Real Time Visual Tracking System Mimicking Saccadic Movements

Yoshinori Takeuchi[†], Zeng-Fu Wang[‡], Noboru Ohnishi[†‡] and Noboru Sugie[‡‡]
[†]Graduate School of Engineering, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-01 Japan
[‡]Bio-Mimetic Control Research Center, RIKEN,
8-31, Rokuban 3-chome, Atsuta-ku, Nagoya, 456 Japan
[‡‡]Faculty of Science and Technology, Meijo University,
1-501, Shiogamaguchi, Tenpaku-ku, Nagoya, 468 Japan

Abstract

This paper presents a real-time visual tracking system that can follow a moving object by controlling the pan and tilt of a camera based on the object position obtained by video-rate image processing. The system consists of a camera, an image processing system, a pan/tilt platform and a motor driver. The developed platform can control the pan and tilt angle of a camera with accuracy of 0.002 deg and speed of 50 deg/sec. We have developed a tracking algorithm using both an edge image and a frame subtraction image so as to be robust against noise. The algorithm, implemented on a general purpose image processor, can detect a moving object in less than 33 msec. As an experimental result, the system can track a moving object in real-time.

1 Introduction

Humans can easily detect moving objects, select one of them, and track the object to locate it in the center of visual field. It is a very interesting subject in computer vision to artificially realize a real-time visual tracking system, which is as perfect as the human vision system. Visual tracking systems have various applications such as security system, crime prevention and television broadcasting. One of the basic problems of a tracking system working in a real-environment is its ability to maintain an object within its field of view. If the target moves quickly, the system must also move the camera quickly to locate it within its field of view. Therefore, the tracking system requires fast image processing and quick camera moving.

Several systems were proposed for real-time tracking by using a moving camera. Ozawa et al.[1, 2] presented a on-line tracking systems with functions of both tracking and zooming. The system sampling period is 1.60 seconds. Eklundh et al.[3] introduced a partial real-time binocular active vision system. They particularly focus on occlusions of other targets, both stationary and moving, and integrate three cues, such as ego-motion, target motion and target disparity, to achieve an overall robust behavior. Ferrier et al.[4] proposed a camera control scheme based on the human oculomotor control described by Robinson, and implemented the control scheme. The implemented system could track simple blob about 10 Hz. Basu et al.[5] proposed a method for real-time motion detection using an active camera mounted on a pan/tilt platform. Sharkey et al. [6] demonstrated an active vision system which performed monocular pursuit in real-time.

This paper presents a real-time visual tracking system that can follow a moving object by controlling the pan and tilt of the camera based on the object position obtained by video-rate image processing. In Section 2, we describe the eyes movement of humans that is mimicked in the system. In Section 3, we present our developed system. Experimental results in Section 4 shows that the system can robustly track a moving object in real-time. Finally, the paper is summarized in Section 5.

2 The Eye Movements

The principle of the system is based on the eyes movement. There are two kinds of the eye movements of a human in following a moving target. One is a stepwise movement, called the saccadic movement and the other is a comparatively slow one, called the smooth pursuit movement. Figure 1 shows an eye movement that is consisted of the two types of movements [7].

The smooth pursuit is observed when the eyes track a moving object. We cannot accomplish the smooth pursuit movement if the object is stationary. The pursuit speed is slow, with maximum speed of $25-30$deg/s. On

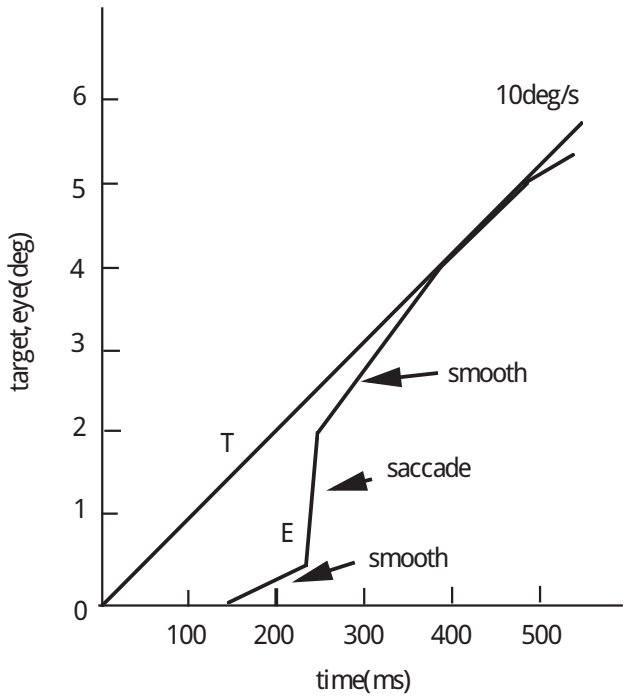

Figure 1: Saccade component and smooth pursuit component of the eye movements: T shows the target movement, E dose the eye movement.

the contrary, when we look at a stationary object or read books, the saccadic movement is observed. The saccade speed is very fast, with maximum speed of $600 - 900$deg/s. From a functional point of view, the saccade is useful for correcting the positional difference between the eye and a target, while smooth pursuit is useful for keeping relative speed between the eye and a target zero.

In humans, physiological evidences indicate that the saccade is controlled with a sampled data system, while the smooth pursuit movement is continuously controlled. The latency of the saccadic system has been estimated to be about 200 msec. This latency is considered the time, that it takes from the change in retinal position of an attended feature to the moment of a saccade command generation. This time is needed to process retinal image and detect the position of the feature. During this period the oculomotor system is insensitive to further changes in the retinal position of the feature [8]. If the feature moves during this period, the saccade will cause a position error. From this observation, the saccade system is thought the sampled data model of the oculomotor control system.

The smooth pursuit movement is thought to use a continuous time data system, or at least a sampled data system in which the sampling rate is much higher than that of the saccadic system [7]. It has been observed

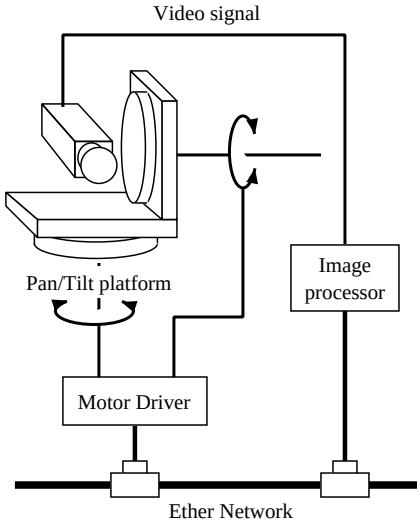

Figure 2: System overview

that smooth pursuit movements are not always smooth, but will include saccadic components if a target moves with a large angular speed. The reason is that if the speed of the target is larger than the maximum speed of the smooth pursuit movement, and the pursuit system cannot track it. In this case, position errors accumulate to reach a certain threshold, and a saccade is generated in order to reduce the position error.

The function of saccade, reducing the positional difference of the eye and the target, is useful for the tracking vision system, and we have taken notice of the saccade component in the eye movements to develop a vision system that captures a moving object in the field of view by mimicking saccade.

3 Tracking System

The system working in a real world cannot have any restrictions on object's movement. Therefore the system requires robustness and a very fast calculation with a very quick camera's motion. To do so, we developed a fast speed platform and a tracking algorithm that is well suited for hardware implementation. The developed system tracks only one object moving in a stable environment. Therefore, the system cannot select and track a desirable or salient object from multiple objects in the field of view. Such functions can be realized by incorporating object recognition. But the system should discriminate objects by pattern matching, binocular disparity, segmentation based on optical flow and so on. It may need much computation to recognize object.

3.1 System Overview

The system consists of a camera, an image processing system, a pan/tilt platform and a motor driver. It

can pursue an object that moves under a real environment in real-time. Figure 2 shows the system overview. We use a small color CCD (1/3") camera (SONY EVI-310). The lens system has a zoom of eight times (5.9 - 47.2mm) and inner focus. The image is processed by a general purpose image processor (Datacube MaxVideo 200) of VME base and a host computer (MVME167). The operating system (OS) of the host computer is a real-time OS (LynxOS). The pan/tilt platform uses a stepping motor for a rotation mechanism, and it can rotate 0.002 deg by one pulse, with maximum speed of 50 deg/sec. The motor driver is linked to the host computer by an Ether network. It receives a value of the motor rotation angle and sends a series of pulses to the motor.

The system tracks a moving object as follows:

1. Image sequence is inputed from the camera on the pan/tilt platform to the general purpose image processor (30 frames/sec).
2. The general purpose image processor estimates the position of a moving object in real-time.
3. Based on the estimated position of the moving object, the camera rotation angle is calculated.
4. The rotation angle is transmitted to the personal computer (PC) via Ether network.
5. The PC generates the pulse sequence and drives the motors.

3.2 Estimating the Position of a Moving Object

The purpose of this system is to capture a moving object at the center of the field of view. The system must measure the displacement of a moving object from the center of the visual field under a real environment. Therefore, initially we separate the input image into a moving object and a background. To do so, we use temporal differences between successive input images. The absolute value of the temporal difference depends on the contrast between a moving object and its background. This method is, however, sensitive to noises under a real environment such as change of illumination. To avoid this problem, we use edge images, because large values of temporal difference are observed near the object edges.

Figure 3 illustrates a process of estimating the displacement of a moving object from the center of visual field. A series of images $I(x,y;t)$ is system input. First, the frame subtraction image $I_s(x,y;t)$ is calculated by frame subtraction, and the edge image by Laplacian operator ∇^2 as follows:

$$I_s(x,y;t) = I(x,y;t) - I(x,y;t-\Delta t) \qquad (1)$$

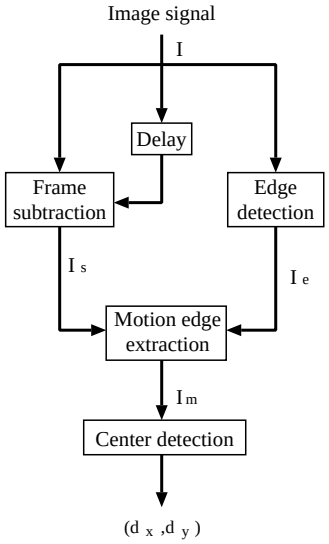

Figure 3: The process of estimating the position of the moving object

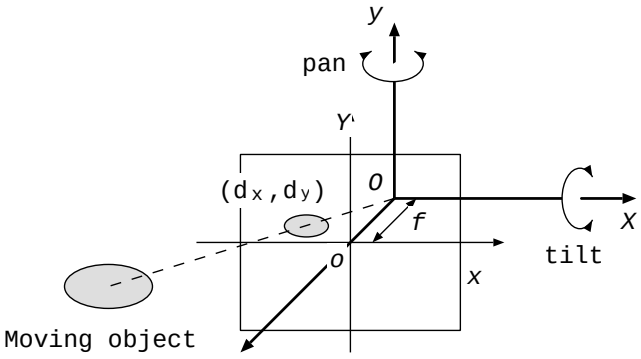

Figure 4: A coordination for moving object tracking.

$$I_e(x,y;t) = \nabla^2 I(x,y;t), \qquad (2)$$

where Δt is the frame interval (33 msec at NTSC). Then, the motion edge image $I_m(x,y)$ is extracted by integrating the obtained frame subtraction image and the edge image as

$$I_m(x,y) = \begin{cases} 1 & \text{if } I_e(x,y) > Th \text{ and} \\ & I_s(x,y) > K I_e(x,y) \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where Th and K are constants determined in the experiments. Eq.(3) means that a pixel of the motion edge image I_m is 1 if the pixel has high contrast and large temporal difference.

1 2 3 4 5 6

7 8 9 10 11 12

Figure 5: Successive images viewed from the camera (time interval is 0.5s).

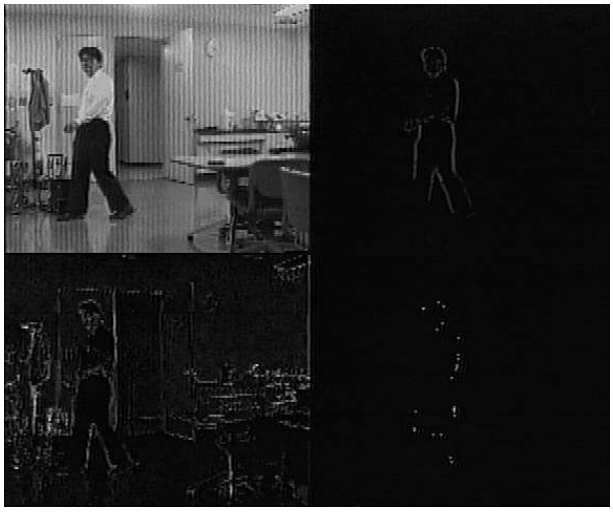

Figure 6: A snap shot of the view from the camera: the left upper is one of a series of input images, the right upper the frame subtraction, the left lower the detected edges and the right lower the integration of temporal and spatial differences.

Next, the object displacement (d_x, d_y) is determined by the center of the motion edge image by

$$\begin{aligned} d_x &= \frac{\sum_{(x,y)} x I_m(x,y)}{\sum_{(x,y)} I_m(x,y)} - x_c \\ d_y &= \frac{\sum_{(x,y)} y I_m(x,y)}{\sum_{(x,y)} I_m(x,y)} - y_c \end{aligned} \tag{4}$$

where x_c and y_c are the coordinates of the optical center.

We have implemented this algorithm as a pipe-line process in a general purpose image processor. The control period was 30 Hz by a fine synchronization between the pipe-line image processor and the real-time OS.

3.3 Tracking Strategy with Active Camera

In order to track a moving object, the pan and tilt angles of the camera are controlled so that its optical axis directs toward the center of an object. Both pan angle θ and tilt angle φ centering an object in the camera view are easily given by the object displacement d_x, d_y (see Figure 4).

$$\begin{aligned} \theta &= \tan^{-1} \frac{\varepsilon_x d_x}{f} \\ \varphi &= \tan^{-1} \frac{\varepsilon_y d_y}{f}, \end{aligned} \tag{5}$$

where ε_x and ε_y are scaling factors, and f is a focal length of camera lens. $\frac{\varepsilon_x}{f}$ and $\frac{\varepsilon_y}{f}$ are calibrated in advance.

3.4 Controlling Camera Rotation

As the result of image processing, pan and tile angles corresponding to a next fixation point is sent to the motor driver via Ether network. The motor driver consists of a personal computer (PC-9821Ap2) and a power amplifier. An Ether board and pulse generator are connected to the personal computer. The personal computer receives a quantity of motor rotation from Ether board, converts it into the number of pulses and sends it to the pulse generator. The pulse generator generates pulse signals performing acceleration or deceleration. The pulses are amplified enough to drive the stepping motor. Even though the control is feedforward, the system can perform a very fine positioning so as to be used in the recognition of an object moving at high speed.

4 Experimental Results

We have conducted an experiment in a real environment (indoor) to check the system ability to track a moving object in real-time. The moving object was a walking person. The person moved back and forth keeping a distance of about 2 meter from the camera. The speed was about 1 m/s and corresponds to about 30 deg/sec of angular velocity of the camera. We have recorded the camera images into a video tape recorder and analyzed them. Figure 5 shows successive images viewed from the camera and Figure 6 a snap shot of recorded images. The left upper is one of a series of input images; the right upper the frame subtraction; the left lower the detected edges; and the right lower the integration of subtraction and edge images.

The system succeeded in capturing the walking person within the camera view. Seeing the right lower image in Figure 6, we find that the edges of the moving object are robustly extracted. This shows that the integration of the obtained frame subtraction and the edge image based on Eq. (3) works well in real-environment. From Figure 5, we find that the moving object is not always in the center of view. The reason why the moving object is not located in the center is that the system tracks an object depending on only the current position. It is necessary to incorporate predictive control by using information about velocity and regularity of motion.

5 Conclusion

We developed a real-time tracking system that runs under real-environment in real-time. First, we have described the eyes movement of a human in tracking a target. The saccade movement is useful for correcting the positional difference of the eye and the target. To mimic the saccade movement, we developed a fast speed and fine precision pan/tilt platform. Using this platform, most of objects in real-environment may be tracked within its maximum speed rotation. To estimate the moving object's position, we developed a technique integrating temporal and spatial differences, and showed that this technique could detect the edges of the moving object robustly in real-environment. We presented experimental results demonstrating robust performance of the technique. From these results, we have confirmed that the tracking system works well even under various noises. Moreover the good point of the system is that it realizes accurate and high-speed tracking by simple estimation of object motion. The estimation of three-dimensional rotation of a real object costs a lot of calculation and is sensitive to noise. Rotation, however, can be neglected at very short time interval Δt because rotation is generally much smaller than translation in a real tracking environment.

The system assumes that there is only one moving object in a scene. When there are more than one objects moving to different directions, the system should discriminate objects by pattern matching, binocular disparity, segmentation based on optical flow and so on. Such tasks cost generally much computation and prevent real-time tracking. Then it is a future subject to develop a fast method for discrimination of moving objects. Currently, we are developing a tracking system that combines saccade and smooth pursuit.

Acknowledgement

The work was supported in part by Real World Computing Partnership.

References

[1] Junghyun Hwang, Yoshiteru Ooi and Shinji Ozawa, "An Adaptive Sensing System with Tracking and Zooming a Moving Object," IEICE Transactions Information and Systems. Vol E76-D, No. 8, pp. 926–934, August 1993.

[2] Junghyun Hwang, Yoshiteru Ooi and Shinji Ozawa, "An Advanced On-Line Visual Tracking System," Transactions of the Society of Instrument and Control Engineers. Vol. 30, No. 12, pp. 1427–1435, December 1994.

[3] T. Uhlin, P. Nordlund, A. Maki and J.-O. Eklundh, "Towards an Active Visual Observer," Proceedings of Fifth International Conference on Computer Vision, pp. 679–686, MIT, Cambridge, Massachusetts, USA, June 1995.

[4] James J. Clark and Nicola J. Ferrier, Active Vision. Massachusetts: MIT Press, 1983, Chap. 9, pp. 137–154.

[5] Don Murray and Anup Basu, "Motion Tracking with an Active Camera," IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 16, No. 5, pp. 449–459, May 1994.

[6] David W Murray, Philip F McLauchlan, Ian D Reid and Paul M Sharkey, "Reaction to Peripheral Image Motion Using a Head/Eye Platform," Proceedings of Fourth International Conference on Computer Vision, pp. 403–411, Berlin, Germany, May 1993.

[7] Robinson, D. A., "The mechanics of human smooth pursuit eye movement," Journal of Physiology. 180, pp. 569–591, 1965.

[8] Noboru Sugie, "Investigation of Visual Perception of Position Based on the Reafferent Theory," Biological Cybernetics 21, pp.17–22, 1976.