Tracking a Variable Number of Human Groups in Video Using Probability Hypothesis Density

Ya-Dong Wang[1,2], Jian-Kang Wu[1], Ashraf A. Kassim[2], and Wei-Min Huang[1]
[1]*Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613*
[2]*Dept. of Electrical & Computer Engineering, National Univ. of Singapore, Singapore 119260*
{stuwy, jiankang, wmhuang}@i2r.a-star.edu.sg, ashraf@nus.edu.sg

Abstract

We apply a multi-target recursive Bayes filter, the Probability Hypothesis Density (PHD) filter, to a visual tracking problem: tracking a variable number of human groups in video. First, we use background subtraction to detect human groups which appear as foreground blobs. The PHD filter is implemented using sequential Monte Carlo methods; and the centroids of the foreground blobs are used as the measurements to update the PHD filter. Our experimental results show that when human groups appear, merge, split, and disappear in the field of view of a camera, our method can track them correctly.

1. Introduction

Tracking multiple targets remains a challenge [17]. Tracking problems are usually modeled as a dynamic system [1] whose order is fixed when there is the fixed number of targets. However, the problem becomes challenging when the number of targets is unknown and variable. The following works are some attempts to meet this challenge. Miller et al. generated the conditional mean estimates of an unknown number of targets and target types via *jump-diffusion* process [13]. Musicki et al. proposed *integrated probabilistic data association* (IPDA) [15] as a recursive formula for both data association and probability of target existence. Khan et al. used the *trans-dimensional Markov chain Monte Carlo* method to track a variable number of interacting targets [10]. Mori and Chong gave a *point process* formalism for multi-target tracking problems [14].

The *finite set statistics* (FISST) proposed by Mahler is the first systematic treatment of multi-sensor multi-target tracking. It contributes to a unified framework of data fusion [6, 11, 12]. The problem of FISST is its combinational complexity when dealing with multiple sensors and multiple targets. To reduce the complexity, Mahler devised the *Probability Hypothesis Density* filter as an approximation of multi-target filter [12]. The PHD filter was implemented using particle filters by Sidenbladh [19] and Vo et al. [21]. The *particle filter* or the *sequential Monte Carlo method* is a Monte Carlo simulation based recursive Bayes filter and can be applied to solve nonlinear and non-Gaussian problems [5, 7]. Sidenbladh used the observation of human observers rather than automatic sensors as their measurements [19] while Vo et al. demonstrated their algorithm by simulations [21]. Tobias and Lanterman [18] applied PHD for radar application while Clark and Bell [3] used PHD in tracking in sonar images. There is no prior work involving the use of PHD for tracking human groups using a camera automatically.

There are some works on tracking multiple targets using particle filters. These works can mainly be divided into two categories: 1) one particle filter with the joint state space for multiple targets [8, 9, 10]; 2) one mixture particle filter, where each component (mode or cluster) is modeled with one individual particle filter that forms part of the mixture [20, 16, 2]. The PHD filter is similar with the second approach but the PHD filter has an important property that the integral of the PHD over a region in a state space is the expected number of targets within this region, which differs from the other multi-target particle filters.

In this article, we combine the object detection with the PHD filter to automatically track an unknown number of human groups in video without human intervention. When human groups appear, merge, split, and disappear in the field of view of a camera, our method can track the number and positions of human groups correctly.

2. Finite set statistics and the PHD filter

Mahler proposed random set theory as a theoretical framework for multi-sensor multi-target data fusion. Under this theory, the state of a target (e.g., position and velocity) is represented by a state vector x; a state set of multiple targets at time t is represented as a *random finite set* (RFS) $X_t = \{x_1, \dots, x_{N(t)}\}$ and $N(t)$ is the variable target number at time t. A measurement of a sensor is represented by a measurement vector y.

The measurement set at time t is also represented as a random finite set $Y_t = \{y_1, \ldots, y_{M(t)}\}$ and $M(t)$ is the variable measurement number at time t. Y^t is used to denote the time sequence $Y_1, \ldots, Y_t$, i.e., the measurement sets accumulated from time 1 to time t.

The 1st moment, or *probability hypothesis density*, of a random finite set is the analogue of the expectation of a random vector [12]. The integral of the PHD over a region in a state space is the expected number of targets within this region. Consequently, the peaks of PHD are points the highest local concentration of expected number of targets and can be used to generate estimates for the states of targets.

Let $D(x_t \mid Y^t)$ denote the probability hypothesis density associated with the multi-target posterior $p(X_t \mid Y^t)$ at time t. The PHD filter consists of two steps: prediction and update. The PHD prediction equation is:

$$D(x_{t+1} \mid Y^t) = b(x_{t+1}) + \int (p_S(x_t) p(x_{t+1} \mid x_t) + b(x_{t+1} \mid x_t)) D(x_t \mid Y^t) dx_t \quad (1)$$

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

$$D(x_{t+1} \mid Y^{t+1}) \cong F(Y_{t+1} \mid x_{t+1}) D(x_{t+1} \mid Y^t) \quad (2)$$

$$F(Y_{t+1} \mid x_{t+1}) = 1 - p_D(x_{t+1}) + \sum_{y \in Y_{t+1}} \frac{p_D(x_{t+1}) p(y \mid x_{t+1})}{\lambda c(y) + D[p_D(x_{t+1}) p(y \mid x_{t+1})]}$$

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

3. Detecting the foreground human groups in video

We detect human groups in video using the adaptive background subtraction method [4] because the camera is fixed and the background image is easily obtained. Let $P_t(a,b)$ and $B_t(a,b)$ represent pixel intensity and the background intensity at position (a, b) at time t, respectively. Pixel (a, b) belongs to the foreground region if:

$$| P_t(a,b) - B_t(a,b) | > Threshold \quad (3)$$

A pixel-based update method periodically updates the background model. When the pixel (a, b) belongs to the background region, its background value is updated as follows:

$$B_{t+1}(a,b) = (1-\gamma) B_t(a,b) + \gamma P_t(a,b) \quad (4)$$

where γ is the learning rate for the background model. When the pixel (a, b) belongs to the foreground region, its background value is not updated. The foreground image obtained using background subtraction is usually noisy and morphological operations are performed to remove noise. Erosion and dilation are applied to the binary foreground images. The resulting foreground objects are our detected objects.

4. Tracking a variable number of human groups using a particle filter based PHD filter

In this work, the state of each target x_t at time t is its position in frame t. We assume that the motion model of each target is a constant velocity model, i.e.,

$$x_{t+1} = x_t + v_t + u_t \quad (5)$$

where v_t is the target's velocity and can be approximated by $v_t = x_t - x_{t-1}$, and u_t is a zero-mean Gaussian white process noise. This model is used as our proposal density of the particle filter for the existing targets. For the spontaneous birth targets, the proposal density is a uniform distribution on the whole image region, i.e.,

$$b(x_{t+1}) \sim U[1, width] \times U[1, height] \quad (6)$$

where *width* and *height* are the size of the image and $U[c, d]$ is a uniform distribution function on the interval [c, d]. For each target, we use the centroid of its foreground blob as our measurement y to update the PHD filter. The likelihood function is:

$$p(y \mid x_t) = \frac{1}{2\pi |M|^{1/2}} \exp[-\frac{1}{2}(y - x_t)^T M^{-1} (y - x_t)] \quad (7)$$

where M is the covariance matrix of the measurement noise. The PHD filter is implemented using the particle filter proposed by Vo et al. [21]. Let L_t denote the particle number at time t, J_t denote the new particle number for the spontaneous birth targets at time t, and w denote a particle's weight. The basic algorithm is summarized as follows:

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

1. *Detection*

Detecting the foreground objects using background subtraction described in Section 3. The centroids of all foreground blobs are the measurement set Y_{t+1} at time t+1.

2. *Prediction*

For $i = 1, \ldots, L_t$, generate a sample $\widetilde{x}_{t+1}^{(i)}$ according to (5) and compute the predicted weights

$$\hat{w}_{t+1}^{(i)} = w_t^{(i)}$$

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

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

3. *Update*

For each $y \in Y_{t+1}$, use the likelihood (7) and compute

$$C_{t+1}(y) = \sum_{i=1}^{L_t + J_{t+1}} p_D(x_{t+1}^{(i)}) p(y \mid x_{t+1}^{(i)}) \hat{w}_{t+1}^{(i)}$$

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

$$\widetilde{w}_{t+1}^{(i)} = \left[1 - p_D(\widetilde{x}_{t+1}^{(i)}) + \sum_{y \in Y_{t+1}} \frac{p_D(x_{t+1}^{(i)}) p(y \mid x_{t+1}^{(i)})}{\lambda c(y) + C_{t+1}(y)} \right] \hat{w}_{t+1}^{(i)}$$

4. *Resampling*

Compute the target number at time t+1

$$\hat{N}_{t+1} = \sum_{i=1}^{L_t + J_{t+1}} \widetilde{w}_{t+1}^{(i)}$$

Initialize the cumulative probability $c_1 = 0$, $c_i = c_{i-1} + \dfrac{\widetilde{w}_{t+1}^{(i)}}{\hat{N}_{t+1}}$, $i = 2, \ldots, L_t + J_{t+1}$.

Draw a starting point $u_1 \sim U[0, L_{t+1}^{-1}]$.

For $j = 1, \ldots, L_{t+1}$,

$$u_j = u_1 + L_{t+1}^{-1}(j-1)$$

While $u_j > c_i$, i=i+1. End while.

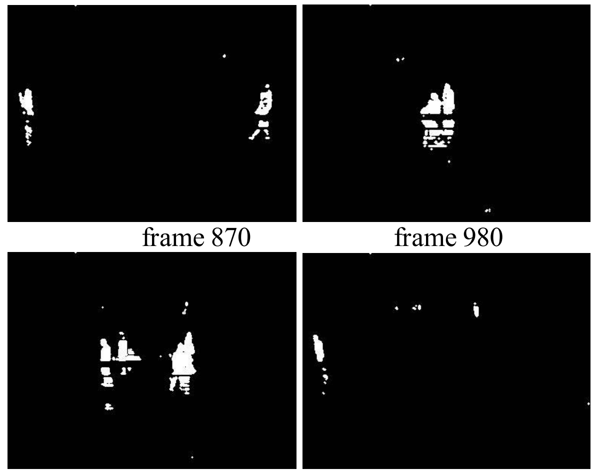
$$x_{t+1}^{(j)} = \widetilde{x}_{t+1}^{(j)}$$

$$w_{t+1}^{(j)} = L_{t+1}^{-1}$$

Rescale (multiply) the weights by $\hat{N}_{t+1}$ to get $\{x_{t+1}^{(i)}, \dfrac{\hat{N}_{t+1}}{L_{t+1}}\}_{i=1}^{L_{t+1}}$.

5. Experimental results

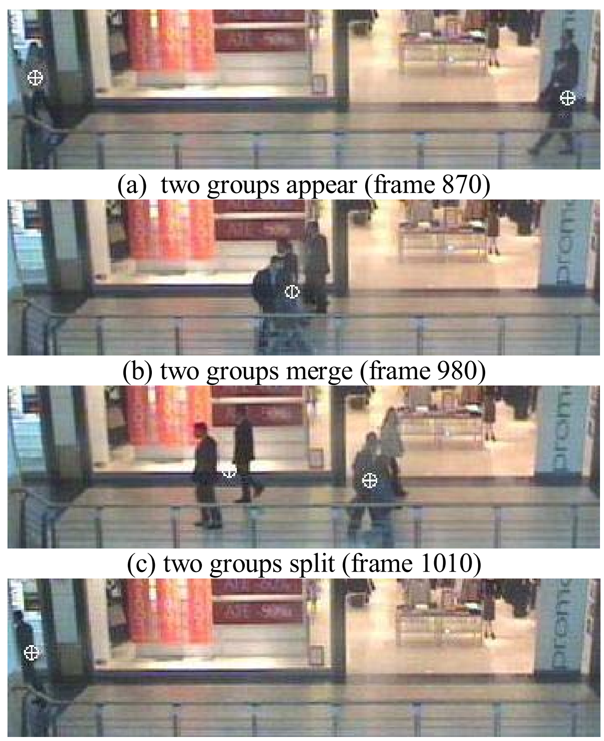
We tested our method using the dataset of the EC Funded CAVIAR project [22]. Video "OneStopMoveEnter1front" has 1589 frames. There were two human groups appearing, merging, splitting, or disappearing in the field of view of the camera. The detection results using background subtraction are shown in Fig. 1.



frame 870 frame 980

frame 1010 frame 1110

Figure 1. Our detection results



(a) two groups appear (frame 870)

(b) two groups merge (frame 980)

(c) two groups split (frame 1010)

(d) one group disappears (frame 1110)

Figure 2. The two human groups appear, merge, split, and disappear in the filed of view of the camera. The white circles are the centroids of human groups.

Fig. 2 shows 4 video frames with white circles indicating the tracking results: the centroids of human groups. Because the PHD filter explicitly models the processes of birth, surviving, death of targets and false alarms of clutter, as shown by our experimental results, our algorithm is able to track the variable number of human groups.

Figure 3 provides the tracking results for the first 1000 frames of Video "OneStopMoveEnter1front". The correct frame number is 744 for these 1000 frames. The errors mainly come from two factors: 1) the inaccuracy of measurements; 2) the importance sampling for birth targets (6) does not generate samples near the birth target's position.

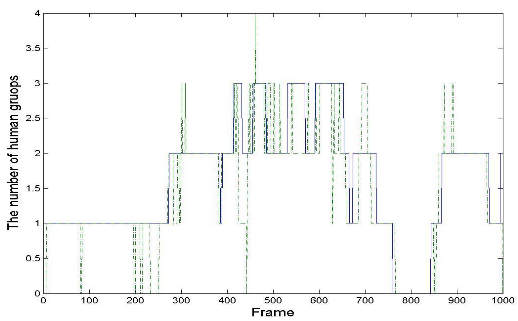


Figure 3. The solid line is the ground truth number of human groups. The dashed line is the tracking result of the PHD filter.

6. Conclusions

In this paper, we apply the probability hypothesis density filter to a visual tracking problem. Foreground objects are detected using the background subtraction method, and a variable number of human groups are tracked using a particle filter based implementation of the PHD filter. The results demonstrate that our method can track a variable number of human groups in video.

Acknowledge

The authors thank Dr. Ronald P.S. Mahler for providing his literatures and Dr. Ba-Ngu Vo for providing the programs of the PHD filter.

References

[1] Y. Bar-Shalom and W. Dale Blair (editors), *Multitarget-Multisensor Tracking: Applications and Advances, vol. III*, Artech House, 2000.

[2] C. Chang, R. Ansari, and A. Khokhar, "Multiple object tracking with kernel particle filter", *in Proc. CVPR*, 2005, pp. 566-573.

[3] D.E. Clark and J. Bell, "Bayesian multiple target tracking in forward scan sonar images using the PHD filter", *IEE Proc. Radar, Sonar, Navig.*, 152(5), 2005, pp. 327-334.

[4] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson, "A System for Video Surveillance and Monitoring: VSAM Final Report", *Technical report CMU-RI-TR-00-12*, Carnegie Mellon University, May, 2000.

[5] A. Doucet, N. de Freitas, and N. J. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, 2001.

[6] I. Goodman, R.P.S. Mahler, and H. Nguyen, *Mathematics of Data Fusion*, Kluwer Academic Publishers, Dordrecht, 1997.

[7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proc. F*, 140(2), 1993, pp.107-113.

[8] C. Hue, J-P. Le Cadre, and P. Perez, "Tracking multiple objects with particle filtering", *IEEE Trans. AES*, 38(3), 2002, pp. 791-812.

[9] M. Isard and J. MacCormick, "Bramble: a Bayesian multiple-blob tracker", *in Proc. ICCV*, 2001, pp. 34-41.

[10] Z. Khan, T. Balch, F. Dellaert, "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets", *IEEE Trans. PAMI*, 27(11), 2005, pp. 1805-1819.

[11] R.P.S. Mahler, An Introduction to Multisource-Multitarget Statistics and Its Applications, Technical Monograph, Lockheed Martin, Regan MN, 2000.

[12] R.P.S. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments", *IEEE Trans. AES*, 39(4), 2003, pp. 1152-1178.

[13] M.I. Miller, A. Srivastava, and U. Grenander, "Conditional-Mean Estimation Via Jump-Diffusion Processes in Multiple Target Tracking/Recognition", *IEEE Trans. SP*, 43(11), 1995, pp. 2678-2690.

[14] S. Mori, and C.-Y. Chong, "Point Process Formalism for Multiple target Tracking", *in Proc. FUSION*, 2003, pp. 10-17.

[15] D. Musicki, R. Evans, and S. Stankovic, "Integrated Probabilistic Data Association", *IEEE Trans. AC*, 39(6), 1994, pp. 1237-1241.

[16] K. Okuma, A. Taleghani, N. de Freitas, J. Little and D. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking**",** *in Proc. ECCV*, 2004, pp. 28-39.

[17] G.W. Pulfold, "Taxonomy of multiple target tracking methods", *IEE Proc. Radar, Sonar, Navig.*, 152(5), 2005, pp. 291-304.

[18] M Tobias and A.D. Lanterman, "Probability hypothesis density-based multitarget tracking with bistatic range and Doppler", *IEE Proc. Radar, Sonar, Navig.*, 152(3), 2005, pp. 195-205.

[19] H. Sidenbladh, "Multi-Target Particle Filtering for the Probability Hypothesis Density", *in Proc. FUSION*, Cairns, Australia, 2003, pp. 800-806.

[20] J. Vermaak, A. Doucet, and P. Perez, "Maintaining Multi-Modality through Mixture Tracking", *in Proc. ICCV*, 2003, pp. 1110-1116.

[21] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo Methods for Multi-Target Filtering with Random Finite Sets", *IEEE Trans. AES*, 41(4), 2005, pp. 1224- 1245.

[22] http://groups.inf.ed.ac.uk/vision/caviar/caviardata1/.