The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Multimodal MSEPF for visual tracking

Masahiro Yokomichi and Yuki Nakagama

University of Miyazaki, Miyazaki 889-2155, Japan
(Tel: 81-985-58-7420, Fax: 81-985-58-7420)

yokomich@cs.miyazaki-u.ac.jp

Abstract: Recently, particle filter has been applied to many visual tracking problems and it has been modified in order to reduce the computation time or memory usage. The one of them is the Mean-Shift embedded particle filter (MSEPF, for short) and Randomized MSEPF. These methods can decrease the number of the particles without the loss of tracking accuracy. However, the accuracy may depend on the definition of the likelihood function (observation model) and of the prediction model. In this paper, the authors propose an extension of these models in order to increase the tracking accuracy. Furthermore, the expansion resetting method, which was proposed for mobile robot localization, and the changing the size of the window in Mean-Shift search are also selectively applied in order to treat the occlusion or rapid change of the movement.

Keywords: Expansion Resetting, Mean-Shift, Multiple Model, Particle Filter, and Visual Tracking.

1 INTRODUCTION

Visual tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of application areas, such as robot vision, human-computer interaction, security and surveillance, video communication, and so on. Visual tracking requires high accuracy tracking and real-time processing. To achieve high accuracy tracking, many approaches have been studied. Particle filter [1-3] is one of the robust tracking approaches in visual tracking, which has recently been developed. It performs a random search guided by a stochastic motion model to obtain an estimate of the posterior distribution describing the object's configuration. However, it is known that the degeneracy is one of the difficult problems inherent in particle filter. The degeneracy problem is a phenomenon of the tracking accuracy's decreasing because most particles may have very low likelihood. One of approaches that deal with it is to use very large number of particles, but it is hard to implement it to real-time systems because it requires a lot of computation times and resources. Shan and coworkers proposed the Mean-Shift embedded particle filter (MSEPF) in order to keep the accuracy with small number of particles [4]. In their approach, the state of each particle moves to the point in the window with the highest likelihood value.

In general, MSEPF overcomes the degeneration problem because each particle has higher likelihood. In addition, the accuracy of estimation depends on the size of the window, but the larger window size makes the computation slower. In the previous work [5], the authors modified the mean value calculation part of MSEPF by Monte-Carlo approximation and the likelihood function by adding the term about frame difference. It was shown that the computation time can be reduced without the loss of tracking accuracy.

In this paper, the authors extend the method in [5] such that the tracking accuracy can be improved. The first is application of the multiple prediction models for the case where the precise model for the movement of the tracked object cannot be obtained. The second is adaptation to the case where the tracked object is occluded or its velocity changes suddenly. In order to track the object robustly, two methods are switched according to the mean value of the likelihood function. The effectiveness of the proposed approach is examined by real video examples.

2 MEAN-SHIFT EMBEDDED PARTICLE FILTER

2.1 Particle Filter

Particle filter is an approach for Bayes Estimation by Random Sampling. A continuous state vector of a target object and the observed feature vector at time step t are denoted by x_t and z_t, respectively. The dynamic model is assumed to be represented as a temporal Markov chain

$$p(x_t|x_1, \dots, x_t) = p(x_t|x_{t-1}), \quad (1)$$

and the observation model is denoted as

$$p(z_1, \dots, z_t|x_1, \dots, x_t) = \prod_{i=1}^{t} p(z_i|x_i). \quad (2)$$

Particle filter aims to estimate the sequence of hidden parameters x_t based only on the observed data $\{z_1, \dots, z_t\}$. According to the Bayes rule, the prior and the posterior are given by

$$p(x_t|z_1,\dots,z_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|z_1,\dots,z_{t-1})dx_{t-1} \quad (3)$$

$$p(x_t|z_1,\dots,z_t) = k_t p(x_t|z_1,\dots,z_{t-1}), \quad (4)$$

where k_t is the normalization term.

In the particle filter, by using a set of samples and the corresponding weights $S_t \coloneqq \{(s_t^n, \pi_t^n)\}_{n=1}^N$ at time step t (where N is the number of particles), the posterior is approximated as

$$p(x_t|z_1,\dots,z_t) \approx \sum_{n=1}^N \pi_t^n \delta(x_t - s_t^n), \quad (5)$$

where $\delta(\cdot)$ is the Dirac's delta function. Then, the prior can be approximated as

$$p(x_t|z_1,\dots,z_{t-1}) \approx \sum_{n=1}^N \pi_{t-1}^n p(x_t|s_{t-1}^n). \quad (6)$$

The weights π_t^n are determined such that $\pi_t^n \propto p(z_t|s_t^n) =: w_t^n$ (this probability is called as *likelihood*) and $\sum_{n=1}^N \pi_t^n = 1$. If sufficiently large number of particles can be prepared, Eq. (5) and (6) are accurate. In reality, however, using an infinite number of particles is not allowed, especially for real-time processing.

2.2 MSEPF

MSEPF is proposed in [4] which incorporates Mean-Shift into particle filter. Mean-Shift is another approach for visual tracking that climbs the gradient of a probability distribution to find the nearest dominant mode (peak). In the search window, the mean position of the target object is computed and the search window is centered at that position. Position of the target object is tracked by iterating this mean position calculation until the shift length converges.

In MSEPF, Mean-Shift analysis is applied to each particle based on observation density, after each particle was measured by likelihood function. MSEPF can keep the accuracy using fewer particles because particles converge to the local maximum. Therefore, MSEPF can reduce particles than particle filter. It is known that the accuracy of estimation depends on the size of the window. However, the larger window size requires additional computation time.

It was pointed out in [6] that the computational cost for mean value calculation is $O(n^2)$, where n denotes the window size. This shortcoming can be improved by replacing the mean value calculation by its Monte-Carlo approximation [5]. It is called as Randomized MSEPF (RMSEPF, for short). By means of this approximation, it was shown that the cost can be reduced to $O(n)$. Furthermore, in [5], the likelihood function for MSEPF was also modified using the edge detection and the frame difference.

3. EXTENSION OF RMSEPF

3.1 Multiple prediction models

In the particle filter-based state estimation, the state of each particle is updated by the prediction model Eq. (1). This is based on the property of the kinematics of the object to be tracked. However, many objects in the real world may not obey a simple kinematics, so precise modeling by single model is difficult. If insufficient prediction model is used, the tracking accuracy may decrease. Thus, in this paper, multiple prediction models are used in order to adapt to the various movement of the object.

3.1.1 Three types of prediction models

In this paper, three types of prediction models are adopted. They are *Static model*, *Drift model*, and *Statistical model*. The state of each model is (x, y) coordinates of the object on the image plane. These models are defined as follows:

1. Static model

$$\begin{aligned} x_t &= x_{t-1} + l_u \cos\theta_u, \\ y_t &= y_{t-1} + l_u \sin\theta_u. \end{aligned} \quad (7)$$

2. Drift model

$$\begin{aligned} x_t &= x_{t-1} + (x_{t-1} - x_{t-2}) + l_u \cos\theta_u, \\ y_t &= y_{t-1} + (y_{t-1} - y_{t-2}) + l_u \sin\theta_u. \end{aligned} \quad (8)$$

3. Statistical model

$$\begin{aligned} x_t &= x_{t-1} + \hat{x}_{t-1}, \\ y_t &= y_{t-1} + \rho_{t-1}\hat{x}_{t-1} + \sqrt{1-\rho_{t-1}^2}\hat{y}_{t-1}, \end{aligned} \quad (9)$$

where $\hat{x}_t \sim N(\mu_t^x, (\sigma_t^x)^2)$ and $\hat{y}_t \sim N(\mu_{t-1}^y, (\sigma_{t-1}^y)^2)$. In Eq. (7), $l_u \sim \mathcal{U}(0, l_{max})$ and $\theta_u \sim \mathcal{U}(0, 2\pi)$ are random variables with some positive l_{max}. In Eq. (9), μ_t^x and $(\sigma_t^x)^2$ denote the median and the variation of the estimated velocity of the object in the x-direction up to time t. In addition, ρ_{t-1} denotes the correlation efficient of estimated velocity in the image plane.

The static model is based on the random walk movement. This model corresponds to the case where the object suddenly changes the direction. The drift model is used to describe the movement of the object with linear velocity. Finally, the statistical model predicts the nonlinear movement of the object by learning the change of the velocity and the direction from the sequence of the estimated state.

3.1.2 Model selection

One approach to use the multiple prediction model for the particle filter is applying the interacting multiple models (IMM, for short) [7, 8]. In this approach, each particle possesses additional information about which model is applied to it. However, the transition probability between models should be determined in advance. In this paper, a

simple adaptive model selection algorithm is proposed. In this algorithm, the probability that each model is selected is proportional to their median of likelihood function. Suppose that the number of the models and the particles are $M > 0$ and $N > 0$ respectively. Furthermore, the models are denoted as$F^1, \dots, F^M$. For any particle, the probability that it adopts the prediction model F^i at time t is denoted asn_t^i. The algorithm is summarized as follows:

0. Initialize the time as $t = 0$ and the model selection probabilities $\mathcal{N}_t := \{n_t^1, \dots, n_t^M\}$ such that $\forall i \in \{1, \dots, M\}, n_t^i > 0, \sum_{i=1}^M n_t^i = 1$.
1. For each particle, select a model according to $\mathcal{N}_t$ and update its state. For each model, define the group of particles X_t^i whose elements use F^i.
2. Compute the likelihood for all particles and obtain the median of likelihood $\widehat{W}_t^i =: med_{s_t^n \in X_t^i}(w_t^n)$.
3. For each group X_t^i. For some small positive constant ε, define $W_t^i := \widehat{W}_t^i + \varepsilon$.
4. Resample the particles and calculate each n_{t+1}^i such as $n_{t+1}^i = W_t^i / \sum_{j=1}^M W_t^j$.
5. Set $t \leftarrow t + 1$ and back to Step 1.

In the above algorithm, small constant ε should be added in order to avoid the case where certain models are not selected.

3.2 Escape from the Kidnapped state

Particle filters can approximate the probability of the existence of the object by means of many particles. However, if the particles lost the object by some reasons, the accuracy becomes very low. In this paper, the authors call such a situation as *kidnapped state*. The kidnapped state may occur when the object moves with very large velocity or the object is occluded by other objects (Fig. 1).

Fig. 1. Example of kidnapped state.

In this paper, two methods are selectively used in order to escape the particles from kidnapped state. The former is Expansion Resetting method [9] (ER method, for short) which was proposed for mobile robot localization problem. The latter is making the window size in the Mean-Shift search variable.

3.2.1 ER method

In the ER method, the particles are re-configured when they lost the object to be tracked. If the target was lost again, the particles are diffused to larger region (see Fig. 2). The decision whether particles lost the target or not is based on the mean $\bar{W}_t$ of likelihood. Suppose that the estimate of the position of the object by particles S_t is given by $\tilde{s}_t$. In addition, a PDF $p_0(x; \mu_x, \sigma_i^2)$ on the image plane, whose increasing variances are given by $\sigma_i^2, \forall j > i, \sigma_j^2 > \sigma_i^2$, is

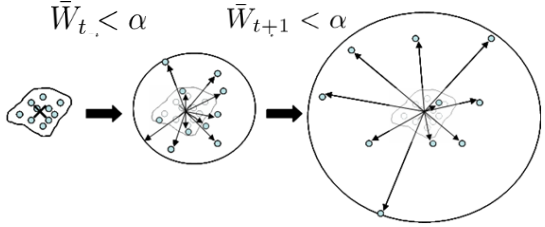

Fig. 2. Expansion resetting in 2D space.

also given. If $\bar{W}_t$ is smaller than a pre-defined threshold $\alpha > 0$ at certain time step t, then the particles are resampled based on $p_0(x; \tilde{s}_t, \sigma_0^2)$. Next, if $\bar{W}_{t+1}$ is also smaller than α, the particles are resampled again based on $p_0(x; \tilde{s}_{t+1}, \beta\sigma_1^2)$.

3.2.2 Changing the window size for Mean-Shift

In the visual tracking by only Mean-Shift search, large size of searching window is preferred because it is possible to robustly track the object even if it moves rapidly. However, in MSEPF, if large window in Mean-Shift search step is used, the particles may gather in local maxima. This weakens the variety of the particles, which is one of the important features of particle filter. Thus, small size of search window should be used, and we can see that there exists a trade-off.

To overcome this difficulty, the variable size of search window is adopted. When tracking is succeeded, it is made small. On the other and, if the object is lost, it becomes large.

3.2.3 Switching two methods

The ER method described in 3.2.1 can re-capture the object even if the object is occluded temporary. However, while this method is being used, tracking accuracy decreases. In contrast, changing the window size for Mean-Shift cannot track the object when the object is occluded, but it can track accurately, if the kidnapped state is rather weak. Thus, in this paper, two methods are switched according to the severity of the kidnapped state.

At first, two thresholds $0 < \alpha_0 < \gamma$ and the minimal (maximal) window sizes $W_{min}(W_{max})$ should be chosen. If $\bar{W}_t < \alpha_0$, then ER method is applied. Else if $\alpha_0 \leq \bar{W}_t < \gamma$, change the window size for Mean-Shift as $\frac{W_{min} - W_{max}}{\alpha_0 - \gamma}(\bar{W}_t - \alpha_0) + W_{max}$.

4. EXPERIMENTS

This section illustrates the performance of proposed algorithm by real video sequences in lab environment. Tracked object is a pink toy shown in the left of Fig. 4 and the environment is shown in the left of Fig. 3.

Fig. 3. Tracked object (left) and lab environme nt with an occluding object

In Fig. 3 (left), there exists an object (yellow bottle) by which the tracked object may be occluded. Furthermore, there exists another object (tree) in the left side of image, which has decollations with several colors. In each video sequence, the object moves from the left to the right in the image plane. The image size is 320x240 (pixel) and the flame rate is 15(fps).

In the ER method, the particles are resampled on circle region whose center is $\tilde{s}_t$ and the radius is $5.18m + 0.5$, where m denotes how many times the heavily kidnapped state has been continue. The window size foe Mean-Shift varies in the range $[10 \times 10 : 30 \times 30]$. The threshold values are set as $(\alpha_0,\ \gamma) = (2.8, 20.6)$. On each experiment, the number of the particles is set to 50, and that of the samples in Mean-Shift search is also set to 50.

Fig. 4. Escape from the occluded state.

Fig. 5. Escape from the light kidnapped state (top: fixed window size; bottom: variable window size).

At first, Fig. 4 shows the results for the case where the object is occluded and this case corresponds to the heavily kidnapped state. We can see that,after diffusing, the particles can catch the object, and then track it.

Next, Fig. 5 shows results for the case where the object moves rapidly. When the constant (small) size of search window is used (top), the particles lose the object and they are caught by decollated tree which has similar color feature. On the other hand, the proposed method can track the object successfully.

5. CONCLUSIONS

In this paper, some extensions for Mean-Shift embedded particle filter are proposed. In order to improve the tracking performance when the tracked object changes the velocity rapidly or is occluded, two methods are selectively applied. In addition, multiple prediction models are used to take the complex and unpredictable movement of the object into account. Their effectiveness was examined by experiment with real video sequence and it was shown that the filter does not lose the object even when the object was occluded.

Evaluating the validity of the prediction models and finding more suitable models for each application are some of the future issues.

REFERENCES

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

[2] N. J. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation, "IEEE Proc. F, Vol.140, No.2, pp.107-113, 1993.

[3] M. Isard and A. Blake, "CONDENSATION: condi tional density propagation for visual tracking," Int. J. Computer Vision, Vol. 29, No. 1, pp.5-28, 199 8.

[4] C. Shan, "Real time hand tracking by combining particle filter and mean shift," IEEE, Vol.88, No.12, pp.989-994, 2005.

[5] Y. Nakagama and M. Yokomichi, An improvemen t of MSEPF for visual tracking," Int. J. Artificial Life and Robotics, Vol. 15, No. 4, pp.534-538, 2 011.

[6] C. Yang, et al., "Efficient Mean-Shift tracking via a new similarity measure," Proc. of IEEE Comp uter Society Conference on Computer Vision and Pattern Recognition (CVPR'05) Vol. 1, pp.176-183, 2005.

[7] Y. Boers, J. N. Driessen, "Interacting multiple model particle filter," IEEE Proc. Radar, Sonar and Navigation, Vol. 150, No. 5, pp.344-349, 2003.

[8] J. Wang, D. Zhao, and W. Gao, and S. Shan, "Interacting multiple model particle filter to adaptive visual tracking," Proc. Int. Conference on Image and Graphics, pp.568-571, 2004.

[9] R. Ueda, T. Arai, K. Sakamoto, T. Kikuchi, a nd S. Kamiya, "Expansion resetting for recovery from fatal error in Monte Carlo localization - co mparison with sensor resetting methods," Proc. of IROS 2004, Vol. 3, pp.2481-2486. 2004.