

TRACKING MULTIPLE OBJECTS USING PROBABILITY HYPOTHESIS DENSITY FILTER AND COLOR MEASUREMENTS

Nam Trung Pham, Weimin Huang

Institute for Infocomm Research
Singapore

S. H. Ong

Department of ECE
National University of Singapore
Singapore

ABSTRACT

Most methods for multiple object tracking in video represent the state of multi-objects in a high dimensional joint state space. This leads to high computational complexity. This paper presents a method using the probability hypothesis density (PHD) filter to estimate the state of multiple objects in video. The method operates on the single object state space instead of the joint state space. A PHD recursion for visual observations with color measurements is proposed. Our method can track varying number of objects.

1. INTRODUCTION

Tracking moving objects in the video sequence is important in many applications, e.g. tracking players in sport sequences [1], [2], surveillance [3], and many more. Challenges in multiple object tracking are varying number of objects, occlusions.

There are many methods for multiple object tracking from video. Among them, methods that are based on Bayesian filter framework with sequential Monte Carlo implementation are attracting substantial interest. Some of these methods operate on the single object state space [1], [2]. These methods approximated the mixture filtering distribution by particle filtering to maintain multi-modality. However, a common limitation of these methods is that if objects are close to each other and particles from one specified object have very high weight, the particles representing the remaining objects are often suppressed. In addition, there are methods using a joint state space for tracking [4], [5]. However, sampling particles from a joint state space can become inefficient as the dimension of the space increases. Although there are some attempts to reduce the number of particles such as [3], it is still computational demanding.

Recently, there has been increasing research interest on using random set theory to solve the multiple object tracking problem. Here, the states of objects and measurements are represented as random finite sets (RFS). Mahler [6] presented a probability hypothesis density filter (PHD) that operates on

the single target state space. Vo [7], [8] proposed implementations of the PHD filter.

In this paper, we propose a method for tracking multiple objects from video data using probability hypothesis density filter on color measurements. A method to obtain the PHD with color histogram measurements is presented. We assumed that we had color histogram models of objects under tracking which can be obtained from detection methods. Then, the proposed tracking can be efficiently applied for tracking varying number of objects. The proposed method can be used for the analysis of different type of video, such as sport video, home video and surveillance video.

2. PROBLEM FORMULATION IN RANDOM SET

2.1. Random finite set formulation

By assuming that the object state does not change so much between frames, each object in multi-object tracking is evolved from a dynamic moving equation as follows

$$x_k = x_{k-1} + w_k \quad (1)$$

where the state $x_k = \{x_c, y_c, H_x, H_y\}$ is a rectangle with center $\{x_c, y_c\}$ and size $\{H_x, H_y\}$, w_k is a process noise. Let the multi-object state be $X_k = \{x_{1,k}, x_{2,k}, \dots, x_{N_k,k}\} \in \mathcal{F}(\mathcal{X})$, where $\mathcal{F}(\mathcal{X})$ denotes the collection of all finite subsets of the single object state space \mathcal{X} . Let Z_k be the image frame at time k . Color tracking is to track objects described by specified color representation q^* e.g. histograms. However, under the RFS framework, it is difficult to represent color histogram as a RFS directly. In the section 3.1, we propose a way to obtain the RFS from color histogram.

The multiple object tracking problem is to find multiple object state \hat{X}_k that have color histograms similar to q^* from the posterior density function $p_k(X_k|Z_{1:k})$.

2.2. Probability hypothesis density approach

It is not easy to obtain the posterior density function $p_k(X_k|Z_{1:k})$ when the state space is too large. Fortunately, it can be ap-

proximately recovered from the first moment of this distribution, the probability hypothesis density (PHD). PHD is defined as follows. For a RFS X on \mathcal{X} with probability distribution P , the PHD is the density $v(x)$ such that for each region $S \subseteq \mathcal{X}$, the integral of v over region S gives the expected number of elements of X that are in S ,

$$\int |X \cap S| P(dX) = \int_S v(x) dx, \quad (2)$$

Thus, we can estimate the states of objects by investigating peaks of PHD.

3. PHD FILTER FOR COLOR OBJECT TRACKING

3.1. Color measurement random set

In this section, we propose a method to obtain the color measurement random set. From [6], we have

$$v_k(x) = \frac{\partial G_k}{\partial x} [1] \quad (3)$$

where $v_k(x)$ is a PHD, G_k is the generating function of posterior density at time k and the derivative of G_k is

$$\frac{\partial G_k}{\partial x} [h] = \frac{\partial}{\partial x} \left(\int h^X p_k(X|Z_{1:k}) \delta X \right) \quad (4)$$

where $h^X = h(x_1) \dots h(x_{|X|})$ and h is some function. Now, we propose a density $\tilde{v}_k(x)$ that is a proportion to $v_k(x)$. So, peaks in $\tilde{v}_k(x)$ are also the peaks in $v_k(x)$. Then, we take these peaks to obtain the color measurement random set.

From Bayesian formula, we have

$$p_k(X|Z_{1:k}) = K^{-1} p_k(Z_k|X) p_{k|k-1}(X|Z_{1:k-1}) \quad (5)$$

where $K^{-1} = \int p_k(Z_k|X) p_{k|k-1}(X|Z_{1:k-1}) \delta X$. So,

$$\frac{\partial G_k}{\partial x} [h] = \frac{1}{K} \frac{\partial}{\partial x} \left(\int h^X p_k(Z_k|X) p_{k|k-1}(X|Z_{1:k-1}) \delta X \right) \quad (6)$$

We follow the assumption in [5]

$$p_k(Z_k|X) \propto \prod_{i=1}^{|X|} f(z_k|x_i) \quad (7)$$

where $f(z_k|x_i)$ is color likelihood for the i th object. Let $l_z(x) = f(z|x)$

$$p_k(Z_k|X) \propto \prod_{i=1}^{|X|} l_z(x_i) \quad (8)$$

Replace formula (8) to (6),

$$\begin{aligned} \frac{\partial G_k}{\partial x} [h] &\propto \frac{1}{K} \frac{\partial}{\partial x} \left(\int (hl_z)^X p_{k|k-1}(X|Z_{1:k-1}) \delta X \right) \\ &\propto \frac{1}{K} \frac{\partial}{\partial x} (G_{k|k-1}[hl_z]) \end{aligned} \quad (9)$$

If assume $p_{k|k-1}(X|Z_{1:k-1})$ is Poisson [6], the generating function $G_{k|k-1}[h]$ has the form

$$G_{k|k-1}[h] = \exp\{\mu s[h] - \mu\} \quad (10)$$

where $s[h] = \int s(x) h(x) dx$, μ is the mean. We have

$$G_{k|k-1}[hl_z] = \exp\{\mu s[hl_z] - \mu\} \quad (11)$$

So,

$$\frac{\partial G_{k|k-1}}{\partial x} [hl_z] = G_{k|k-1}[hl_z] (\mu h(x) l_z(x) s(x)) \quad (12)$$

Let $h = 1$, from formulas (12), (11), (9), (3), we have

$$v_k(x) \propto \frac{1}{K} e^{\mu s[l_z] - \mu} (\mu l_z(x) s(x)) \quad (13)$$

However,

$$v_{k|k-1}(x) = \frac{\partial G_{k|k-1}}{\partial x} [1] = \mu s(x) \quad (14)$$

So, we can conclude

$$v_k(x) \propto \tilde{v}_k(x) = l_z(x) v_{k|k-1}(x) \quad (15)$$

From the joint state space, we found a function $\tilde{v}_k(x)$ in single target state space that is a proportion to true PHD $v_k(x)$

3.2. PHD filter for color multi-object tracking

We cannot apply directly GMPHD filter with color measurements because obtaining the measurement random set from video is challenging. Here, we proposed a PHD recursion for visual observations with color measurements. At the beginning, color histogram models of objects are learnt from template images. These models are used when we evaluate the color likelihood to find the color measurement random set in section 3.1. With the color measurement random set, we employed GMPHD filter [8] to obtain the estimation of PHD. Like GMPHD filter, we have some assumptions. Each target follows a linear Gaussian model, i.e.,

$$f_{k|k-1}(x|x_{k-1}) = N(x; F_{k-1}x_{k-1}, Q_{k-1}), \quad (16)$$

$$g_k(z|x) = N(z; H_k x, R_k), \quad (17)$$

where $F_{k-1} = [I]$, $H_k = [I]$, Q_{k-1} is the process noise covariance, and R_k is the observation noise covariance. The survival (S) and detection (D) probabilities are $p_{S,k}$ and $p_{D,k}$ respectively. The intensity of the spontaneous birth RFS $\gamma_k(x)$ is a Gaussian mixture. The posterior intensity at time $k-1$, $v_{k-1}(x)$ is also a Gaussian mixture of the form

$$v_{k-1}(x) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} N(x; m_{k-1}^{(i)}, P_{k-1}^{(i)}) \quad (18)$$

where J_{k-1} is the number of Gaussian components of $v_{k-1}(x)$. The algorithm is described as follows:

- Step 1. Prediction

The predicted intensity to time k is given by

$$v_{k|k-1}(x) = v_{S,k|k-1}(x) + \gamma_k(x) \quad (19)$$

where

$$\begin{aligned} v_{S,k|k-1}(x) &= p_{S,k} \sum_{j=1}^{J_{k-1}} w_{k-1}^{(j)} N(x; m_{S,k|k-1}^{(j)}, P_{S,k|k-1}^{(j)}), \\ m_{S,k|k-1}^{(j)} &= F_{k-1} m_{k-1}^{(j)}, \\ P_{S,k|k-1}^{(j)} &= Q_{k-1} + F_{k-1} P_{k-1}^{(j)} F_{k-1}^T. \end{aligned}$$

$v_{S,k|k-1}(x)$ and $\gamma_k(x)$ are Gaussian mixtures, so $v_{k|k-1}(x)$ can be expressed as a Gaussian mixture of the form

$$v_{k|k-1}(x) = \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{(i)} N(x; m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)}) \quad (20)$$

We propose a way to integrate Monte Carlo technique with GMPHD filter for video tracking to obtain the color measurement random set. From $v_{k|k-1}$, we obtain \tilde{v}_k by the Monte Carlo technique. First, we draw samples $\{x^i, \psi^i\}_{i=1}^N$ from $v_{k|k-1}$, where ψ^i is the weight for sample x^i . Then, \tilde{v}_k will be represented by $\{x^i, \xi^i\}$, where

$$\xi^i = l(x^i) \psi^i \quad (21)$$

Next, $\{x^i, \xi^i\}$ are re-sampled and clustered. Center of these clusters will form the measurements for next update step. That means

$$Z_k = \{z_1, \dots, z_m\} \quad (22)$$

where z_i is the center of i -th cluster.

- Step 2. Update

The posterior intensity at time k , $v_k(x)$ is also a Gaussian mixture that is obtained from the predicted intensity $v_{k|k-1}(x)$ and the color measurement random set Z_k by using the updating step in GMPHD filter (see details in [8]).

From PHD $v_k(x)$, we find Gaussian components whose weights are larger than a threshold (0.5). The set of means of these Gaussian components are state estimates. We have noticed that we associated each Gaussian with each label. These labels are also the object identifications. Gaussian components that are near each other and having the same label are merged after the updating step. Another notice is that if a new object appears, the peak from this object will be in the color measurement random set with assuming that we knew the color model of this object before. Then, after the updating step, the estimation of this new object is obtained. In a similar manner, when an object disappears, the color measurement from this object is not in the color measurement random set and the weight of Gaussian component represented for this

object is very small. Hence, this Gaussian component is removed. Moreover, when objects are occluded, peaks can also be detected and used to update Gaussian components. Hence, this method can estimate states of objects when occlusions occur.

4. EXPERIMENTAL RESULTS

We tested our method in many sequences from [3], [5] and [9]. There are total about 9500 frames. The errors of estimations that are measured by Wasserstein distance [7] are shown in table 1

Table 1. Error of estimation

Sequences	Mean error (pixel)
seq24-2p-0111-cam1	7.2
seq24-2p-0111-cam2	4.8
seq35-2p-1111-cam1	4.8
seq35-2p-1111-cam2	3.9
seq44-3p-1111-cam1	8.2
seq44-3p-1111-cam2	6.1
football	7.2
seq16	9

Figure 1 shows the comparison between our method and boosted particle filter [2] that we implemented. For boosted particle filter, we assumed that we have very good detections (from ground truth) and the proposal coefficient $\alpha = 0.8$ (80% particles from detection distribution which means majority of particles are around the real state). However, because the likelihood when particles near the black person are too high, boosted particle filter was ambiguous between two persons and lost tracking. The results showed that our method can maintain the tracking through the occlusion. This is because we can detect the peak caused by the white person. If detected peaks are not caused by persons, these peaks will be false alarms. GMPHD filter can handle these false alarms. Otherwise, if peaks are caused by persons, state estimates of these persons will be obtained. Hence, in this case, our method is better than boosted particle filter.

Figure 2 shows the results of tracking white football players. In this sequence, the number of white players changes during the tracking period. The camera is moving when capturing. Hence, segmentation methods are difficult to apply. In this sequence, we used 400 particles. It is less than 5000 joint state space particles in [5].

Figure 3 shows some results of the tracking in the seq16 from [3]. In this sequence, at the beginning, there is no one in the scene. At frame 34, 78, 135 and 141, the first, second, third and fourth person enter the tracking area, respectively. They walk in two opposite direction and occlusions may occur. The results showed that our method can track varying number of people in this sequence.

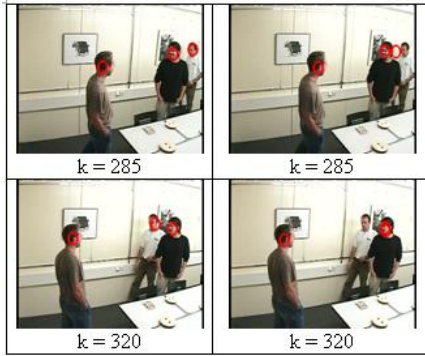


Fig. 1. Comparison of our method (left) and boosted particle filter (right)

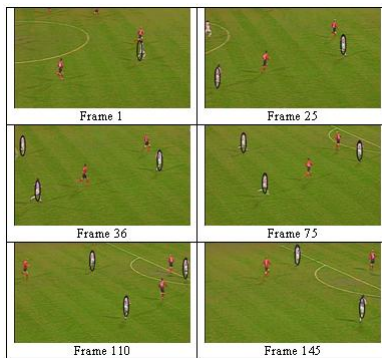


Fig. 2. Tracking multiple players in football sequence

5. CONCLUSION

The paper described a method of using GMPHD filter to track multiple objects by incorporating the color representation. We showed that the PHD is proportional to our approximated density from color likelihood, which helps to define the color measurement random set. A PHD recursion for visual observations with color measurements is proposed. With this approach, the experiments showed that the video tracking works for varying number of targets in single target state space.

6. ACKNOWLEDGMENT

The authors would like to specially thank Prof Ba Ngu Vo for his helps and fruitful discussions, and thank Dr J. Czyz, Dr K. Smith for their data. This work is partially supported by EU project ASTRALS (FP6-IST-0028097).

7. REFERENCES

[1] J. Vermaak, A. Doucet, and P. Perez, "Maintaining multimodality through mixture tracking," in *ICCV*, France, May 2003.

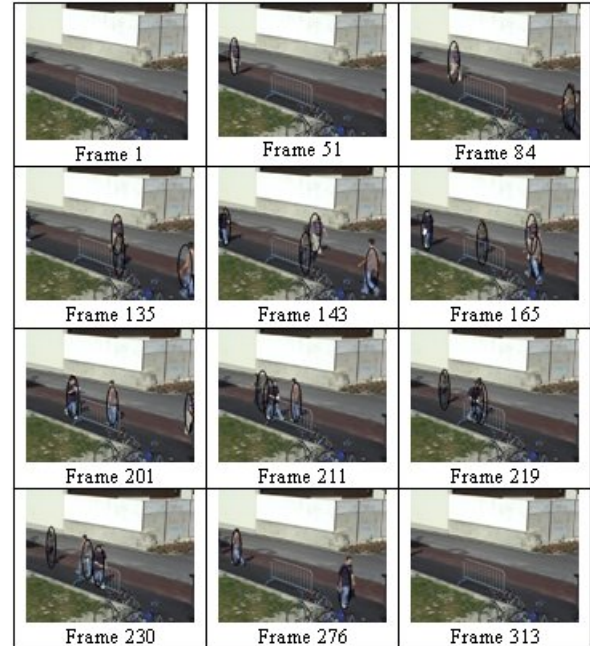


Fig. 3. Tracking multiple persons in second sequence

- [2] K. Okuma, A. Taleghani, N. de Freitas, J.J. Little, and D.G. Lowe, "A boosted particle filter: multitarget detection and tracking," in *ECCV*, Prague, Czech Republic, May 2004.
- [3] K. Smith, D. Perez, and J. M. Odobez, "Using particles to track varying numbers of interacting people," in *CVPR*, San Diego, June 2005.
- [4] M. Isard and J. MacCormick, "Bramble: A bayesian multiple-blob tracker," in *ICCV*, Vancouver, Canada, 2001.
- [5] J. Czyz, B. Ristic, and B. Macq, "A color-based particle filter for joint detection and tracking of multiple objects," in *ICASSP*, 2005.
- [6] R. Mahler, "Multi-target bayes filtering via first-order multi-target moments," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.
- [7] B. N. Vo, S. Singh, and A. Doucet, "Sequential monte carlo methods for bayesian multi-target filtering with random finite sets," *IEEE Trans. Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1224–1245, Oct 2005.
- [8] B. N. Vo and W. K. Ma, "The gaussian mixture probability hypothesis density filter," *IEEE Transaction Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.
- [9] G. Lathoud, J. Odobez, and D. Perez, "Av16.3: an audio-visual corpus for speaker localization and tracking," in *Proceedings of the MLMI'04 Workshop*, 2004.