Abstract – This paper presents our work which involves the application of a recursive Bayesian filter, the Gaussian mixture probability hypothesis density (GMPHD) filter, to a visual tracking problem. Foreground objects are detected using statistical background modeling to obtain measurements which are input into the filter. The GMPHD filter explicitly models the birth, survival and death of objects by managing the number of Gaussian components and jointly estimates the time-varying number of objects and their states. A scene-driven method is proposed to initialize the GMPHD filter and model the birth of new objects. The results show when a person or a group appeared, merged, split, and disappeared in the field of view, the GMPHD filter can track the number and positions at the most time. The scene-driven GMPHD filter can track the birth of new objects faster than the particle PHD filter.

Keywords: People tracking, Bayesian filtering, probability hypothesis density, Gaussian mixture

1 Introduction

Tracking an unknown and variable number of objects remains a challenge. Mahler proposed finite set statistics (FISST) [1, 2] as the first systematic treatment of multisensor multitarget tracking problems and presented probability hypothesis density (PHD) [3] as a first-order approximation of the random finite set of objects. The PHD filter can jointly estimate the time-varying number of objects and their states from a sequence of measurement sets. The PHD filter has been implemented by particle filter (Sidenbladh [4] and Vo et al. [5]) and Gaussian mixture (Vo and Ma [6]).

There have been some applications of the particle filter based FISST and PHD. Sidenbladh [7] tracked vehicles in terrain using the FISST particle filtering. Tobias et al. [8] applied the particle PHD filter for radar tracking problem. Clark et al. [9] used the particle PHD filter in tracking in sonar images. Ikoma et al. filtered trajectories of feature points in images using the particle PHD filter [10]. Haworth et al. presented a system to detect and track metallic objects concealed on people in sequences of millimetre-wave images [11]. Our previous work applied the particle PHD filter to track a variable number of human groups in video [12].

In contrast to the particle PHD filter, there are few applications of the Gaussian mixture PHD filter. Clark et al. developed the GMPHD multitarget tracker [13] and demonstrated it on forward-looking sonar data [14].

In this paper, we combine object detection with the GMPHD filter to automatically track an unknown number of people or groups in image sequences without human intervention. The procedure is outlined in Fig. 1.

![Image data detection GMPHD filtering](image)

Fig. 1. GMPHD visual tracking implementation

Our results show that this method could track the variable number of people or groups and their positions when people or groups appeared, merged, split, and disappeared in the field of view of a camera. We also propose a scene-driven method to initialize the GMPHD filter and to model the appearance/birth of new objects. The scene-driven method can track the birth of new objects faster than the particle PHD filter.

2 Detection

Detection methods for visual tracking include background subtraction with a mixture of Gaussian as background model [15] and statistical background modeling [16]. In our work, we use the statistical background modeling which incorporates spectral, spatial, and temporal features to characterize the background appearance. The experimental dataset used in our work is from the European Commission funded CAVIAR project [17]. Some detection results for video OneStopMoveEnter1front are shown in Fig. 2.

The foreground blobs with their area under a specific threshold are removed to reduce noise. The centroids of remaining foreground blobs are chosen as the measurements and are input to the following GMPHD filter.
3 Tracking model

The linear Gaussian dynamic model with the constant velocity [pp. 273, 18] is used:
\[ x_{t+1} = F_t x_t + \Gamma_t u_t \]
where the state of a target \( x_t \) consists of its position and velocity
\[ x_t = [x_t, \dot{x}_t, y_t, \dot{y}_t]^T \]
and the system noise \( u_t = (u_{t,1}, u_{t,2})^T \) is mutually independent zero-mean Gaussian white noise with covariance \( Q_t = \text{diag}(\sigma_u^2, \sigma_u^2) \). Only position measurements are available and the measurement model is
\[ y_t = H_t x_t + v_t \]
the measurement matrix is
\[ H_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \]
the measurement noise the measurement noise \( v_t = (v_{t,1}, v_{t,2})^T \) is mutually independent zero-mean Gaussian white noise with covariance \( R_t = \text{diag}(\sigma_v^2, \sigma_v^2) \).

4 Probability hypothesis density

Mahler proposed random set theory as a theoretical framework for multisensor multitarget 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) \( X_t = \{x_1, \ldots, x_{N(t)}\} \) 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 \( Y_t, \ldots, Y_{t+1} \), 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. 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 \( D_{\theta}(x | Y) \) denote the probability hypothesis density associated with the multitarget posterior \( p(X_t | Y^t) \) at time \( t \). The PHD filter consists of two steps: prediction and update. The PHD prediction equation is:
\[ D_{\theta}(x | Y) = b(x_t) + \int (p_3(x_t)p(x_{t+1} | x_t) \]
\[ + b(x_{t+1} | x_t) D_{\theta}(x | Y) dx \]
where \( b(x_{t+1}) \) denotes the intensity function of the spontaneous birth RFS, \( b(x_{t+1} | x_t) \) denotes the intensity function of the RFS of targets spawned from the previous state \( x_t \), \( p_3(x_t) \) is the probability that the target still exists at time \( t+1 \) given it has previous state \( x_t \), and \( p(x_{t+1} | x_t) \) is the transition probability density of individual targets. The PHD update equation is:
\[ D_{\theta}(x | Y) \equiv F(Y_{t+1} | x_{t+1})D_{\theta}(x | Y) \]
\[ F(Y_{t+1} | x_{t+1}) = 1 - p_3(x_{t+1}) \]
\[ + \sum_{y_{t+1} \in Y_{t+1}} p_2(x_{t+1}) p(y_{t+1} | x_{t+1}) \]
where \( p_2(x_{t+1}) \) is the probability of detection, \( p(y_{t+1} | x_{t+1}) \) is the likelihood of individual target, \( \lambda \) is the average number of clutter points per scan, \( c(y) \) is the probability distribution of each clutter point, and \( D[h] = \int h(x)D_{\theta}(x | Y) dx \).

5 The Gaussian mixture probability hypothesis density filter

The GMPHD filter is initialized in Step 0 and iterates through Step 1 to 5.
Step 0: Initialization
Initialize the algorithm with the weighted sum of $J_0$ Gaussians,
\[
D_{q_0} = \sum_{i=1}^{J_0} w_0^{(i)} N(x; m_0^{(i)}, P_0^{(i)})
\]
where $N(x;m,P)$ is a Gaussian distribution with the mean $m$ and the variance $P$. The sum of weights,
\[
\sum_{i=1}^{J_0} w_0^{(i)} = \hat{T}_0
\]
is the expected number of objects at the beginning.

Step 1: Prediction
The prediction density at time $t+1$ is
\[
D_{x_{t+1}}(x) = b_{t+1}(x) + D_{x_{t}}(x)
\]
The intensity of the new-born objects is
\[
b_{t+1}(x) = \sum_{i=1}^{J_t} w_{t+1}^{(i)} N(x; m_{t+1}^{(i)}, P_{t+1}^{(i)})
\]
The intensity of the surviving objects is
\[
D_{x_{t+1}}(x) = P_{t+1}^{(i)} = F_t m_t^{(i)}
\]

Step 2: Update
When the measurements $Y_{t+1} = \{y_{t+1,1}, \ldots, y_{t+1,J_{t+1}}\}$ at time $t+1$ are available, the posterior intensity is computed as follows:
\[
D_{x_{t+1}}(x) = (1 - p_d) D_{x_{t}}(x) + \sum_{j=1}^{J_{t+1}} \sum_{i=1}^{J_t} w_j^{(i)}(y) N(x; m_j^{(i)}, P_j^{(i)})
\]
\[
w_j^{(i)}(y) = \frac{P_{t+1}^{(i)} q_j^{(i)}(y)}{\alpha_c(y) + p_d \sum_{j=1}^{J_{t+1}} w_j^{(i)} q_j^{(i)}(y)}
\]
\[
q_j^{(i)}(y) = N(y; H_{t+1} m_{t+1}^{(i)}, R_{t+1} + H_{t+1} P_{t+1} H_{t+1}^T)
\]
\[
m_j^{(i)}(y) = m_j^{(i)} + K_j^{(i)}(y - H_{t+1} m_{t+1}^{(i)})
\]
\[
P_j^{(i)}(y) = [I - K_j^{(i)} H_{t+1}] P_j^{(i)}
\]
\[
K_j^{(i)}(y) = P_{t+1}^{(i)} H_{t+1}^T (H_{t+1} P_{t+1} H_{t+1}^T + R_{t+1})^{-1}
\]

Step 3: Pruning
In the pruning stage, the Gaussian components with low weights are eliminated. Let the weights $w_1^{(i)}, \ldots, w_{N_t}^{(i)}$ be those which are below the eliminated threshold, and the intensity after pruning is
\[
\bar{D}_{x_{t+1}} = \sum_{i=1}^{J_t} \sum_{j=N_t+1}^{J_t} w_j^{(i)} N(x; m_j^{(i)}, P_j^{(i)})
\]

Step 4: Merging
In the merging stage, Gaussian components whose distance between the means falls within a threshold $U$ are merged. For example, if the means of components $i$ and $j$ satisfies
\[
(m_i^{(i)} - m_j^{(j)})^T P_j^{-1} (m_i^{(i)} - m_j^{(j)}) < U
\]
these components are merged into a single Gaussian. Given $\{w_{t+1}^{(i)}, m_{t+1}^{(i)}, P_{t+1}^{(i)}\}_{i=N_t+1}^{J_t}$, a merging threshold $U$, and a maximum allowable number of Gaussian terms $J_{max}$, the merging procedure is as follows:

Set $l = 0$, and $I = \{i = 1, \ldots, J_{t+1} | w_{t+1}^{(i)} > r\}$
Repeat
\[
l = l + 1
\]
\[j = \arg \max_{i \in I} w_{t+1}^{(i)}\]
\[
L = \{i \in I | (m_{t+1}^{(i)} - m_{t+1}^{(j)})^T (P_{t+1}^{(j)})^{-1} (m_{t+1}^{(i)} - m_{t+1}^{(j)}) \leq U\}
\]
\[
\bar{w}_{t+1}^{(i)} = \sum_{i \in L} w_{t+1}^{(i)}
\]
\[
\bar{m}_{t+1}^{(i)} = \frac{1}{\bar{w}_{t+1}^{(i)}} \sum_{i \in L} w_{t+1}^{(i)} m_{t+1}^{(i)}
\]
\[
\bar{P}_{t+1}^{(i)} = \frac{1}{\bar{w}_{t+1}^{(i)}} \sum_{i \in L} w_{t+1}^{(i)} (P_{t+1}^{(i)} + (m_{t+1}^{(i)} - m_{t+1}^{(j)})(\bar{m}_{t+1}^{(i)} - m_{t+1}^{(j)})^T)
\]
\[
I = I \setminus L
\]
Until $I = \phi$
If $I > J_{max}$, replace $\{\bar{w}_{t+1}^{(i)}, \bar{m}_{t+1}^{(i)}, \bar{P}_{t+1}^{(i)}\}_{i=1}^{I}$ by the $J_{max}$ Gaussians with largest weights.
Output $\{\bar{w}_{t+1}^{(i)}, \bar{m}_{t+1}^{(i)}, \bar{P}_{t+1}^{(i)}\}_{i=1}^{I}$ as the merged Gaussian components.

Step 5: State estimation
The states of objects are determined from the posterior intensity by taking the components whose weights are above a specific threshold, which represents the expectation of the object. For example, if the weight is greater than 0.5, the expectation of an object which falls within the region defined by component $i$ is greater than 0.5. The state set estimates at time $t+1$ is
\[
\hat{X}_{t+1} = \{m_i^{(i)} : w_{t+1}^{(i)} > 0.5\}
\]

6 Scene-driven method for new objects
From Fig. 2 we found that new objects can only enter the field of view of the camera at 3 positions, i.e., position A, B, and C in Fig. 3.
We use this prior scene knowledge in tracking. For the initialization of Gaussian mixture (9) and the model of new-birth objects (12), we model them with 3-components Gaussian mixture whose means are the locations of position A, B, and C.

7 Results

The GMPHD filter is tested using the CAVIAR dataset. Fig. 4 shows 4 frames (frame 275, 391, 459, and 484) of video OneStopMoveEnter1Front with white squares indicating the tracking results: the centroids of individual person and human group.

Because the PHD filter explicitly models the processes of birth, surviving, death of targets and false alarms of clutter, as shown by the experimental results, this method is able to track the variable number of people or groups. It is noted that our method considered the 2 people on the right in Fig. 4a as a group. The explanation for this is that the 2 Gaussian components of a Gaussian mixture merge into 1 component when the 2 components are very close.

We compare the scene-driven GMPHD filter with the particle PHD filter [12]. The scene-driven GMPHD filter can track the birth of new objects faster than the particle PHD filter. Fig. 5 shows the first frame when the new object was tracked. The white squares in Fig. 5a are the results of the GMPHD filter and the red circles in Fig. 5b are the results of the particle PHD filter. Because the particle PHD filter uses a uniform distribution as the proposal density of particle filter for new-birth objects and the sample number of particle filter is limited in practice, it is possible that it does not generate samples near the positions of new birth objects. While the GMPHD filter uses the prior scene knowledge and can track the new objects quickly.

Fig. 6 provides the estimates of object number of the GMPHD filter. The estimates are correct for 1148 out of the 1588 frames of video OneStopMoveEnter1Front. For the estimated positions, the Wasserstein distance is used as a metric to measure the performance because it defines a metric for multitarget distance which penalises when the estimated number of targets is incorrect [19]. The above figure of Fig. 7 is the the Wasserstein distance between the estimated positions of the GMPHD filter and the ground-truth positions. While the below figure of Fig. 7 is the the Wasserstein distance between the observed positions of detection and the ground-truth positions. The results show that the tracking errors mainly came from the inaccuracy of measurements.
The results confirm that the Gaussian mixture probability hypothesis density filter can track the variable number of targets and their positions. Our results are consistent with Clark’s work [13, 14]. This property of the GMPHD filter is suitable for multisensor multitarget tracking under complex environments. The results can be explained by the reason that the PHD filter (both particle filter implementation and Gaussian mixture implementation) explicitly models the processes of birth, survival, death of targets and false alarms of clutter and is a model-driven method.

It is worth noting that the GMPHD filter differs from the traditional visual tracking methods. The traditional visual tracking methods rely on only detection results to determine birth or death of targets. Therefore, they are data-driven methods. While the GMPHD filter explicitly models the birth, survival, or death of targets in its dynamics and is a model-driven method. Our work is to combine the data-driven method (detection) with the model-driven method (PHD) and the scene-driven method (Gaussian mixture for prior knowledge).

8 Conclusion

In this paper, the Gaussian mixture probability hypothesis density filter is applied to a visual tracking problem. Foreground objects are detected using the background subtraction method and a variable number of individual person and human group are tracked using a scene-driven GMPHD filter. The results demonstrate that the GMPHD filter can track a variable number of objects and their positions and can track new-birth objects quickly in image sequences. In future work, we will study scene-driven particle PHD filter and compare it with the scene-driven Gaussian mixture PHD filter.

Acknowledgements

The authors thank Dr. Ba-Ngu Vo for providing the programs of the GMPHD filter.

References


