

# Optical Engineering

[SPIDigitalLibrary.org/oe](http://SPIDigitalLibrary.org/oe)

## **Particle probability hypothesis density filtering for multitarget visual tracking with robust state extraction**

Jingjing Wu  
Shiqiang Hu  
Yang Wang

# Particle probability hypothesis density filtering for multitarget visual tracking with robust state extraction

Jingjing Wu,<sup>a</sup> Shiqiang Hu,<sup>a</sup> and Yang Wang<sup>b</sup>

<sup>a</sup>Shanghai Jiao Tong University, School of Aeronautics and Astronautics, Shanghai, 200240 China

E-mail: wjjlady720@sjtu.edu.cn

<sup>b</sup>Shanghai Jiao Tong University, School of Electronic, Information and Electrical Engineering, Shanghai, 200240 China

**Abstract.** Particle probability hypothesis density (PHD) filter-based visual trackers have achieved considerable success in the visual tracking field. But position measurements based on detection may not have enough ability to discriminate an object from clutter, and accurate state extraction cannot be obtained in the original PHD filtering framework, especially when targets can appear, disappear, merge, or split at any time. To meet the limitations, the proposed algorithm combines a color histogram of a target and the temporal dynamics in a unifying framework and a Gaussian mixture model clustering method for efficient state extraction is designed. The proposed tracker can improve the accuracy of state estimation in tracking a variable number of objects. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3638121]

Subject terms: probability hypothesis density; motion detection; Gaussian mixture; expectation maximum.

Paper 110772LR received Jul. 4, 2011; accepted for publication Aug. 24, 2011; published online Sep. 20, 2011.

## 1 Introduction

Probability hypothesis density (PHD) filter-based<sup>1</sup> trackers have enjoyed growing popularity in recent years, particularly in the field of nonlinear non-Gaussian multitarget visual tracking. The original PHD filter-based visual tracker usually uses outputs of detectors, such as a motion detector to establish the observation model, whose efficiency relies on the accuracy of the detection.<sup>2</sup> In addition, due to the potential nonlinearity and non-Gaussianity of target models in most visual trackers, a particle PHD filter<sup>3</sup> is used to implement the PHD recursion. However, the intersections of multiple targets like occlusion and clutter often lead to the complex multimodality distribution of the resampled particles, which obviously increase the complexity of state extraction. The classical  $k$ -means clustering algorithm may present serious degradation in state extraction performance.

In this paper, to avoid inaccurate detection generating estimation errors in the original PHD filter-based visual tracker,<sup>2</sup> color histogram with position constraints<sup>4</sup> is incorporated into the PHD filtering framework, which combines the appearance model of the target with its temporal dynamics in a

unifying framework. Moreover, to obtain more accurate state estimates, a new state extraction method based on Gaussian mixture model (GMM) clustering is proposed. Hence, a robust visual tracking framework is obtained.

The multitarget visual tracking problem can be formulated as multitarget Bayes filter in a random finite set (RFS) framework by propagating the multiple-target posterior in time. The particle PHD filter<sup>3</sup> is a sequential Monte Carlo implementation for the multitarget Bayes filter, which approximates the PHD with a set of random samples (weighted particles). The particle PHD filter involves prediction and update steps. Let posterior PHD at time  $k - 1$  be approximated by  $\{w_{k-1}^{(i)}, x_{k-1}^{(i)}\}_{i=1}^{L_{k-1}}$  of  $L_{k-1}$  particles and their corresponding weights. The predicted PHD  $v_{k|k-1}(x_k)$  can be approximated by  $\{\tilde{w}_{k|k-1}^{(i)}, \tilde{x}_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$  after applying importance sampling below

$$v_{k|k-1}(x_k) = \sum_{i=1}^{L_{k-1}+J_k} \tilde{w}_{k|k-1}^{(i)} \delta_{\tilde{x}_k^{(i)}}(x_k), \quad (1)$$

where

$$\tilde{w}_{k|k-1}^{(i)} = \begin{cases} \frac{\phi_{k|k-1}(\tilde{x}_k^{(i)}, x_{k-1}^{(i)}) w_{k-1}^{(i)}}{q_k(\tilde{x}_k^{(i)} | x_{k-1}^{(i)}, Z_k)}, & i = 1, \dots, L_{k-1} \\ \frac{\gamma_k(\tilde{x}_k^{(i)})}{J_k p_k(\tilde{x}_k^{(i)} | Z_k)}, & i = L_{k-1} + 1, \dots, L_{k-1} + J_k \end{cases} \quad (2)$$

Here,  $q_k(\cdot | x_{k-1}^{(i)}, Z_k)$  and  $p_k(\cdot | Z_k)$  are the importance functions for targets at time  $k - 1$  and new targets at time  $k$ ,  $\phi_{k|k-1}(\cdot, \cdot)$  denotes the intensity of survived and spawned targets from time  $k - 1$ , and  $\gamma_k(\cdot)$  is the intensity of new target birth RFS. Once the observation likelihood  $p(z_k | \tilde{x}_k^{(i)})$  is obtained, the weights in Eq. (2) are updated by

$$\tilde{w}_k^{(i)} = \left[ P_M(\tilde{x}^{(i)}) + \sum_{z_k \in Z_k} \frac{P_D(\tilde{x}_k^{(i)}) P(z_k | \tilde{x}_k^{(i)})}{\kappa_k(z) + C_k(z)} \right] \tilde{w}_{k|k-1}^{(i)}, \quad (3)$$

where  $P_M(\tilde{x}^{(i)}) = 1 - P_D(\tilde{x}^{(i)})$  with  $P_D(\tilde{x}^{(i)})$  denoting the detection probability,  $\kappa_k(\cdot)$  is the clutter intensity, and  $C_k(z) = \sum_{j=1}^{L_{k-1}+J_k} P_D(\tilde{x}_k^{(j)}) p_k(z_k | \tilde{x}_k^{(j)}) w_{k|k-1}^{(j)}$ .

## 2 Tracking Model

In the proposed tracker, the target candidate in an image is approximated with a  $w \times h$  rectangle. Let the state of a target at time  $k$  be  $x_k = (p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}, w, h)^T$  with the centroid  $\mathbf{p}_k = (p_{x,k}, p_{y,k})$  and the target speed. Assume that each target follows a linear Gaussian constant velocity model, i.e.,

$$x_k = \mathbf{F}x_{k-1} + v_k, \quad (4)$$

where  $\mathbf{F}$  is the state, transition matrix and  $v_k$  is the zero-mean Gaussian white process noise.

To incorporate the appearance model into the tracking framework, we design the observation model by a color histogram.<sup>4</sup> Let  $\{\mathbf{s}_i\}_{i=1 \dots n_h}$  be the pixel locations of the target centered at  $\mathbf{p}_k = (p_{x,k}, p_{y,k})$ , and the window radius be  $\mathbf{h} = (w, h)$ . Define a function  $\mathbf{b} : R^2 \rightarrow \{1 \dots m\}$  associating the pixel at location  $\mathbf{s}_i$  to the index  $\mathbf{b}(\mathbf{s}_i)$  of the histogram bin corresponding to the color of that pixel. The color

histogram of a target candidate  $\hat{\mathbf{q}}(\mathbf{p}_k)$  and the probability of the feature  $u = 1, \dots, m$  are defined by Eqs. (5) and (6),

$$\hat{\mathbf{q}}(\mathbf{p}_k) = \{\hat{q}^{(u)}(\mathbf{p}_k)\}_{u=1, \dots, m}, \sum_{u=1}^m \hat{q}^{(u)}(\mathbf{p}_k) = 1, \quad (5)$$

$$\hat{q}^{(u)}(\mathbf{p}_k) = C_h \sum_{i=1}^{n_h} k \left( \left\| \frac{\mathbf{p}_k - \mathbf{s}_i}{\mathbf{h}} \right\| \right) \delta[b(\mathbf{s}_i) - u], \quad (6)$$

where  $u$  denotes the color histogram bins,  $k$  is a spatially weighting function and  $C_h$  is a normalization term. Similarly, the reference target model can be represented by  $\hat{\mathbf{q}}_c = \{\hat{q}_c^{(u)}\}_{u=1, \dots, m}$ . Then the observation likelihood is defined by the similarity between a target candidate  $\hat{\mathbf{q}}(\mathbf{p}_k)$  and the reference target model  $\hat{\mathbf{q}}_c$ , i.e.,

$$p(z_k | x_k) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp \left\{ -\frac{d^2(\hat{\mathbf{q}}(\mathbf{p}_k), \hat{\mathbf{q}}_c)}{2\sigma_c^2} \right\}, \quad (7)$$

where  $d(\hat{\mathbf{q}}(\mathbf{p}_k), \hat{\mathbf{q}}_c) = \sqrt{1 - \rho[\hat{\mathbf{q}}(\mathbf{p}_k), \hat{\mathbf{q}}_c]}$  is the similarity computed by Bhattacharyya coefficient  $\rho[\hat{\mathbf{q}}(\mathbf{p}_k), \hat{\mathbf{q}}_c] = \sum_{u=1}^m \hat{q}^{(u)}(\mathbf{p}_k)\hat{q}_c^{(u)}$ , and  $\sigma_c$  is the standard deviation of noise which is determined experimentally.

### 3 GMM Clustering

In the particle PHD filter, a clustering algorithm is required to detect the peaks of the PHD defining candidate states of targets from the resampled particles. We propose a GMM clustering method for state extraction. First, GMM is used to fit the underlying distribution of a resampled particle  $x_k$  as

$$S_k(x_k | \Theta_k) = \sum_{l=1}^{G_k} \pi_k^l \mathcal{N}(x_k | \mu_k^l, \Sigma_k^l) \text{ with } \sum_{l=1}^{G_k} \pi_k^l = 1, \quad (8)$$

where  $G_k$  is the number of Gaussian components, and  $\Theta_k = \{\pi_k^l, \mu_k^l, \Sigma_k^l\}$  with weight, mean, and covariance is the parameter set of a Gaussian item. Assume that the state vectors of all particles are independent, the resulting density for the resampled particles  $\tilde{X}_k = \{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_k}$  is

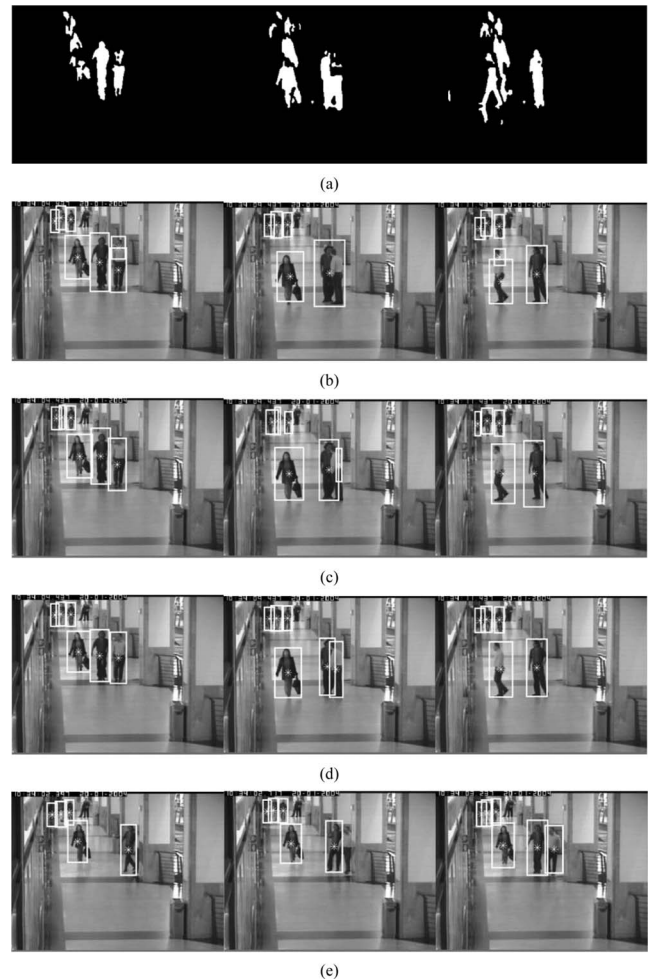
$$S_k(\tilde{X}_k | \Theta_k) = \prod_{i=1}^{L_k} S_k(x_k^{(i)} | \Theta_k) = \prod_{i=1}^{L_k} \sum_{l=1}^{G_k} \pi_k^l \mathcal{N}(x_k^{(i)} | \mu_k^l, \Sigma_k^l). \quad (9)$$

Then the maximum-likelihood estimate of the parameters  $\hat{\Theta}_k = \{\hat{\pi}_k^l, \hat{\mu}_k^l, \hat{\Sigma}_k^l\}_{l=1}^{\hat{L}_k} = \arg \max_{\Theta_k} S_k(\tilde{X}_k | \Theta_k)$  can be computed by the expectation maximum (EM) algorithm, where each Gaussian component  $\hat{\Theta}_k^{(l)} \in \hat{\Theta}_k$  indicates a cluster of particles, the target number estimate is  $\hat{L}_k$  and  $\{\hat{\mu}_k^l\}_{l=1}^{\hat{L}_k}$  are state estimates. Since EM requires a known component number of the mixture density, a component management procedure is proposed to estimate initial clusters for EM from cluster set  $\hat{\Theta}_{k-1}$  at previous time  $k-1$  and new clusters generated by position observations at time  $k$ . The detailed algorithm is presented in Sec. 4.

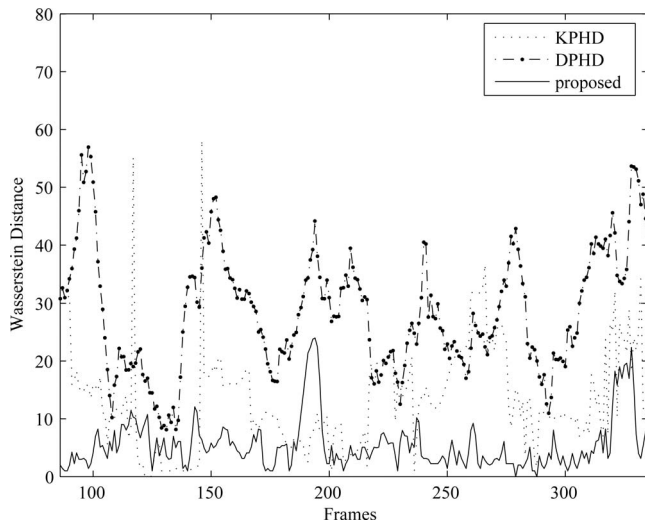
### 4 Particle PHD Filter-Based Visual Tracker with Robust State Extraction

When tracking starts, the target's initial state RFS is input into the proposed algorithm and extract reference models of targets using Eq. (5) at time  $k=0$ . Then the tracking starts from time  $k \geq 1$  as follows.

1. Prediction: according to Eq. (1), draw particles  $\tilde{x}_k$  and compute the predicted weights  $\{\tilde{w}_{k|k-1}^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ .
2. Compute the observation likelihood: for  $i=1, \dots, L_{k-1}+J_k$ , compute  $p(z_k | \tilde{x}_k^{(i)})$  using Eq. (7).
3. Update: update the weights  $\{w_{k|k-1}^{(i)}\}_{i=1}^{L_{k-1}+J_k}$  using  $p(z_k | \tilde{x}_k^{(i)})$  according to Eq. (3).
4. Resampling: resample the updated particles to get  $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_k}$  using multinomial resampling algorithm.
5. GMM clustering: cluster the resampled particles  $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_k}$  using the proposed GMM clustering method.



**Fig. 1** Detecting and tracking results of frames 155, 280, and 330 and an example of failure modality in frames 104, 113, and 125: (a) detection results of background subtraction; (b) tracking by DPHD; (c) tracking by KPHD; (d) the proposed method; (e) an example of failures of the proposed method.



**Fig. 2** Comparison results of Wasserstein distance for DPHD, KPHD, and the proposed method.

Step 1: generate observations (positions)  $Z_k$  of candidate targets by detectors like background subtraction.

Step 2:  $\forall$  each observation  $z \in Z_k$ , associate  $z$  to  $\hat{\Theta}_{k-1}$  by the nearest neighborhood algorithm, discard  $z$  if it can be associated to an old cluster, otherwise add it to  $\theta_k$ . Then  $\forall$  each remained observation  $c \in \theta_k$ , initialize a new cluster  $\{1/\hat{N}_{k|k}, [c, 0, 0], \sum(c)\}$  and add it to new cluster set  $\Theta'$ .

Step 3: augment  $\hat{\Theta}_{k-1}$  with  $\Theta'$ , and update their parameters using EM algorithm on  $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{L_k}$  to obtain  $\tilde{\Theta}_k$ .

Step 4: remove the small clusters in  $\tilde{\Theta}_k$  with  $\tilde{\pi}_k < 0.2$ , where 0.2 is set experimentally and merge similar clusters using pruning method in Ref. 5 to obtain  $\hat{\Theta}_k$ .

6. State output: extract  $\hat{X}_k = \{\hat{u}_{k,i} | \hat{\pi}_{k,i} > 0.5\}_{i=1}^{N_s}$  from  $\hat{\Theta}_k$  where 0.5 is set experimentally.

## 5 Results

The pedestrian sequence from the CAVIAR data set is used as a test video. Figure 1 indicates that PHD filter-based visual trackers can deal with a variable number of targets tracking problems without data association. Figure 1(a) presents the detection results by a background subtraction detector. Figure 1(b) shows the particle PHD filter directly using detection results as measurements (denoted as DPHD) would like to generate some false state estimates due to inaccu-

rate detection such as person detection splitting into several blobs. Figure 1(c) shows the particle PHD filter with observation likelihood based on color histogram and  $K$ -means clustering (denoted as KPHD) can avoid failures due to inaccurate detection but output state estimates without satisfying accuracy. Figure 1(d) demonstrates that more accurate state estimates can be filtered and extracted effectively by our method. Figure 1(e) shows an example of a slower response of the proposed tracker due to color histogram variation of the candidate target region suffering occlusion. Moreover, it can be derived that the appearance variation of targets due to illumination change and occlusion, as well as regions of background with similar color histograms to targets would mislead the tracker using color histograms only. To improve the tracker additional information is needed.

The Wasserstein distance<sup>5</sup> is introduced here to evaluate the performance of trackers. In Fig. 2, the comparison of Wasserstein distance of the three trackers is provided and it demonstrates that our tracker is the best.

## 6 Conclusions and Discussion

In this paper, we have presented a robust multitarget visual tracking framework based on the PHD filter which stabilizes the tracker by incorporating color histograms of targets and their temporal dynamics in a unifying framework and improving the accuracy of state extraction using the proposed GMM clustering method. Experiments show the proposed framework can effectively track a varying number of targets with more accurate state estimates. Possible topics of future work include the incorporation of brightness gradient into the appearance model for more robust observation likelihood and the development of a more efficient particle clustering method.

### Acknowledgments

This paper is jointly supported by the National Natural Science Foundation of China (61074106) and China Aviation Science Foundation (2009ZC57003).

### References

1. R. Mahler, "Multitarget filtering using a multitarget first-order moment statistic," *Proc. SPIE* **4380**, 184–195 (2001).
2. E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," *IEEE Trans. Circuits Syst. Video Technol.* **18**(8), 1016–1027 (2008).
3. B. N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo implementation of the PHD filter for multi-target tracking," in *Proceedings of the 6th International Conference Infor. Fusion*, Queensland, Australia, pp. 792–799 (2003).
4. D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference Computer Vision Patt. Recog.*, Hilton Head Island, SC, USA, pp. 142–149 (2000).
5. B. N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.* **54**(11), 4091–4104 (2006).