19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Multi-Features Particle PHD Filtering for Multiple Humans Tracking

Tassaphan Suwannatat and Krisana Chinnasarn
Knowledge and Smart Technology Research Laboratory
Faculty of Informatics, Burapha University
Sansuk, Muang, Chonburi, Thailand
Email:tassaphan_su@rmutto.ac.th, krisana@buu.ac.th

Nakorn Indra-Payoong
Faculty of Logistics
Burapha University
Sansuk, Muang, Chonburi, Thailand
Email: nakorn.ii@gmail.com

***Abstract*—This paper proposes multi-features visual tracking algorithm based on the particle Probability Hypothesis Density filter, which allows accurate and robust tracking under the circumstance of visual tracking. We apply a particle PHD filter implementation to the multiple humans tracking using multi-features observation that exploits skin and head-and-shoulder boundary as its prior density. The relevance of our approach to the problem of multiple humans tracking is then investigated using a tracker which is able to follow the state according to the humans' motion. The accuracy and robustness are evaluated and compared using real visual tracking experiments.**

I. Introduction

One of the challenges in multiple targets tracking is to estimate the positions and number of targets in the image sequence. Unpredictably is due to occlusion, clutter and complex background. The follow researchs are some visual tracking of multiple targets. A possible way to solve the problem of tracking each target independently is to use the single algorithms [1], [2], [3]. The method of tracking a variable number of interacting targets using trans-dimensional Markov chain Monte Carlo [4]. A more elegant solution to solve nonlinear problem is particle filter which is a Monte Carlo simulation based recursive Bayes filter [5]. In this solution, Random Finite Set (RFS) represents the multi-target state and the dimensionalilty of the target grows exponentially with the number of targets. Also a number of samples from Monte Carlo sampling grows exponentially, therefore making the propagation of the full posterior are impractical. To reduce the computational complexity, Mahler created the Probability Hypothesis Density filter (PHD) as an estimation of multi-target filter [6]. The samples generated by a SMC method can estimate the integrals of the PHD recursion [7]. The PHD filter is a tractable alternative to the optimal multi-target filter. Under the assumption that the predicted multi-target density is Poisson, this recursion is exact and completely characterizes the statistics of the dynamic Poisson point process of interest.

There are some works on multiple targets tracking using particle filters which can be divided into two categories: 1) one particle filter with the joint state space [8], [4]; 2) one mixture particle filter with one individual particle filter [9], [2]. The PHD filter is similar to the second group but the integral of the PHD filter is the expected number of targets within this region, which differs from the other multiple target particle filters.

978-1-4673-7825-3/15/$31.00 © 2015 IEEE

The PHD filter have been tested on synthetic data [7], 3D sonar data [10], feature point filtering [11], multi-sensor vehicles tracking [12], groups of human detection [13] and multi-target visual tracking [14]. None of the above techniques are applied to track multiple humans by using multi features filtering. This paper considers the problem of tracking unexpected moving humans. The ideal behavior of moving humans tracking system is to provide a set of track that is in one-to-one correspondence with the object. The aims of the system are: 1) to realize the task of multiple humans detection with the aim of a real-time effective implementation; 2) to determine the trajectory of moving humans within the environment.

In this paper we present a multi-features probabilistic framework for tracking multiple humans. This technique uses a single camera and a model of the objects being tracked in order to estimate their positions. In particular, we introduce the PHD-based multi-features probability framework which combines the advantages of color feature-based and contour feature-based method. By taking advantage of the unique feature and properties of the color and contour feature of humans, we hope to overcome the inherent disadvantages of each, resulting in a combined feature which is more effective than either feature is used individually.

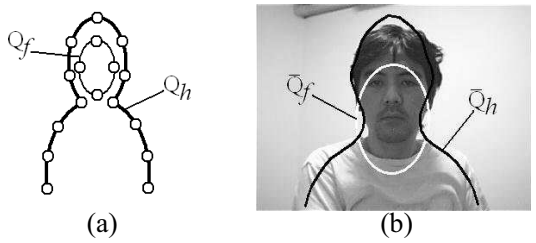
This paper is organized as follows. After introducing the learning of human's multi-features model in the next section, we introduce the general concept of particle PHD filter in Section III. In Section IV, we present the multi-features particle PHD filter and the implementation. Section V describes several experiments. In this section, our proposed method is evaluated and compared on a challenging synthetic tracking problem. Finally, Section VI summarizes this paper and makes the suggestions for future research.

II. Learning Human's Multi-Features Model

An essential ingredient in the proposed method is a prescription for choosing posterior density. This will require a prior density which could be specified by the user but for our application is learnt automatically by the following method.

A. A Shape Vector of Face Model

In order to model a shape of human's face area; the skin-color detection [15] is applied to the image sequence. After the skin-color boundary has been detected, then the ellipse shape is fit to the skin-color contour. The four points of ellipse



(a) (b)

Fig. 1. (a) Human's multi-features model and (b) Mean shape model.

(as shown in Fig. 1(a)) are used to form mean shape of face model $\bar{\mathbf{Q}}_f$ (as shown with white line in Fig. 1(b)) by using the Generalised Procrusts Analysis, GPA [16].

B. A Shape Vector of Head-and-Shoulder Model

The system first segments moving human from a sequence by simple differencing from a background image and thresholding. After the region of human is detected, the points around head-and-shoulder area (as shown in Fig. 1(a)) are chosen. Then the mean shape model of head-and-shoulder (as shown with black line in Fig. 1(b)), $\bar{\mathbf{Q}}_h$, is formed by GPA [16].

C. Representation of Shape Space

For our tracking system, we represent the model of human as a *shape space*. We parameterize the *shape space vector* of both face and head-and-shoulder models according to:

$$\mathbf{Q} = W\mathbf{X} + \bar{\mathbf{Q}} \tag{1}$$

where

$$W = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \bar{\mathbf{Q}}^x & -\bar{\mathbf{Q}}^y \\ \mathbf{0} & \mathbf{1} & \bar{\mathbf{Q}}^y & \bar{\mathbf{Q}}^x \end{pmatrix} \tag{2}$$

$\mathbf{1}$ and $\mathbf{0}$ are column vectors 1's and 0's of length n, respectively. The *shape space*, $\mathbf{X}$, defines the parameterization the set of Euclidean similarities (translation, rotation, and scaling). $\bar{\mathbf{Q}}$ is the template model which is derived from mean shape model as described in previous section where $\bar{\mathbf{Q}}^x$ and $\bar{\mathbf{Q}}^y$ are components of coordinate x and y, respectively. $\mathbf{Q}$ is the stacked coordinates x and y of the tracking object. The usage of *shape space* is described in Sect. 4.

III. Overviews of The Particle PHD Filter

We approximate the target center in an image at $\boldsymbol{s} = (s_1, s_2)$, rotation φ and scaling γ with the speed $\boldsymbol{v} = (v_1, v_2)$. The state of a single target at time t be $x_t = (\boldsymbol{s}, \boldsymbol{v}, \varphi, \gamma)$. And the single target observation $y_t = (\mathbf{s}, \varphi, \gamma)$ be generated by the observation. The corresponding multi-targets state x_t and y_t are the measurements of each target. If $M(t)$ is the number of targets at time t, then $X_t = x_1, x_2, ..., x_{M(t)}$ is the multi-targets state. $Y_t = y_1, y_2, ..., y_{N(t)}$ is the multiple target measurement formed by the observations $N(t)$, where some of these observations may drop into clutter background.

The Probability Hypothesis Density [6] (PHD) is the first order moment of a Random Finite Sets (RFS) whose integral on any region S of the state space and gives the number of targets in S. The PHD is a function in the single target state space. Peaks of the PHD identify the probable position of the targets. Consequently, the peaks of PHD are the highest local concentration of expected number of targets. And it can be used to generate the states of the targets. The iterative prediction and update of the Baysian is known as the PHD filter.

Let $D(x_t|Y^t)$ stand for the PHD associated with the multi-target posterior $p(X_t|Y^t)$ at time t. The original PHD filter consists of prediction step and update step. The PHD prediction equation is:

$$\begin{aligned} D(x_{t+1}|Y_t) &= b(x_{t+1}) + \int (p(x_t)p(x_{t+1}|x_t) \\ &+ b(x_{t+1}|x_t))D(x_t|Y_t)dx_t \end{aligned} \tag{3}$$

where $b(x_{t+1})$ signify the intensity function of the spontaneous birth RFS, $b(x_{t+1}|x_t)$ signify the intensity function of the RFS of targets position from the previous state x_t, $p(x_t)$ is the probability that the target still exists at time $t+1$ given it has previous state x_t. $p(x_{t+1}|x_t)$ is the transition probability density of individual target. The PHD update equation is:

$$D(x_{t+1}|Y^{t+1}) \cong F(Y_{t+1}|x_{t+1})D(x_{t+1}|Y^t) \tag{4}$$

$$\begin{aligned} F(Y_{t+1}|x_{t+1}) = 1 - p_D + \sum_{y \in Y_{t+1}} \\ \frac{p_D.p(y|x_{t+1})}{\lambda c(y) + D[p_D.p(y|x_{t+1})]} \end{aligned} \tag{5}$$

where p_D is the probability of detection $p(x_{t+1})$, $p(y|x_{t+1})$ is the likelihood of target, λ is the average number of clutter points, $c(y)$ is the probability distribution of each clutter point, and $D[h] = \int h(x_{t+1})D(x_{t+1}|Y^t)dx_{t+1}$.

IV. Multi-Features Particle PHD Filter

Let us assume that the motion model of each target is a constant velocityl,

$$x_{t+1} = x_t + v_t + u_t \tag{6}$$

where v_t is the velocity of the target approximated by $v_t = x_t - x_{t-1}$ and u_t is a zero-mean Gaussian white process noise at time t. For the existing target, this model is used as our proposal density of the particle filter. For birth targets, the proposal density is a uniform distribution on the centroid of each face regions,

$$b(x_t) \sim U[c] \tag{7}$$

where $U[c]$ is a uniform distribution function at region c. For each target, we use the centroid of its face region as our measurement y to update the PHD filter. The likelihood function is:

$$p(y|x_t) = \exp[-\left|(y - x_t)^T(y - x_t)\right|^{1/2}] \tag{8}$$

Let L_t denote the particle number at time t, J_{t+1} denote the new particle number for the birth targets at time $t+1$, and w denote a particle's weight. Our proposed multi-featured PHD particle filter consists of 6 steps as describe as follows.

At time $t \geq 0$, let $x_t^{(i)}, w_t^{(i)}$, for $i = 1, ..., L_t$ denote a particle approximation of the PHD.

A. Prediction

For $i = 1, ..., L_t$, generate a sample $\tilde{x}_{t+1}^{(i)}$ according to Eq. (6) and compute the predicted weights

$$\widehat{w}_{t+1}^{(i)} = w_t^{(i)} \tag{9}$$

For $i = L_t + 1, ..., L_t + J_{t+1}$, generate sample $\tilde{x}_{t+1}^{(i)}$ according to Eq. (6) and compute the predicted weights

$$\widehat{w}_{t+1}^{(i)} = 1/J_{t+1} \tag{10}$$

B. Update

For each $y \in Y_{t+1}$, use the likelihood Eq. (8) and compute

$$C_{t+1}(y) = \sum_{i=1}^{L_t+J_{t+1}} p_D.p(y|\tilde{x}_{t+1}^{(i)})\hat{w}_{t+1}^{(i)} \tag{11}$$

For $i = 1, ..., L_t + J_{t+1}$ update weights

$$\tilde{w}_{t+1}^{(i)} = \hat{w}_{t+1}^{(i)} \times \left[1 - p_D + \sum_{y=Y_{t+1}} \frac{p_D.p(y|\tilde{x}_{t+1}^{(i)})}{\lambda c(y) + C_{t+1}(y)}\right] \tag{12}$$

C. Re-sampling

Compute the target number at time $t + 1$

$$M_{t+1} = \sum_{i=1}^{L_t+J_{t+1}} \widetilde{w}_{t+1}^{(i)} \tag{13}$$

Initialize the cumulative probability $c_1 = 0$,

$$c_i = c_{i-1} + \frac{\tilde{w}_{t+1}^{(i)}}{M_{t+1}}, for\ i = 2, ..., L_t + J_{t+1} \tag{14}$$

Then the samples $x_{t+1}^{(i)}$ representing possible future states are sampled using a factored sampling algorithm as shown in Fig. 2 and the weights $w_{t+1}^{(i)}$ are rescale by M_{t+1}/L_{t+1}, for $i = 1, ..., L_t$.

For $i = 1, 2, ..., L_t$
  $r = Random(0,1)$
  $selected = Search(\mathbf{c},r)$
  $x_{t+1}^{(i)} \leftarrow \tilde{x}_{t+1}^{(selected)}$
End
where
  $c = cumulative\ distribution\ of\ weights$
  $Search(\mathbf{a},b) = find\ smallest\ j\ such\ that\ \mathbf{a}[j] \geq b$
  $Random(a,b) = uniform\ random\ number \in [a, b]$

Fig. 2. The factored sampling algorithm.

D. Observation from PHD Samples and Clustering

1) Observation from PHD Samples: To observe from the PHD samples, first an estimate for the *shape space*, $\mathbf{X}$, is expanded into a vector $\mathbf{Q}_h$ according to Eq. (15).

$$\mathbf{Q}_h = W\mathbf{X} + \bar{\mathbf{Q}}_h \tag{15}$$

Then the entire current image is then run through a Canny edge detection and the vector $\mathbf{Q}_h$ is deformed into $\mathbf{Q'}_h$ according to the edges. To do this, for each point of $\mathbf{Q}_h$, two rays are cast from the point in direction normal to the curve (in opposite direction to search both side of the curve). If either or both of the rays hit an edge, the (x, y) coordinate of the closest one replaces the current point. A result of *shape space* fitting is shown in Fig. 3(a).

After the vector $\mathbf{Q}$ is deformed into $\mathbf{Q'}_h$, it is required to come up with the closest *shape space*, $\mathbf{X}$, to the deformed vector produced by displacing vector points to the nearest detected edge. According to Eq. (1), this problem can be expressed as a minimization problem:

$$\min_{\mathbf{X}} ||W\mathbf{X} + \bar{\mathbf{Q}}_h - \mathbf{Q'}_h|| \tag{16}$$

Now, because the parameterization is a proper subset of W, we can project a deformed vector onto W to get the closest *shape space*, $\mathbf{X}$. Thus, we can make an approximation as

$$\mathbf{X} = (W^{\mathrm{T}}W)^{-1}W^{\mathrm{T}}(\mathbf{Q'}_h - \bar{\mathbf{Q}}_h) \tag{17}$$

Then, how well a *shape space* fits with the edges is calculated by expanding the *shape space*, $\mathbf{X}$ into the points vector by using Eq. (15), and then the error between transformed points and edge points is calculated as follows:

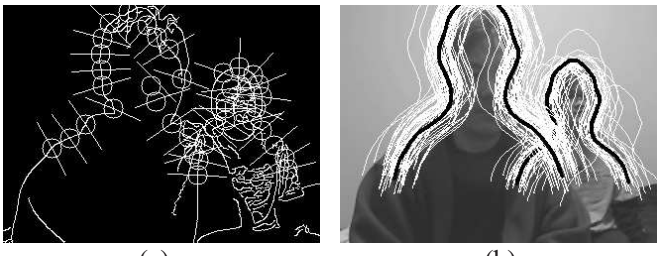
$$err = \sum_i |p_i - e_i|^2 \tag{18}$$

where p_i and e_i are the image coordinates of the transformed points and the edge points respectively. If it found no edge for this point, then the term is replaced with the fixed penalty. Finally, the error is converted into a weight according to:

$$w = e^{-\sqrt{err}} \tag{19}$$

The results of the observation from PHD samples are shown with white lines in Fig. 3(b).

2) Clustering: In our system, PHD particles are clustered with *K-means* where M_{t+1} is used as the number of clusters. After convergence of *K-means*, the samples which have maximum weight in each cluster (as shown with black line in Fig. 3(b)) are recorded as $\left\{u^{(m)}, \beta^{(m)}\right\}$, for $m = 1, ..., M_{t+1}$ where u is the *shape space* and β is the weight cost.



(a) (b)

Fig. 3. (a) shape space fitting and (b) The mean of sampling distribution

E. Observation from a New Sample

To do this, first, the skin-color detection is applied to the current frame. After the face boundary on the current frame has been detected, the ellipse shape is fitted to the face contour. Then, the four points of ellipse are stored as $\mathbf{Q'}_f$. Together with the template of trained face shape, $\bar{\mathbf{Q}}_f$, they are fed into Eq. (20) to compute the *shape space* $\mathbf{X}$.

$$\mathbf{X} = (W^{\mathrm{T}}W)^{-1}W^{\mathrm{T}}(\mathbf{Q'}_f - \bar{\mathbf{Q}}_f) \tag{20}$$

Similar to the observation from PHD samples, the *shape space*, $\mathbf{X}$, is expanded into a spline $\mathbf{Q}_h$ by using Eq. (15). Then, the error between a new sample and the nearest edge is calculated by Eq. (18). Finally, a weight of new observation is calculated by Eq. (19). The samples and weights of new observation are stored as $\left\{v^{(n)}, \alpha^{(n)}\right\}$, for $n = 1, ..., N_{t+1}$ where v is the *shape space*, α is the weight cost and N_{t+1} is the number of samples from a new observation.

F. Measurement

Firstly, we use the Hungarian algorithm [17] to find maximum weight matching of $u^{(m)}$ and $v^{(n)}$. The algorithm is formulated by using a bipartite graph. We have a complete bipartite graph with m vertices from samples $u^{(m)}$ and n vertices from samples $v^{(n)}$, and each edge has a weight cost which is computed by expanding $u^{(m)}$ and $v^{(n)}$ into vectors and calculate the likelihood between them. For each maximum matching, we perform a measurement for deciding which sample is most likely to be the target as

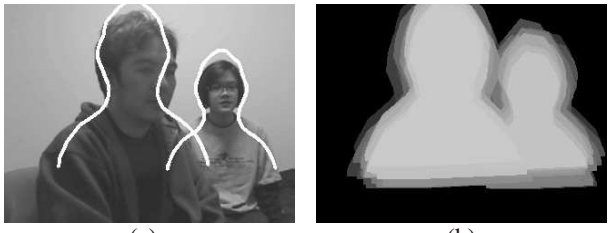
$$D_{m,n} = \frac{\alpha_m}{\beta_n} \tag{21}$$

This likelihood ratio shows that the sample discriminant can be used for two inferences:

- If $D_{m,n} > 1$, a sample n which is observed from a new sampling is more "target-like" than a PHD sample with maximum weight in cluster m . Then update $x_{t+1}^{(i)} = v^{(n)}$, for all PHD samples i in cluster m.
- Otherwise, a PHD sample with maximum weight in cluster m is more "target-like" than a new sample n.

G. Report

After any time step, it is possible to report the result of tracking by displaying the PHD samples which have maximum weight in each cluster. Fig. 4(a) shows a sample output of report step, where white lines correspond to maximum weight of the PHD samle in each cluster. In additional, we also report the likelihood map of the humans tracking (as shown in Fig. 4(b)) by displaying the weight of each PHD sample by grey scale color.



(a) (b)

Fig. 4. (a) Tracking output and (b) Likelihood map.

TABLE I. COMPARISON OF TRACKING ACCURACY BETWEEN THE SKIN-PHD AND MF-PHD.

	SFDA-D	STDA-D
Skin-PHD	0.42	118.53
MF-PHD	0.47	133.18

V. EXPERIMENTAL RESULTS

A. Performance Comparison

In this section, we conducted the experiment to compare the result of our tracking method with the original PHD filter of skin color tracking. We used the real video sequence with a resolution of 320×240 pixels. In the image sequence, the left person is moving turnaround with approximately constant velocity while the right person stays fixedly. The length of sequence was 777 frames and had duration of 25s. For the tracking parameter, we used the number of particles $L_t = 500$, the particle number of birth targets $J_{t+1} = 500$ and a trained head-and-shoulder model with 19 control points to maintain real-time quality. From our implementation we found that the translation noise with magnitude ± 2.5, scaling/rotation noise with magnitude ± 0.005 yield acceptable results for humans motion with slightly movement.

Figure 5 shows the comparative results of our multi-features PHD particle filtering method (MF-PHD) and the PHD particle filtering of face detection using skin-color (Skin-PHD) that we implemented. Figure 5(a) shows sub-sampled frames of the result of Skin-PHD. The result show that after clustering of Skin-PHD particles, the maximum weight of each cluster are right around the face area of humans. But when face area of the left person disappeared from the scene (as shown in frame 172 of Fig. 5(a)) then the state of the left person cannot be tracked by the Skin-PHD filter. By comparing the result of our proposed method (as shown in Fig. 5(b) and Fig. 5(c)), the state of maximum weight of the MF-PHD are correctly detected on each human. The result showed that our method can maintain the tracking through the occlusion of the skin area. Hence, in this case, our method is better than Skin-PHD filter.

To evaluate the performance of our method, we used the VACE protocol [18], to show the comparative performance of Skin-PHD and MF-PHD. We compared the performance by evaluating the Frame Detection Accuracy-Distance (FDA-D). This makes the distinction between the individual objects in the frame and requires a unique one-to-one mapping of the ground truth data. To simplify our evaluation, we used only x-coordinate value to compute the error distances between the i^{th} mapped pair of the ground truth and detected output objects and call this the d_i' in Eq. (22). N_{mapped} is the number of mapped object sets, N_G and N_D are the number of ground truth objects and detected objects respectively. Using the assignment sets, we computed FDA-D for each frame t as:

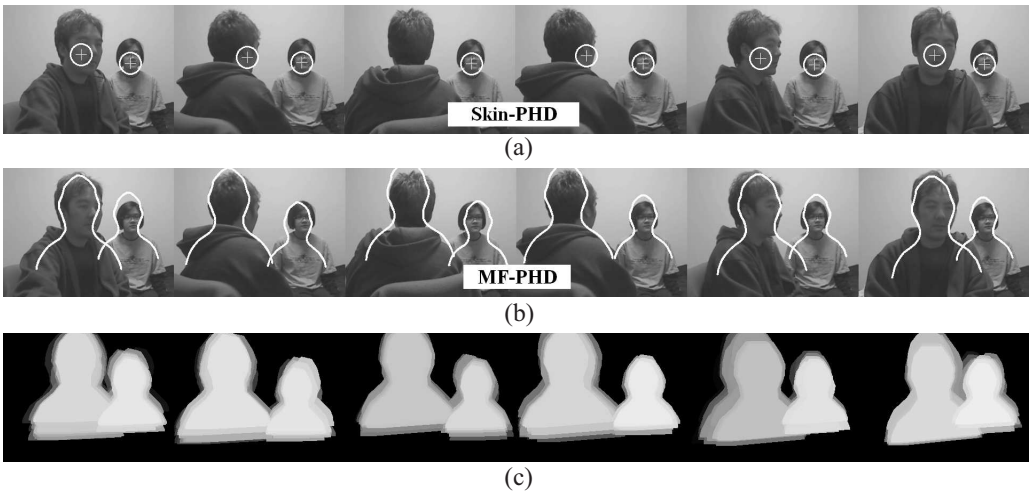


Fig. 5. The sub-sampled output at frame 7, 110, 172, 280, 334 and 390. (a) Tracking output of Skin-PHD (b) Tracking output of MF-PHD and (c) Likelihood map.

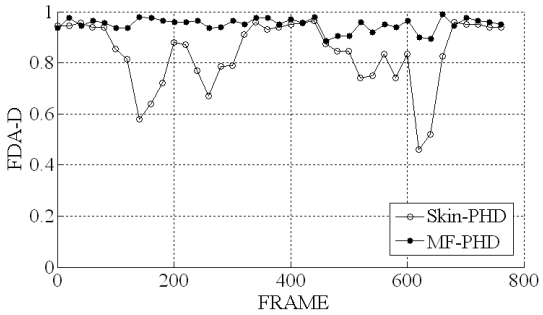


Fig. 6. The evaluation of Frame Detection Accuracy-Distance (FDA-D) between Skin-PHD and MF-PHD.

$$FDA{-}D(t) = \frac{\sum_{i=1}^{N_{mapped}} (1 - d_i')}{\left[\frac{N_G^{(t)} + N_D^{(t)}}{2}\right]} \tag{22}$$

This evaluation is then averaged the Sequence Frame Detection Accuracy-Distance (SFDA-D), which is defined as the ratio of the sum total of the FDA-D over the sequence to the number of frames in the sequence where either ground truth or detected object exists.

$$SFDA{-}D = \frac{\sum_{t=1}^{N_{frames}} FDA{-}D(t)}{\sum_{t=1}^{N_{frames}} \exists(N_G^{(t)} \mathrm{OR} N_D^{(t)})} \tag{23}$$

Furthermore, we evaluate the Sequence Tracking Detection Accuracy-Distance (STDA-D) which is based on how well a system can track the person in the entire sequence. The normalization of STDA-D is denoted by the $N_{(G_i \cup D_i \neq 0)}$, which indicates the total number of frames in which either a ground truth object or detected object or both are present. The summation runs from 1 through to N_{mapped} which indicates the number of mapped objects.

$$STDA{-}D = \sum_{i=1}^{N_{mapped}} \frac{\sum_{t=1}^{N_{frames}} (1 - d_t')}{N_{(G_i \cup D_i \neq 0)}} \tag{24}$$

The results of FDA-D is shown in Fig. 6, the result shows that MF-PHD maintained a better tracking accuracy throughout the sequence. In contrast, Skin-PHD tracking is distracted when its lack of skin area such as frame 172. SFDA-D and STDA-D are shown in Table. 1. It is possible to notice that MF-PHD outperform Skin-PHD for both scores.

B. Robust Testing

To investigate the robustness of our tracking method, we performed experiments on the tracking of multiple humans on oclusion environment. In this experiment, the ability of our tracking method is demonstrated in the real-time tracking of two humans who occlude each other as they walk past. The hardware used in this experiment is the conventional camera with the resolution of 320×240 pixels. Our method was implemented on a 2.20 GHz Intel based PC. The parameter of tracker were set same as in Sect. V.A. The overall system runs about 40 Hz when tracking. In this experiment, two humans walked left and right in front of a cluttered room. Figure 8(a) shows the sub-sampled output at frame 100, 125, 130, 133, 135, 140, 150, 160 and 200 of the total 435 frames. In this experiment, tracking output are drawn into the output frame image by thier face's bounding ellipse and head-and-shoulder's bounding spline. Figure 8(b) depicts a plot of state density on horizontal image axis, to demonstrate the occlusion condition, only the state densities among frame 100 to 200 are plotted. From the results, most of the outputs are correctly detected but there were some frames (such as frame 150 of Fig. 8(a)) the distribution is attracted to background clutter, but rapidly converges back onto a right human at a few consecutive frames. In occlustion condition (frame 133), we see that the man's black shirt is occluded by the person in front of him. In this condition, the tracker reports only one output for frame 133. But in a few frame later, frame 135 of Fig. 8(a) shows that

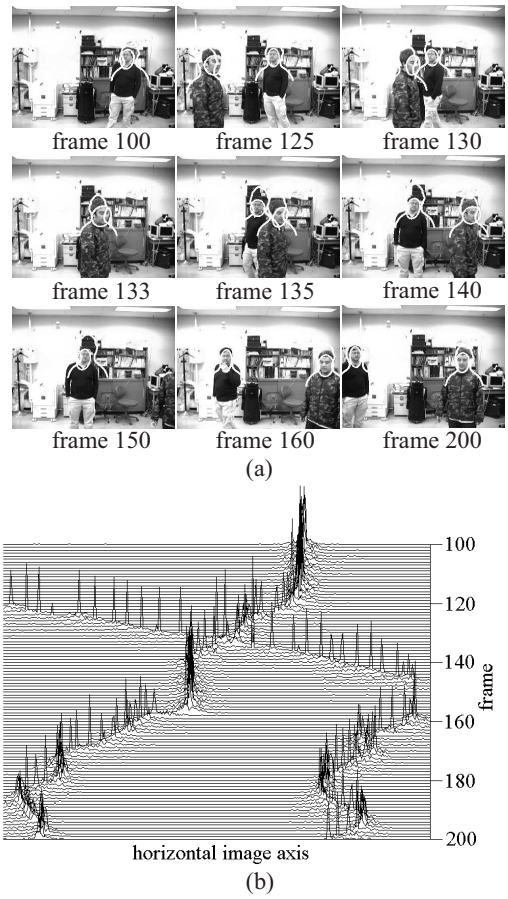


Fig. 7. The experiment on occlusion condition (a) Sub-sampled outputs (b) The plot of state density on horizontal translation axis

the two humans have been detected and keeps track on the humans correctly.

VI. Conclusions and Future work

In this paper, we proposed a new particle PHD filtering method called multi-features particle PHD filter. Although the original PHD filter was successful implementation for multiple targets tracking but when the filter is unable to assign the observed features, it causes a crude approximation of the posterior distribution. This problem often occurs in real-time visual tracking tasks due to the target's feature is slightly changed in model or the target is fallen into cluttered background. Our proposed method emphasizes the detection of target likelihood that used the measurement of samples discriminant, a likelihood ratio which is derived from a probability model of current observation and prediction density. The method requires no previous knowledge of background, is efficient in clutter scense and not affected by lighting change. We also proposed multi-features model of human as a priori distribution for the tracking which combined the advantages of skin color feature-based and contour feature-based of a human. Impressive results have been demonstrated in several experiments. Promising comparative experimental results of the original PHD filter and our new tracking method are demonstrated, in which we examined their accuracy and robustness. The result demonstrated that our method can track correct positions of multiple humans throughout the sequence. To apply our method in various problems, however, improvement of the observation model should be based on the tasks. An improvement in the observation model is one of the most important issues for our future work. For example, the problem will become challenging when using multi-modal as the observation model of the filter.

References

[1] C. Li, Y. Xu, X. Shi, and S. Wu, "Human body detection and tracking from moving cameras," *Proc. of BioMedical Engineering and Informatics*, pp. 278–281, 2012.

[2] K. Okuma, A. Taleghali, N. D. Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," *Proc. of the European Conf. on Computer Vision*, vol. 1, pp. 28–39, 2007.

[3] W. Zhang and S. M. Bhandarkar, "Face detection and tracking using a boosted adaptive particle filter," *Journal of Visual Commu. and Image Representation*, vol. 20, no. 1, pp. 9–27, 2009.

[4] Z. Khan, T. Balch, and F. Dellaert, "Mcmc-based particle filtering for tracking a variable number of interacting targets," *IEEE Tranc. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1805–1819, 2005.

[5] A. Doucet, N. de Freitas, and N. J. Gordon, "Sequential monte carlo methods in practice," 2003, new York, Springer-Verlag.

[6] R. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Trans. Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[7] B. N. Vo, S. R. Singh, and A. Doucet, "Sequential monte carlo implementation of the phd filter for multi-target tracking," *Proc. of Int. Conf. on Information Fusion*, pp. 792–799, 2003.

[8] C. Hue, J. L. Cadre, and P. Perez, "Tracking multiple objects with particle filtering," *IEEE Trans. on Aerospace and Electronic System*, vol. 38, no. 3, pp. 791–812, 2002.

[9] C. Chang, R. Ansari, and A. Khokhar, "Multiple object tracking with kernel particle filter," *Proc. of Computer Vision and Pattern Recognition.*, vol. 1, pp. 20–25, 2005.

[10] D. E. Clark and J. Bell, "Bayesian multiple target tracking in forward scan sonar images using the phd filter," *IEE Proc. of Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 327–334, 2005.

[11] H. Maeda, N. Ikoma, and T. Uchino, "Tracking of feature points in image sequence by smc implementation of phd filter," *Proc. of Annual Conf. on SICE*, vol. 2, pp. 1696–1701, 2004.

[12] M. Maehlisch, R. Schweiger, W. Ritter, and K. Dietmayer, "Multisensor vehicle tracking with the probability hypothesis density filter," *Proc. of Int. on Information Fusion*, pp. 1–8, 2006.

[13] Y. D. Wang, J. K. Wu, A. A. Kassim, and W. M. Huang, "Trackin a variable number of human groups in video using probability hypothesis density," *Proc. of IEEE Int. Conf. on Pattern Recoqnition*, vol. 3, pp. 1127–1130, 2006.

[14] E. Maggio, E. Piccardo, C. Regazzoni, and A. Cavallaro, "Particle phd filtering for multi-target visual tracking," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1101–1104, 2007.

[15] H. Quan, M. Meguro, and M. Kaneko, "Skin-color-based image segmentation and its application in face detection," *Proc. of IAPR Workshop on Machine Vision Applications*, pp. 48–51, 2002.

[16] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Training models of shape from sets of examples," *Proc. of British Machine Vision*, pp. 9–18, 1992.

[17] G. A. Mills-Tettey, A. T. Stentz, and M. B. Dias, "The dynamic hungarian algorithm for the assignment problem with changing costs," Robotics Institute, Carnegie Mellon University, Technical Report CMU-RI-TR-07-27, 2007.

[18] R. Kasturi, "Performance evaluation protocol for face, person and vehicle detection&tracking in video analysis and content extraction," 2006, computer Science and Engineering, University of South Florida.