Active Frame Subtraction from Sequential Images Obtained from Moving Camera

Tomonori HASHIYAMA
Grad. School of Information Systems
The Univ. of Electro-Communications
1–5–1, Chofugaoka, Tokyo 182–0026
Email: hashiyama@is.uec.ac.jp

Mitsuru IWATA
Grad. School of Information Systems
The Univ. of Electro-Communications
1–5–1, Chofugaoka, Tokyo 182–0026
Email: miwata@is.uec.ac.jp

Shun'ichi TANO
Grad. School of Information Systems
The Univ. of Electro-Communications
1–5–1, Chofugaoka, Tokyo 182–0026
Email: tano@is.uec.ac.jp

***Abstract*— To detect the movements of the object from sequential images, subtraction of each image frame is useful to eliminate background image when the camera is fixed. However, when the camera is moving with the vehicle or mobile robot, etc., some compensation techniques are required. It is useful to calculate optical flow in the images, but this needs much computational cost to use in real-time systems. We have developed active frame subtraction to compensate the rotation and translation of axes using the sensory information of the camera movements. This is simple and easy to implement on the real systems. In this paper we apply this technique to the autonomous mobile robot.**

I. Introduction

The development of portable computers enables us to bring out a lot of industrial equipments. As a result, we receive much benefit to do outside what we were limited to do within the laboratory. For instance, image processing technique have developed in the lab, because some limitations exists such that lightning must be controlled, the camera should be fixed, and so forth. With the development of computer technologies, novel and useful image processing application have been proposed. One of the main topics within this area is to detect and track the moving object such as human movements. These applications are useful for security device, intelligent transportation systems and so on. Wren et. al. [1] developed Pfinder system. Pfinder models human head, body and legs as oval blob filters and detect human beings. In Ptracker [2] systems, the system mounted on the autonomous mobile robots detects and track human movements using particle filters.

The authors have developed the system to detect the pedestrian from the sequential images by the camera on board using active frame subtraction technique [3]. The system uses affine transformation to compensate camera movements. Using data from acceleration sensor on the vehicle, camera movements and rotations are calculated. This method is simple enough to implement and run in the real-time processing systems. Mukai [4] also proposed the system to compensate the camera movements with transformations and acceleration sensors, but the applications are limited to recovering the 3D structures from the sequential images.

In this paper, we introduce our system and implement into autonomous mobile robot with poor sensors. Some experimental results and future extensions will be shown.

II. Active Frame Subtraction

A. System Definition

In this section, we mention about the Active Frame Subtraction method. In this method, we use two kinds of coordinates, as shown in Fig.1.

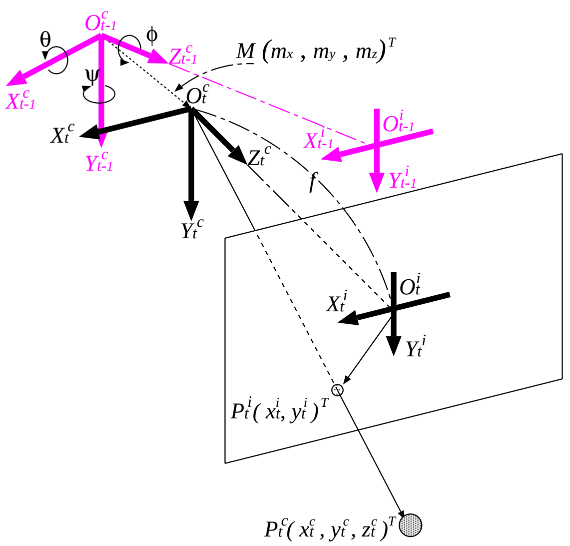


Fig. 1. Coordinate Systems

The first one is a camera coordinate which is attached to and moves with the camera. The camera coordinate at time t is denoted as $O_t^c - X_t^c Y_t^c Z_t^c$ where Z_t^c is along with the optical axis. A posision vector P represented in this coordinate is denoted as $P_t^c = (x_t^c, y_t^c, z_t^c)^T$.

The second one is a image plane coordinate which is located at $Z^c = f$. The image plane coordinate at time t is denoted as $O_t^i - X_t^i Y_t^i$. Let $P_t^i = (x_t^i, y_t^i)^T$ be a projection of P_t^c.

The projection of P_t^c to P_t^i is a perspective projection. The relationship between P_t^c and P_t^i is given by

$$\begin{bmatrix} x_t^i \\ y_t^i \end{bmatrix} = \frac{f}{z_t^c} \begin{bmatrix} x_t^c \\ y_t^c \end{bmatrix} \tag{1}$$

And the relationship between P^c_{t-1} and P^i_{t-1} is similarly given by

$$\begin{bmatrix} x^i_{t-1} \\ y^i_{t-1} \end{bmatrix} = \frac{f}{z^c_{t-1}} \begin{bmatrix} x^c_{t-1} \\ y^c_{t-1} \end{bmatrix} \tag{2}$$

When the object P remain still, the relationship between P^c_{t-1} and P^c_t is given by

$$P^c_t = RP^c_{t-1} + M \tag{3}$$

$$R = \begin{bmatrix} c_\phi c_\psi - s_\phi s_\theta s_\psi & -s_\phi c_\theta & c_\phi s_\psi + s_\phi s_\theta s_\psi \\ s_\phi c_\psi + c_\phi s_\theta s_\psi & c_\phi c_\theta & s_\phi s_\psi - c_\phi s_\theta c_\psi \\ -c_\theta s_\psi & s_\theta & c_\theta c_\psi \end{bmatrix} \tag{4}$$

$$\begin{cases} s_\alpha = \sin\alpha \\ c_\alpha = \cos\alpha \end{cases} \tag{5}$$

where ϕ , θ , ψ and $M = (m_x, m_y, m_z)^T$ denote the angles of roll, pitch, yaw and position vector from the $O^c_{t-1} - X^c_{t-1}Y^c_{t-1}Z^c_{t-1}$, respectively. The angles are derived from the outputs of the gyrosensor. The position vector is calculated using the outputs and the speed of the vehicle.

B. Affine Transformation Among Frame Images

From Eqs.(1), (2), (3), (4) and (5), $P^i_{t-1} = (x^i_{t-1}, y^i_{t-1})^T$ is represented as functions of $P^i_t = (x^i_t, y^i_t)^T$ and z^c_t by

$$\begin{cases} x^i_{t-1}(x^i_t, y^i_t, z^c_t) & = & \frac{(R_{11}x^i_t + R_{12}y^i_t + R_{13}f)z^c_t + fm_x}{(R_{31}x^i_t + R_{32}y^i_t + R_{33}f)z^c_t + fm_z} f \\ y^i_{t-1}(x^i_t, y^i_t, z^c_t) & = & \frac{(R_{21}x^i_t + R_{22}y^i_t + R_{23}f)z^c_t + fm_y}{(R_{31}x^i_t + R_{32}y^i_t + R_{33}f)z^c_t + fm_z} f \end{cases} \tag{6}$$

where R_{ij} denotes the i th row and j th column of matrix R. Given $P^i_t = (x^i_t, y^i_t)^T$ which is the projection of the object P at t, the projection P at $(t-1)$, P^i_{t-1}, is represented as functions of z^c_t by (6). $z^c_t(z_{\text{th}} < z^c_t < \infty)$ is the z-coordinate of the object P in the camera coordinate, where $z^c_t < z_{\text{th}}$ is neglected because objects in the domain aren't detected in the view.

In case z^c_t were available, the relationship between P^c_{t-1} and P^c_t was clearly defined. In general, z^c_t will not be available *a priori*. We have to estimate the optimal z^c_t to obtain the relationship. However z^c_t is hard to deal with, because it is continuous and have the infinite domain. Thus we remove z^c_t from Eq.(6) to derive Eqs.(7),(8) and (9) .

$$A(x^i_t, y^i_t)x^i_{t-1} + B(x^i_t, y^i_t)y^i_{t-1} + C(x^i_t, y^i_t)f = 0 \tag{7}$$

$$\begin{cases} A(x^i_t, y^i_t) & = & (R_{31}x^i_t + R_{32}y^i_t + R_{33}f)m_x \\ & & -(R_{11}x^i_t + R_{12}y^i_t + R_{13}f)m_z \\ B(x^i_t, y^i_t) & = & -(R_{31}x^i_t + R_{32}y^i_t + R_{33}f)m_y \\ & & +(R_{21}x^i_t + R_{22}y^i_t + R_{23}f)m_z \\ C(x^i_t, y^i_t) & = & -(R_{21}x^i_t + R_{22}y^i_t + R_{23}f)m_x \\ & & +(R_{11}x^i_t + R_{12}y^i_t + R_{13}f)m_y \end{cases} \tag{8}$$

$$\left(\begin{array}{ll} & x^i_{t-1}(z^c_t = z_{\text{th}}) < x^i_{t-1} < \lim_{z^c_t \to \infty} x^i_{t-1}(z^c_t) \\ \text{or} & \lim_{z^c_t \to \infty} x^i_{t-1}(z^c_t) < x^i_{t-1} < x^i_{t-1}(z^c_t = z_{\text{th}}) \end{array} \right) \tag{9}$$

Where $P^i_{t-1} = (x^i_{t-1}, y^i_{t-1})^T$ which represents the same object with $P^i_t = (x^i_t, y^i_t)^T$ exists on the trajectory of the line segment determined by Eqs.(7), (8) and (9). With that, we define the Active Frame Subtraction $F_t(x^i_t, y^i_t)$ between the images on $(t-1)$ and t as follows:

$$F_t(x^i_t, y^i_t) \equiv \min_{(u,v) \in \mathcal{S}_{(x^i_t, y^i_t)}} |I_t(x^i_t, y^i_t) - I_{t-1}(u, v)| \tag{10}$$

where $\mathcal{S}_{(x^i_t, y^i_t)}$ denotes the group of coordinates on the trajectory determined with (x^i_t, y^i_t) and $I_t(x, y)$ denotes the pixel value at position (x, y) and time t.

As well as a standard frame subtraction, given a threshold α, P^i_t which satisfy

$$F_t(x^i_t, y^i_t) > \alpha \tag{11}$$

is considered as a projection of object in a foreground. At the same time, P^i_t which is not less α considered as a projection of object in a background. Note that objects which appear from behind other objects and disappear behind other objects owing to the movement of the camera are involved in foreground regions .

III. Experimental Results

In this section, we denote two results using Active Frame Subtraction. The first one is the case where the camera and the precise gyro-sensor are settled on the vehicle. The next one is the case amusement autonomous mobile robot, Sony AIBO(ERS-7) with built-in camera and acceleration sensors.

A. Images from Camera Mounted on the Vehicle

1) Conditions: In this experiments, we denote some results on detecting pedestrians from a moving vehicle using our proposed method [3]. CCD camera(SONY DFW-VL500) and a gyrosensor(DATATEC GU3020) are used. 256 gray scaled images of 640 $\times$ 480 pixels are captured from the camera with 33 msec sampling time. The angle of roll, pitch and yaw of the camera are measured by the gyrosensor at the same time.

In the experiments, we removed the bottom of images and used 640 $\times$ 360 pixels because the body is taken there.

2) Results: Figs.2 show set of sequential frame images. Here, a pedestrian was crossing a street in front of the vehicle running at the speed of 33 km/h.

(a) Time t-1

(b) Time t

Fig. 2. Scene images

Active Frame Subtraction on the images were processed to eliminate background. One sample of this process is shown in Fig.3(a). The background and foreground are painted black and white, respectively. Here $\alpha = 20$ and $z_{\text{th}} = 5.0[m]$. In Fig.3(a), pedestrian was clearly separated from the background. Part of foliage which are hidden in the image at $(t-1)$ are involved in foreground. The total time in this process was 0.1 [sec/frame] on a Athlon XP 1700+. We also show a result of simple frame subtraction with the same threshold in Fig.3(b).

As we can see from this Figs, Active Frame Subtraction method is quite effective to eliminate the background while the camera is moving. When we focused on the pedestrian detection using pattern matching method, the processed images have two main advantages,;

1) Only small number of matching schemes are needed.
2) It will reduce the misdetection rate of pedestrians.

B. Images from the Autonomous Mobile Robot

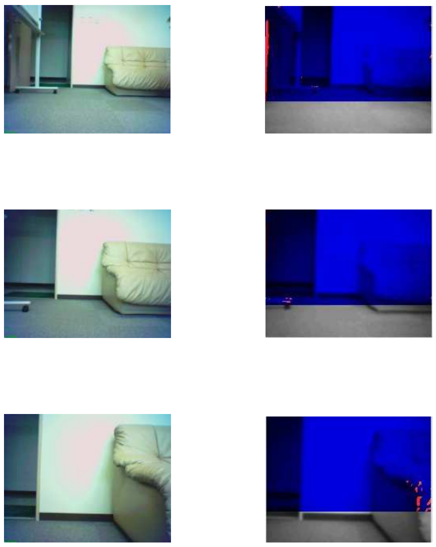
1) Conditions: In this experiment, we show the result with the case of using the autonomous mobile robot Sony AIBO(ERS-7). This robot has built-in CMOS image sensors and 3 acceleration sensors. We use the upper 100 pixels of 176×144 obtained images. The robot was settled on the cart and stand still. The cart was pushed and moved with a walking speed. Other conditions are the same with the previous experiments.

(a) Active Frame Subtraction

(b) Simple Frame Subtraction

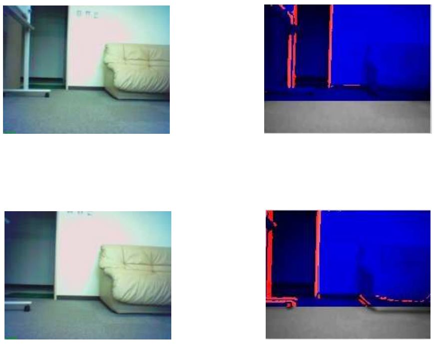
Fig. 3. Subtracted Images

2) Results: In Fig.4, we show the results of this experiments. Figs. 4(a) are the original images obtained from the robot. Figs. 4(b) are the images applied our Active Frame Subtraction method. If object movements have been detected, white(red when in color printed paper) is plotted on the images. As we can see from these figures, background is well eliminated because there are no moving object.



(a)Original Images (b)Processed Images

Fig. 4. Result of Obtained and Processed Images



(a)Original Images (b)Processed Images

Fig. 5. Images with accumulative error of acceleration sensor

On the other hand, when accumulative errors of acceleration sensor are big, some edge of the still object have been mis-detected. These results are shown in the Figures. 5. We have to take care of these problems in the near future.

IV. CONCLUSIONS

This paper introduced some implementation results of our Active Frame Subtraction Systems. When this method is applied to the pedestrian detection system from a moving vehicle with precise sensors, it showed the good performance to eliminate a background of sequential images. The elimination is effective to reduce the cause of misdetection at the template matching. This method is simple enough to process in the real time systems.

Another experiments with the autonomous mobile robot have been carried out. It also showed good results to eliminate a background of sequential images. But in the case where the accumulative error of acceleration sensor are big, some edges of still object were mis-detected. We have to take care of these problems in the near future. We also aims to realize self navigation system using these autonomous mobile robots.

REFERENCES

[1] C. R. Wren, A. Azabayejani, T. Darrell and A. Pentland, “ Pfinder: Real-Time Tracking of the Human Body“, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.19, No. 7, pp. 780-785, (1997)

[2] D. Schulz, W. Burgard, D. Fox and A. Cremers, “Tracking multiple moving targets with a mobile robot using particle filters and statistical data association“, Proc. of IEEE International Conference on Robotics and Automation(ICRA), pp.1656-1670, (2001)

[3] T. Hashiyama, D. Mochizuki, Y. Yano and S. Okuma, “Active frame subtraction for pedestrian detection from images of moving camera“, Proc. of IEEE International Conference on Systems, Man and Cybernetics 2003, Vol.1, pp. 480-485, (2003)

[4] T. Mukai and N. Ohnishi, “Sensor Fusion of a CCD Camera and an Acceleration-Gyro Sensor for the Recovery of Three-Dimensional Shape and Scale“, Proc of the 2nd International Conference on Information Fusion (FUSION‘99), pp.221-228, (1999)