TRACKING OF FEATURE POINTS IN A SCENE OF MOVING RIGID OBJECTS BY BAYESIAN SWITCHING STRUCTURE MODEL WITH PARTICLE FILTER

Norikazu Ikoma, Yasutake Miyahara, and Hiroshi Maeda
Faculty of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kita-Kyushu, Fukuoka 804-8550, JAPAN
E-mail: ikoma@comp.kyutech.ac.jp

Abstract. Causal estimation of multiple feature points trajectories by using a switching state space model is proposed. The state vector of the model consists of the position of each feature point, the velocity of each rigid object, and some indicator variables for each feature point. There are two types of indicator variables: an object indicator representing the association between the feature point and rigid object, and an aperture indicator representing the attribute of the point e.g. aperture or not. By estimating the state vector using a Rao-Blackwellized particle filter, smooth trajectories of feature points, velocity of objects, object indicators, and aperture indicators are obtained simultaneously. Performance on a real image sequence is presented by comparing to a Kalman filter being given true indicators.

Key words: feature point tracking, image sequence, particle filter, Rao-Blackwellization, state space model.

INTRODUCTION

The bottom-up approach to image processing is an attractive topic in the field of computer vision. It uses 2D trajectories of feature points detected and tracked locally in an image sequence, without specific knowledge about the objects in the scene. A typical example of this approach is reconstruction of 3D structure and motion from the 2D trajectories, called "Structure from Motion(SfM)" [10].

Although this general approach is attractive, there are several problems with it. Firstly, correct detection of corners is a difficult task in a real scene when based only on local features of the image. Thus incorrect feature points such as edges of the object are unavoidable. Secondly, tracking of the feature point might involve some difficulties due to the simple template matching using the local feature. The most important one among them is the aperture problem, i.e., incorrect feature points such as edges lead to a matching ambiguity. Consequently, the resulting trajectories of feature points are not necessarily as smooth as the actual motion of objects.

Some conventional approaches to mitigate these problems apply "filtering" technique by using a state space model for tracking feature points. Among them, [9] uses a heavy-tailed distribution instead of a Gaussian distribution for measurement noise in order to avoid the undesired effects of feature points of aperture. The idea is to consider the aperture point as an outlier, and perform robust estimation based on the property of heavy-tailed distributions. However, the actual trajectory of a feature point of aperture is often much worse than an outlier, i.e., the observation error distribution is neither centered at the true position nor symmetric.

We approach these problems by using multiple feature points rather than a single feature point. Then, based on the fact that all feature points on the same rigid object share a common velocity, position estimation of aperture points becomes possible. Here, we also use a large variance of observation noise for aperture points. To perform estimation in this approach, we should know two types of associations: (1) between feature points and rigid objects, and (2) which points are of an aperture. However, these associations are unknown.

To overcome this, we introduce a Bayesian switching structure into the state space model. Then, we can estimate these associations simultaneously with the position of feature points and the velocity of objects. The model can be written as a jump Markov linear system where the transition matrix of the system equation and the coefficients matrix of the observation noise term depend on a part of the state vector. Although the model is highly nonlinear, a recent computer intensive estimation method called "particle filter", or Sequential Monte Carlo, [6] [12] can be applied to state estimation in this case [5]. A numerical variance reduction method called Rao-Blackwellization [1] is also applied to make state estimation more precise than for the conventional particle filter. As a novelty of this paper, we propose a sub-optimal importance function for the particle filters.

For the first attempt to use particle filters to this problem, we introduce several assumptions as follows: a known number of rigid objects are moving in parallel with the image plane, without rotating. Velocity of each object is almost constant over short periods and changes smoothly. The number of feature points is fixed throughout the observations, i.e., there is no occlusion. These assumptions are just introduced to illustrate a simple case of dynamic image processing to evaluate the performance of particle filters, and we have a plan to relax these assumptions in future researches. We will show an example of real image analysis under this condition at the end of this paper.

PROBLEM STATEMENT

Let p be the fixed and known number of feature points in the image. The j-th feature point at discrete time k (integer) is denoted by a row vector

$$\mathbf{x}_{k,j} = [X_{k,j}\ ,\ Y_{k,j}]\,,\quad j = 1, \cdots, p. \tag{1}$$

Let q be the fixed and known number of objects in the scene. The velocity of i-th object at time k is denoted by a row vector

$$\mathbf{s}_{k,i} = \left[S^x_{k,i}\ ,\ S^y_{k,i}\right]\,,\quad i = 0, 1, \cdots, q, \tag{2}$$

where $\mathbf{s}_{k,0}$ is a null vector that corresponds to the case of background objects in the image. More precisely, it is the velocity of an object for which the relative velocity to the observer is always zero.

Dynamics of feature point is assumed to satisfy

$$\mathbf{x}_{k,j} = \mathbf{x}_{k-1,j} + \mathbf{s}_{k,I(j)}\,, \tag{3}$$

where $I(j) \in \{0, 1, \cdots, q\}$ denotes association of j-th feature point to object.

Observation of feature point in eq. (1) is denoted by

$$\mathbf{y}_{k,j} = [x_{k,j}\ ,\ y_{k,j}]\,,\quad j = 1, \cdots, p. \tag{4}$$

It depends on the physical observation process. Typically, it is obtained from the true one $\mathbf{x}_{k,j}$ with additive Gaussian observation noise $\mathbf{w}_{k,j}$. However, in aperture case, the noise could be more complex.

Suppose we can only observe $\mathbf{y}_{k,j}$; see eq. (4). Consequently, feature points in eq. (1), velocity of objects in eq. (2), association between feature points and objects $I(j)$ in eq. (3) are unknown. It is also assumed that we would not know which points are of aperture in the observation process eq. (4). Then, the problem is to estimate these unknown variables from the observation series causally (i.e. using up to current observation), thus sequentially.

MODEL

To solve the problem stated in the previous section, we propose to use a switching state space model.

We assume that the velocity of each object is almost constant over short time periods. This can be represented by a random walk model with Gaussian white noise $\mathbf{v}_{k,j} \sim N(0, \tau^2)$ as

$$\mathbf{s}_{k,j} = \mathbf{s}_{k-1,j} + \mathbf{v}_{k,j}. \tag{5}$$

Let the state vector consists of the positions of all feature points and the velocities of all objects,

$$\mathbf{x}_k{}^T = [\mathbf{x}_{k,1}\ , \cdots,\ \mathbf{x}_{k,p}\ ,\ \mathbf{s}_{k,1}\ , \cdots,\ \mathbf{s}_{k,q}]\,, \tag{6}$$

where $\mathbf{x}^T$ denotes $\mathbf{x}$ transposed. The observation vector is defined as

$$\mathbf{y}_k{}^T = [\mathbf{y}_{k,1}\ , \cdots ,\ \mathbf{y}_{k,p}] \,. \tag{7}$$

Then, we can formulate a state space model as a pair of system and observation equations

$$\begin{cases} \mathbf{x}_k &= \mathbf{F}(\mathbf{m}_k)\mathbf{x}_{k-1} + \mathbf{G}\mathbf{v}_k, \\ \mathbf{y}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{E}(\mathbf{a}_k)\mathbf{w}_k, \end{cases} \tag{8}$$

where $\mathbf{v}_k$ is a vector of system noise with element $\mathbf{v}_{k,j}$ in eq. (5), and $\mathbf{w}_k$ is a vector of observation noise in which each element of $\mathbf{w}_{k,i}$ is assumed to be zero mean Gaussian white with variance σ^2. $\mathbf{F}(\cdot)$ and $\mathbf{G}$ are matrices properly defined to satisfy eq. (3) and (5), and matrix $\mathbf{H}$ is defined as identity $\mathbf{y}_{k,j}$ to be $\mathbf{x}_{k,j}$ plus corresponding noise term. The role of matrix $\mathbf{E}(\cdot)$ is to control the variance of observation noise depending on $\mathbf{a}_k$.

Actually, in eq. (8), $\mathbf{F}(\cdot)$ and $\mathbf{E}(\cdot)$ depend on the object indicator vector $\mathbf{m}_k$ and aperture indicator vector $\mathbf{a}_k$,

$$\mathbf{m}_k = [m_{k,1}, \cdots, m_{k,p}] \,, \mathbf{a}_k = [a_{k,1}, \cdots, a_{k,p}] \,, \tag{9}$$

where $m_{k,j} \in \{0, 1, \cdots, q\}$ and $a_{k,j} \in \{-1, 0, 1\}$. These dependencies define the switching structure of the model. $\mathbf{F}(\cdot)$ and $\mathbf{E}(\cdot)$ are defined as follows. For $\mathbf{F}(\cdot)$, if $m_{k,j} = i$, then $I(j) = i$ is used in eq. (3). For $\mathbf{E}(\cdot)$, if $a_{k,j} = 1$, then j-th feature point is considered as of aperture and a large variance of observation noise is applied. $a_{k,j} = 0$ corresponds to the ordinary feature points so σ^2 is used, and $a_{k,j} = -1$ corresponds to the background feature points so a small variance of observation noise is used.

If we knew the true value of indicator variables in eq. (9), then the state space model, eq. (8), would be simply linear Gaussian, so the Kalman filter could be successfully applied to perform state estimation. However we do not know these values in our setting, so we have to estimate these variables simultaneously with the state $\mathbf{x}_k$. To achieve this, indicator variables are assumed to be Markov chains

$$\Pr\{m_{k,l} = i | m_{k-1,l} = j\} = p^m{}_{i,j}, \tag{10}$$

$$\Pr\{a_{k,l} = i | a_{k-1,l} = j\} = p^a{}_{i,j}, \tag{11}$$

with high probability for diagonal ($i = j$) elements. Then, we have a form of jump Markov linear system by eqs.(8) (9), (10), and (11).

A formal description of the model is as follows; Let $\theta_{k,j} \equiv [m_{k,j}\ ,\ a_{k,j}]$. By augmenting the state vector $\mathbf{x}_k$ into

$$\mathbf{z}_k{}^T = \left[\mathbf{x}_k{}^T\ ,\ \theta_k\right], \quad \theta_k \equiv [\theta_{k,1}, \cdots, \theta_{k,p}] \,, \tag{12}$$

we can formulate a nonlinear state space model as

$$\begin{cases} \mathbf{z}_k &= \mathbf{f}(\mathbf{z}_{k-1}, \mathbf{v}_k), \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{z}_k, \mathbf{w}_k). \end{cases} \tag{13}$$

ESTIMATION

State estimation, more specifically, "filtering", consists of obtaining the probability distribution $p(\mathbf{z}_k|\mathbf{y}_{1:k})$, where $\mathbf{y}_{1:k} \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_k\}$. By obtaining this distribution, we can calculate characteristic values of our interest, such as mean, mode, median, etc. Thus, in the model eq. (13), we can obtain the estimation of positions of feature points and velocities of objects, as well as indicator variables of object and aperture.

To perform state estimation in this nonlinear state space model eq. (13), we employ a computer intensive method called "particle filter" , or sequential Monte Carlo [6],[12]. For more detail than written here, see [7] for a good survey of this field.

The key idea is to approximate the state distribution by a large number (say M) of weighted random samples called particles

$$\left\{\mathbf{z}_k^{(l)}, \omega_k^{(l)} \;\mid\; l = 1, \cdots, M\right\} \tag{14}$$

where $\mathbf{z}_k^{(l)}$ denotes l-th instance of particle (a realization in state space), and $\omega_k^{(l)}$ denotes the corresponding weight (non-negative, assumed to be normalized). The approximation of the state distribution is formally written as

$$p(d\mathbf{z}_k|\mathbf{y}_{1:k}) \simeq p_M(d\mathbf{z}_k|\mathbf{y}_{1:k}) \equiv \sum_{l=1}^{M} \delta_{\mathbf{z}_k^{(l)}}(d\mathbf{z}_k)\omega_k^{(l)} \tag{15}$$

where, $\delta_{z_k^{(l)}}(dz)$ denotes delta-mass, i.e. its integral on region Z is equal to 1 if $z_k^{(l)} \in Z$, 0 otherwise.

Then, filtering becomes a task to obtain weighted particles at time k given weighted particles at time $k-1$ and observation $\mathbf{y}_k$. It proceeds as follows. First, generate particles at time k according to an importance function $\pi(\mathbf{z}_k|\mathbf{z}_{0:k-1}, \mathbf{y}_{1:k})$ as

$$\tilde{\mathbf{z}}_k^{(l)} \sim \pi(\mathbf{z}_k|\mathbf{z}_{0:k-1}^{(l)}, \mathbf{y}_{1:k}), l = 1, \cdots, M. \tag{16}$$

Second, the weight is calculated by modifying the corresponding weight at $k-1$ by

$$\tilde{\omega}_k^{(l)} \propto \omega_{k-1}^{(l)} \frac{p(\mathbf{y}_k|\tilde{\mathbf{z}}_k^{(l)})p(\tilde{\mathbf{z}}_k^{(l)}|\mathbf{z}_{k-1}^{(l)})}{\pi(\tilde{\mathbf{z}}_k^{(l)}|\mathbf{z}_{0:k-1}^{(l)}, \mathbf{y}_{1:k})}, \tag{17}$$

where $p(\mathbf{y}_k|\mathbf{z}_k)$ is the likelihood function that is defined by the observation equation in eq. (13), and $p(\mathbf{z}_k|\mathbf{z}_{k-1})$ is the state transition density defined by the system equation in eq. (13).

Next, resampling of particles $\mathbf{z}_k^{(l)} (l = 1, \cdots, M)$ from $\{\tilde{\mathbf{z}}_k^{(l)}|l = 1, \cdots, M\}$ is performed if required (typically when the variance of weights is large). Here, the probability that l-th particle is sampled is equal to $\omega_k^{(l)}$. If resampling has been performed, then set $\omega_k^{(l)} = 1/M$. Otherwise, let $\mathbf{z}_k^{(l)} = \tilde{\mathbf{z}}_k^{(l)}$ and

$\omega_k^{(l)} = \tilde{\omega}_k^{(l)}$ for $l = 1, \cdots, M$. Thus, we have obtained the weighted particles at time k.

The idea of Rao-Blackwellization [1] is to use the conditional linear-Gaussian property, i.e., given $\theta_{0:k}$, the model (8) is linear Gaussian. Thus, $\mathbf{x}_k$ is estimated by Kalman filter given $\theta_{0:k}$. Then, a Rao-Blackwellized version of the filtering procedure proceeds as follows; Eq.(16) is applied only for θ_k and we obtain $\tilde{\theta}_k^{(l)}$ for $l = 1, \cdots, M$. Then, for each particle $\tilde{\theta}_{0:k}^{(l)} \equiv \{\tilde{\theta}_k^{(l)}, \theta_{0:k-1}^{(l)}\}$, the Kalman filter is applied to obtain $p(\mathbf{x}_k|\mathbf{y}_{1:k}; \tilde{\theta}_{0:k}^{(l)})$ and $p(\mathbf{y}_k|\mathbf{y}_{1:k-1}; \tilde{\theta}_{0:k}^{(l)})$. In eq. (17), with replacing all $\mathbf{z}$ by θ, use $p(\mathbf{y}_k|\mathbf{y}_{1:k-1}; \tilde{\theta}_{0:k}^{(l)})$ as the likelihood(1st term of numerator), instead.

There are several choices of importance function. If we choose it as $p(\mathbf{z}_k|\mathbf{z}_{k-1})$, i.e., prior(before observe $\mathbf{y}_k$) density of $\mathbf{z}_k$, then weight modification eq. (17) becomes simpler one. It is called bootstrap filter[8], or Monte Carlo Filter[11]. However, the prior importance function often fails except simple model cases, so we need to use clever importance function that effectively use the information of current observation, $\mathbf{y}_k$.

Although the best way is to use the optimal importance function $\pi(\theta_k|\theta_{0:k-1}, \mathbf{y}_{1:k}) = p(\theta_k|\theta_{0:k-1}, \mathbf{y}_{1:k})$ in the sense of minimum variance of the weights [5], however in our problem, the computation of it is not tractable due to the huge number of combinations for the state θ_k. Instead of the optimal one, we propose to use a sub-optimal importance function as follows. Fortunately in this problem, the likelihood can be divided into each feature point, so we can construct a sub-optimal importance function as

$$\pi(\theta_k|\theta_{0:k-1}, \mathbf{y}_{1:k}) \propto \prod_{j=1}^{p} p(\mathbf{y}_{k,j}|\theta_{k,j}, \theta_{0:k-1}, \mathbf{y}_{1:k-1}) p(\theta_{k,j}|\theta_{k-1,j}). \qquad (18)$$

EXAMPLE

A sequence of image (size 512×440 [pixels]), in which two objects(books) are moving, is taken with 30 frames of about 0.1[sec] sampling time. The 1st frame is shown in **Fig.1** with feature points extracted by [2], denoted by rectangles. These are manually selected feature points among all (more than 100) in order to illustrate the performance of the proposed method: 5 points including one aperture for lower book (moving toward right), 5 points for upper book (moving toward left), and 5 points for background. By tracking the feature points in the subsequent images, we have obtained trajectories of feature points shown in **Fig.2** with dashed line.

Estimated trajectories by the proposed model are shown in **Fig.2** with solid line. Smooth trajectories are obtained for all points, including the aperture one at the center of the lower book. Comparison with a Kalman filter being given true indicators has been done and results are as follows; **Fig.3** shows velocity of two moving objects, our method ("RBPF") can estimate results close to the Kalman filter ("KF") despite indicator values being un-

Figure 1: Selected feature points on 1st image.

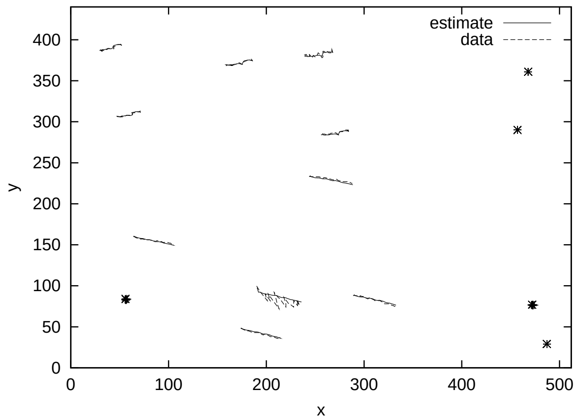

Figure 2: Trajectories of feature points(with estimate).

known. **Fig.4** shows representative results of position estimation of a feature point (not of aperture) on the lower object. Close results to Kalman filter ("KF") are also obtained by our method ("RBPF"). Here, "KF:a" shows the result without taking into account aperture points (i.e. not robust), and it is far from both "RBPF" and "KF" due to the influence of aperture points.

Table 1 shows the estimation result of object indicator vector, where the most frequent value appeared in the particles is shown for each element. First 5 columns are points on the lower object, subsequent 5 columns are on the upper object, and the last 5 columns are on the background. Most of results are correct, but, a few associations are wrong. Although no table is presented here, results for aperture indicators are similar and better than for object indicator.

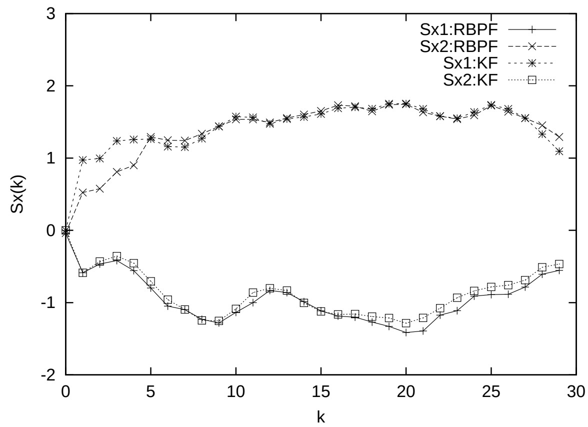

Figure 3: Estimated velocity of a object(x-axis).

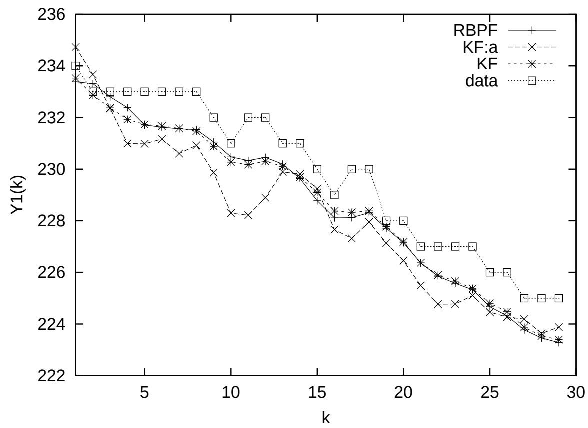

Figure 4: Estimated position of a feature point(y-axis).

CONCLUSION

A new approach to detect moving object labeling and aperture label by taking state vector with these labels as well as feature points positions and objects' velocities. Particle filter is used to estimate the state. A two-dimensional simple experiment demonstrates how the method works.

For future researches, we are planning to deal with more realistic cases such as including rotation, using 3-dimensional motion. In the latter case, reconstruction of 3-dimensional information will be involved. There are few researches about causal estimation to address this case, while there are several conventional methods for non causal case, e.g.[4]. There is a paper dealing with the causal estimation, [3], however this work treats only single object case using an extended Kalman filter. So it is a challenging problem of

Table 1: Estimated object indicator

k	$m_{k,j}$														
1	0	0	2	2	0	0	0	0	2	0	2	2	0	2	0
2	2	2	0	2	2	0	1	1	0	1	2	0	2	0	0
3	2	2	0	2	2	1	1	1	0	1	2	0	2	0	0
4	2	2	2	2	2	1	1	1	0	1	2	0	2	0	0
5	2	2	2	2	2	1	1	1	0	1	2	0	2	0	0
6	2	2	2	2	2	1	1	1	1	1	1	0	1	0	0
7	2	2	2	2	2	1	1	1	1	1	1	0	1	0	0
8	2	2	2	2	2	1	1	1	1	1	1	0	1	0	0
9	2	2	2	2	2	1	1	1	1	1	2	0	2	0	0
10	2	2	2	2	2	1	1	1	1	1	2	0	2	0	0
11	2	2	2	2	2	1	2	1	1	1	2	0	0	0	0
12	2	2	2	2	2	1	2	1	1	1	0	0	0	0	0
13	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
14	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
15	2	2	2	2	2	1	1	1	1	1	1	0	0	0	0
16	2	2	2	2	2	1	1	1	1	1	2	0	0	0	0
17	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
18	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
19	2	2	2	2	2	1	1	2	1	1	0	0	0	0	0
20	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
21	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
22	2	2	2	2	2	1	2	1	2	1	0	0	0	0	0
23	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
24	2	2	2	2	2	1	1	1	2	1	0	0	0	0	0
25	2	2	2	2	2	1	1	1	1	1	0	0	2	0	0
26	2	2	2	2	2	1	1	2	1	1	0	0	0	0	0
27	2	2	2	2	2	1	1	1	1	1	1	0	0	0	0
28	2	2	2	2	2	1	1	1	1	1	1	0	0	0	0
29	2	2	0	2	2	1	1	1	2	1	1	0	1	0	0
30	2	2	0	2	2	1	0	1	2	1	0	0	1	0	0

causal estimation and is also very interesting from both a methodological and a practical points of view.

Acknowledgements

This study was carried out under the ISM Cooperative Research Program (2002-ISM·CPR-1005). Authors also thank to Dr.A.Doucet, Signal Processing lab. in CUED for useful comments.

REFERENCES

[1] G.Casella and C.P.Robert, "Rao-Blackwellisation of Sampling Schemes", *Biometrika*, **83**, 1, pp.81-94, 1996.

[2] F. Chabat, G.Z. Yang and D.M. Hansell, "A corner orientation detector", *Im. and Vis. Comp.*, **17**, pp.761-769, 1999.

[3] A.Chiuso, P.Favaro, H.Jin, and S.Soatto: "Structure from motion causally integrated over time", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **24**, 4, pp.523-535, 2002.

[4] J. Costeria and T. Kanade, "A multi-body factorization method for motion analysis", *Technical report CMU-CS-TR-94-220*, Computer Science Department, Carnegie Mellon University, 1994.

[5] A.Doucet, N.J.Gordon, and V.Krishnamurthy, "Particle Filter for State Estimation of Jump Markov Liner Systems", *IEEE Trans.on Signal Processing*, Vol.**49**, No.3, pp.613-624, 2001.

[6] A.Doucet, J.F.G.de Freitas, and N.J.Gordon (eds), *Sequential Monte Carlo Methods in Practice*, New York, Springer, 2001.

[7] A.Doucet, S.J.Godsill, and C.Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, Vol.**10**, No.3, 197-208, 2000.

[8] N.J.Gordon, D.J.Salmond, and A.F.M.Smith: "Novel approach to nonlinear/non-Gaussian Bayesian State Estimation", *IEE Proceedings-F*, **140**, No.2, pp.107-113, 1993.

[9] N.Ichimura and N.Ikoma, "Filtering and smoothing for motion trajectory of feature point using non-Gaussian state space model", *IEICE Trans. on Information and Systems*, **E84D**,(6): pp.755-759, 2001.

[10] T. Kanade and D. D. Morris, "Factorization methods for structure from motion", *Phil. Trans. R. Soc. Lond. A,* **356**, pp.1153-1173, 1998.

[11] G.Kitagawa: "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of Computational and Graphical Statistics*, **5**, No.1, pp.1-25, 1996.

[12] J.S.Liu and R.Chen: "Sequential Monte Carlo methods for dynamic systems", *Journal of the American Statistical Association*, **93**, pp.1032-1044, 1998.