a PC for both image processing and mirror control. The detailed specifications are as follows:

- PC: DELL Precision T7400.
- CPU: Xeon, 2.83 GHz.
- RAM: 3.25 GB.
- OS: Windows XP Professional, Service Pack 2.



Fig. 11. System setup for target tracking

B. Tracking algorithm

The tracking algorithm is described here. An image at a certain time is first captured by the high-speed image processor and transferred to the PC. Next, the target is distinguished precisely from the background using adequate thresholding of the HSV color image to obtain a binarized image, expressed as:

$$I(x,y) = \begin{cases} 0\\ 1 \end{cases} \tag{10}$$

Using I(x, y), the (p+q)-th image moment is defined as:

$$m_{pq} = \sum_{x} \sum_{y} x^p y^q I(x, y). \tag{11}$$

The center of mass in the image can be calculated as

$$\begin{pmatrix} x_{cm} \\ y_{cm} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{x} \sum_{y} xb(x,y)}{\sum_{x} \sum_{y} I(x,y)} \\ \frac{\sum_{x} \sum_{y} yb(x,y)}{\sum_{x} \sum_{y} I(x,y)} \end{pmatrix}.$$
 (12)

Then, the mirrors are controlled to reduce the distance between $G(x_{cm}, y_{cm})$ and the center of the image to close to zero. This process is repeated every 1 ms to realize target tracking [10].

C. Frequency Response

The frequency response of the target tracking system was measured using a marker on a rotating fan as the target. A Bode plot is shown in Fig. 12. The cutoff frequency (-3 dB) was found to be around 100 Hz (pan gazing) or more

than 100 Hz (tilt gazing). The phase delay at 100 Hz was approximately 180 deg (5 ms). In a previous study [3], the frequency response of a target tracking system implemented with a mechanical pan/tilt camera was obtained. The cutoff frequency was 10 Hz (pan gazing), and the phase delay at the cutoff frequency was 90 deg (25 ms). Thus, compared with the previous system, our target tracking system successfully attained a performance level almost ten times higher, in terms of the cutoff frequency. However, the phase delay at the cutoff frequency in our system was larger than that in the previous system.

The reasons for the differences can be considered as follows. If a Windows PC is used, a delay (2–3 ms) from the PC will inevitably appear regardless of the gazing method. This delay is actually not so long for conventional systems, whose response time is ≥ 20 ms. However, because the response time of our millisecond-order pan/tilt camera using the Saccade Mirror was 3.5 ms, as mentioned above, the delay of 2–3 ms from PC became comparatively large.



Fig. 12. Bode plot of the target tracking system: (a) gain, and (b) phase delay.

D. Observation of Moving Ball

Our millisecond-order pan/tilt camera should be useful for observing or evaluating high-speed dynamic phenomena with high resolution and no motion blur. If only a highspeed camera is used, the entire range of the dynamic motion must be let into the visual field of the camera using a wideangle lens. Thus, the resolution of the desired area becomes relatively low. Moreover, motion blur also appears because the target moves at high speed.

We focused on applications in the field of sports. Balls in baseball, football or tennis are difficult to observe continuously because the velocity vectors of the balls fluctuate extremely rapidly during games. The ability to acquire dynamic images of the fast-moving balls would make significant



Fig. 13. Image sequences of a dynamic rubber ball: The ball is (a) falling from a hand, (b) bouncing against a table, (c) approaching a racket, (d) being hit by the racket, and (e) flying away.

contributions to broadcasting techniques, practice for players, product development for sports goods manufacturers, and so on.

Therefore, as an exploratory experiment, we observed a dynamic rubber ball using our millisecond-order pan/tilt camera. The target suddenly came into the visual field, bounced against a table, and was finally hit by a racket.

The image sequences are shown in Fig. 13. To keep the images of the target ball stationary, they are corrected so that the center of mass in each image is placed at the center. Although the instant where the ball bounces or is hit (change of the velocity vector) is conventionally difficult to capture, clearly observable image sequences could be successfully obtained with our system. The ball's estimated trajectory is also shown in Fig. 14.



Fig. 14. The ball's estimated trajectory.

VI. CONCLUSION

In this paper, we proposed an optical high-speed gaze controller, called the Saccade Mirror. The developed prototype successfully demonstrated a response time as low as 3.5 ms. We also developed a millisecond-order pan/tilt camera using this prototype and a high-speed image processor, and we used it to implement a target tracking application. The performance (in terms of cutoff frequency) was almost ten times higher than that of a conventional mechanical pan/tilt camera. We demonstrated the utility of the system for the field of sports with an exploratory experiment in which a fast-moving rubber ball was observed.

In future work, we intend to make improvements to the prototype and other systems to be developed, such as construction of a higher performance prototype, implementing other practical applications, and so on.

REFERENCES

- T. Komuro, I. Ishii, M. Ishikawa and A. Yoshida: "A digital vision chip specialized for high-speed target tracking," IEEE Trans. Electron Devices, Vol. 50, pp. 191-199, 2003.
- [2] J. Aloimonos, I. Weiss and A. Bandyopadhyay: "Active Vision," International Journal of Computer Vision, Vol. 1, No. 4, pp. 333-356, 1988.
- [3] Y. Nakabo, M. Ishikawa, H. Toyoda and S. Mizuno: "1ms column parallel vision system and its application of high speed target tracking," Proc. of IEEE Int. Conf. Robotics and Automation, pp. 650-655, 2000.
- [4] J. Taylor, J.-A. Beraldin, G. Godin, L. Cournoyer, R. Baribeau, F. Blais, M. Rioux and J. Domey: "NRC 3D imaging technology for museum and heritage applications," The Journal of Visualization and Computer Animation, Vol. 14, pp. 121-138, 2003.
- [5] H. Kim, C. Lin, C. Yoon, H. Lee and H. Son: "A Depth Measurement System with the Active Vision of the Striped Lighting and Rotating Mirror," Progress in Pattern Recognition, Image Analysis and Applications, Vol. 3287, pp. 108-115, 2004.
- [6] R. M. Haralick and L. G. Shapiro: Computer and Robot Vision, 1st edition, Addison-Wesley Longman Publishing Co., Inc, 1992.
- [7] M. A. Paesler and P. J. Moyer: Near-Field Optics: Theory, Instrumentation, and Applications, John Wiley & Sons, Inc, 1996.
- [8] H. Gross, W. Singer and M. Totzeck: Handbook of Optical Systems, Vol. 2: Physical Image Formation, WILEY-VCH Verlag GmbH & Co. KGaA, 2005.
- [9] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki and K. Tajima: "2000 fps Real-time Vision System with High-frame-rate Video Recording," Proc. of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010), pp. 1536-1541, 2010.
- [10] B. K. P. Horn: ROBOT VISION, The MIT Press, 1986.