International Journal of Innovative
Computing, Information and Control
Volume **15**, Number **3**, June **2019**

ICIC International ©2019 ISSN 1349-4198
pp. 1131–1143

ADAPTIVE VISUAL MULTI-OBJECT TRACKER BASED ON MULTI-BERNOULLI FILTER AND REGION FEATURE COVARIANCE

Guangnan Zhang[1,2], Jinlong Yang[3,*], Weixing Wang[1,*], Shi Qiu[4] and Yu Tang[3]

[1]College of Information Engineering
Chang'an University
Middle-section of Nan'er Huan Road, Xi'an 710064, P. R. China
zgn_2003@163.com; *Corresponding author: wxwang@chd.edu.cn

[2]School of Computer Science and Technology
Baoji University of Arts and Science
No. 44, Baoguang Road, Baoji 721007, P. R. China

[3]School of Internet of Things Engineering
Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
*Corresponding author: yjlgedeng@163.com

[4]Xi'an Institute of Optics and Precision Mechanics
Chinese Academy of Sciences
No. 17, Xinxi Road, New Industrial Park, Xi'an 710119, P. R. China

Received November 2018; revised February 2019

Abstract. *Multi-Bernoulli (MB) filter has been demonstrated as a promising algorithm for tracking multiple point targets with unknown and time-varying number of targets. However, for the visual multi-object tracking (VMOT), the tracking accuracy will decrease severely due to the problems of closely-spaced object, occlusion and scale variation. To solve these problems, an adaptive VMOT algorithm based on the framework of the MB filter is proposed in this paper. First, the region feature covariance (RFC) is employed to enhance the ability of anti-interference, and then the discrimination strategy for the closely-spaced objects and the adaptive estimate strategy for scale variation are developed according to the constraints of tracking boxes. Finally, the particle labeling technique is introduced to identify the track of each target. Experimental results validate superior performance of the proposed algorithm over the traditional MB-based VMOT algorithm in scenarios that include occlusion, background clutter, scale changes, and closely-spaced objects.*

Keywords: Visual multi-object tracking (VMOT), Multi-Bernoulli (MB) filter, Region feature covariance, Closely-spaced objects

1. **Introduction.** Visual object tracking (VOT) [1-5] is a key enabling technology for numerous emerging computer vision applications including video surveillance, navigation, human-computer interactions, augmented reality, higher level scene understanding and action recognition among many others. Although VOT has been researched for several decades, and much progress has been made in recent years [6-10], it is still a challenging task because the visual observations often suffer from disturbances due to occlusions, scale variation, illumination variation, complex background, etc. Especially, visual multi-object tracking (VMOT) has become one of the most difficult problems due to the extra influence of object number variation, closely-spaced objects and mutual disturbance [11-16].

DOI: 10.24507/ijicic.15.03.1131

For the early VMOT, tracking is mainly achieved through association technologies with target detection, such as linear programming [11], belief propagation [12], network flow [13,14], subgraph decomposition [15], and multiple hypothesis tracking (MHT) [16]. However, these algorithms are quite complicated in computation. Especially, as the target number increases, the computational cost increases exponentially, seriously affecting the real-time performance of VMOT.

Recently, there is a great concern for random finite set (RFS) theory [17-30]. The probability hypothesis density (PHD) filter [17], the cardinalized PHD (CPHD) filter [18] and cardinality-balanced MeMBer (CBMeMBer) filter [19] have proved to be promising approaches based on the RFS theory for multi-target tracking [20-24]. Especially, the multi-Bernoulli (MB) representation allows reliable and inexpensive extraction of state estimates. Some of improved versions are proposed in [25-28], such as variational Bayesian MB (VB-MB), labeled MB (LMB), generalized LMB, and poisson MB (PMB). However, these methods are mainly used for point target tracking. In [29-36], RFS-based filters have been applied to VMOT by using detectors to generating target position and box (scale) as the tracking feature. In [29], adaptive PHD tracker is proposed for VMOT by introducing both scale invariant feature and color distribution feature. In [31], Hoseinnezhad et al. extended the MB filter to track the visual targets and successfully demonstrated tracking of sport players. In [32], an MB-based algorithm is proposed to construct a multi-target likelihood function by using kernel density estimation and background subtraction technique, which can make the multi-target posterior be efficiently propagated forward by using the MB filter. In [35], LMB is employed for VMOT. Generally, the abovementioned methods cannot efficiently handle the VMOT with some severe disturbance factors, such as illumination variation, complex background and occlusion.

Taking the abovementioned problems into account, we propose an improved VMOT algorithm based on MB framework which can oppose occlusions and get adapt to the target appearance variation. First, the multiple features covariance matrices (MFCM) [37] are employed to analyze the local region of the target using block-partition technique, which can express the target accurately alleviating the occlusion disturbance. Moreover, the closely-spaced objects handling method and adaptive template update strategy are proposed to improve the robustness of the proposed algorithm. Finally, the MB based particle label technique is employed to identify the track of each object.

The remainder of the paper is organized as follows. Section 2 summarizes the MB filter and the MFCM method. Section 3 proposes VMOT algorithm based on MFCM and an adaptive template scheme under the MB filter framework, and PF is employed to extract the multi-target states. Particle label technique is implemented to identify the track of each object. Experimental results are presented in Section 4. Finally, the conclusions are given in Section 5.

2. Multiple Feature Express of Target.

In order to cope with occlusions, illumination changes, etc., the block-partition technique is employed for expressing the target accurately, i.e., we partition the target into five parts as shown in Figure 1, including local and global parts. Simultaneously, each part is expressed as a feature covariance matrix (FCM) by using multiple features, and we can analyze the change degree of the local region by comparing the local FCM between the target temple and the candidate target.

2.1. Feature covariance matrix.

Assume $F_I(m,n)$ is a feature image extracting from a gray (intensity) image $I(m,n)$ with size $M \times N$, according to the gradient operator technique, i.e.,

$$F_I(m,n) = \phi(I(m,n)) = \left[I(m,n) \quad \frac{\partial I(m,n)}{\partial m} \quad \frac{\partial I(m,n)}{\partial n} \quad \frac{\partial^2 I(m,n)}{\partial m^2} \quad \frac{\partial^2 I(m,n)}{\partial n^2} \right]^{\mathrm{T}}, \quad (1)$$
$$1 \le m \le M,\ 1 \le n \le N$$

where $\phi(\cdot)$ denotes the mapping function which can be any mapping, such as gray intensity and gradients. For the color images, we transfer them from RGB space into HSV space and gray space, and then three extra features (hue, saturation and value) are added in (1) to construct the feature image of color images. For simplicity, we generally discuss the feature covariance matrix (FCM) of the intensity image here. FCM of intensity image is defined as

$$R_I = \frac{1}{C-1} \sum_{p=1}^{C} (f_p - \mu)(f_p - \mu)^{\mathrm{T}} \quad (2)$$

where f_p denotes the feature vector corresponding to the pth pixel of the intensity image. $f_p \in F_I$, $p = 1, 2, \ldots, C$, $C = M \times N$, $\mu = \frac{1}{C} \sum_{p=1}^{C} f_p$ denotes the mean feature vector of the intensity image and it can be fast computed by using the integral images method [38].

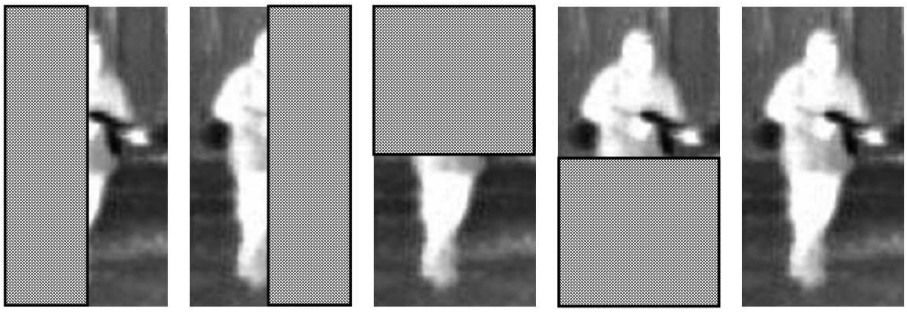


FIGURE 1. Block partition of the target

2.2. **Feature divergence.** In Figure 1, the target is partitioned into five parts, and they are expressed as five FCMs. To measure the similarity between the target temple and the candidate target, the Log-Euclidean distance [39] is introduced to measure the divergence between two covariance matrices, i.e.,

$$d(R_C, R_T) = \|\log(R_C) - \log(R_T)\| \quad (3)$$

where R_C and R_T denote the candidate target covariance and the target model, respectively. For the five FCMs of the target, we discard the least matching region covariance, and the fusion feature divergence (distance) can be expressed as [39]

$$D\left(R_T^i, R_C^j\right) = \min_{\gamma} \left(\sum_{\xi=1}^{5} d\left(R_{T_\xi}^i, R_{C_\xi}^i\right) - d\left(R_{T_\gamma}^i, R_{C_\gamma}^i\right) \right) \quad (4)$$

where $R_{C_\gamma}^i$ denotes the γ^{th} block covariance of the i^{th} candidate (particle) target. The feature covariance $R_{T_\xi}^i$ of the ξ^{th} block of the i^{th} target template is defined as the Log-Euclidean mean [39] that blends the s matrices before time k, i.e.,

$$R_{T_\xi}^i = \exp\left(\frac{1}{s} \sum_{\tau=k-1}^{k-s} \log\left(R_{C_\xi}^i\right) \right), \quad \xi = 1, 2, 3, 4, 5 \quad (5)$$

where $R^i_{C_\xi}$ denotes the feature covariance of the ξ^{th} block of the i^{th} estimated target $\hat{x}^i_\tau$ at time τ.

3. **Adaptive-Particle-MB Filter.** As shown in Figure 2, two parallel tasks should be completed in the initial stage, newborn multi-target prediction and initial multiple targets. New targets are generated in some given regions according to the prior knowledge, but their scales are not known. Notice that the initial targets are considered as the surviving targets in the next frame, and latter the estimated targets are considered as the surviving targets. Then particle filter is applied to generating potential positions of newborn targets and surviving targets at $(k+1)^{\text{th}}$ frame according to the representation of the state transition probability $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$, such that $\boldsymbol{E}(\mathbf{x}_{k+1}|\mathbf{x}_k) = \mathbf{x}_k$, where $\mathbf{x}_k = \{m_k, n_k, w_k, h_k\}$ is the state vector of the target at the k^{th} frame. The assumption is that the target motion may be described by an affine transformation such that (m_k, n_k) is the target position, w_k, and h_k are respectively the width and height of the tracking box. We also assume $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ has a Gaussian distribution where the covariance matrix is selected based on prior knowledge of the tracking task.

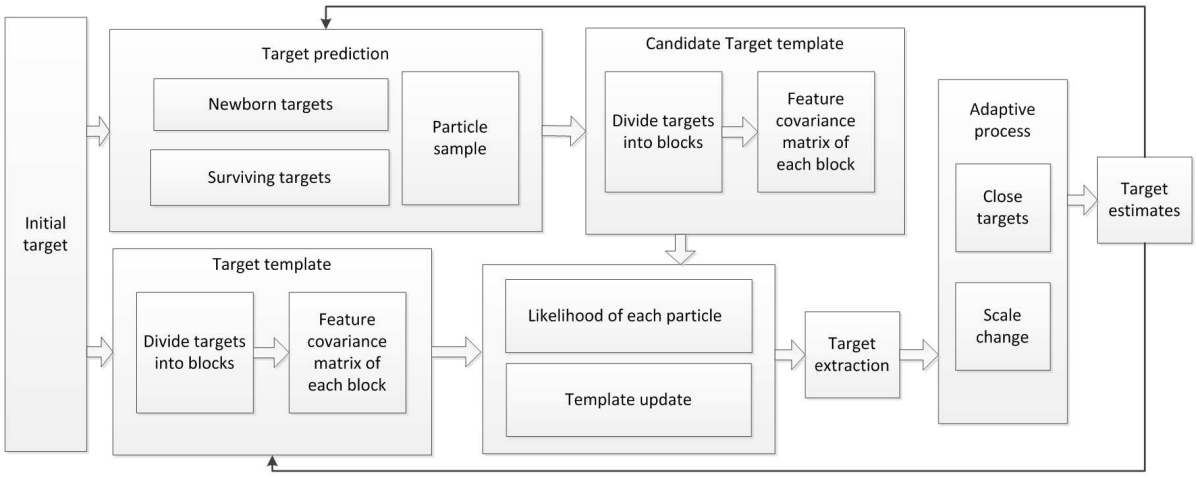


FIGURE 2. Block of the proposed algorithm

Each particle corresponds to a candidate target and we divide each particle into several blocks shown in Figure 1, and calculate its feature covariance which is used to estimate the likelihood between the candidate target template and the target template. Target states can be extracted according to the existence probability of each target. Notice that the existence probability depends on the likelihood. Finally, the adaptive scale process scheme and closely-spaced targets process scheme are proposed and implemented to obtain the optimal states of the estimated multiple targets. The steps of the proposed algorithm are as follows.

3.1. **Steps of proposed algorithm.** At time $k = 0$, assume that the initialized target state set is $X_0 = \{x^i_0\}_{i=1}^{M_0}$, and $x^i_0 = \left[m^i_0, n^i_0, w^i_0, h^i_0\right]$ denotes the i^{th} target state by using a rectangular box, where $\left(m^i_0, n^i_0\right)$ denotes the position coordinates of the box center. w^i_0 and h^i_0 denote the width and height of the box. M_0 denotes the target number. The FCM of the i^{th} target is defined as $R^{0,i}_T$. The newborn target can be initialized with the same method. The FCM of the i^{th} newborn target is defined as $R^{\Gamma,i}_T$.

(1) Prediction. The CBMeMBer filter [24] can eliminate the posterior cardinality bias existing in the MeMBer filter by modifying the measurement-updated tracks parameters. Assume the dynamic process of the target state is $x(k+1) = x(k) + e(k)$, where $e(k)$ is

the Gaussian white noise with zero mean and covariance $\Sigma_k = diag(\sigma_m^2, \sigma_n^2, \sigma_w^2, \sigma_h^2)$. The state transfer function is defined as $f_{k|k-1}(x|x_{k-1}) = N(x; x_{k-1}, \Sigma)$.

Assume at time $k-1$, the posterior multi-target density can be presented by the multi-Bernoulli parameter set, i.e., $\pi_{k-1} = \left\{\left(r_{k-1}^{(i)}, p_{k-1}^{(i)}\right)\right\}_{i=1}^{M_{k-1}}$, where $r_{k-1}^{(i)}$ and $p_{k-1}^{(i)}$ denote the existence probability and probability density of the i^{th} Bernoulli component, respectively. $p_{k-1}^{(i)}$ can be expressed by a group weighted particle $\left\{\omega_{k-1}^{(i,j)}, x_{k-1}^{(i,j)}\right\}_{j=1}^{L_{k-1}^{(i)}}$, i.e.,

$$p_{k-1}^{(i)}(x) = \sum_{j=1}^{L_{k-1}^{(i)}} \omega_{k-1}^{(i,j)} \delta_{x_{k-1}^{(i,j)}}(x) \tag{6}$$

where $\omega_{k-1}^{(i,j)}$ denotes the weight of the particle $x_{k-1}^{(i,j)}$ of the i^{th} target and $L_{k-1}^{(i)}$ denotes the number of the particles. $\delta(\cdot)$ is the Dirichlet function.

In the process of prediction, newborn particles can be obtained from the parameters $\left\{\left(r_{\Gamma,k}^{(i)}, p_{\Gamma,k}^{(i)}\right)\right\}_{i=1}^{M_{\Gamma,k}}$ of newborn targets. $r_{\Gamma,k}^{(i)}$ and $p_{\Gamma,k}^{(i)}$ denote the existence probability and probability density of the i^{th} newborn target, respectively. $M_{\Gamma,k}$ denotes the number of the newborn targets.

The predicted parameters of the surviving targets can be expressed as

$$r_{p,k|k-1}^{(i)} = r_{k-1}^{(i)} \sum_{j=1}^{L_{k-1}^{(i)}} \omega_{k-1}^{(i,j)} p_{S,k}\left(x_{k-1}^{(i,j)}\right) \tag{7}$$

where $p_{S,k}$ denotes the existence probability of the targets at time k,

$$x_{p,k|k-1}^{(i,j)} \sim f_{k|k-1}\left(\cdot \middle| x_{k-1}^{(i,j)}\right) \tag{8}$$

$$f_{k|k-1}\left(\cdot \middle| x_{k-1}^{(i,j)}\right) = N\left(\cdot | x_{k-1}^{j}, \Sigma_{k-1}^{j}\right) \tag{9}$$

$$\tilde{\omega}_{p,k|k-1}^{(i,j)} = \omega_{p,k|k-1}^{(i,j)} \Big/ \sum\nolimits_{j=1}^{L_{k-1}^{(i)}} \omega_{p,k|k-1}^{(i,j)} \tag{10}$$

$$\omega_{p,k|k-1}^{(i,j)} = \omega_{k-1}^{(i,j)} \tag{11}$$

(2) Update. Assume the predicted multi-target density at time k can be expressed by a known multi-Bernoulli parameter set as $\pi_{k|k-1} = \left\{\left(r_{k|k-1}^{(i)}, p_{k|k-1}^{(i)}\right)\right\}_{i=1}^{M_{k|k-1}}$, where $p_{k|k-1}^{(i)}$ is composed of a set of weighted samples $\left\{\omega_{k|k-1}^{(i,j)}, x_{k|k-1}^{(i,j)}\right\}_{j=1}^{L_{k|k-1}^{(i)}}$, i.e.,

$$p_{k|k-1}^{(i)} = \sum_{j=1}^{L_{k|k-1}^{(i)}} \omega_{k|k-1}^{(i,j)} \delta_{x_{k|k-1}^{(i,j)}}(x) \tag{12}$$

where $\omega_{k|k-1}^{(i,j)}$ denotes the predicted weight of the predicted particle $x_{k|k-1}^{(i,j)}$. The number of the predicted Gaussian complements is $M_{k|k-1} = M_{k-1} + M_{\Gamma,k}$.

Update the parameters of the multi-Bernoulli as follows,

$$r_k^{(i)} = \frac{r_{k|k-1}^{(i)} \varepsilon_k^{(i)}}{\left(1 - r_{k|k-1}^{(i)} + r_{k|k-1}^{(i)} \varepsilon_k^{(i)}\right)} \tag{13}$$

$$\omega_k^{(i,j)} = \frac{1}{\varepsilon_k^{(i)}} \omega_{k|k-1}^{(i,j)} g_{y_k}\left(x_{k|k-1}^{(i,j)}\right) \tag{14}$$

$$p_k^{(i)} = \frac{1}{\varepsilon_k^{(i)}} \sum_{j=1}^{L_{k|k-1}^{(i)}} \omega_{k|k-1}^{(i,j)} g_{y_k}\left(x_{k|k-1}^{(i,j)}\right) \delta_{x_{k|k-1}^{(i,j)}}(x) \tag{15}$$

where $\varepsilon_k^{(i)} = \sum_{j=1}^{L_{k|k-1}^{(i)}} \omega_{k|k-1}^{(i,j)} g_{y_k}\left(x_{k|k-1}^{(i,j)}\right)$, $g_{y_k}\left(x_{k|k-1}^{(i,j)}\right)$ denotes the likelihood of the latest observation (candidate target or particle in k^{th} frame from the i^{th} target), which can be obtained according to the fusion feature distance $D^{(i,j)} = D\left(R_T^i, R_C^j\right)$ of Equation (4), i.e.,

$$g_{y_k}\left(x_{k|k-1}^{(i,j)}\right) = \frac{1}{\sqrt{2\pi}} e^{-\lambda D(i,j)/2} \tag{16}$$

where $\lambda \in [10, 30]$.

(3) Target extraction. To alleviate the effect of the particle degeneracy, the updated particles are resampled with the number of particles reallocated in proportion to the probability of existence as well as restricted between a minimum $L_{\min}$ and maximum $L_{\max}$. The resampling step can effectively eliminate the particles with low weights and multiply the particles with high weights to focus on the important zones of the state space. The resampling process is similar to that of the CBMeMBer filter [24]. Notice that the number of the particles increases due to the spontaneous births in the prediction and the averaging of the hypothesized tracks in the update. Therefore, the hypothesized tracks need to be pruned by discarding those with existence probabilities below a threshold η (e.g., 0.01), which can reduce the number of particles effectively.

Finally, the number of targets and their states are estimated via finding the multi-Bernoulli parameters with existence probabilities larger than a threshold (set at 0.5 in our experiments). Each target state estimate is then given by the weighted average of the particles of the corresponding density, i.e.,

$$x_k^i = \sum_{j=1}^{L_{k|k-1}^{(i)}} \omega_k^{(i,j)} x_{k|k-1}^{(i,j)} \tag{17}$$

3.2. **Adaptive process of closely-spaced objects.** When two targets are close to each other, it is easy to be estimated as one target, which causes underestimate. To solve this problem, we propose an effective scheme to detect the close targets and separate them adaptively.

Taking two targets shown in Figure 3 as an example, they are close in the horizon direction (x direction). Assume the states of the two targets are $x_i = [m_i, n_i, w_i, h_i]$, $x_j = [m_j, n_j, w_j, h_j]$, and the horizon distance between the two targets is $d_{ij}^h = |m_j - m_i|$, s.t. $n_j = n_i$. If d_{ij}^h is less than a distance threshold T_h^1, then we consider the two targets are superposition without special process. If $T_h^1 \le d_{ij}^h \le T_h^2$, then we consider that the two targets are close, and we separate the tracking box apart each other according to the scheme in Table 1. In this paper, we set $T_h^1 = (w_i + w_j)/5$ and $T_h^2 = (w_i + w_j)/2$. If the targets are close in the vertical direction (y direction), the similar method can be used. In Table 1, α is an empirical parameter and set as 0.15, which means that the tracking box will be moved apart 0.15 times of the box width when the targets are close to each other.

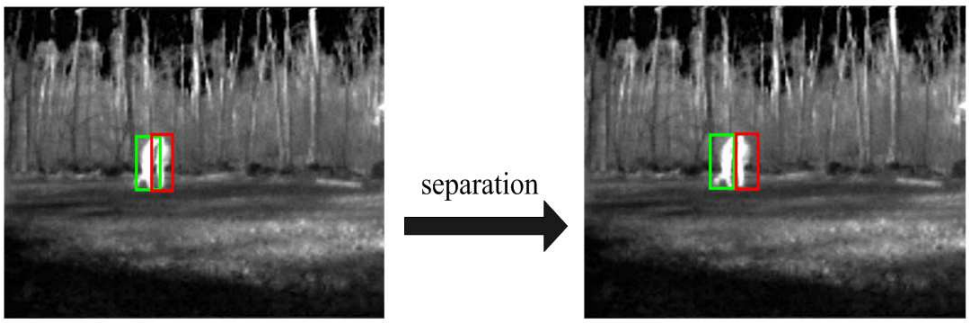


FIGURE 3. Separation of the close targets

TABLE 1. The scheme to separate close targets

Initial parameter $\alpha = 0.15$
if $T_h^1 < d_{ij}^h < T_h^2$ then
if $m_i \leq m_j$ then
$m_i \leftarrow m_i - w_i\alpha$
$m_j \leftarrow m_j + w_j\alpha$
else
$m_i \leftarrow m_i + w_i\alpha$
$m_j \leftarrow m_j - w_j\alpha$
end if
end if

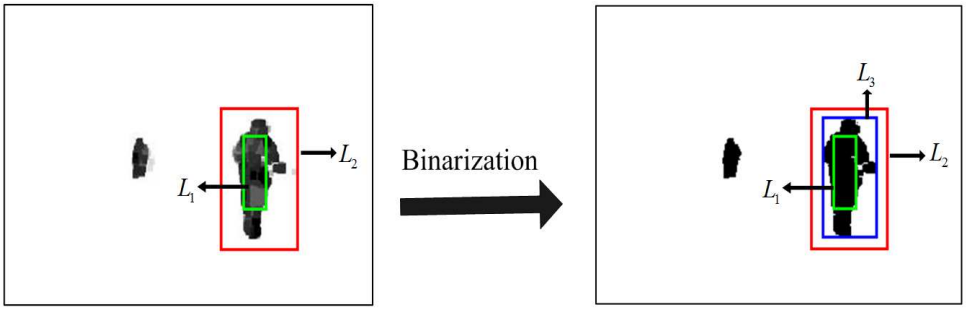


FIGURE 4. Adaptive tracking box

3.3. **Adaptive tracking box.** Assume the state of the i^{th} extracted target at time k is expressed as $x_{k,i}^R = [m_k^r, n_k^r, w_k^r, h_k^r]$, and the tracking box is L_1 shown in Figure 4. When the scale variation of the target is big, L_1 may not describe the target well, i.e., it cannot contain the whole target; therefore, we should adjust the size of the tracking box. We propose to enlarge L_1 as L_2, and the enlarged state is described as $x_{k,i}^E = [m_k^r, n_k^r, (1+\lambda)w_k^r, (1+\lambda)h_k^r]$, where λ is the scale factor and it is set as 0.4 according to the experience in this paper. Then image binarization is implemented in the box of L_2, and the maximal connected region of the target can be obtained. Finally, the minimal box L_3 containing the maximal connected region can be extracted as the estimated tracking box of the target, and the state is described as $x_{k,i}^{\min}$. If $|C_{L_3} - C_{L_1}| > \eta$, where η is a threshold of the difference value between two boxes and set as $0.4C_{L_1}$, C_{L_3} and C_{L_1} denote the perimeter of the boxes L_3 and L_1, then the estimated state is updated as

$x_{k,i}^F = \beta * x_{k,i}^{\min} + (1-\beta)x_{k,i}^R$, where $\beta = 0.1$, otherwise, $\beta = 0.9$. It means that if the difference value is bigger than the threshold, i.e., the variation of the target state $x_{k,i}^{\min}$ is big corresponding to $x_{k,i}^R$, then we more believe $x_{k,i}^R$; therefore, we select smaller value of β.

3.4. **Track maintenance.** In order to obtain the target tracks, the particle labeling technique [40,41] is employed to identify the targets. In prediction stage, the prediction labels of the Bernoulli components keep the same labels with the previous components. It is noticed that all the particles from the same Bernoulli component have the same label. While in the update stage, the measurement-updated components are assigned the labels of the predicted tracks. Moreover, the resampled particles keep the same label as their father particles. Finally, the track continuity can be completed according to the particle labels by the data association technique [40,41], i.e., the track can be obtained by comparing the number of particles with the same labels in each component.

4. **Experimental Results and Analysis.** To verify the effectiveness of the proposed algorithm, some challenging sequences from the public dataset of VMOT are used to evaluate the performance of the proposed algorithm. The video data (Data 1, 2 and 3) of the experiment are from the Terravic Research Infrared Database [42], CAVIAR Database 05 and Dataset 03 [43]. The main challenging features of these data include target number variation, scale changes, close target, occlusion and background clutter. The proposed algorithm is compared with the original MB-based tracking algorithm [31]. The experiments are implemented on computer with Intel Core 2.4 GHz, i7-4700HQ processor with 8GB RAM. The software tool is MATLAB 2014a. For each sequence, the location of the target is manually labeled in the first frame.

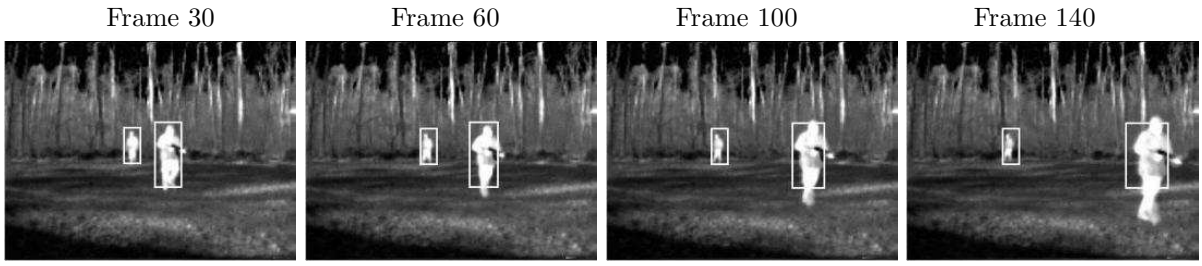
In the experiments, the probabilities of target survival $p_{S,k} = 0.99$, and at each time, a maximum of $L_{\max} = 1200$ and minimum of $L_{\min} = 300$ particles per-hypothesized track are imposed so that the number of particles representing each hypothesized track is proportional to its existence probability after resampling in the update stage. In addition, the hypothesized tracks pruning is performed with a weight threshold of $\eta = 10^{-3}$ and the maximum number of hypothesized tracks is $T_{\max} = 100$.

To evaluate the performance of the tracking algorithms, four evaluation criteria are employed to quantitatively assess the performance of the proposed algorithm. They are average number estimate of targets, optimal subpattern assignment (OSPA) distance [19], average miss tracking rate (AMTR) and average running time. AMTR is defined as the rate between the average number of the miss estimated targets and the real number of targets. Experimental results are shown in Figures 5-10 and Table 2.

4.1. **Qualitative evaluation.** Figures 5, 6 and 7 show the tracking results of the visual targets with different algorithms, where the video Data 1 in Figure 5 is infrared and Data 2 and 3 in Figures 6 and 7 are colorized.

(1) Scale change. As we can see that the targets in Figure 5 undergo heavy scale changes. The proposed algorithm can adaptively estimate the scale of target accurately, which demonstrates that the adaptive scale process scheme in the proposed algorithm can effectively work. From Figures 6 and 7, we can also conclude that the proposed algorithm has a better performance for estimating the target scale.

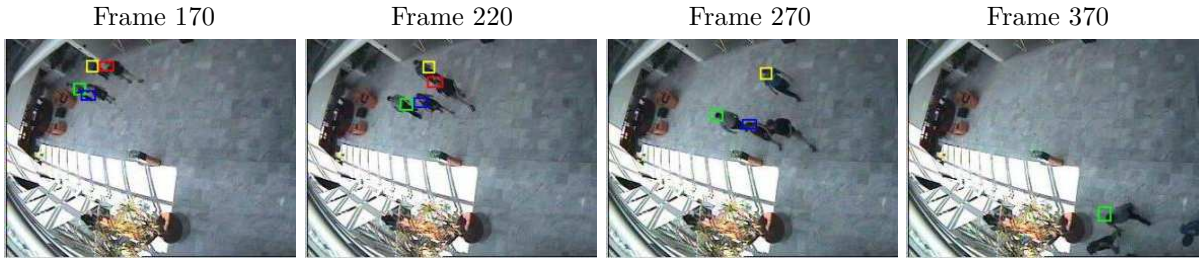
(2) Closely-spaced targets. In Figure 6, some of the targets are close to each other, and partial occlusion occurs in this scenario. However, it is clear that the proposed algorithm can identify each target and extract the targets accurately, which illustrates that the adaptive process scheme of closely-spaced targets in the proposed algorithm can work well. Moreover, the multiple features of the local blocks and global of targets are



(a) Algorithm in [31]

(b) Proposed algorithm

FIGURE 5. Experimental results of Data 1 (Infrared Database)



(a) Algorithm in [31]

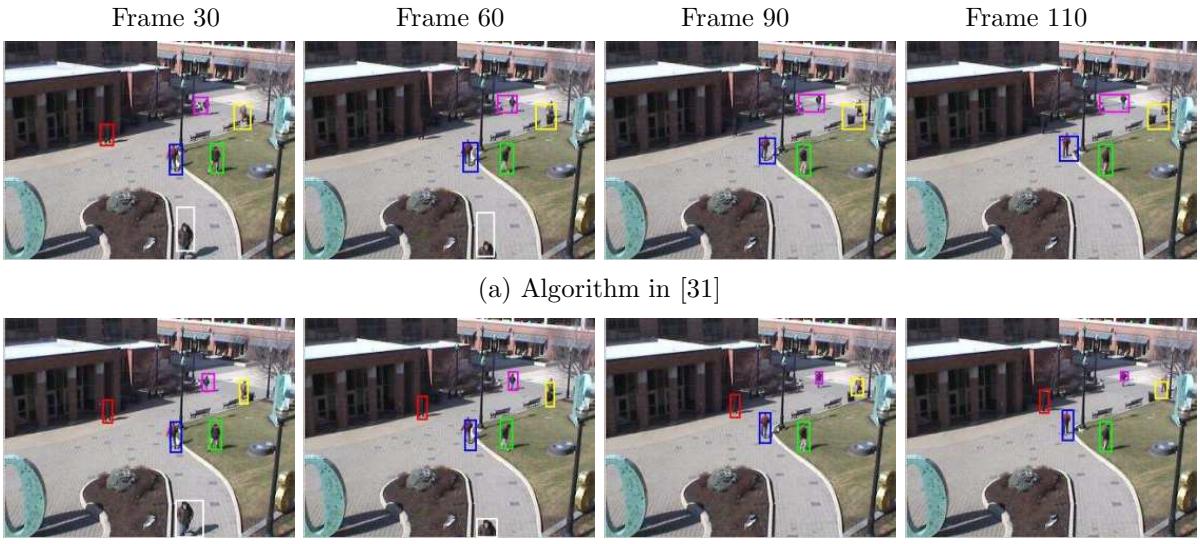
(b) Proposed algorithm

FIGURE 6. Experimental results of Data 2 (CAVIAR Database 05)

extracted to form the FCMs which can more accurately express the targets and alleviate the disturbance of the occlusion. While part of the targets are missed by using the original MB-based algorithm due to the close target and the occlusion, such as the targets in the 270^{th} frame.

(3) Disturbance of complex background. In Figure 7, some of the targets are disturbed in the complex background, such as the shadow, building, lawn, and tree. We also draw the conclusion that the proposed algorithm has a better tracking performance of position and scale estimation than the original MB-based algorithm.

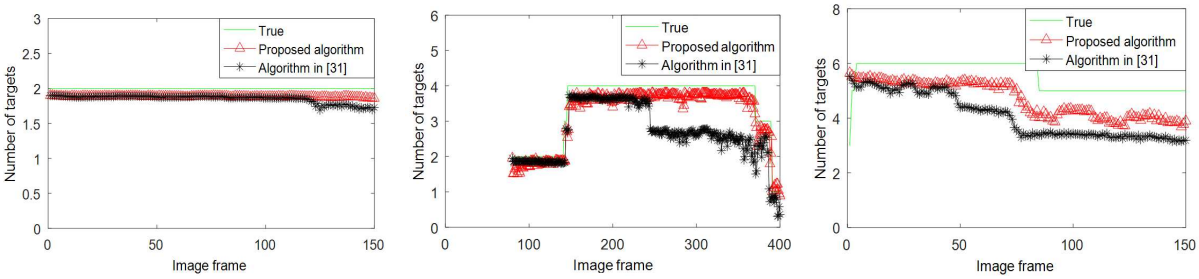
4.2. **Quantitative evaluation.** Figure 8 shows the average target number estimates of the three different video. We can see that the proposed algorithm provides more accurate target number estimates than the original MB-based method. The reason is that the feature covariance technique of the proposed algorithm can effectively present the targets. Moreover, the tracking box and the close targets are adaptively processed by the proposed



(a) Algorithm in [31]

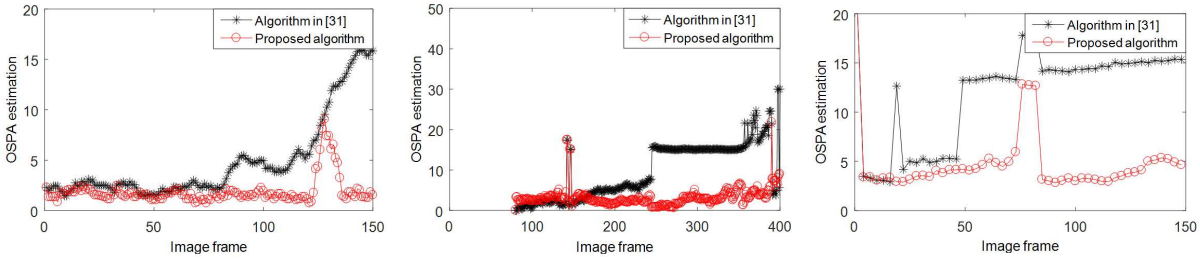
(b) Proposed algorithm

FIGURE 7. Experimental results of Data 3 (CAVIAR Dataset 03)



(a) Data 1 (b) Data 2 (c) Data 3

FIGURE 8. The comparisons of target number estimates



(a) Data 1 (b) Data 2 (c) Data 3

FIGURE 9. OSPA distance estimation

scheme. While for the original MB-based method, it is clear that some of the targets are missed.

Figure 9 shows the comparisons of the OSPA distances between the two algorithms, and it is clear that the proposed algorithm again outperforms the original MB-based algorithm. This is also due to the fact that the proposed method can adapt to the scale changes and has a good performance for tracking the close targets.

Table 2 shows the average OSPA distance and miss tracking rate of the two algorithms. It can be seen that the proposed algorithm has a better performance than the originated

(a) Data 1 (b) Data 2 (c) Data 3

FIGURE 10. (color online) Results of the track estimation (different colors denote different tracks)

TABLE 2. Average OSPA distance and miss tracking rate

Data	Algorithm	Average OSPA	Miss tracking rate
Data 1	Algorithm in [31]	5.3963	0
	Proposed algorithm	1.8763	0
Data 2	Algorithm in [31]	12.7337	0.5275
	Proposed algorithm	3.4233	0.0125
Data 3	Algorithm in [31]	23.7528	0.9733
	Proposed algorithm	5.1358	0.0333

MB-based method. However, the average run time of the proposed algorithm is slightly higher than that of another algorithm. The reason is that the proposed algorithm includes the procedure of adaptive close targets and scale variation.

4.3. **Trajectory estimates.** Figure 10 shows the estimated trajectories of the targets, it is clear that the proposed algorithm can effectively estimate the trajectory of each target, although some of the targets are close in Figure 10(b), and partial targets are occluded by the background in Figure 10(c). Therefore, the proposed algorithm has a good ability of track estimation by employing the particle labeling technique.

5. **Conclusions.** In this paper, we present an adaptive visual multiple object tracking based on MB filter. The region feature covariance (RFC) is employed to construct the target template which has a better ability of anti-interference. To robustly decide the final tracking state, the discrimination strategy for the closely-spaced objects and the adaptive estimate strategy for scale variation are proposed according to the image fusion technique. Finally, the particle labeling technique is introduced to identify the track of each target. The effectiveness of the proposed algorithm is experimentally demonstrated by being compared with the original MB-based algorithm on challenging video sequences, and the experimental results show that the proposed algorithm has a better tracking performance for visual multi-object tracking in the scenarios with the interference of occlusion, background clutter, scale change, shadow as well as the closely-spaced targets. However, the compute cost is high; in future, we would like to improve the computational efficiency by considering some optimal schemes.

Acknowledgment. This work is partially supported by the Natural Science Foundation of Jiangsu Province (Nos. BK20181340, BK20130154), the National Natural Science

Foundation of China (Nos. 61305017, 61772237). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] T. Z. Zhang, B. Ghanem, S. Liu et al., Robust visual tracking via multi-task sparse learning, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.2042-2049, 2012.

[2] D. A. Ross, J. Lim, R. S. Lin et al., Incremental learning for robust visual tracking, *International Journal of Computer Vision*, vol.77, nos.1-3, pp.125-141, 2008.

[3] X. Mei and H. B. Ling, Robust visual tracking using L1 minimization, *Proc. of IEEE International Conference on Computer Vision*, Japan, pp.1436-1443, 2009.

[4] S. P. Zhang, H. X. Yao, X. Sun et al., Sparse coding based visual tracking: Review and experimental comparison, *Pattern Recognition*, vol.46, no.7, pp.1772-1788, 2013.

[5] C. L. Bao, Y. Wu, H. B. Ling et al., Real time robust L1 tracker using accelerated proximal gradient approach, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.1830-1837, 2012.

[6] Y. Wu, J. Lim and M. H. Yang, Object tracking benchmark, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.37, no.9, pp.1834-1848, 2015.

[7] M. Kristan, A. Eldesokey, Y. Xing et al., The visual object tracking VOT2017 challenge results, *Proc. of IEEE International Conference on Computer Vision Workshop*, Italy, pp.1949-1972, 2017.

[8] L. Jiang, C. Bai, Y. Wang and H. Sun, Modeling visual appearance as a global optimization problem in visual tracking, *ICIC Express Letters, Part B: Applications*, vol.9, no.2, pp.163-170, 2018.

[9] K. Chen, W. Tao and S. Han, Visual object tracking via enhanced structural correlation filter, *Information Sciences*, vol.394, pp.232-245, 2017.

[10] T. Zhang, C. Xu and M. H. Yang, Learning multi-task correlation particle filters for visual tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2018.

[11] W. Choi and S. Savarese, A unified framework for multi-target tracking and collective activity recognition, *Proc. of European Conference on Computer Vision*, Italy, pp.215-230, 2012.

[12] A. V. Segal and I. Reid, Latent data association: Bayesian model selection for multi-target tracking, *Proc. of IEEE International Conference on Computer Vision*, Australia, pp.2904-2911, 2014.

[13] A. Dehghan, Y. Tian, P. H. S. Torr et al., Target identity-aware network flow for online multiple target tracking, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.1146-1154, 2015.

[14] S. H. Ben, J. Berclaz, F. Fleuret et al., Multi-commodity network flow for tracking multiple people, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.36, no.8, pp.1614-1627, 2014.

[15] S. Tang, B. Andres, M. Andriluka et al., Subgraph decomposition for multi-target tracking, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.5033-5041, 2015.

[16] H. Ergezer and K. Leblebicioglu, Visual tracking of objects via rule-based multiple hypothesis tracking, *Proc. of Signal Processing, Communication and Applications Conference*, Turkey, 2008.

[17] R. Mahler, Multi-target Bayes filtering via first-order multi-target moments, *IEEE Trans. Aerospace and Electronic Systems*, vol.39, no.4, pp.1152-1178, 2003.

[18] R. Mahler, PHD filter of higher order in target number, *IEEE Trans. Aerospace and Electronic Systems*, vol.43, no.3, pp.1523-1543, 2007.

[19] B. T. Vo, B. N. Vo and A. Cantoni, The cardinality balanced multi-target multi-Bernoulli filter and its implementations, *IEEE Trans. Signal Processing*, vol.57, no.2, pp.409-423, 2009.

[20] B. N. Vo and W. K. Ma, The Gaussian mixture probability hypothesis density filter, *IEEE Trans. Signal Processing*, vol.54, no.11, pp.4091-4104, 2006.

[21] B. N. Vo, S. Singh and A. Doucet, Sequential Monte Carlo methods for multi-target filtering with random finite sets, *IEEE Trans. Aerospace and Electronic Systems*, vol.41, no.4, pp.1224-1245, 2005.

[22] J. Yang and H. Ge, An adaptive PHD filter based on variational Bayesian approximation for multi-target tracking, *IET Radar, Sonar & Navigation*, vol.7, no.9, pp.959-967, 2013.

[23] C. Li, W. Wang, T. Kirubarajan et al., PHD and CPHD filtering with unknown detection probability, *IEEE Trans. Signal Processing*, vol.66, no.14, pp.3784-3798, 2018.

[24] M. E. Choi and S. W. Seo, Robust multitarget tracking scheme based on Gaussian mixture probability hypothesis density filter, *IEEE Trans. Vehicular Technology*, vol.65, no.6, pp.4217-4229, 2016.

[25] J. Yang and H. Ge, An improved multi-target tracking algorithm based on CBMeMBer filter and variational Bayesian approximation, *Signal Processing*, vol.93, no.9, pp.2510-2515, 2013.
[26] B. N. Vo, B. T. Vo and D. Phung, Labeled random finite sets and the Bayes multi-target tracking filter, *IEEE Trans. Signal Processing*, vol.62, no.24, pp.6554-6567, 2014.
[27] F. Papi, B. N. Vo, B. T. Vo et al., Generalized labeled multi-Bernoulli approximation of multi-object densities, *IEEE Trans. Signal Processing*, vol.63, no.20, pp.5487-5497, 2015.
[28] Y. Xia, K. Granstrcom, L. Svensson et al., Performance evaluation of multi-Bernoulli conjugate priors for multi-target filtering, *Proc. of International Conference on Information Fusion*, China, pp.1-8, 2017.
[29] J. Wu, S. Hu and Y. Wang, Adaptive multifeature visual tracking in a probability-hypothesis-density filtering framework, *IEEE Trans. Signal Processing*, vol.93, no.11, pp.2915-2926, 2013.
[30] X. Zhou, Y. Li, B. He et al., GM-PHD-based multi-target visual tracking using entropy distribution and game theory, *IEEE Trans. Industrial Informatics*, vol.10, no.2, pp.1064-1076, 2014.
[31] R. Hoseinnezhad, B. N. Vo, B. T. Vo et al., Visual tracking of numerous targets via multi-Bernoulli filtering of image data, *Pattern Recognition*, vol.45, no.10, pp.3625-3635, 2012.
[32] R. Hoseinnezhad, B. N. Vo and B. T. Vo, Visual tracking in background subtracted image sequences via multi-Bernoulli filtering, *IEEE Trans. Signal Processing*, vol.61, no.2, pp.392-397, 2013.
[33] Y. K. Du, Multi-Bernoulli filtering for keypoint-based visual tracking, *Proc. of International Conference on Control, Automation and Information Sciences*, South Korea, pp.37-41, 2016.
[34] T. Rathnayake, A. K. Gostar, R. Hoseinnezhad et al., Labeled multi-Bernoulli track-before-detect for multi-target tracking in video, *Proc. of International Conference on Information Fusion*, USA, pp.1353-1358, 2015.
[35] S. Reuter, B. T. Vo, B. N. Vo et al., Multi-object tracking using labeled multi-Bernoulli random finite sets, *Proc. of International Conference on Information Fusion*, Spain, pp.1-8, 2014.
[36] M. Jiang, Z. Pan and Z. Tang, Visual object tracking based on cross-modality Gaussian-Bernoulli deep Boltzmann machines with RGB-D sensors, *Sensors*, vol.17, no.1, 2017.
[37] F. Porikli, O. Tuzel and P. Meer, Covariance tracking using model update based on lie algebra, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.728-735, 2006.
[38] F. Porikli, Integral histogram: A fast way to extract histograms in Cartesian spaces, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp.829-836, 2005.
[39] Z. H. Fan, H. B. Ji and Y. Q. Zhang, Iterative particle filter for visual tracking, *Image Communication*, vol.36, pp.140-153, 2015.
[40] K. Panta, B. N. Vo and S. Singh, Novel data association schemes for the probability hypothesis density filter, *IEEE Trans. Aerospace and Electronic Systems*, vol.43, no.2, pp.556-570, 2007.
[41] J. L. Yang, H. B. Ji and H. W. Ge, Multi-model particle cardinality-balanced multi-target multi-Bernoulli algorithm for multiple manoeuvring target tracking, *IET Radar, Sonar & Navigation*, vol.7, no.2, pp.101-112, 2013.
[42] *Dataset 05: Terravic Motion IR Database*, OTCBVS Benchmark Dataset Collection [DB/OL], http://vcipl-okstate.org/pbvs/bench/Data/05/download.html, 2018.
[43] *CAVIAR Test Case Scenarios [DB/OL]*, http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1, 2018.
[44] *Dataset 03: OSU Color-Thermal Database*, OTCBVS Benchmark Dataset Collection [DB/OL], http://vcipl-okstate.org/pbvs/bench/Data/03/download.html, 2018.