Image motion tracking by FRS state space model using SMC implementation of PHD filter

Norikazu Ikoma, Tsuyoshi Uchino, and Hiroshi Maeda

Faculty of Engineering, Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kita-Kyushu, Fukuoka 804-8550, JAPAN

ABSTRACT

This study is aimed at obtaining smoothed trajectory of feature points in dynamic image, where the feature points are extracted at each image frame with missing and false detection. The image scene contains independently moving multiple objects with occlusion and appearance. We develop a state space model using Finite Random Set (FRS) to cope with this situation since FRS is a set of random variables with the number of the variables also being random (integer) variable so it is suitable for representing the variable number of feature points caused by appearance/occlusion and missing/false detection. By estimating the state of the model using Sequential Monte Carlo (SMC) implementation of Probability Hypothesis Density (PHD) filter, we obtain smoothed trajectory of the feature points. PHD is 1st order moment of FRS, and it has a property that its integration yields the expected number of feature points in the integrated region. The SMC implementation gives approximated solution by weighted particles, where the number of particles varies depending on the number of feature points in the scene. Experiment of dynamic image demonstrates that proposed method successfully smoothed the trajectory of feature points responding to appearance and occlusion of objects without being affected by missing and false detection.

Keywords: dynamic image, image motion tracking, Probability Hypothesis Density (PHD) filter, Sequential Monte Carlo (SMC).

1. INTRODUCTION

Tracking feature points in dynamic image and its use to reconstruct the information involved in the scene is one of the most important topic in the research field of computer vision. Pioneering works have been done by using singular value decomposition (SVD) to reconstruct 3 dimensional structure and motion, which is called 'factorization'. Those researches are; factorization under orthographic projection by Tomasi and Kanade (1992),[1] it extension to multiple objects case by Costeira and Kanade(1994),[2] extension to paraperspective projection case by Poleman and Kanade (1993),[3] and that to sequential estimation case by Morita and Kanade (1997).[4] In this paper, we focus on the fundamental part of the topic, that is, to extract the feature points in the image and tracking them in successive image frames.

Notice that it is important to have 'good' feature points from dynamic image to obtain 'good' reconstruction of 3 dimensional information. Although SVD is not too sensitive to errors involved in the feature points, it is necessary to have feature point with small error to obtain sufficient precision of the reconstructed information. We are interested in removing the error of the feature points using the idea of filtering based on a state space modeling. There are some researches related to this approach. One example is a pre-process for the factorization by Ichimura and Ikoma (2001),[5] which assumes smoothness prior to the motion of feature points with formulation of linear Gaussian state space modeling and uses Kalman filter for state estimation. Another example is comprehensive method including the filtering of feature points' motion and reconstruction of structure and motion in 3 dimensional space by Chiuso et al. (2002),[6] which uses extended Kalman filter with perspective projection in the observation process of the state space modeling. There is also a series of researches that uses particle filters for the state estimation with a novel state space modeling using index variables to determine the belongings of feature points into objects in the scene done by Ikoma et al (2003).[7] In addition to this, 3

Further author information: (Send correspondence to Norikazu Ikoma)
Norikazu Ikoma: E-mail: ikoma@comp.kyutech.ac.jp

dimensional information reconstruction by introducing the perspective projection to the observation process is done by Miyahara et al (2003)[8] and Ikoma et al (2004).[9]

Basically, feature points are extracted by using local feature of the image, such as large value of 1st order spatial difference of intensity and a kind of uniqueness to identify the feature point from the other part of the image. Corner of object is typical feature of the image and it can be detected using the idea of corner detector by Chabat et al. (1999).[10] In dynamic image case, the feature points are necessary to be tracked in successive image frames. Here, methods based on local feature are also used to track them. Typical one of the methods is using normalized correlation to measure the similarity between two small image regions of feature point in the previous image frame and the current image frame. Searching in the current image around the area of the small region of the previous image, the highest matched position is employed as the feature point in the current image frame. Thus the extraction process of feature point is deterministic, and it mostly uses primitive image processing techniques due to demand of real time computation.

Probabilistic approach to track the feature points is one of the promising methods to achieve robust tracking. It is possible for the probabilistic approaches to cope with missing and false detection of feature points. Some examples of the approaches are as follows; Ikoma and Godsill (2003)[11] and Vermaak, Ikoma and Godsill (2004)[12] deal with association problem between observed points and tracking points under the situation with missing and false detection in the context of target tracking by particle filters. They have proposed a clever proposal, which draws particles effectively using the current observation, to solve the association problem. Furthermore, it is possible to use the probabilistic approach in the context of data fusion such as tracking object in dynamic image with auxiliary use of sound signal by Ito, Ikoma and Maeda (2003),[13] Pérez, Vermaak and Blake (2004),[14] and Ikoma et al. (2004).[15] Also the probabilistic approach can be used for stereo matching problem in dynamic image situation as in Ikoma et al. (2003),[16]

There are two common problems for all the above approaches; the first is the assumption of the fixed number of feature points, and the second is huge dimension of state vector. The state vector is defined as p-tuples of feature point's state, where p is the number of feature points. The first problem concerns with the fixed p formulation while actual scene may have variable p due to occlusion and appearance. Even if we can vary p in some manner, we will be faced with the second problem as follows. Dimension of the state vector is $p \times d$ with d being the dimension of feature point's state, and it is large due to the large number of feature points p (say p being hundreds). Additionally, we need to use M particles for state estimation, where M is usually large such as $1,000$ at least. Thus the total number of real numbers appearing in the computation process becomes $p \times d \times M$, and it is obviously huge.

To overcome the above problems, we propose to use Finite Random Set (FRS) state space model with estimation of the state using Sequential Monte Carlo (SMC) implementation (Vo, Singh and Doucet, 2003)[17] of Probability Hypothesis Density (PHD) filter (Mahler, 2000).[18] FRS is a set of random variables with variable number of elements, where the number of element is also random (integer) variable. It is suit for representing the state and the observation for feature points occluding and appearing in the scene, as well as for representing random occurrence of missing and false detection in observation process. PHD filter is one realization for solving the state estimation problem of FRS state space model by using 1st order moment of the FRS, which is called Probability Hypothesis Density (PHD). It circumvent the problem of huge dimension of state vector mentioned above. SMC implementation of PHD filter is an elaborated way to implement the approximate computation of PHD filter by controlling the number of particles and their weight suit for estimating PHD while ordinary SMC is designed for estimating pdf.

In the following sections, we formulate the FRS state space model for smoothing of feature points' motion and explain the state estimation method using PHD filter and its SMC implementation. Then we have applied the model to a dynamic image with scene of multiple objects' motions to demonstrate the performance of proposed method. It has successfully smoothed the trajectory of feature points responding to appearance and occlusion of feature points on the objects without being affected by missing and false detection. That is, the estimated PHDs of feature points on the objects are smoothly moving along with the actual objects' motion without affected noises due to quantization of digital image, errors caused by image processing, and missing/false detection of feature point extraction. The estimated PHD becomes relatively lower value when the corresponding object occludes behind the obstacle, and it returns to higher value when the object appears again. Finally, we conclude this paper with consideration of future works.

2. MODEL

2.1. Measurement and state

Measurement at discrete time k is represented by a finite set of positions of feature points

$$\mathbf{Z}_k = \left\{ \mathbf{z}_{k,1}, \mathbf{z}_{k,2}, \cdots, \mathbf{z}_{k,m(k)} \right\} \subseteq E_o \tag{1}$$

where $m(k)$ is the number of feature points measured at time k, $\mathbf{z}_{k,j} \in E_o$ represents position of j-th feature point, and E_o is the set of measurement space consisting of all possible measurements. In this paper, we deal with image plane as observation space so $E_o = [x_{min}, x_{max}] \times [y_{min}, y_{max}]$ with x_{min} and x_{max} respectively being minimum and maximum values of image plane in x-coordinate, as well as y_{min} and y_{max} being the same notation for y-coordinate. We suppose a situation that not necessarily all the feature points of interest are in the element of the measurement eq.(1), and some elements could be a spurious measurement. Thus the measurement process may have missing and the measurement may contain false detection of feature point. Note that we have no information to discriminate missing and false detection, i.e., we know neither which elements coming from the actual feature point nor which elements being false detection.

What we need to estimate is state of the feature points. The state of a feature point could be position, position and velocity, or up to certain order of differential of its location. Let $\mathbf{x} \in E_s$ be a general notation for the state with E_s being the set of state space consisting of all possible values for the state. In case the state of a feature point being position, E_s is equivalent to E_o. By considering the situation with occlusion and appearance of feature points in dynamic image, we represent the state vector of feature points at time k by

$$\mathbf{X}_k = \left\{ \mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \cdots, \mathbf{x}_{k,n(k)} \right\} \subseteq E_s \tag{2}$$

where $n(k)$ is the number of feature points actually existing in the image at time k. We can observe only $\mathbf{Z}_k$ not $\mathbf{X}_k$. Thus we need to make inference on $\mathbf{X}_k$ with given series of observations up to current time k, which is denoted by $\mathbf{Z}_{1:k} \equiv [\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_k]$.

2.2. System model

As we employ Bayesian inference explained later, we need to define likelihood and prior for the Bayesian inference. First, as a part of the prior, we give a model for dynamics of feature points using a finite random set (FRS) in conditional distribution form

$$\Xi_k \sim F(\mathbf{X}_k | \mathbf{X}_{k-1}) \tag{3}$$

where Ξ_k is a finite random set on E_s representing the uncertain state. The uncertainty comes from three reasons in general; 1) survival of feature points from previous time step $k-1$ with random disturbance, 2) spawned (or branched) points from feature points at previous time $k-1$, and 3) spontaneously generated points at time k. It can be shown by a decomposed form of the FRS

$$\Xi_k = S(\mathbf{X}_{k-1}) \cup B(\mathbf{X}_{k-1}) \cup \Gamma_k, \tag{4}$$

where above three reasons 1)-3) correspond to three terms of right hand side respectively. In this paper, as we are considering the feature points in dynamic image with occlusion and appearance, unsurvived feature points in $S(\mathbf{X}_{k-1})$ correspond to the occlusion, and generated feature points by Γ_k correspond to the appearance. And we suppose that there is no feature point generated by $B(\mathbf{X}_{k-1})$, thus $B(\mathbf{X}_{k-1})$ is empty set.

Two FRSs in eq.(4), $S(\mathbf{X}_{k-1})$ and Γ_k, are defined as follows. Survival of feature points from previous time step $k-1$ with random disturbance $S(\mathbf{X}_{k-1})$ is written as union of subsets

$$S(\mathbf{X}_{k-1}) = S(\mathbf{x}_{k-1,1}) \cup S(\mathbf{x}_{k-1,2}) \cup \cdots \cup S(\mathbf{x}_{k-1,n(k-1)}) \tag{5}$$

where $S(\mathbf{x}_{k-1,j})$ is survival of j-th feature point from previous time step k, and it is defined as

$$S(\mathbf{x}_{k-1,j}) = \begin{cases} \{\mathbf{X}_{k,j}\} & \text{with probability} \quad p_s \\ \emptyset & \text{with probability} \quad 1-p_s \end{cases} \tag{6}$$

with p_s being survival probability. Here $\mathbf{X}_{k,j}$ is according to the conditional distribution of single feature point such that

$$\mathbf{X}_{k,j} \sim f(\mathbf{x}_{k,j}|\mathbf{x}_{k-1,j}). \tag{7}$$

In eq.(7), random walk model is typically used as the conditional distribution.

As FRS for spontaneously generated points at time k, which is denoted by Γ_k, the number of feature points to generate is according to Poisson distribution and their locations are according to uniform distribution over the state space. It is formally denoted by

$$\Gamma_k = \left\{\mathbf{X}^{\Gamma}_{k,1}, \mathbf{X}^{\Gamma}_{k,2}, \cdots \mathbf{X}^{\Gamma}_{k,N_{\Gamma}(k)}\right\}, \tag{8}$$

where

$$N_{\Gamma}(k) \sim Poisson(\lambda_{\Gamma}), \quad \lambda_{\Gamma} = \nu\,|E_s| \tag{9}$$

with ν being spatial density of feature point generation and $|E_s|$ being the volume of the set of state space E_s, and

$$\mathbf{X}^{\Gamma}_{k,j} \sim U\left[E_s\right] \tag{10}$$

with $U\left[E_s\right]$ being uniform distribution over E_s.

2.3. Observation model

We define a likelihood model by conditional distribution form

$$\mathbf{\Sigma}_k \sim H(\mathbf{Z}_k|\mathbf{X}_k) \tag{11}$$

where $\mathbf{\Sigma}_k$ is a finite random set representing the measurement with uncertainty. It can be rewritten in a decomposed form of FRSs as

$$\mathbf{\Sigma}_k = E(\mathbf{X}_k) \cup C(\mathbf{X}_k) \tag{12}$$

where $E(\mathbf{X}_k)$ is FRS of measurement with missing, and $C(\mathbf{X}_k)$ is FRS of false detection which may depend on the state $\mathbf{X}_k$.

FRS of measurement with missing, $E(\mathbf{X}_k)$, is similarly defined as $S(\mathbf{X}_{k-1})$ such that

$$E(\mathbf{X}_k) = E(\mathbf{x}_{k,1}) \cup E(\mathbf{x}_{k,2}) \cup \cdots \cup E(\mathbf{x}_{k,n(k-1)}) \tag{13}$$

where $E(\mathbf{x}_{k,j})$ is measurement of j-th feature point with missing, which is defined as

$$E(\mathbf{x}_{k,j}) = \begin{cases} \{\mathbf{Z}_{k,j}\} & \text{with probability} \quad p_d \\ \emptyset & \text{with probability} \quad 1-p_d \end{cases} \tag{14}$$

with p_d being detection probability. Here $\mathbf{Z}_{k,j}$ is according to the conditional distribution of single feature point measurement such that

$$\mathbf{Z}_{k,j} \sim h(\mathbf{z}_{k,j}|\mathbf{x}_{k,j}). \tag{15}$$

In eq.(15), Gaussian observation noise is typically used as the conditional distribution.

As FRS of false detection, which is denoted by $C(\mathbf{X}_k)$, it is similarly defined as FRS for spontaneously generated points Γ_k. The number of feature points to generate is according to Poisson distribution and their locations are according to uniform distribution over the observation space. The use of the uniform distribution is the case when the generated locations does not depend on $\mathbf{X}_k$, so we use notation C_k for the FRS instead of $C(\mathbf{X}_k)$. It is formally denoted by

$$C_k = \left\{\mathbf{Z}^{C}_{k,1}, \mathbf{Z}^{C}_{k,2}, \cdots \mathbf{Z}^{C}_{k,M_C(k)}\right\}. \tag{16}$$

where

$$M_C(k) \sim Poisson(\lambda_C), \quad \lambda_C = \mu\,|E_o| \tag{17}$$

with μ being spatial density of feature point generation and $|E_o|$ being the volume of the set of measurement space E_o, and

$$\mathbf{Z}^{\Gamma}_{k,j} \sim U\,[E_s] \tag{18}$$

with $U\,[E_s]$ being uniform distribution over E_s.

3. STATE ESTIMATION

3.1. Formal solution

Formal solution of state estimation to the state space model defined in previous section is represented by two formulae of one-step-ahead prediction and filtering such that

$$p(\mathbf{X}_k|\mathbf{Z}_{1:k-1}) = \int F(\mathbf{X}_k|\mathbf{X}_{k-1})p(\mathbf{X}_{k-1}|\mathbf{Z}_{1:k-1})d\mathbf{X}_{k-1}, \tag{19}$$

$$p(\mathbf{X}_k|\mathbf{Z}_{1:k}) \propto p(\mathbf{X}_k|\mathbf{Z}_{1:k-1})H(\mathbf{Z}_k|\mathbf{X}_k). \tag{20}$$

These formulae do not have analytical solution since they involve very complicated form as defined in previous section. Thus, to obtain the solution to the state estimation, we need to employ some approximation methods.

One way to do this is to use particle filters.[19, 20] Particle filters use many weighted particles to approximate the target distribution, and filtering process proceeds with computation of updating the particles and corresponding weights. It consists of three steps, 1) draw of particles of current time step, 2) weight update, and 3) resampling. The third step, resampling, is not mandatory, it is conducted when the variance of the weights is large by judging it with certain criterion. See related articles for more details, e.g., Doucet, et al (2000).[21]

The formal solution, by eq(19) and (20), is too complicated because of the state vector being finite random set, not fixed element set. So let us consider an easier case for a while that the state vector is just ordinary vector (i.e., a set of fixed number of elements having the order of the elements). Let p be the number of feature points in the state vector $\mathbf{X}_k$, and assume that it is fixed and known. Thus the survival probability p_s is zero, and FRS for spontaneously generated points Γ_k is empty set. It is the same case as Ikoma and Godsill (2003)[11] and Vermaak, Ikoma and Godsill (2004)[12] in the context of target tracking. Then, according to their methods the solution is obtained by introduction index vector which associates the measured points and the points in the state with the use of clever way of drawing the candidate particles for the index vector.

However, as explained in the introduction section, the number of elements in the state vector is $p \times d$ with d being the dimension of single feature point (typically $d = 2$ for position random walk case, $d = 4$ for velocity random walk case), and p is typically the order of hundreds. So the state dimension $p \times d$ is very large, and it is impractical to have efficient estimate. To relax this difficult situation, we employ Probability Hypothesis Density (PHD) filter which has been proposed by Mahler,[18, 22] as explained in next section.

3.2. PHD filter

Probability Hypothesis Density (PHD) is 1st order moment of finite random set (FRS), and it is uniquely defined by a property that its integration yields the expected number of feature points in integrated region (note: 'feature point' is specific for this paper, generally it is 'point'). Let $D(\mathbf{x})$ be a general notation of PHD, where $\mathbf{x} \in E$ and $D(\mathbf{x}) : E \mapsto [0, \infty)$, then $\int_{\mathbf{S}} D(\mathbf{x})d\mathbf{x}$ is equal to the expected number of feature points in $\mathbf{S}$. See related books[23] for more details.

The exact solution to the state estimation as formally shown in eq.(19) and (20) is in a form of finite random set, i.e., distributions appeared at left hand sides of these two formulae are of their FRSs'. Let us define the notations of PHD for these FRSs. Let $D_k(\mathbf{x}|\mathbf{Z}_{1:k-1})$ be the PHD for $p(\mathbf{X}_k|\mathbf{Z}_{1:k-1})$ and $D_k(\mathbf{x}|\mathbf{Z}_{1:k})$ for $p(\mathbf{X}_k|\mathbf{Z}_{1:k})$, where subscript k of D denotes the time step.

Note that these PHDs are defined on the space E_s, not over the space of ${E_s}^p$ with p being the number of feature points (suppose it is fixed only here, again). Thus the PHD is much more simple than the distribution of $\mathbf{X}_k$, so computation of it using particle could be easier than the case of using the distribution itself. Later we will mention about the method to estimate the PHD by particles.

Now we summarize the method for obtaining the PHDs $D_k(\mathbf{x}|\mathbf{Z}_{1:k-1})$ and $D_k(\mathbf{x}|\mathbf{Z}_{1:k})$ based on eq.(19) and (20) according to Mahler.[18,22] First for the one-step-ahead prediction, we use the formula

$$D_k(\mathbf{x}|\mathbf{Z}_{1:k-1}) = p_s \int f(\mathbf{x}|\mathbf{y}) D_{k-1}(\mathbf{y}|\mathbf{Z}_{1:k-1}) d\mathbf{y} + D_{\Gamma_k}(\mathbf{x}) \tag{21}$$

where $D_{\Gamma_k}(\mathbf{x})$ is PHD of FRS Γ_k, which is for spontaneously generated points at time k.

For the filtering, use the formula

$$D_k(\mathbf{x}|\mathbf{Z}_{1:k}) = \{(1 - p_d) + p_d \Psi(\mathbf{x}; \mathbf{Z}_k)\} D_k(\mathbf{x}|\mathbf{Z}_{1:k-1}) \tag{22}$$

where

$$\Psi(\mathbf{x}; \mathbf{Z}_k) = \sum_{\mathbf{z} \in \mathbf{Z}_k} \frac{h(\mathbf{z}|\mathbf{x})}{\lambda_C p_C(\mathbf{z}) + p_d \langle h(\mathbf{z}|\cdot), D_k(\cdot|\mathbf{Z}_{1:k-1}) \rangle} \tag{23}$$

with $\langle f(\cdot), g(\cdot) \rangle$ denoting the convolution of $f(\cdot)$ and $g(\cdot)$ with respect to $\cdot$. $p_C(\mathbf{z})$ is pdf of false detection for single feature point and it is of uniform distribution over E_o as we have assumed in eq.(18). Note that λ_C is the expected number of generated feature points as false detection, and it is equivalent to the parameter of distribution in Poisson case.

3.3. SMC implementation

We explain here Sequential Monte Carlo (SMC) implementation of PHD filter according to Vo,Singh and Doucet (2003).[17] SMC is originally developed for the estimation of pdf (distribution), not PHD, so the weights of particles has property of sum up to 1, which corresponds to the fact that integration of pdf is equal to 1. For PHD, its integration yields the expected number of points, so the main difference between original SMC and the SMC for PHD is how to manage the weights of the particles. That is, in the SMC implementation of PHD filter, the sum of weight is equal to the expected number of points over the domain. Note that we suppose a constraint on the number of particles that each feature point has the same number of particles in principle.

We begin with the explanation about the approximation of PHD by weighted particles. Let us use the notation of the weighted particles as a set of particle and weight pairs

$$\left\{ \left(\mathbf{x}^{(i)}, w^{(i)} \right) \right\}_{i=1}^{L} \tag{24}$$

where particle $\mathbf{x}^{(i)}$ is defined on E, i.e., $\mathbf{x}^{(i)} \in E$, and weight is non-negative such that $w^{(i)} \geq 0$. With the weighted particles in eq.(24), we approximate PHD as

$$D(\mathbf{x}) \simeq \hat{D}(\mathbf{x}) \equiv \sum_{i=1}^{L} w^{(i)} \delta_{\mathbf{x}^{(i)}}(\mathbf{x}) \tag{25}$$

where $\delta_{\mathbf{y}}(\mathbf{x})$ is Dirac delta function with its center being $\mathbf{y}$, which has property that its integration is equal to 1 when the integrating region $\mathbf{S}$ includes $\mathbf{y}$ (otherwise, is equal to zero). This property is represented in general form with any integrable function $f(\mathbf{x})$ by

$$\int_{\mathbf{S}} f(\mathbf{x}) \delta_{\mathbf{y}}(\mathbf{x}) d\mathbf{x} = \begin{cases} f(\mathbf{y}) & \mathbf{y} \in \mathbf{S}, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Next, consider the property of PHD that its integration yields the expected number of points. For the approximated PHD of eq.(25), it leads to

$$\int \hat{D}(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{L} w^{(i)} \int \delta_{\mathbf{x}^{(i)}}(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{L} w^{(i)} \simeq p \tag{27}$$

where p is the number of feature points.

Let ρ be the number of particles per one feature point, and it is assumed to be fixed. Then, the number of particles for all the feature points should be $L = \rho \times p$. This suggests that the total number of particles L should vary depending on the number of feature points p (p is given). On the other hand, consider the case when we do not know how many number of feature points p actually exist, and assume that we have just obtained the pairs of particle and weight through SMC procedure (explained later in detail). In this case, the relation (though it is approximated relation) of most right side in eq.(27) is possible to use for estimating the number of feature points p, and thus, we have the desired number of particles L by relation $L = \rho \times p$.

Now, let us explain the procedures of SMC implementation of PHD filter.[17] Initially, at time $k-1$, a set of pairs of particle and weight is given by

$$\left\{ \left(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)} \right) \right\}_{i=1}^{L_{k-1}} \tag{28}$$

as approximation of PHD $D_{k-1}(\mathbf{x}|\mathbf{Z}_{1:k-1})$.

First, for the one-step-ahead prediction of eq.(21), we use two procedures, one is for particles of prediction represented by the first term of eq.(21), and the other is for particles of spontaneously generated points represented by the second term of eq.(21). Both procedures consist of two steps; draw of particles from proposal and weight calculation. Specifically, for particles of prediction, draw particles from proposal by

$$\tilde{\mathbf{x}}_k^{(i)} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{Z}_k) \tag{29}$$

for $i = 1, 2, \cdots, L_{k-1}$, and update the weight by

$$\tilde{w}_{k|k-1}^{(i)} = p_s \frac{f(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{q(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{Z}_k)} w_{k-1}^{(i)} \tag{30}$$

for $i = 1, 2, \cdots, L_{k-1}$. Then, for particles of spontaneously generated points, draw J_k particles with

$$J_k = \rho \int D_{\Gamma_k}(\mathbf{x}) d\mathbf{x} \tag{31}$$

by

$$\tilde{\mathbf{x}}_k^{(i)} \sim p(\mathbf{x}_k|\mathbf{Z}_k) \tag{32}$$

for $i = L_{k-1}+1, L_{k-1}+2, \cdots, L_{k-1}+J_k$, and set the weight by

$$\tilde{w}_{k|k-1}^{(i)} = \frac{D_{\Gamma_k}(\mathbf{x}_k^{(i)})}{J_k p(\mathbf{x}_k^{(i)}|\mathbf{Z}_k)} \tag{33}$$

for $i = L_{k-1}+1, L_{k-1}+2, \cdots, L_{k-1}+J_k$. Note that eq.(33) is interpreted as uniform weight $1/\rho$ times importance weight (the importance weight is ratio of two values of target density and the proposal density). There is only one different point from the ordinary method that the target distribution is constructed from PHD, $D_{\Gamma_k}(\mathbf{x}_k^{(i)})$, by normalizing it with eq.(31) (ρ in this equation is used for the uniform weight, though).

Then obtained set of pairs of particle and weightm is

$$\left\{ \left(\tilde{\mathbf{x}}_k^{(i)}, \tilde{w}_{k|k-1}^{(i)} \right) \right\}_{i=1}^{L_{k-1}+J_k} \tag{34}$$

which approximates PHD $D_k(\mathbf{x}|\mathbf{Z}_{1:k-1})$.

Next, we will explain the procedure for the filtering of eqs.(22) and (23). Before it, we show an approximate calculation of convolution appearing in eq.(23) by replacing the PHD in the equation with its approximation

$$\int h(\mathbf{z}|\mathbf{x}) \hat{D}_k(\mathbf{x}|\mathbf{Z}_{1:k-1}) d\mathbf{x} = \sum_{i=1}^{L_{k-1}+J_k} \tilde{w}_{k|k-1}^{(i)} h(\mathbf{z}|\tilde{\mathbf{x}}_k^{(i)}) \equiv C_k(\mathbf{z}). \tag{35}$$

Note that computation of $C_k(\mathbf{z})$ requires the set of weighted particles in eq.(34), as shown in the second term in eq.(35). Then, the procedure for the filtering is derived from eq.(22) with (23) resulting as updating of weight

$$\tilde{w}_k^{(i)} = \left\{ (1-p_d) + p_d \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{h(\mathbf{z}|\tilde{\mathbf{x}}_k^{(i)})}{\lambda_C p_C(\mathbf{z}) + p_d C_k(\mathbf{z}).} \right\} \tilde{w}_{k|k-1}^{(i)} \tag{36}$$

for $i = 1, 2, \cdots, L_{k-1}, L_{k-1}+1, L_{k-1}+2, \cdots, L_{k-1}+J_k$.

So obtained set of pairs of particle and weight

$$\left\{ \left(\tilde{\mathbf{x}}_k^{(i)}, \tilde{w}_k^{(i)} \right) \right\}_{i=1}^{L_{k-1}+J_k} \tag{37}$$

approximates the PHD $D_k(\mathbf{x}|\mathbf{Z}_{1:k})$.

Finally, we explain the resampling procedure for SMC implementation of PHD filter. Using the property in eq.(27), we can estimate the number of feature points by summing up the weights. Thus we can determine the reasonable number of particles required for approximating the PHD $D_k(\mathbf{x}|\mathbf{Z}_{1:k})$ by

$$L_k = \rho \sum_{i=1}^{L_{k-1}+J_k} \tilde{w}_k^{(i)}. \tag{38}$$

Let $M_k = L_{k-1} + J_k$ and $p_k = \sum_{j=1}^{M_k} \tilde{w}_k^{(j)}$, then, resampling procedure of particle is

$$\mathbf{x}_k^{(i)} = \begin{cases} \tilde{\mathbf{x}}_k^{(1)} & \text{with probability} \quad \tilde{w}_k^{(1)}/p_k, \\ \tilde{\mathbf{x}}_k^{(2)} & \text{with probability} \quad \tilde{w}_k^{(2)}/p_k, \\ \vdots & \qquad \vdots \qquad\qquad \vdots \\ \tilde{\mathbf{x}}_k^{(M_k)} & \text{with probability} \quad \tilde{w}_k^{(M_k)}/p_k, \end{cases} \tag{39}$$

for $i = 1, 2, \cdots, L_k$. Then we set the weights uniform as

$$w_k^{(i)} = p_k / L_k \tag{40}$$

for $i = 1, 2, \cdots, L_k$. Thus we obtain the set of pairs of particle and weight

$$\left\{ \left(\mathbf{x}_k^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^{L_k} \tag{41}$$

to approximate the PHD $D_k(\mathbf{x}|\mathbf{Z}_{1:k})$.

3.4. Tracking of feature points

We propose a method to track the feature points based on the estimated PHD. Before tracking, we calculate the PHD function from the obtained set of pairs of particle and weight by using kernel density estimation method, which is modified to the PHD case. Here, let the PHD be $D(\mathbf{x}) : \Re^d \mapsto [0, \infty)$, and the set of particle and weight pairs used for the estimation is $\left\{ \left(\mathbf{x}^{(i)}, w^{(i)} \right) \right\}_{i=1}^{L}$. Then, the kernel density estimation for the PHD is

$$\hat{D}(\mathbf{x}; \mathbf{h}) = \frac{1}{L} \sum_{i=1}^{L} \frac{w^{(i)}}{h_1 \cdot h_2 \cdots h_d} K_d \left(\frac{x_1^{(i)} - x_1}{h_1}, \frac{x_2^{(i)} - x_2}{h_2}, \cdots, \frac{x_d^{(i)} - x_d}{h_d} \right) \tag{42}$$

with $\mathbf{h} = [h_1, h_2, \cdots, h_d]$ being bandwidth vector, and K_d being the kernel function of d dimensional case. For K_d, typically the one independent for each element, i.e., $K_d(\mathbf{x}) = K(x_1)K(x_2)\cdots K(x_d)$, is used.

The method for tracking the feature points is as follows. For each time step k (corresponding to image frame), we apply the method below. First, search for a region of high value in the obtained kernel estimation of PHD,

$\hat{D}$. It start from an initial threshold high enough and letting the threshold going down with certain step size, while deciding the region when the value of PHD estimate $\hat{D}$ is greater than the threshold. This is the procedure for detecting a candidate feature point.

Next, prepare a circle region with center being the candidate feature point above and radius being r. If the number of particles in the region is large enough, such that it is greater than $c\rho$ with $c < 1$ and $c \simeq 1$, then we decide that there is a feature point around that region and let the center be a landmark point of the feature point. This procedure repeats until the number of detected feature points is equal to the expected one (p_k at time step k).

Now, we form a set of particle where the particles are in the circle region (this is done for each landmark point). Let I be the index set of the particles in the region. Next, calculate the center of gravity for particles in the set by

$$\mathbf{c}_k = \frac{1}{\#(I)} \sum_{i \in I} \mathbf{x}_k^{(i)} \tag{43}$$

where $\#(I)$ is the number of elements in I. Calculate eq.(43) both for current time step k and previous step $k-1$, then we have $\mathbf{c}_k$ and $\mathbf{c}_{k-1}$. Note that to calculate $\mathbf{c}_{k-1}$, we need to keep particles of previous time step. Thus the set of particle and weight pairs is $\left\{ \left(\mathbf{x}_{k-1:k}^{(i)}, w_k^{(i)} \right) \right\}_{i=1}^{L_k}$, though particles spontaneously generated at time k do not have their previous particle (which is denoted by $\mathbf{x}_{k-1}^{(i)}$ in above set).

Finally, by connecting $\mathbf{c}_{k-1}$ and $\mathbf{c}_k$, we obtain tracking path of feature point between time steps $k-1$ and k. By repeating this until the number of paths suffices (as mentioned above) and continuing it over all frames, we complete the tracking of feature points with the result of obtained connections.

4. EXPERIMENT

We have conducted an experiment using dynamic image of scene with human's leg motion and car motion. Some frames of the dynamic image are shown in the left panel of Fig.1. In this figure, feature points extracted for each frame are also displayed on the images by rectangles. Feature points for all frame are shown in the right panel of the same figure. For the extraction of feature points, we have used a corner detector by Chabat, et al (1999).[10]

We have applied the proposed method to track the feature points for the dynamic image with conditions as follows. Due to the image size of 512×440, the set of observation space is $E_o = [1, 512] \times [1, 440]$. As we have employed 1st order (velocity random walk) model, in which the state consists of position and velocity, the set of state space E_s is a product of two sets E_o (for position) and set for velocity. As the set for velocity, we have used $[-5, 5] \times [-6, 6]$.

The conditions for the FRS state space model are as follows. First, for the system model, survival probability is set to $p_s = 0.9$, the conditional distribution of single feature point $f(\mathbf{x}_k|\mathbf{x}_{k-1})$ is set to normal distribution with mean $\mathbf{x}_{k-1}$ and diagonal covariance matrix with its diagonal elements being identical to $\tau^2 = 2.0$, and the spatial density is set to $\nu = 0.1$. Second, for the observation model, detection probability is set to $p_d = 0.9$, the conditional distribution of single feature point $h(\mathbf{z}_k|\mathbf{x}_k)$ is set to normal distribution with mean being the position part of $\mathbf{x}_k$ and diagonal covariance matrix with its diagonal elements being identical to $\sigma^2 = 1.0$, and the spatial density is set to $\mu = 0.1$.

Initial distribution of the state is set as follows; Position part of it is according to Gaussian mixture with components' mean being the observed feature points' positions at 1st frame and their covariance matrices being identical to diagonal matrix with its diagonal elements identical to $\sigma_0^2 = 2.0$. Velocity part of the initial distribution is according to uniform distribution over the velocity part of E_s.

Conditions for the SMC implementation of PHD filter are as follows. Particles per feature point is set to $\rho = 300$, the proposal $q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{Z}_k)$ in eq.(29) is equivalent to the conditional distribution of single feature point $f(\mathbf{x}_k|\mathbf{x}_{k-1})$, and the proposal $p(\mathbf{x}_k|\mathbf{Z}_k)$ in eq.(32) is the same to the initial distribution of the state (the Gaussian mixture) except the time step of the observed feature points' positions user for components' mean being k, while it is $k = 1$ for the initial distribution. Number of feature points at first time step $k = 1$ is set to $p_1 = 9$, which controls the total number of particles at the initial step.

Estimated PHDs are shown in Fig.2 for the same frames as the figure of dynamic image (Fig.1), where marginal PHD with respect to the position of feature point is calculated by eq.(42), and its contour is shown in the figure. Tracking of feature points has been done using the estimated PHD and the set of weighted particles as explained in 3.4. Result of the tracking is shown in Fig.3.

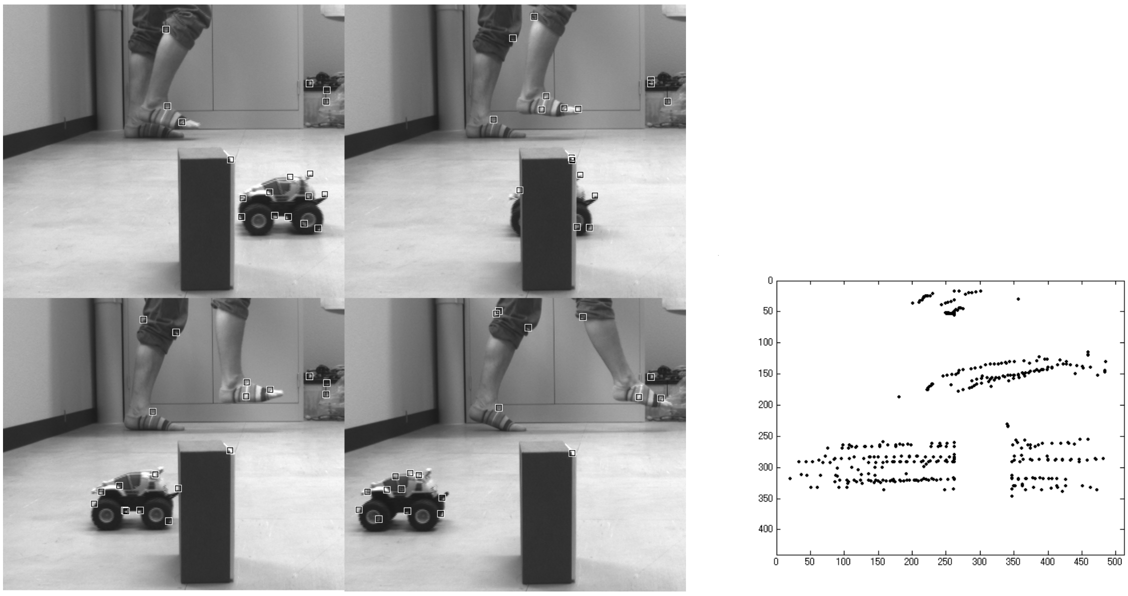

Figure 1. Dynamic image of leg and car motions with extracted feature points (left panel), and plot of the extracted feature points (right panel). In the left panel, 1st frame (upper left), 10th frame (upper right), 20th frame (lower left), and 30th frame (lower right) are shown with rectangles being the feature points. The right panel shows the plots of feature points for all frames.

5. CONCLUSION

We have proposed a method using FRS state space model with estimation by SMC implementation of PHD filter for the purpose to track feature points in dynamic image robustly under existence of missing/false detection of feature points and occlusion/appearance of them in the scene. One of the novelty in this paper is a method to obtain the tracking of feature points from the estimation result of SMC implementation of PHD filter, which is a set of weighted particles. The method uses kernel density estimation of PHD calculated from the resulting set of weighted particles of SMC implementation. In the method, we first calculate the center of gravity of particles in high PHD region, next we connect the centers of gravity between current time step and previous time step, then we obtain the tracking result. Performance of the method has been demonstrated by an experiment with dynamic image of humans' leg and car motions scene.

We here mention about future works. At first, reconstruction of 3 dimensional information is important. The result of proposed method is tracking of feature point between two consecutive frames and it has occlusion and appearance of point. So direct application of factorization method is not possible. Instead, an approach using state space model to reconstruct the 3D information is possible for this case by introducing the projection model at the observation process in the model. Second for the future works, classification of feature point into individual objects based on the estimated PHD is interesting. Although we have not shown the result of PHD for velocity, it forms some peaks corresponding to the objects moving with different velocities. So using the PHD of velocity, it is possible to classify the feature points into objects in principle. Combining these two possible future works, 3D information reconstruction in the scene with multiple objects is considered to be achieved. Additionally, PHD approach only with its position may suffer a problem that it is difficult to discriminate two feature points placed at very close positions. Solution to this by introducing the velocity of PHD is also interesting as the future work.

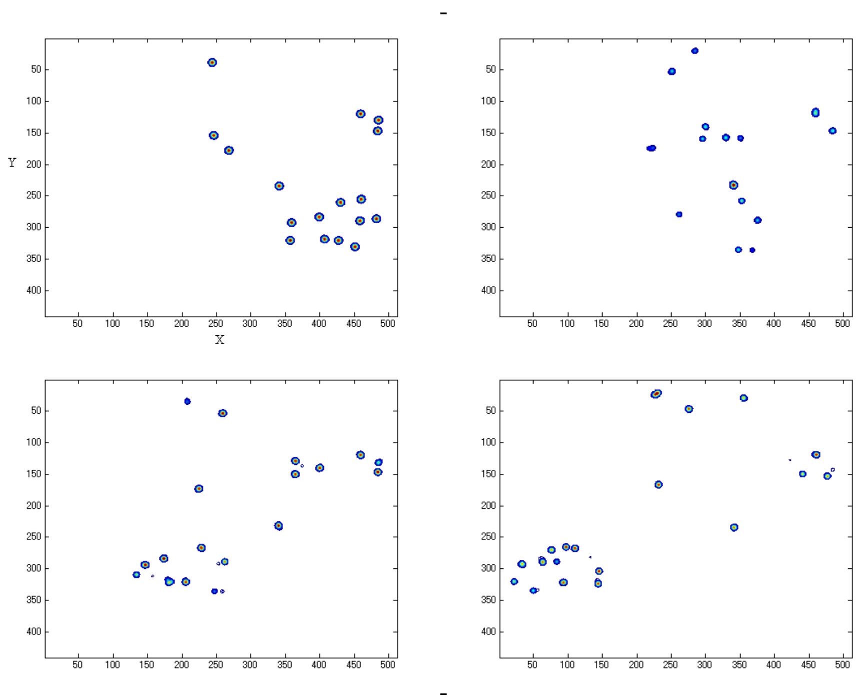

Figure 2. Estimated PHD for dynamic image of leg and car motions. Contour plot of PHD, which is obtained by kernel density estimation method modified for the PHD case, is shown for each frame. Marginal PHD with respect to position of feature point is shown.

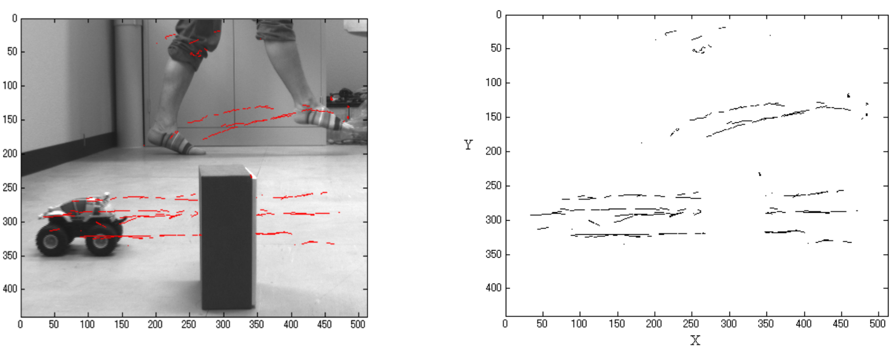

Figure 3. Tracking result for dynamic image of leg and car. Left panel shows the tracking result displayed on the final frame of the images, and right panel shows the tracking result only.

ACKNOWLEDGMENTS

This study was carried out under the ISM Cooperative Research Program (2004-ISM·CRP-2020).

REFERENCES

1. C.Tomasi and T.Kanade, "Shape and motion from image streams: A factorization method," *International Journal of Computer Vision* **9**(2), pp. 137–154, 1992.
2. J.Costeira and T.Kanade, "A multi-body factorization method for motion analysis," Tech. Rep. CMU-CS-TR-94-220, 1994.

3. C.J.Poleman and T.Kanade, “A paraperspective factorization method for shape and motion recovery,” Tech. Rep. CMU-CS-TR-93-219, 1993.
4. T.Morita and T.Kanade, “A sequential factorization method for recovering shape and motion from image streams,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**(8), pp. 858–867, 1997.
5. 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.
6. 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.
7. N.Ikoma, Y.Miyahara, and H.Maeda, “Tracking of feature points in a scene of moving rigid objects by Bayesian switching structure model with particle filter,” *Proc. of 2003 IEEE International Workshop on Neural Networks for Signal Processing (NNSP2003)*, pp. 719–728, 2003.
8. Y.Miyahara, T.Uchino, N.Ikoma, and H.Maeda, “Structure from motion for multiple objects by Rao-Blackwellized particle filters,” *Proc. of - the 30th Anniversary of the Information Criterion(AIC) - Science of Modeling*, pp. 376–377, 2003.
9. N.Ikoma, Y.Miyahara, T.Uchino, and H.Maeda, “Tracking of feature points in dynamic image with classification into objects and 3d reconstruction by particle filters,” *Proc. of International Workshop on Fuzzy Systems and Innovation Computing 2004 (FIC 2004)*, pp. 320–325, 2004.
10. F.Chabat, G.Z.Yang, and D.M.Hansell, “A corner orientation detector,” *Image and Vision Compting* **17**, pp. 761–769, 2003.
11. N.Ikoma and S.Godsill, “Extended object tracking with unknown association, missing observations, and clutter using particle filters,” *Proc. of the 2003 IEEE Workshop on Statistical Signal Processing*, pp. 485–488, 2003.
12. J.Vermaak, N.Ikoma, and S.J.Godsill, “Extended object tracking using particle techniques,” *Proc. of IEEE Aerospace Conference 2004*, p. 1015, 2004.
13. W.Ito, N.Ikoma, and H.Maeda, “On association problem for sensor fusion in dynamic situation using particle filters,” *Proc. of -The 30th Anniversary of the Information Criterion (AIC)- Science of Modeling*, pp. 414–415, 2003.
14. P.Pérez, J.Vermaak, and A.Blake, “Data fusion for visual tracking with particles,” *Proceedings of the IEEE* **92**(3), pp. 495–513, 2004.
15. N.Ikoma, W.Ito, M.Kawanishi, and H.Maeda, “Sensor fusion under unknown associations by particle filters with clever proposal,” *Proc. of IEEE TENCON 2004*, pp. 673–676, 2004.
16. N.Ikoma, W.Ito, T.Irie, and H.Maeda, “3D reconstruction from stereo camera dynamic image based on particle filter,” *Proc. of International Conference on Intelligent Technologies 2003 (InTech 2003)*, pp. 394–403, 2003.
17. B. Vo, S. Singh, and A. Doucet, “Sequential Monte Carlo implementation of the PHD filter for multitarget tracking,” Tech. Rep. CUED/F-INFENG/TR 451, 2003.
18. R.Mahler, “A theoretical foundation for the Stein-Winter Probability Hypothesis Density (PHD) multitarget tracking approach,” *Proc. 2002 Mss Nat. Symp. On Sensor and Data Fusion* **1(Unclassified)**, 2000.
19. A.Doucet, J.F.G.de Freitas, and N.J.Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.
20. J.S.Liu and R.Chen, “Sequential Monte Carlo methods for dynamic systems,” *Journal of the American Statistical Association* **93**, pp. 1032–1044, 1998.
21. A.Doucet, S.J.Godsill, and C.Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering,” *Statistics and Computing* **10**(3), pp. 197–208, 2000.
22. R.Mahler, “Multitarget bayes filtering via first-order multitarget moments,” *IEEE Trans. on Aerospace and Electronic Systems* **39**(4), pp. 1152–1178, 2003.
23. I.R.Goodman, R. P.S.Mahler, and H. T. Nguyen, *Mathematics of data fusion*, Kluwer Academic Publishers, 1997.