5—3
Object Tracking Method Based on Local Moments

R.Takamatsu H.Kawarada M.Sato
Precision and Intelligence Lab., Tokyo Institute of Technology
{rtakamat,kawarada,msato}@pi.titech.ac.jp

Abstract: This paper proposes a object tracking method based on the local moments, or moment of some restricted area, in which the idea of the viewpoint and the visual field corresponding to the local area of an image is introduced. Using local moment with the optimally controlled viewpoint and visual field, the target position and its breadth are estimated robustly. By two experiments, the validity of the proposed method is shown.

1 Introduction

To track moving target by vision is one of an important issue in robot vision research. Tracking process is often divided into two stages: extraction of the target object from each frame of image sequence (extraction process) and matching between extracted targets from successive frames (matching process). In particular, various methods for extraction process, such as optical flow, have been proposed as a basis of target tracking procedure. However, those "extraction based" methods generally requires relatively high computational cost and has problem of robustness for noise.

In general, if the aim of the system is limited to target tracking, extraction process is not necessary, and the aim can be sufficiently achieved by observing the basic feature of the target such as the center position and breadth. Therefore, we adopt low degree moments of local area to observe a target as the feature. Moments of an image, such as the center of gravity and the breadth around the center of gravity, are robust for noise because they are non-local features in contrast to local features which is usually used for extraction process.

However, moment feature has two defects. Firstly, if an image contains object(s) other than the target, value of moment contains the information other than the target. Secondly, it is difficult to apply moment to real-time target tracking system due to high computational cost[2].

This paper proposes a tracking system based on the local moment, in which the idea of viewpoint and visual field corresponding to local area of an image is introduced. Since the local moment is defined as moment of this local area, it has the capability to suppress unnecessary information outside of the area, and can be applied effectively to an image containing much noises and obstacle objects. Comparing with the ordinary moment, computational cost of the local moment is fairly small, since the area used for calculation is restricted locally.

To apply local moment to actual tracking task, it is indispensable to clarify the procedure to determine viewpoint and visual field . We formulate this determination problem as optimization of viewpoint and visual field by which the error of estimated position of the tracking target is minimized, under the condition that some stochastic properties of the target and noise are given as a priori information.

In brief, the optimality of viewpoint and visual field can be explained as follows. Let us assume that the system tracks a target by iterative process, and the target is most likely to move to the position where the system estimated at previous time period. Then clearly an optimal viewpoint should be that position. What about visual field? In case that the system restricts its visual field to very small area, signal to noise ratio is fairly high, and this yields precise estimation of the target position. However, the system may miss the target even by small fluctuation of the target movement. Conversely, if the system observes the target with broad visual field, it is free from this kind of failure, but accuracy of position estimation decreases. There should be the optimal visual field somewhere between these two extremes.

Using local moment with the optimal viewpoint and visual field stated above, the proposed system observes the image, and estimates the position and breadth of the target. The system is also able to estimate error variance of estimated parameters of the target. Those estimated values will be utilized as a priori information of stochastic properties of the target at next estimation.

2 The Local Moment of Image

Let $\mathbf{r} = [r_1, r_2]^T$ be two dimensional vector, and $f(\mathbf{r})$ be a gray-scale image. Then (i,j)th $(i,j \geq 0)$ moment of the image can be defined as

$$m_{ij} = \int r_1^i \, r_2^j \, f(\mathbf{r}) \, d\mathbf{r}. \tag{1}$$

In general environment, the input image $f(\mathbf{r})$ possibly contains not only the target but also other objects and noises. In such a case, moments would contain components other than the target, and precise tracking may be difficult. One way to solve this problem is to weight the image so that local area around the target is emphasised and information surrounding this area is suppressed. In the following discussion, Gaussian function G written below is used as the weight function.

$$G(\mathbf{r} - \mathbf{r}_c, \tau) = \frac{1}{4\pi\tau} \exp\left[-\frac{\|\mathbf{r} - \mathbf{r}_c\|^2}{4\tau}\right]$$

where $\mathbf{r}_c = [r_{c1}, r_{c2}]^T$ is the center of the weight and τ represents the breadth around the center. Considering this weight as "focus of attention " which plays a crucial role in actual visual system of animals, we name the center of the weight c viewpoint, and the breadth around the center τ visual field.

The local moment m_{ij} is a moment with viewpoint and visual field as stated above, and is defined as follows.

$$m_{ij}(\mathbf{r}_c, \tau) = \int (r_1 - r_{c1})^i (r_2 - r_{c2})^j \, G(\mathbf{r} - \mathbf{r}_c, \tau) f(\mathbf{r}) d\mathbf{r}$$

Without loss of generality, we can set $\mathbf{r}_c$ to be zero vector. Therefore, the (i,j)th local moment can be written as

$$m_{ij} = \int r_1^i \, r_2^j \, G(\mathbf{r}, \tau) \, f(\mathbf{r}) \, d\mathbf{r}.$$

If the task of the system is limited to target tracking, it is not necessary to use higher local moments which represents detailed information of target form. Instead, target position and breadth should be estimated for tracking the target and controlling visual field. Hence, the system makes use of the local moments up to the second degree.

For the purpose of simplifying the following calculation, we orthonormalize kernels of the local moments $r_1^i \, r_2^j \, G(\mathbf{r}, \tau)$ Then, we have four local moments:

$$m_k = \int f(\mathbf{r}) q_k(\mathbf{r}) d\mathbf{r}, \quad (k = 0, 1, 2, 3) \tag{2}$$

where

$$\begin{aligned} q_0 &= \sqrt{8\pi\tau}\, G(\mathbf{r}, \tau) \\ q_1 &= \sqrt{8\pi}\, r_1 G(\mathbf{r}, \tau) \\ q_2 &= \sqrt{8\pi}\, r_2 G(\mathbf{r}, \tau) \\ q_3 &= \sqrt{\frac{2\pi}{\tau}}(r_1^2 + r_2^2 - 2\tau) G(\mathbf{r}, \tau). \end{aligned}$$

Considering the geometrical characteristics of kernels, we can see that each local moments $m_i (i = 0, 1, 2, 3)$ represent the following features.

- As q_0 has the "bell" like form, m_0 represents the average gray level within visual field.
- m_1, m_2 stand for gravity center of r_1 and r_2 direction within visual field respectively.
- m_3 shows a degree of how the image within visual field concentrates around the origin, because q_3 has the "Mexican hat" like form.

3 Target Estimation by Image Model

3.1 Image Model

The local moments can be affected by the image of obstacle objects within visual field. To avoid this problem, the system estimates the position and breadth of the target by using a simple model of a target object and noise. The system can achieve robustness to obstacle and noise which do not fit to the model is attained by this model based estimation.

As described before, the proposed system does not extract the target. Instead, it achieves tracking task by using basic feature of the target such as position or breadth. Moreover, one of the most essential characteristics of the target in general as image pattern is that they are distributed locally. Therefore, desirable target model must have position and breadth as the parameters, and must be distributed locally as image pattern. As the model of the target which satisfies those requirements, we adopt isotropic Gaussian function and formulate the model as follows.

$$g(\mathbf{r}) = a_T \, G(\mathbf{r} - \mathbf{r}_T, \tau_T) \tag{3}$$

where parameter of object model is height, position and breadth of Gaussian function, which are represented as a_T, $\mathbf{r}_T$ and τ_T respectively.

Generally, input image $f(\mathbf{r})$ can be represented by the sum of object model and noise component as follows.

$$f(\mathbf{r}) = g(\mathbf{r}) + n(\mathbf{r}) \tag{4}$$

Accordingly, as the result of observation of image $f(\mathbf{r})$ by local moment $m_k (k = 0, 1, 2, 3)$, the following four observation formulae are obtained.

$$m_k = g_k + n_k \qquad (k = 0, 1, 2, 3), \tag{5}$$

where g_k and n_k are determined by target model $g(\mathbf{r})$ and noise component $n(\mathbf{r})$ respectively, and can be written as,

$$g_k = \int g(\mathbf{r})\, q_k(\mathbf{r})\, d\mathbf{r} \tag{6}$$

$$n_k = \int n(\mathbf{r}) q_k(\mathbf{r})\, dr \tag{7}$$

Then, model parameters can be expressed by $g_k (k = 0, 1, 2, 3)$ as follows.

$$\begin{cases} a_T = \sqrt{2\pi\tau}\dfrac{4{g_0}^3}{\gamma}\exp\left[\dfrac{{g_1}^2 + {g_2}^2}{\gamma}\right] \\ \mathbf{r}_T = \dfrac{4g_0}{\gamma}\sqrt{\tau}[g_1, g_2]^T \\ \tau_T = \dfrac{4{g_0}^2 - \gamma}{\gamma}\tau \end{cases} \tag{8}$$

where $\gamma = 2{g_0}^2 + {g_1}^2 + {g_2}^2 - 2g_0g_3$

3.2 Stochastic Properties of Image

The proposed system tracks the target by iterative process. At each time period of iteration, the system estimates probability density of the target position and breadth first. Variance and average of these densities are updated every time by target parameters estimated at previous time period. This updating procedure will be discussed in the later section.

We assume that probability density of object position $\mathbf{r}_T = [r_{T1}, r_{T2}]^T$ is normal distribution and is given as

$$p(\mathbf{r}_T) = \frac{1}{2\pi s^2} exp(-\frac{\|\mathbf{r}_T - \bar{\mathbf{r}_T}\|^2}{2s^2}),$$

where s stands for the degree of uncertainty of target position, and $\bar{\mathbf{r}_T}$ stands for the average *position of the target*.

Hence, the average and covariance matrix of $\mathbf{r}_T$ become

$$E[\mathbf{r}_T] = \bar{\mathbf{r}_T} \equiv [\bar{r}_{T1}, \bar{r}_{T2}]^T$$

$$E[(\mathbf{r}_T - \mathbf{r}_T)(\mathbf{r}_T - \bar{\mathbf{r}_T})^T] = \begin{bmatrix} s^2 & 0 \\ 0 & s^2 \end{bmatrix} \equiv P_\eta.$$

As for probability density of object breadth τ_T, we assume gamma distribution which has the average of $\bar{\tau_T}$ and the variance of $t^2 \equiv P_\zeta$ shown below.

$$p(\tau_T) = \frac{\vartheta^\iota}{\Gamma(\iota)}(\tau_T)^{\iota-1}\exp(-\vartheta\tau_T)$$

where $\iota = \frac{(\bar{\tau_T})^2}{t^2} > 0$, and $\vartheta = \frac{\bar{\tau_T}}{t^2} > 0$.

Assuming that $n(\mathbf{r})$ is an additive noise which has average of 0 and variance of Δ^2, and is independent of $g(\mathbf{r})$, we have

$$E[n(\mathbf{r})] = 0, \tag{9}$$

$$E[n(\mathbf{r})n(\mathbf{r}')] = \Delta^2\delta(\mathbf{r} - \mathbf{r}'), \tag{10}$$

where $\delta(\mathbf{r})$ denotes for delta function.

3.3 Observation of Image

Based on Eq.(8), we introduce observation vector $\eta = [\eta_1, \eta_2]^T$ and value ζ as follows.

$$\eta \equiv \frac{4m_0}{\mu}\sqrt{\tau}[m_1, m_2]^T$$

$$\zeta \equiv \frac{4{m_0}^2 - \mu}{\mu}\tau$$

where $\mu = 2{m_0}^2 + {m_1}^2 + {m_2}^2 - 2m_0m_3$

If noise factor n_i are sufficiently small, above equations can be written as

$$\eta = \mathbf{r}_T + \varepsilon$$

$$\zeta = \tau_T + \rho$$

where, $\varepsilon = [\varepsilon_1, \varepsilon_2]^T$ and ρ are observation noise expressed as follows.

$$\begin{aligned} \varepsilon_1 &= \frac{4\sqrt{\tau}}{\gamma^2}\{(-2g_0^2 + g_1^2 + g_2^2)g_1n_0 \\ &\quad +\gamma g_0n_1 - 2g_0g_1g_2n_2 \\ &\quad +2g_0^2g_1n_3\} \\ \varepsilon_2 &= \frac{4\sqrt{\tau}}{\gamma^2}\{(-2g_0^2 + g_1^2 + g_2^2)g_2n_0 \\ &\quad -2g_0g_1g_2n_1 + \gamma g_0n_2 \\ &\quad +2g_0^2g_2n_3\} \\ \rho &= \frac{16\tau g_0}{\gamma^2}\{(g_1^2 + g_2^2 - 2g_0g_3)n_0 \\ &\quad -g_0g_1n_1 - g_0g_2n_2 + g_0^2n_3\} \end{aligned}$$

where $\gamma = 2g_0^2 + g_1^2 + g_2^2 - 2g_0g_3$.

Average and covariance matrix of observation noise are as follows.

$$E[\boldsymbol{\varepsilon}] = 0$$

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \begin{bmatrix} E[\varepsilon_1^2] & E[\varepsilon_1\varepsilon_2] \\ E[\varepsilon_2\varepsilon_1] & E[\varepsilon_2^2] \end{bmatrix} \equiv R_\eta$$

$$R_\eta = \begin{bmatrix} \kappa_1^2 & \kappa_3^2 \\ \kappa_3^2 & \kappa_2^2 \end{bmatrix}$$

$$E[\rho] = 0 \quad E[\rho^2] \equiv R_\zeta \quad R_\zeta = \lambda^2$$

Note that as shown in the following equations, $\boldsymbol{\varepsilon}$ *and* ρ *depend on but have no correlation to* $\mathbf{r}_T$ and τ_T respectively.

$$E[\boldsymbol{\varepsilon}\mathbf{r}_T^T] = 0 \qquad E[\rho\tau_T] = 0$$

3.4 Target Estimation

Based on minimum variance estimate[3], estimated object position $\mathbf{r}_T$ and object breadth τ_T can be written as

$$\begin{aligned}\mathbf{r}_T &= P_\eta(P_\eta + R_\eta)^{-1}(\eta - \bar{\mathbf{r}_T}) + \bar{\mathbf{r}_T} \\ &= \frac{s^2}{(s^2+\kappa_1^2)(s^2+\kappa_2^2)-\kappa_3^4} \\ &\quad \begin{bmatrix} s^2+\kappa_2^2 & -\kappa_3^2 \\ -\kappa_3^2 & s^2+\kappa_1^2 \end{bmatrix} \begin{bmatrix} \eta_1 - \bar{x_T} \\ \eta_2 - \bar{y_T} \end{bmatrix} \\ &\quad + \begin{bmatrix} \bar{x_T} \\ \bar{y_T} \end{bmatrix} \qquad (11)\end{aligned}$$

$$\begin{aligned}\tau_T &= P_\zeta(P_\zeta + R_\zeta)^{-1}(\zeta - \bar{\tau_T}) + \bar{\tau_T} \\ &= t^2(t^2+\lambda^2)^{-1}(\zeta - \bar{\tau_T}) + \bar{\tau_T} \\ &= \frac{t^2}{t^2+\lambda^2}(\zeta - \bar{\tau_T}) + \bar{\tau_T} \qquad (12)\end{aligned}$$

The error variance of estimated position and breadth can be obtained as the following forms which have no relation to observation vectors. This means that the system can evaluate the variance of estimation error prior to actual observation.

$$\begin{aligned}J &= tr\{P_\eta - P_\eta(P_\eta + R_\eta)^{-1}P_\eta\} \\ &= 2s^2 - \frac{s^4(2s^2+\kappa_1^2+\kappa_2^2)}{(s^2+\kappa_1^2)(s^2+\kappa_2^2)-\kappa_3^4} \qquad (13) \\ L &= tr\{P_\zeta - P_\zeta(P_\zeta + R_\zeta)^{-1}P_\zeta\} \\ &= \frac{t^2\lambda^2}{t^2+\lambda^2} \qquad (14)\end{aligned}$$

Here J and L denote the error variance of position estimation and breadth estimation respectively. Note that the matrix R_η and thus J are a function of viewpoint and visual field.

4 Optimal Control of Viewpoint and Visual Field

As shown in previous section, the system can evaluate the variance of estimation error prior to actual observation. Therefore, the system can determine the optimal viewpoint and visual field which minimizes the error variance of estimated position J.

4.1 Optimal Viewpoint

Averaging Eq.(13) with respect to object position and noise, J becomes even function with respect to $\bar{r}_{T1}$ and $\bar{r}_{T2}$. Therefore, it is clear that J has minimum when viewpoint is the average position of the object $\bar{\mathbf{r}_T}$, which is the optimal viewpoint $\mathbf{r}_c^*$.

4.2 Optimal Visual Field

If the model sets its viewpoint to be optimal, each components of matrix R_η become

$$\begin{aligned}\kappa^2 &= \kappa_1^2 = \kappa_2^2 \\ &= \frac{\Delta^2\pi(\tau_T+\tau)^4}{a_T^2 \quad \tau^3}\{\frac{3\tau^2 s^6}{4(\tau_T+\tau-s^2)^4} \\ &\quad + \frac{(\tau_T+\tau)(\tau_T-\tau)s^2}{(\tau_T+\tau-s^2)^2} + \frac{2\tau(\tau_T+\tau)}{\tau_T+\tau-s^2}\} \qquad (15)\end{aligned}$$

$$\kappa_3^2 = 0. \qquad (16)$$

Then we have

$$J = \frac{2s^2\kappa^2}{s^2+\kappa^2}. \qquad (17)$$

Because J is a monotonic function of κ, the optimal visual field τ^* is the one which minimizes κ. If the breadth of the target changes slowly, or $t \simeq 0$, we can approximate Eq.(15) by substituting $\bar{\tau_T}$ for τ_T. Fig.1 shows the relationship between κ^2 and τ calculated from Eq.(15) with this substitution.

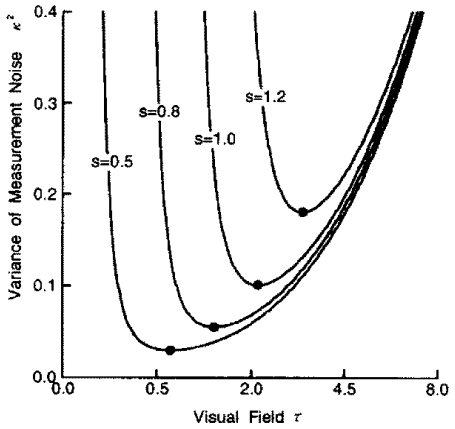

Figure 1: Change of κ^2 with respect to visual field τ

($\Delta = 0.03$, $a_T = 1$, $\bar{\tau_T} = 0.5$)

As in this figure, the graphs are downwards convex, and have their minima κ^{2*} at certain visual fields τ^* indicated as dots. Therefore, we can conclude that if the average of object breadth and the variance of object position are given, the optimal visual field τ^* can be determined. That is,

$$\tau^* = F(\bar{\tau_T}, s),$$

where F is the function which maps $(\bar{\tau_T}, s)$ to τ^*. The system has this function as a table, items of which are calculated beforehand.

Fig.1 also indicates that κ^{2*} increases as s increases. In other words, estimation of position becomes inaccurate if the uncertainty of position

of appearance increases. Note that Eq.(15) shows that the optimal visual field τ^* is determined by s and $\bar{\tau}_T$ and does not depend on Δ or a_T.

Fig.2 shows the change of the optimal visual field with respect to variance of object position s.

This figure indicates the fact that with the increase of s^2, τ^* also increases monotonously. In other words, the larger uncertainty of object position is, the wider visual field should be, in order to prevent the image of an object from being outside of visual field.

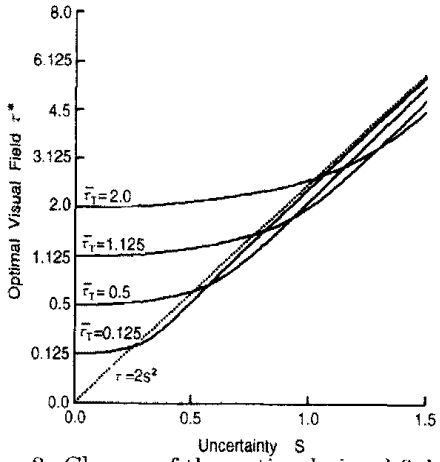

Figure 2: Change of the optimal visual field with respect to variance of object position s ($\bar{\tau}_T = 0.5$)

5 Structure of Tracking System

An overview of whole structure of the tracking system is indicated in Fig.3. As shown in the figure, the tracking system is divided into four procedures. Note that in the following discussion, suffix (t) denotes time period t.

First, viewpoint and visual field are determined by given a priori information of the object as follows.

$$\begin{aligned} \mathbf{r}_c(t) &\equiv \bar{\mathbf{r}}_T(t) \\ \tau(t) &\equiv F(\bar{\tau}_T(t), s(t)) \end{aligned}$$

Secondly, as shown in Eq.(2) and Eq.(11),an image is observed by the local moments.

Thirdly, the target position and breadth are estimated by image model, as indicated in Eq.(11), Eq.(12), Eq.(15) and Eq.(16).

Finally, the estimated target parameters and its error variances are adopted as a priori information of the object at the next time period.

As shown in the following equations, the averages of object position and breadth of next time period are updated by using object position and breadth currently estimated.

$$\begin{aligned} \bar{\mathbf{r}}_T(t+1) &\equiv \hat{\mathbf{r}}_T(t) \\ \bar{\tau}_T(t+1) &\equiv \hat{\tau}_T(t) \end{aligned}$$

Simultaneously, the variance of object position and breadth of next time period are updated by using object position and breadth currently estimated.

$$\begin{aligned} s^2(t+1) &\equiv J(t) \\ t^2(t+1) &\equiv L(t) \end{aligned}$$

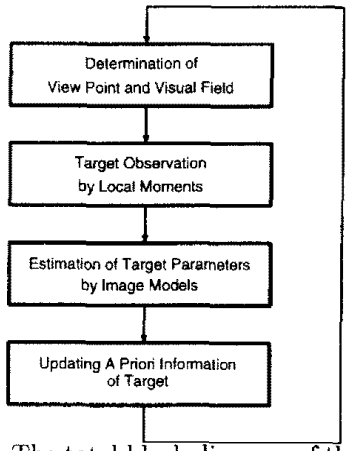

Figure 3: The total block diagram of the system.

6 Tracking Experiments

α - β filter [4] is the widely used for object tracking, by which target position is iteratively predicted from given object position, assuming that the target motion is uniform and linear with some maneuverability. Combining this filter with proposed estimation system, some tracking experiments are executed. As shown in figure4, in which model image is used, the system precisely tracks the target, while it keeps visual field around the target stably and narrows visual field gradually.

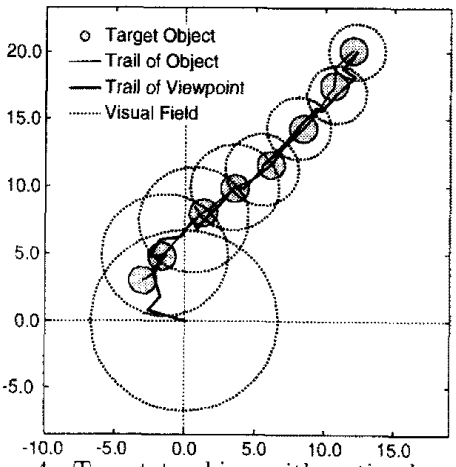

Figure 4: Target tracking with optimal control of viewpoint and visual field
($a_T = 1$, $\Delta = 0.03$, $\bar{\tau}_T = 0.5$)

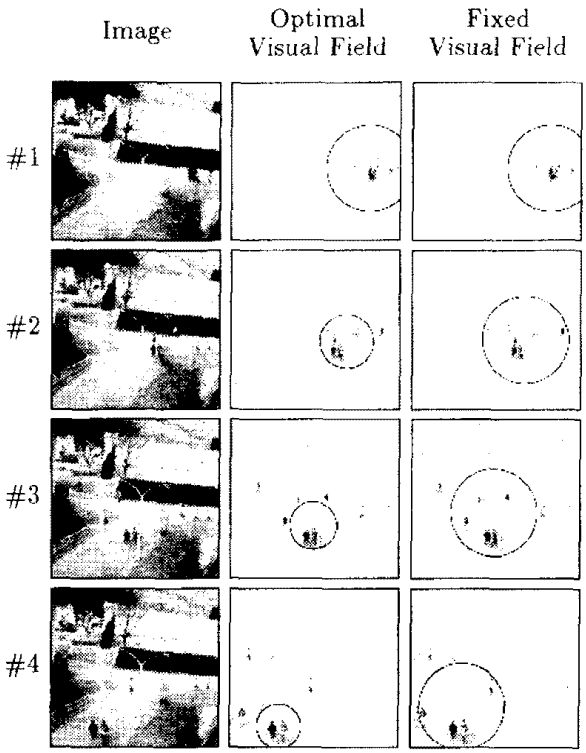

Figure 5: Tracking experiment by real image sequence

The result of tracking experiment by real image sequence is shown in Fig. 5. In this experiment, time differences between adjacent frames are used as input image sequence, and the system tracks the area of motion enhanced by difference. In this example, two persons walking from right side to lower left corner. If visual field is fixed, as shown in the right row, tracking is affected by other persons, and the system does not observe the target in the center of the visual field after frame #3. Conversely, if visual field is controlled optimally, as shown in the middle row, the system tracks the target steadily.

7 Conclusion

In this paper, we propose the target tracking system in which the observation is made by the local moment, and the parameters of the local moments are controlled to be optimal. Our method has the following properties.

- The local moment introduces the idea of viewpoint and visual field corresponding to local area of an image. Since the local moment is defined as moment of this local area, it has the capability to suppress unnecessary information outside of the area.
- Because the local moments are non-local features, they can be affected by the image of obstacle objects within visual field. To avoid this problem, the system estimates the position and breadth of the target by using a simple model of a target object and noise. The system can achieve robustness to obstacle and noise which do not fit to the model is attained by this model based estimation.
- Determination procedure of viewpoint and visual field is formulate as optimization of observation system in which the error of estimated position of the tracking target is minimized.

As the future work, tracking experiment using real image sequence to evaluate validity and stability of our model and application to real-time robot vision are considered.

References

[1] M-K.Hu: "Visual pattern recognition by moment invariants", IRE Trans. on Information Theory, IT-8, 2, pp. 179–187 (1962).

[2] C.Maggioni: "A novel gestural input device for virtual reality", IEEE Virtual Reality Annual International Symposium, pp. 118–124 (1993).

[3] David G. Luenberger:"Optimization by Vector Space Methods", John Wiley & Sons (1969)

[4] Paul R. Kalata:"The Tracking Index:A Generalized Parameter for α-β and α-β-γ Target Trackers", IEEE Transactions on Aerospace and Electronic Systems, Vol.AES-2, No.2, pp 174-182 (1984)

[5] R. Takamatsu, S. Nakai, M. Sato and H. Kawarada:"Object Detection with the Optimal Viewpoint and Visual Field", Proc. Int. Joint Conf. on Neural Networks, Nagoya, Japan, pp 215–218 (1993)

[6] R. Takamatsu, S. Nakai, M. Sato and H. Kawarada:"Optimal Control System of Viewpoint and Visual Field on Object Tracking", IEICE Tech. Report, PRU95-79, pp1–6 (1995).