Proc. Int. Symp. Robotics, Mechatronics and Manufacturing Systems '92,
pp. 957-962, Kobe, Japan, Sep. 1992.

Tracking a Moving Object by an Active Vision System: PANTHER-VZ

Jun Miura, Hideharu Kawarabayashi, Makoto Watanabe,
Teruaki Tanaka, Minoru Asada, and Yoshiaki Shirai

Dept. Mech. Eng. for Computer-Controlled Machinery,

Osaka University, Suita, Osaka 565, Japan

jun@ccm.osaka-u.ac.jp

Abstract

This paper describes an active vision system and its application to object tracking. The system has two cameras with pan, tilt, and vergence (yaw rotation for each camera) mechanisms. Zoom, focus, and aperture of a lens are controllable for each camera. In the experiment, black hair of human was tracked by the system. Candidate regions for the target in the image are first detected based on the similarity of the size and the shape. Three dimensional information is then used to select the target among the candidates. The system successfully tracked a person who walked around the room.

1. INTRODUCTION

Recently, there has been an increasing interest in *active vision*. In the active vision, a camera moves to another viewpoint or changes the viewing direction in order to acquire more information [2] or to simplify ill-posed problems in early vision [1].

Several active vision systems have been developed. A binocular vision system by Clark et al. [4] used a saliency map to find the most interesting point in the image. A saliency map is a weighted sum of multiple feature images which indicate the presence of a feature such as a specific color at each location in the image. By altering the weights, the focus of attention was shifted. Ballard [3] proposed *animate vision* paradigm based on the task-oriented gaze control of a human. By making cameras fixate on an object in the scene, we can easily get depths in the scene in the object-centered frame using motion parallax. Krotokov [6] constructed an agile stereo camera system. He discussed various ranging method such as focus ranging, stereo with verging cameras, and cooperative ranging by focusing and stereo. Olson et al. [8] proposed a method of real-time control of the vergence. A vergence error was estimated with the cepstral disparity filter. Lumia and his group [7] developed a trinocular vision system. A combination of one lens of short focal length and a pair of lenses of long focal length realized the foveal-peripheral vision of human. Crowley et al. [5] constructed a binocular stereo head and described its layered control system.

This paper describes an active vision system, PANTHER-VZ (**Pan** and **T**ilt **He**ad **R**otator with **V**ergence and **Z**oom), as a testbed for active vision research. In this paper, we realize one bahavior of an active observer, which is to keep a target in the image in good quality. That is, the system tries to adjust viewing direction, zoom, and focus so that the target in the image is kept at the center, constant size, and sharp.

2. SYSTEM CONFIGURATION

Figure 1 shows a photograph of PANTHER-VZ. This section describes 1) the camera platform; 2) the stereo cameras and the motorized lenses; and 3) the control system.

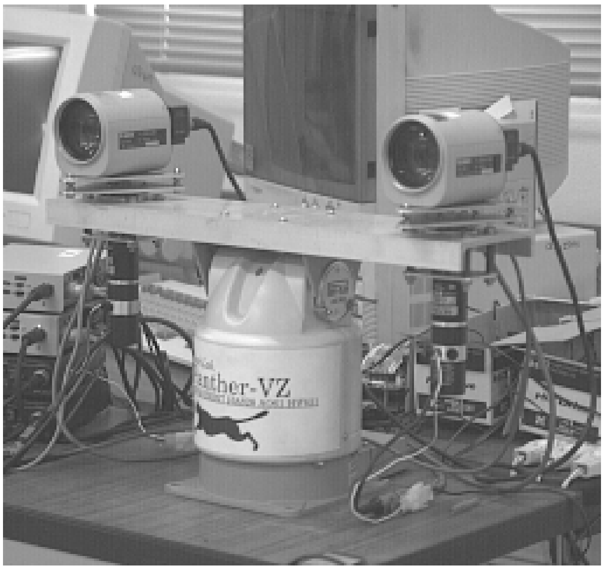


Fig. 1: A photograph of PANTHER-VZ.

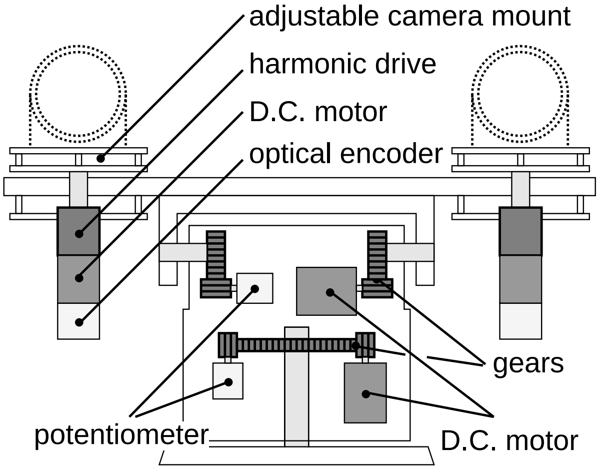


Fig. 2: Mechanism of PANTHER-VZ.

2.1. Camera Platform

Figure 2 shows the mechanism of the camera platform. There are four degrees of freedom; pan, tilt, and vergence. For the vergence mechanism, each camera can rotate independently around the yaw axis. The vergence mechanism which was made by ourselves is on a commercial pan-tilt camera head (by Mikami Co.). The pan and tilt mechanisms are actuated by D.C. servo motors via reduction gears. Pan and tilt positions are sensed by potentiometers. The vergence mechanism is actuated by D.C. servo motors via harmonic drives. The vergence position is sensed by optical encoders. The camera mounts are manually adjustable for the optic axis calibration. Table 1 lists the specifications of the camera platform.

TABLE 1: Specifications of camera platform

Part	*range*	*speed*	*precision*	*actuator*	*position sensor*
Pan	360°	7° /sec	0.276°	D.C. motor	potentiometer
Tilt	-65°∼ 65°	3.5° /sec	0.099°	D.C. motor	potentiometer
Vergence	360°	180° /sec	0.0144°	D.C. motor	optical encoder

TABLE 2: Specifications of motorized lens

lens attribute	*value*
focal length	10 ∼ 100 mm
viewing angle	horizontal: 35.5°∼ 3.7° vertical: 27.0°∼ 2.7°
time for covering the whole range	zoom: 5(s) focus: 8(s) aperture: 3.5(s)
minimum focusing distance	1.0 m
mass (approx.)	1.4 kg

2.2. Cameras and Lenses

The system has two CCD cameras (EC-202II by ELMO Co.). A motorized lens (J10X10R-II by Canon Co.) is attached to each camera. Zoom, focus, and aperture of the lens are controllable with D.C. motors and potentiometers. Table 2 lists the specifications of the lens.

2.3. Control System

Figure 3 shows the configuration of the experimental system. A personal computer (NEC PC9801VX) controls PANTHER-VZ. A high-speed image processing system (Fujitsu FIVIS/VIP) processes video signals from both cameras. A workstation (SUN SPARCstation 330) generates a motion command and sends it to the controller.

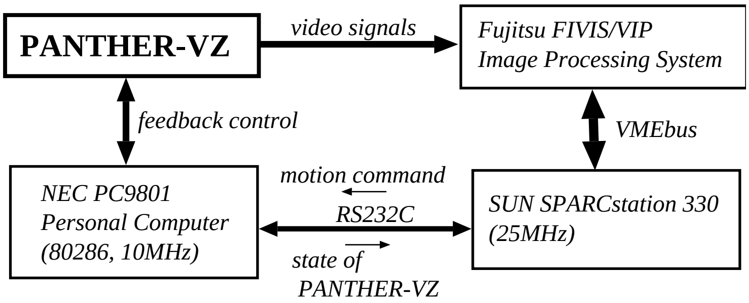
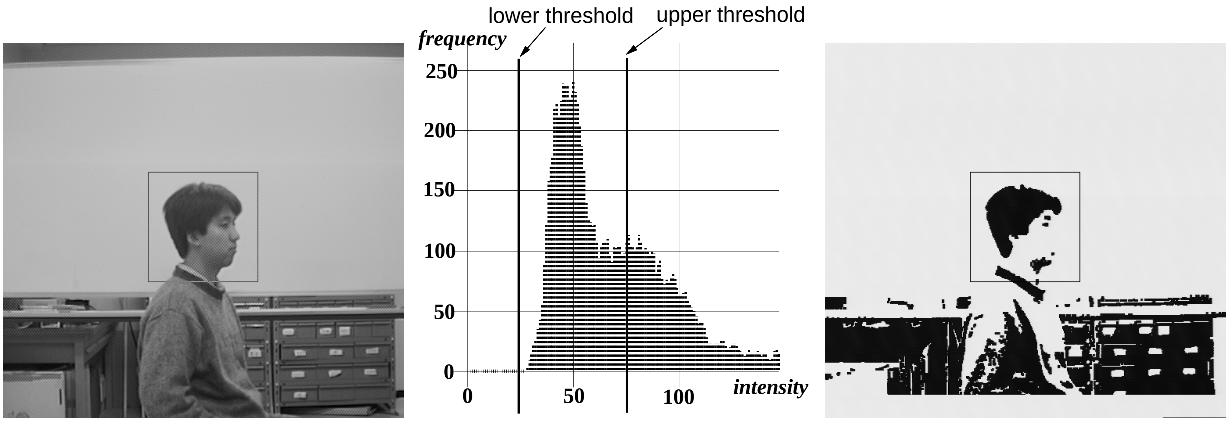


Fig. 3: Configuration of the experimental system.

3. TRACKING

3.1. Detection of Candidate Regions by Image Processing

To realize a fast tracking, simple image features must be used. We used black hair of human as a target. Since black hair usually forms a dark region in the input image, we first collect dark regions as candidates by thresholding. The thresholds are determined by analyzing the first input image: a small rectangle surrounding the target is manually set (see Fig.4(a)); the histogram of intensity values in the rectangle is calculated (see Fig.4(b)); and the upper and the lower thresholds are determined by referring to the position of the largest peak. Fig.4(c) shows remaining regions after thresholding. The biggest region in the rectangle is determined as the target region and its properties are recorded to be used in the similarity judgement described below.



(a) Original image and a rectangle surrounding the target

(b) Histogram for the rectangle

(c) Binary image after thresholding

Fig. 4: First input image.

Each time a pair of image is input, candidate regions of the target are collected among dark regions on the basis of the similarity of the size and the shape. The maximum-minimum ratio of the second moment is used to evaluate the shape. Thresholds for the similarity judgement are empirically determined so that from two to five candidate regions usually remain in the image. If no candidate region remains, the thresholds for detecting dark regions are slightly modified and then the candidate detection is tried again.

3.2. Determination of Target Region using 3D Information

To select the correct region by only the similarity judgement is often difficult especially when the background includes many dark regions. Thus, we use the 3D distance between the position of a candidate and the predicted position in order to select the target among candidates. A target position is predicted by extrapolation with two past positions. For every possible pair of candidate regions in both images, the 3D position is calculated. A pair, the position of which is in a predetermined range and is nearest to the predicted position, is determined as the correct pair of regions of the target.

3.3. Tracking Control

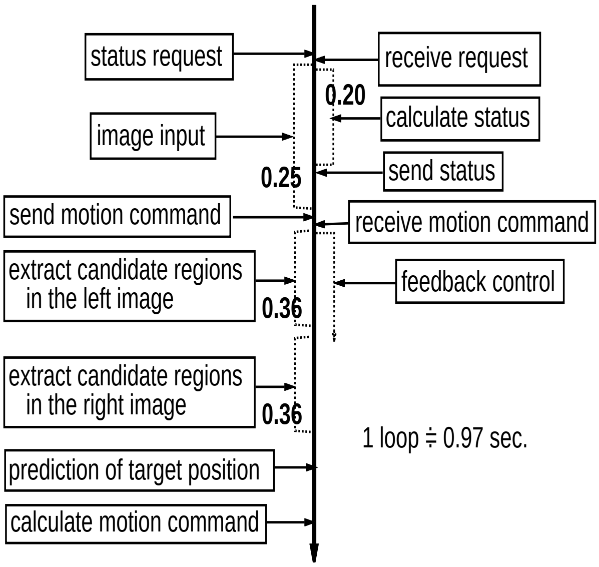
After the target position has been calculated, we again predict the *current* position of the target because the calculated position is the position at the time of the image acquisition. Using such a predicted current position, desired values of pan, tilt, and vergence are calculated so that the target is pointed by the intersection of two optical axes and the rotation angle for the vergence of one camera is equal but opposite in sign to the other. The latter condition is necessary because the positioning mechanism has redundant degrees of freedom. The desired focus and zoom positions are also determined so that the target in the image is kept to be sharp and almost constant size. Since we can calculate the 3D position of the target, we look up the appropriate zoom and focus settings in the precalculated table. From the desired state of PANTHER-VZ, a motion command is generated and is sent to the feedback controller. We currently use a simple PD-controller.

4. EXPERIMENTS

We performed tracking experiments in our laboratory. PATHER-VZ successfully tracked a person by detecting the position of his head. It takes about one second to process a pair of images. Image processing and feedback control are performed in parallel as shown by the timechart in Fig.5. Fig.6 shows the tracking results. Sequences of predicted positions and calculated ones are indicated, in which positions are projected onto the 2D plane. Since we use a linear extrapolation with two past data, prediction sometimes causes overshoots when the target changes its moving direction much.

5. CONCLUSION AND DISCUSSION

We have constructed an active vision system, PANTHER-VZ, as a testbed for active vision research. We applied our system to object tracking. The system successfully tracked a person walking around the room. To track a person who walks more faster, the vergence mechanism must be mainly used since the vergence mechanism is faster than the pan mechanism. In such a case, a good coordination of vergence and pan is necessary as observed in the coordination of head and eyes of human. It is also necessary to investigate the use of other properties in the similarity judgement in order to make the tracking more robust.



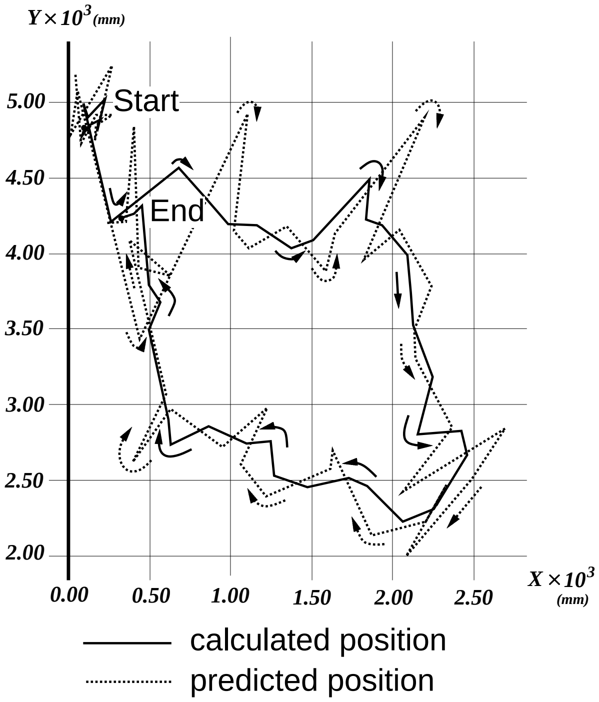


Fig. 5: Time chart in the experiment.

Fig. 6: Predicted and calculated positions.

References

[1] J. Aloimonos and A. Bandyopadhyay. Active Vision. In *Proceedings of the 1st Int. Conf. on Computer Vision*, pp. 35–54, 1987.

[2] R. Bajcsy. Active Perception. *Proceedings of IEEE*, Vol. 76, No. 8, 1988.

[3] D. H. Ballard. Reference Frames for Animate Vision. In *Proceedings of IJCAI-89*, pp. 1635–1641, 1989.

[4] J. J. Clark and N. J. Ferrier. Modal Control of an Attentive Vision System. In *Proceedings of the 2nd Int. Conf. on Computer Vision*, pp. 514–523, 1988.

[5] J. L. Crowley, P. Robet, and M. Mesrabi. Gaze Control for a Binocular Camera Head. In *Proceedings of the 2nd European Conf. on Computer Vision*, pp. 588–596, 1992.

[6] E. P. Krotokov. *Active Computer Vision by Cooperative Focus and Stereo.* Springer-Verlag, New York, 1989.

[7] R. Lumia, et al. TRICLOPS: A Tool for Studying Active Vision. personal communication, 1992.

[8] T. J. Olson and D. J. Coombs. Real-Time Vergence Control for Binocular Robots. In *1990 DARPA Image Understanding Workshop*, pp. 881–888, 1990.