International Journal of Innovative
Computing, Information and Control
Volume **12**, Number **3**, June **2016**
 ISSN 1349-4198
pp. 1021–1026

HUMAN TRACKING USING PARTICLE FILTER BASED ON SWITCHING ADAPTIVE/NONADAPTIVE OBSERVATION MODEL

Keiichi Horio and Toutaro Nagasakiya

Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan
horio@brain.kyutech.ac.jp

Received October 2015; revised March 2016

Abstract. *In this paper, we describe a human tracking method based on a particle filter with the switching observation model. During tracking humans in video scenes, occlusions and shape changes of target human often occur. Adaptive observation model in which the model is updated in every frame is effective for shape changes; however, it causes wrong tracking in occlusion scene. To realize robust tracking, switching observation model is introduced. In the method, adaptive observation is basically used, and the update of observation model is stopped when an occlusion occurs. To detect an occlusion, likelihoods are used. The model is applied to some datasets, and the effectiveness of the method is verified.*
Keywords: Human tracking, Particle filter, Switching observation model, Pose change, Occlusion

1. Introduction.

In recent years, with the increase in crime, awareness of crime prevention is increasing in order to realize a safe and secure social environment [1]. In particular, security measures with surveillance cameras have been already incorporated in shops, stations, streets and so on. According to a survey on the effects of surveillance cameras, crime prevention effect in the parking lot and the hospital is large [2, 3]. In such surveillance systems, videos are used for the early resolution of the crime, and captured images are analyzed at a later date by the human eye. Therefore, it cannot respond quickly if the detection of the red-handed is important, and it is not realistic taking into account the costs, such as personnel expenses for humans to monitor the camera. Human tracking is a technique that a human in the movie image is detected and is tracked, and it is a hot topic in various fields in recent years [4]. By using a human tracking technology in surveillance cameras, suspicious persons can be detected in real time and it is expected that it is possible to prevent a crime. However, in a real environment, there are many objects; a target person is often occluded. Furthermore, pose of the target person usually changes, and illumination condition sometimes varies. In such cases, it is difficult to detect a target person, and it has become a problem in human tracking using movie image.

Various human tracking methods have been proposed so far. As a basic method, a template matching method [5], a mean shift method [6] and Kalman filter [7] are well known and widely used. Template matching can detect a target person by calculating matching degree between template image and local image in whole image of current frame, and accurate tracking can be achieved if an adequate template can be designed. However, it takes much computational costs when size of image becomes large, and it is not suitable for use in real-time application. Mean Shift method is one of the gradient methods by moving the center position to the direction in which the similarity is high in the local region. It is possible to shorten a calculation time by limiting the search area. However, the method often loses a target, and it is not good to apply the method to surveillance camera system. Kalman filter is a kind of time series filter [8] that estimates the state of the subject from past observations. In Kalman filter, distribution of noise generated at the time is assumed to follow a Gaussian distribution, and a position of the object is estimated by using the state transition model. It is very useful and widely used to many applications; however, in the human tracking characteristics of human motion, noise distribution and so on do not follow Gaussian distribution. Particle filter [9] is a kind of the same time series filter, whose efficacy has been demonstrated with respect to the position estimation of the mobile object performing the irregular movement such as a person. In particle filters, an estimate position of the target is represented as a discrete probability density distribution using a set of particles, where each particle is characterized with state of the target, e.g., position. Estimation is based on a distribution of a set of particles, and there is no limitation for distribution. It is the most important difference between Kalman filter and particle filter, and particle filter can robustly track human even if a target irregularly moves. Researches of applying the particle filter in human tracking, are widely reported.

In human tracking methods using the particle filter, a discriminator and features which are used for object detection are often used as an observation model. Furthermore, online learning method in which discriminator is updated during tracking facilitates a robustness of tracking for pose or illumination changes [10]. In the online learning, however, discriminator sequentially updated; thus obstacles are sometimes tracked after occlusion occurs. Although online real boosting method [11], semi-supervised tracking [12], and so on are reported to cope with the problem, high computational cost is required and it is not suitable for real-time application [13].

In this study, we proposed a simple but powerful human tracking method robust for both pose/illumination change and occlusion by adopting the idea of online learning in the particle filter. In the proposed method, occlusion can be detected by observing a likelihood of particles, and update of the discriminator is stopped during the occlusion. Although an additional special calculation is not required in the method, it is expected that robust tracking can be realized. The effectiveness of the proposed method is verified by some simulations in which experimental videos assuming a real environment.

2. **Tracking by Particle Filter.**

The particle filter is a technique of estimating of feature state that is based on recursively computing Bayesian formulation. The Bayesian formulation predicts the sequence of hidden parameters based on only the observed data.

The visual tracking problem is cast an inference task in a Markov model with hidden state variables [14]. The state variables X_t is described target configuration at time t, and observation Y_t extracted from images. We aim to estimate the value of the hidden state variable of X_t based on given all observation $Y_{1:t} = (Y_1, \ldots, Y_t)$ up to time t. The posterior probability $p(X_t|Y_{1:t})$ is estimated with the following Bayesian formulation. Bayesian estimation updates the posterior probability with the following rule

$$p(X_t|Y_{1:t}) \propto p(Y_t|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Y_{1:t-1}) dX_{t-1}, \qquad (1)$$

where, $p(Y_t|X_t)$ denotes the observation model that measures how well the observation fits the predictions, and $p(X_t|X_{t-1})$ represents the motion model that proposes the next state X_t based on the previous state X_{t-1}. Condensation algorithm [9] which is based on factored sampling, approximates an arbitrary distribution of observation with a generated set of weighted samples. The distribution is represented by a set of particles $\{(X_t^{(i)}, w_t^{(i)}), i = 1, \ldots, N\}$, where, $X_t^{(i)}$ and $w_t^{(i)}$ denote a state and an associated weight of the i-th particle at time t, respectively. Using importance sampling, Equation (1) is recursively approximated by

$$p(X_t|Y_{1:t}) \approx p(Y_t|X_t) \sum_i w_{t-1}^{(i)} p(X_t|X_{t-1}). \tag{2}$$

Given a set of particles from the previous time $\left\{\left(X_{t-1}^{(i)}, w_{t-1}^{(i)}\right)\right\}$ configurations at the current time $X_t^{(i)}$, are drawn from a proposal distribution

$$q(X_t) = \sum_i w_{t-1}^{(i)} p\left(X_t|X_{t-1}^{(i)}\right). \tag{3}$$

The weights are then updated as $w_t^{(i)} \propto p\left(Y_t|X_t^{(i)}\right)$. The motion model predicts particle position. The estimate of state of particles is based on a linear extrapolation of the previous state plus Gaussian noise. The current state of target is estimated as weighted average over the states of the particles. The observation model evaluates the weight by likelihood, and is generally established the appearance of target from the image that is utilized color, edges or texture as a feature [15].

In the human tracking, a pose of a target or illumination condition usually changes during tracking. It is difficult to design an observation model which is useful for various conditions. An adaptation method, in which the observation model is updated based on the current observation, has been proposed and has an advantage for tracking under pose/illumination change.

3. **Switching Observation Model for Robust Tracking.** As mentioned in the previous section, adaptive method is very effective in change of shape or illumination. However,

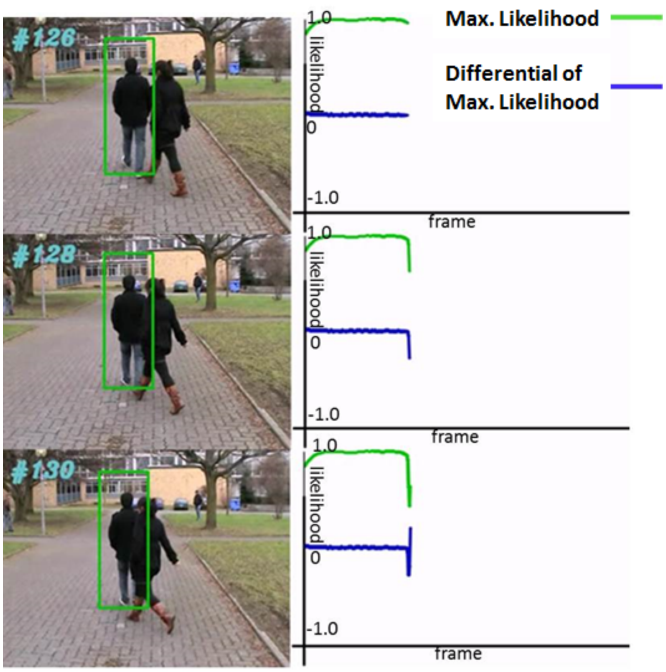
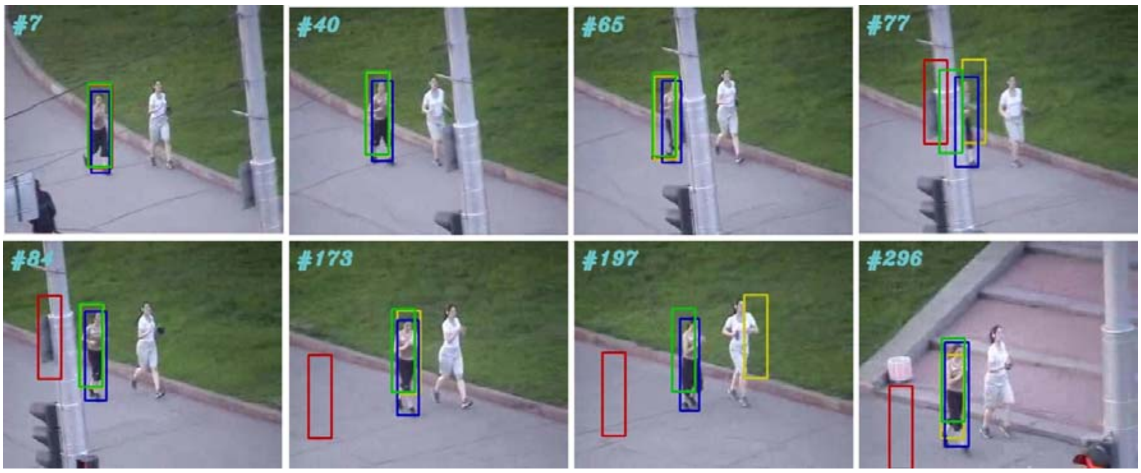


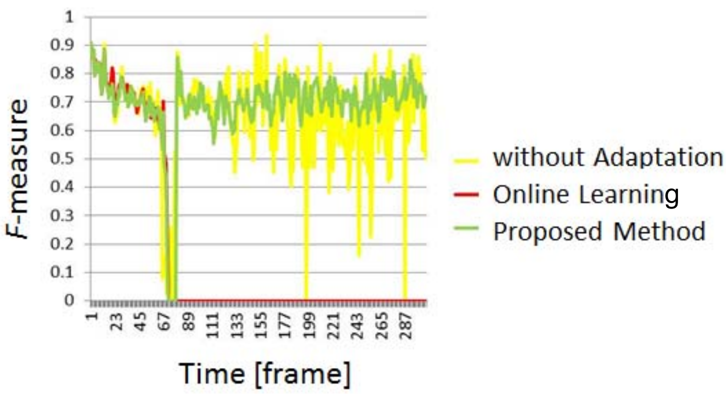
FIGURE 1. Example of change of maximum likelihood and its differential

in case that an object is occluded by obstacles, updated observation model includes obstacles and it causes wrong tracking. In this paper, switching the observation model based on the accuracy of tracking is employed to realize a robustness for both of pose change and occlusion. In this idea, timing of switching the observation model is important. We focus on the maximum likelihood in the particles, and an example of tracking result in which the adaptive observation model is used is shown in Figure 1. In Figure 1, image of some frames, the maximum likelihood in each frame, and differential of maximum likelihood in each frame are shown. It is shown that the maximum likelihood drastically changes when an occlusion occurs. We have confirmed that this phenomenon occurs in various video scenes. Thus, we use the change of maximum likelihood as a trigger for switching the observation model. In other words, adaptive observation model is usually used, and the adaptation is stopped when the differential of maximum likelihood exceeds a threshold value. When the maximum likelihood rises and exceeds the threshold, adaptation of the observation model starts again. The proposed method does not require additional calculation, and then the method is suitable for real time application.

4. **Experimental Results and Discussion.** We applied the proposed method to some benchmark videos [16, 17] in which pose change and/or occlusion occurred in order to confirm the effectiveness of the method. In the experiments, to calculate a likelihood of

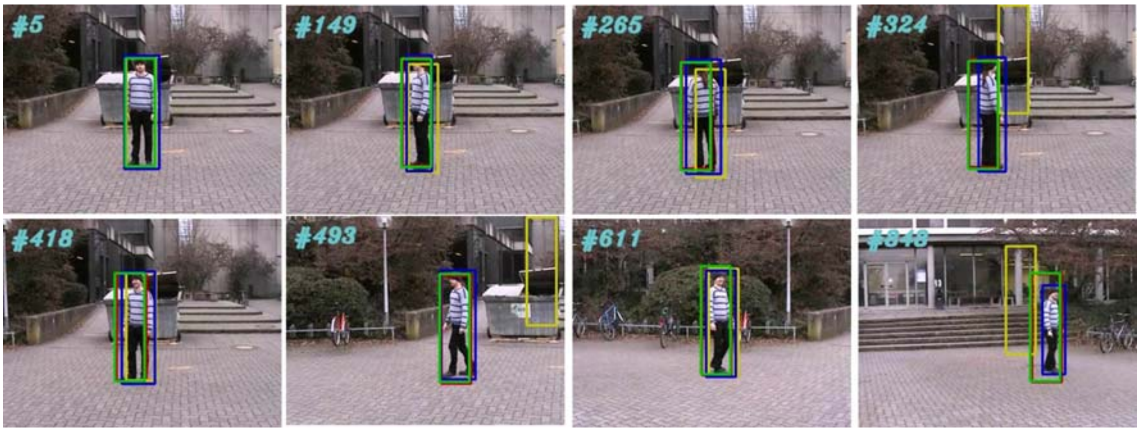


(a)

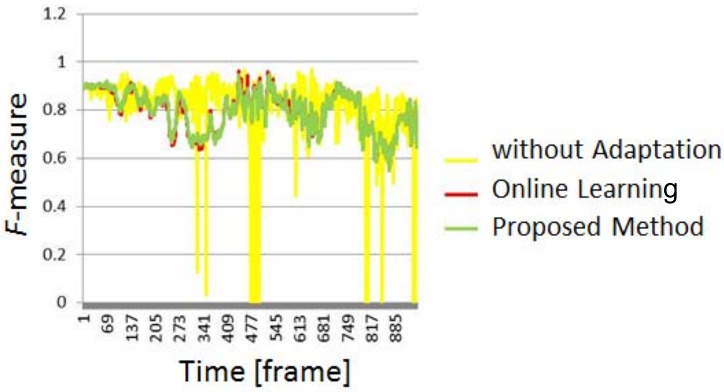


(b)

FIGURE 2. Results for "Jogging". (a) Images of several frames, (b) change of F-measures.



(a)



(b)

FIGURE 3. Results for "Seq.D". (a) Images of several frames, (b) change of F-measures.

each particle, Harr-like feature is adopted as a feature and Bayesian classifier is used. To quantitatively evaluate the accuracy of tracking, F-measure is used, and F-measure is given by

$$F = \frac{2\rho\nu}{\rho + \nu}, \tag{4}$$

where, ρ and ν are called as recall and precision, respectively, and are given by

$$\rho = \frac{A_t^E \bigcap A_t^G}{A_t^G}, \tag{5}$$

$$\nu = \frac{A_t^E \bigcap A_t^G}{A_t^E}, \tag{6}$$

where, A_t^G and A_t^E are correct and estimated regions, respectively. Figure 2 and Figure 3 show the example of results for "Jogging" and "Seq.D", respectively. In each frame image, black box is a correct region, and yellow, red and green boxes represent region estimated by without adaptation, adaptation and proposed method, respectively.

"Jogging" video includes occlusion, and then the adaptive observation model changes to inadequate one during occlusion (see frame 77). On the other hand, "Seq.D" video includes pose change only, and adaptive method succeeds in tracking the target. In both videos, the proposed method exhibits robust tracking.

In the method, change of maximum likelihood is used for deciding that adaptation is stopped or not. Large threshold is useful for improvement of accurate tracking. Although adaptation is sometimes stopped because of some reasons even if occlusion does not occur, target can be found in some frames.

Calculation time of the proposed method is 60 [fps], under Intel i7 870 CPU (2.93GHz). In the contrast, semi-supervised tracking [12] and online real boosting [18] perform 11 [fps] and 20 [fps] under the environment Intel i7 3770 CPU (3.4GHz) and Intel Core 2 Duo CPU (2.4GHz), respectively.

5. **Conclusions.** In this paper, we described a switching method for adaptive/nonadaptive observation model in the particle filter. In the method, adaptive observation model is basically used, and the adaptation is interrupted when the target is occluded. To detect an occlusion, differential maximum likelihood of set of particles. For all movies used in the simulation, all occlusion could be detected by the method, and robust tracking was achieved. Furthermore, the proposed method does not require additional calculation, and then it is expected that the method can be applied to practical problems.

REFERENCES

[1] Y. Yamauchi, H. Fujiyoshi, B. W. Hwang and T. Kanade, People detection based on co-occurrence of appearance and spatiotemporal features, *Proc. of MIRU2007*, pp.1492-1497, 2007 (in Japanese).
[2] M. Gill and A. Spriggs, *Assessing the Impact of CCTV*, Home Office Research Study, 2005.
[3] C. Welsh and D. P. Farrington, *Crime Prevention Effects of Closed Circuit Television: A Systematic Review*, Home Office Research Study, 2002.
[4] Y. Li, H. Ai, S. Lao, T. Yamashita and M. Kawade, Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.30, no.10, pp.1728-1740, 2008.
[5] T. Nihei, Y. Hagihara, Y. Hagihara, T. Miyoshi and A. Yumit, Adaptice template generation in template matching – The method by the genetic algorithm based on a statistical valuation function, *Proc. of SICE Tohoku Chapter*, 2013 (in Japanese).
[6] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *IEEE Trans. Pattern Analysis and Image Intelligence*, vol.25, no.5, pp.603-619, 2002.
[7] C. Masreliez and R. Martin, Robust Bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Trans. Automatic Control*, vol.22, no.3, pp.361-371, 1977.
[8] T. Kato, Introduction of perticle filter and its implementation, *Proc. of CVIM-157*, pp.161-168, 2009 (in Japanese).
[9] M. Isard and A. Blake, Condensation – Conditional density propagation for visual tracking, *Int. J. of Computer Vision*, vol.29, no.1, pp.5-28, 1998.
[10] T. Yamashita, H. Fujiyoshi, S. Lao and M. Kawade, Human tracking based on soft decision feature and online real boosting, *Proc. of Int. Conf. on Pattern Recognition*, pp.1-4, 2008.
[11] T. Yamashita, S. Lao and M. Kawade, Human tracking by online real boosting, *IPSJ Trans. Computer Vision and Image Media*, vol.1, no.1, pp.73-82, 2008.
[12] H. Grabner, C. Leistner and H. Bischof, Semi-supervised on-line boosting for robust tracking, *Lecture Notes in Computer Science*, vol.5302, pp.234-247, 2008.
[13] Y. Wu, J. Lim and M. H. Yang, Online object tracking: A benchmark, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.2411-2418, 2013.
[14] D. A. Ross, J. Lim, R. Lin and M. Yang, Incremental learning for robust visual tracking, *Int. J. of Computer Vision*, vol.77, pp.125-141, 2008.
[15] A. Yilmaz, O. Javed and M. Shah, Object tracking: A survey, *ACM Computing Surveys (CSUR)*, vol.38, no.4, pp.1-45, 2006.
[16] D. A. Klein and D. Schulz, Adaptive real-time video-tracking for arbitrary objects, *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp.772-777, 2010.
[17] *BoBoT Datasets*, http://www.iai.uni-bonn.de/kleind/tracking/index.htm.
[18] T. Yamashita, T. Noda, S. Lao and M. Kawade, Human tracking by online boosting using real adaboost, *Technical Report of CVIM2007 (31(2007-CVIM-158))*, pp.85-92, 2007.