Visual multiple-object tracking for unknown clutter rate

Du Yong Kim

Abstract: In multi-object tracking applications, model parameter tuning is a prerequisite for reliable performance. In particular, it is difficult to know statistics of false measurements due to various sensing conditions and changes in the field of views. In this study, the authors are interested in designing a multi-object tracking algorithm that handles unknown false measurement rate. Recently proposed robust multi-Bernoulli filter is employed for clutter estimation while generalised labelled multi-Bernoulli filter is considered for target tracking. Performance evaluation with real videos demonstrates the effectiveness of the tracking algorithm for real-world scenarios.

1 Introduction

Multi-object tracking is one of the fundamental problems in many applications. There are abundant research works; however, it is still far from practical use. The overwhelming majority of multi-target tracking algorithms are built on the assumption that multi-object system model parameters are known a priori, which is generally not the case in practice [1, 2]. While tracking performance is generally tolerant to mismatches in the dynamic and measurement noise, the same cannot be said about missed detections and false detections. In particular, mismatches in the specification of missed detection and false detection model parameters such as detection profile and clutter intensity can lead to a significant bias or even erroneous estimates [3].

Unfortunately, except for a few application areas, exact knowledge of model parameters is not available. This is especially true in visual tracking, in which the missed detection and false detection processes vary with the detection methods. The detection profile and clutter intensity are obtained by trial and error. A major problem is the time-varying nature of the missed detection and false detection processes. Consequently, there is no guarantee that the model parameters chosen from training data will be sufficient for the multi-object filter at subsequent frames.

In radar target tracking applications, stochastic multi-object tracking algorithms based on online (i.e. recursive) filtering algorithms including Kalman filtering or Sequential Monte Carlo (SMC) method have been widely used [2, 4]. This approach is of great importance in time critical applications due to its recursive structure that is able to provide tracking outputs online. This approach also has been used in visual multi-object tracking research [5–7]. On the other hand, deterministic approach such as network flow [8], continuous energy optimisation [9], has become a popular method for multi-object tracking problem in visual tracking application. This approach is known to be free from tuning parameters, however, it is useful only when reliable object detection is available.

Unknown observation model parameters (i.e. clutter rate, detection profile) in online (i.e. recursive) multi-object filtering was recently formulated in a joint estimation framework using random finite set (RFS) approach [10, 11]. Recently, Mahler et al. [3] and its application to cell microscopy [12] showed that clever use of the cardinalised probability hypothesis density (CPHD) filter can accommodate unknown clutter rate and detection profile. In [13] it was demonstrated that by bootstrapping clutter estimator from [3] to the Gaussian mixture CPHD filter [14], performed very close to the case with known clutter parameter can be achieved. Vo et al. [15] extended it to multi-Bernoulli filter with SMC implementation. The multi-Bernoulli filter was used for visual multi-object tracking in [16]. While the solution for filtering with unknown clutter rate exists, these filters do not provide tracks that identify different objects. In particular, the conference version of this work [16] is seriously extended as a new algorithm that is able to provide track identities with a completely new structure and evaluated using challenging pedestrian tracking and cell migration experiments to the best of our knowledge this paper is the first attempt at handling unknown false measurement information in online tracking. The main contribution of this paper is to design a multi-object tracker that also produces trajectories and estimates clutter rate on the fly.

2 Problem formulation

The goal of visual multiple-object tracking is to accurately localise each object of interest with unique labels in the presence of measurement uncertainty. Measurement uncertainty in visual tracking includes missed detections (i.e. false negatives), clutters (i.e. spurious measurements originated from non-targets) and measurement noises (i.e. errors with respect to the ground truths) due to imperfect detectors. In this paper, we formulate visual multi-object tracking problem based on the state-space model that is widely used in the object tracking literature [1, 2]. Furthermore, the model is extended to take into account clutters that will be regarded as a special type of object as suggested in [3].

Let $X = \mathbb{R}^n$ denote the space of the target kinematic state, and $\{0, 1\}$ denote the discrete space of labels to distinguish clutter and actual targets. Here, clutter is modelled as a special type of targets by denoting with the label $u = 0$.

Then, the augmented state space is given by

$$\tilde{X} = X \times \{0, 1\}$$

(1)

where $\times$ denotes a Cartesian product. Consequently, the state variable contains the kinematic state, and target/clutter indicator. The discrete value $u$ is used for subscripts with $u = 0$ as clutter targets and the label $u = 1$ for actual targets.

Suppose that there are $T_t$ targets including actual targets and clutter objects, and we have $O_t$ observations including target originated detections and clutters. In the RFS framework, the collections of targets (including clutter objects) and measurements can be described as finite subsets of the state and observation spaces, respectively, as
\[ \tilde{x}_k = \{ \tilde{x}_{ik} \}_{i=1}^{T_i} \subset \tilde{x}, \quad Z_k = \{ z_{ij} \}_{j=1}^{O_j} \subset Z, \]  
(2)

where \( \tilde{x}_k \) represents either the kinematic state of the actual target or clutter target; \( z_{ij} \) is a measurement and \( Z \) is the space of measurement, respectively. Considering the dynamic of the state, the RFS model of the multi-target state at time \( k \) consists of surviving targets and new targets entering the scene. This new set is represented as the union
\[ \tilde{x}_k = \bigcup_{i=1}^{T_{i-1}} \tilde{x}_{i-1} \bigcup \Gamma_k, \]  
(3)

where \( \Gamma_k \) is a set of spontaneous birth objects (actual target or clutter targets) and \( \tilde{x}_{i-1} \) is the set of survived object states at a time \( k \) with survival probability \( P_S(x) < 1 \).

The set of observations given the multi-target state is expressed as
\[ Z_k = Z_{T,k} \cup Z_{F,1,k}, \]  
(4)

where \( Z_{T,k} \) and \( Z_{F,1,k} \) are, respectively, sets of clutter and target-originated observations with unknown detection probability \( P_D(x) < 1 \).

With the RFS multi-target dynamic and measurement model, the multi-object filtering problem amounts to propagating multi-target posterior density recursively forward in time via the Bayes rule. Equation (1)–(4) called RMB filter [15] is summarised to make the implementation of the multi-object filtering problem amounts to propagating multi-Bernoulli: multi-Bernoulli: multi-Bernoulli: multi-Bernoulli:

\[ \{ r^0_{ik,k-1}, p^0_{ik,k-1} \}_{i=1}^{M_{ik-1}}. \]  
(6)

A set of predicted Bernoulli components is a union of birth components \( \{ r^B_{ik,k} \}_{i=1}^{M_{ik}} \) and surviving components \( \{ r^P_{ik,k-1}, p^P_{ik,k-1} \}_{i=1}^{M_{ik-1}} \). The birth Bernoulli components are chosen a priori by considering the entrance region of the visual scene, e.g. image border. The surviving components are calculated by
\[ r^P_{ik,k-1} = r^P_{ik-1} \sum_{a=0,1} \langle p^P_{ik,k-1} \rangle_{P_{ik,k-1}} \]  
(7)

where \( x \) is the kinematic state, \( P_{S,x,k} \) is the survival probability to time \( k \) and \( f_{ik,k-1}(x(\cdot)) \) is the state transition density specified by either for actual target \( f_{ik,k-1}(x(\cdot)) \) or for clutter target \( f_{ik,k-1}(x(\cdot)) \).

If at the time \( k \), the predicted multi-target density is multi-Bernoulli of the form (6), then the updated multi-Bernoulli density approximation \( \{ r^U_{ik,k} \} \}_{i=1}^{M_{ik}} \) is composed of the legacy components with the subscript \( L \) and the measurement updated components with the subscript \( U \) as follows: (see (8)). The legacy and measurements updated components are calculated by a series of (9) and (10) as follows:
\[ r^L_{ik,k} = \sum_{a=0,1} r^L_{ik,a,k}, \]  
(8)

\[ r^L_{ik,a,k} = \frac{r^L_{ik-1} \langle p^L_{ik-1} \rangle_{r^L_{ik-1}, a} - p^{D_{ik,a,k}}}{1 - \langle p^{D_{ik,a,k}} \rangle_{r^L_{ik-1}, a}}, \]  
(9)

\[ r^L_{ik,a,k} = \frac{1 - p^{D_{ik,a,k}}}{\sum_{a=0,1} \langle p^{D_{ik,a,k}} \rangle_{r^L_{ik-1}, a}}. \]  
(10)

(see (10)) where \( P_{D_{ik,a,k}} \) is the state dependent detection probability, \( P_{D_{ik,a,k}}(z|x) \) is the ML function that will be defined in the following section. Note that the SMC implementation of summarised (5)–(10) can be found in [15].

3.2 Boosted GLMB filter

The GLMB filter provides a solution of multi-object Bayes filter with unique labels. In this paper, the GLMB filter is used as a tracker that returns trajectories of multi-object given the estimated clutter rate from the RMB. As shown in Fig. 1, GLMB and RMB filters are interconnected by sharing tracking parameters and facilitate feedback mechanism in order for robust tracking against time-varying clutter background. Note that one step RMB filter is used for the estimation of clutter rate. Thus, it is not a parallel implementation of independent filters, however, track parameters are shared between GLMB and RMB filters to save computational resources. We call the proposed tracker as boosted GLMB tracker.

In multi-object tracking with labels, formally, the state of an object at a time \( k \) is defined as \( x_k = (x_k, c_k) \in \mathbb{X} \times \mathbb{L}_k \), where \( \mathbb{L}_k \) denotes the label space for objects at a time \( k \) (including those born prior to \( k \)). Note that \( \mathbb{L}_k \) is given by \( \mathbb{B}_k \cup \mathbb{L}_{k-1} \), where \( \mathbb{B}_k \) denotes the label space for objects born at a time \( k \) (and is disjoint from \( \mathbb{L}_{k-1} \)) and we do not consider a clutter generator in designing GLMB, thus, the label \( u \) is omitted. Suppose that there are \( N_k \) objects at the time \( k \) as (2), but only consider actual target with the label \( x_{1,k}, \ldots, x_{N_k} \), in the context of multi-object tracking
\[ X_k = \{ x_{1,k}, \ldots, x_{N,k} \} \in \mathbb{F}(\mathbb{X} \times \mathbb{L}_k), \]  
(11)

\[ \{ r^L_{ik,k}, p^L_{ik,k} \}_{i=1}^{M_{ik}} \cup \{ r^U_{ik,k}(\cdot; z) \}_{i \in Z_k}. \]  
(8)

IET Comput. Vis., 2018, Vol. 12 Iss. 5, pp. 728-734
© The Institution of Engineering and Technology 2018
where \( \mathcal{F}(\mathbb{X} \times \mathbb{L}) \) denotes the space of finite subsets of \( \mathbb{X} \times \mathbb{L} \). We denote cardinality (number of elements) of \( \mathbb{X} \) by \(|\mathbb{X}|\) and the set of labels of \( \mathbb{X} \), \( \{\ell' : (x, \ell') \in \mathbb{X}\} \), by \( \mathcal{L}_X \). Note that since the label is unique, no two objects have the same label, i.e., \( \delta_X(\mathcal{L}_X) = 1 \). Hence \( \Delta(X) = \delta_X(\mathcal{L}_X) \) is called the distinct label indicator.

In the GLMB the posterior density takes the form of a GLMB

\[
\pi_{k-1}(X) = \Delta(X) \sum_{\ell \in \mathbb{L}} \omega_{k-1}^{(\ell)}(\mathcal{L}_X)[p_{k-1}^{\ell}(X)]^X.
\]

Given the posterior multi-object density of the form (12), the predicted multi-object density to time \( k \) is given by

\[
\pi_{k|k-1}(X) = \Delta(X) \sum_{\ell \in \mathbb{L}} \omega_{k-1}^{(\ell)}(\mathcal{L}_X)[p_{k|k-1}^{\ell}(X)]^X
\]

where

\[
\omega_{k-1}^{(\ell)}(L) = \omega_{k-1}^{(\ell)}(L \cap \mathcal{L}_X)L \omega_{k-1}^{\ell}(L \cap \mathcal{L}_X),
\]

\[
p_{k|k-1}^{\ell}(x, \ell) = 1_{\mathcal{L}_X}(\ell) p_{k-1}^{\ell}(x, \ell) + (1 - 1_{\mathcal{L}_X}(\ell)) p_{k-1}(x, \ell),
\]

\[
\eta_{k}^{(\ell)}(\ell) = \int \{p_{k-1}^{\ell}(\cdot, \ell')|_{\{x|\ell' \in \mathcal{L}_X\}} f_{k-1}(x, \ell')\} dx,
\]

\[
\alpha_{k}^{(\ell)}(J) = \{\eta_{k}^{(\ell)}(\ell')(J) \} \sum_{\ell' \in \mathcal{L}_X} \alpha_{k-1}(J)|_{\ell'} \omega_{k-1}^{\ell}(L),
\]

where \( \ell \) is the index for track hypothesis, \( L \) is an instance of label set, \( \mathcal{L}_X \) is track labels from previous time step.

Moreover, the updated multi-object density is given by

\[
\pi_{k|k}(X|Z_k) = \Delta(X) \sum_{\ell \in \mathbb{L}} \omega_{k}^{\ell}(\mathcal{L}_X)[p_{k|k}^{\ell}(X|Z_k)]^X
\]

where \( \Theta_k \) is the space of mappings \( \theta : \mathbb{L}_k \rightarrow \mathbb{L} \), such that \( \theta(i) = \theta(i') \) if \( i = i' \), and

\[
\omega_{k}^{\ell}(L) \propto \delta_{\alpha^{-1}(\mathcal{L}_X)(L)} \omega_{k-1}^{\ell}(L)[\eta_{k}^{\ell}(\ell)],
\]

\[
r_{U,k}(z) = \sum_{u=0,1} r_{U,k,u}(z),
\]

\[
r_{U,k,u}(z) = \frac{\sum_{u' = 0,1} \{r_{k-1,u}(z) \} \{P_{k|k-1,u}(z, \ell') \} \{P_{D,u,k}(\ell') \} \{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \}}{\{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \} \{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \}},
\]

\[
p_{U,k}(X; z) = \frac{\sum_{u' = 0,1} \{r_{k-1,u}(z) \} \{P_{k|k-1,u}(z, \ell') \} \{P_{D,u,k}(\ell') \} \{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \}}{\{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \} \{1 - r_{k-1,u}(z) \} \{P_{D,u,k}(\ell') \}}.
\]
experiments are to illustrate the efficacy of the proposed algorithm in real-world applications: cell migration analysis and video surveillance applications where the prior information of clutter cannot be easily obtained.

4.1 Performance metric

The proposed algorithm is evaluated via two well-known performance metrics, optimal subpattern assignment (OSPA) distance [21] and visual multi-object tracking index [8, 22], respectively, for effective demonstration.

4.1.1 OSPA distance: The OSPA distance is the widely known performance metric in multi-object filtering research. It is a mathematically rigorous metric that not only reflects the localisation error but also the error in the estimated target number is considered. It is the distance between the estimate and the ground truth, thus, the lower OSPA distance shows the better performance.

The OSPA distance \( \overline{d}_p^n \) is defined as follows. Let \( \overline{d}^c(x, y) = \min(e(\| x - y \|)) \) for \( x, y \in \mathcal{X} \), and \( \Pi \) denotes the set of permutations on \( \{1, 2, \ldots, j\} \) for any positive integer \( j \) where \( c \) denotes the maximum error value. Then, for \( p \geq 1, c > 0, \)

- if \( m \leq n: \)
  \[
  \overline{d}_p^c(X, Y) = \left[ \frac{1}{|\Pi|} \min_{\pi \in \Pi} \sum_{i=1}^{m} c^p(\sigma_{\pi(i)} - \sigma_{\pi(i-1)})^p \right]^{\frac{1}{p}}.
  \]

- if \( m > n: \overline{d}_p^c(X, Y) = \overline{d}_p^c(Y, X); \)

- if \( m = n = 0: \overline{d}_p^c(X, Y) = 0. \)

Note that the OSPA distance is used for point-wise estimates. Thus, the OSPA is applied for the numerical example and cell migration analysis that only requires point localisation.

4.1.2 Visual multi-object tracking index: For visual surveillance experiments, publicly available performance indices in visual multi-object tracking [23] for datasets [24–26] are demonstrated in Table 1. Specifically, each attribute in Table 1 is as follows. Recall (correctly tracked objects over total ground truth), precision (correctly tracked objects over total tracking results), and false positives per frame (FPF). We also report the number of identity switches (IDS) and the number of fragmentations (Frag), ratio of tracks with successfully tracked parts for >80% (mostly tracked (MT)), 20% (mostly lost (ML)), or <80% and >20% (partially tracked (PT)). The up (down) arrows in Table 1 mean that higher (lower) the values indicate better performance.

### Table 1: Comparison with the state-of-the-art trackers

| Dataset                | Method        | Recall % | Precision % | FPF | GT | MT % | PT % | ML % | Frag % | IDS % |
|------------------------|---------------|----------|-------------|-----|----|------|------|------|--------|-------|
| PETS09-S2L1            | boosted GLMB  | 90.2     | 89.5        | 0.03| 19 | 20   | 10   | 0.0  | 0.0    | 23    |
|                        | GLMB [19]     | 82.6     | 81.4        | 0.16| 19 | 82.7 | 17.3 | 0.0  | 23     | 12    |
|                        | Rмот [22]     | 80.6     | 85.4        | 0.25| 19 | 84.7 | 15.3 | 0.0  | 20     | 11    |
|                        | Scea [27]     | 83.6     | 89.4        | 0.30| 19 | 86.7 | 13.3 | 0.0  | 20     | 9     |
| TUD-Stadmitte          | boosted GLMB  | 83.4     | 85.6        | 0.10| 10 | 80   | 20   | 0.0  | 12     | 16    |
|                        | GLMB [19]     | 80.0     | 83.0        | 0.16| 16 | 78.0 | 22.0 | 0.0  | 23     | 12    |
|                        | Rмот [22]     | 82.9     | 86.6        | 0.19| 10 | 80   | 20   | 0.0  | 10     | 16    |
|                        | Scea [27]     | 83.0     | 85.0        | 0.22| 10 | 83.7 | 16.3 | 0.0  | 22     | 10    |
| ETH BAHNHOF and SUNNYDAY | boosted GLMB  | 73.1     | 82.6        | 0.78| 124| 60.4 | 34.6 | 5.0  | 110    | 20    |
|                        | GLMB [19]     | 71.5     | 76.3        | 0.88| 124| 58.7 | 27.4 | 13.9 | 112    | 40    |
|                        | Rмот [22]     | 71.5     | 76.3        | 0.98| 124| 57.7 | 37.4 | 4.8  | 68     | 40    |
|                        | Scea [27]     | 72.6     | 80.4        | 0.99| 124| 59.1 | 35.7 | 5.2  | 68     | 38    |

4.2 Object motion model

In the experiments, we used the following dynamic motion model. The target dynamic is described as a coordinated turn model as

\[
\begin{align*}
 f_{k, \bar{k} + 1} = & \mathcal{N}(x_{\bar{k}}, m_{k, \bar{k} + 1}, (x_{\bar{k} + 1} - x_{\bar{k}}), P_{k, \bar{k} + 1}), \\
 P_{k, \bar{k} + 1} = & \text{diag}(\{\sigma_{\theta, k}^G G^T, \sigma_{r, k}^G\}),
\end{align*}
\]

where

\[
 m_{k, \bar{k} + 1} = \left[ F(\mathbf{0}) \mathbf{X}_{\bar{k}} - 1 \mathbf{X}_{\bar{k}} \right] G (T)^{-1}
\]

\[
 F(\mathbf{0}) = \begin{bmatrix}
 1 & \sin T \omega & 0 & -1 - \cos T \omega \\
 0 & \cos T \omega & 0 & -\sin T \omega \\
 0 & 1 - \cos T \omega & 1 & \sin T \omega \\
 0 & \sin T \omega & 0 & \cos T \omega
\end{bmatrix}
\]

\[
 G = \begin{bmatrix}
 T^2 & 0 \\
 T & 0 \\
 0 & T^2 \\
 0 & T
\end{bmatrix}
\]

\[
 x_{\bar{k}} = \left[ \arctan(\sigma_{\theta, k}^G G^T), \sqrt{P_{\bar{k}}^k + P_{\bar{k}}^k} \right],
\]

where \( T \) is the sampling time, \( \sigma_c \) is the standard deviation of the process noise, \( \sigma_o \) is the standard deviation of the turn rate noise. These standard deviation values are determined by the maximum allowable object motion with respect to the image frame rate. For clutter targets, the transition density \( f_{k, \bar{k} + 1} \) is given as a random walk to describe arbitrary motion [15].

4.3 Numerical example

The proposed algorithm is tested with a nonlinear multi-target tracking scenario in [15, 19]. The actual target is observed from the noisy bearing and range information \( z_k = [\theta_k, r_k]^T \) and its likelihood function is given by

\[
 g_k(\mathbf{z}_k | \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; m_{k,k}, P_{k,k})
\]

where

\[
 m_{k,k} = [\arctan(p_{k,k}^r / p_{k,k}^e), \sqrt{p_{k,k}^r + p_{k,k}^e}]
\]

and

\[
 P_{k,k} = \text{diag}(\{\sigma_{\theta, k}^r, \sigma_{r, k}^r\}).
\]

For RMB implementation, we follow the allowable object motion with respect to the image frame rate. In the experiments, we used the following dynamic motion model.

\[
 x_{\bar{k}} = \left[ \arctan(\sigma_{\theta, k}^G G^T), \sqrt{P_{\bar{k}}^k + P_{\bar{k}}^k} \right],
\]

where \( T \) is the sampling time, \( \sigma_c \) is the standard deviation of the process noise, \( \sigma_o \) is the standard deviation of the turn rate noise. These standard deviation values are determined by the maximum allowable object motion with respect to the image frame rate. For clutter targets, the transition density \( f_{k, \bar{k} + 1} \) is given as a random walk to describe arbitrary motion [15].
4.4 Cell migration analysis in a microscopy image

The proposed algorithm is tested with live-cell microscopy image data for cell migration analysis. The proposed boosted GLMB tracking method is tested on a real stem cell migration sequence as illustrated in Fig. 3. The image sequence is recorded for 3 days, i.e. 4320 min and each image is taken in every 16 min.

Performance comparison is conducted with the state-of-the-art multiple hypothesis tracker (MHT) [28]. The same motion and measurement models are used as in the first experiments and spot detection in [28] is applied for the fair comparison. In Fig. 4, reconstructed cell trajectories from the MHT and the boosted GLMB tracker are displayed. The MHT is tuned to obtain the best tracking results. Note that short tracks represent clutter tracks that should be rejected by the tracking algorithm. The boosted GLMB tracker produces significantly less false tracks and alleviate fragmented tracks because the tracker efficiently manages time-varying clutter information and keeps confident tracks. Quantitatively, as can be seen in Table 2, time averaged OSPA distances [21] for both trackers verify that the boosted GLMB shows effective performance even when the clutter rate is unknown. It is important to note that other existing trackers cannot be applied for cell microscopy image data due to the low quality of detections compared to video surveillance scenarios.

4.5 Video surveillance

We also evaluate the proposed algorithm for the tracking of multiple pedestrians in video surveillance. To detect pedestrians, we apply the state-of-the-art pedestrian detector proposed by Piotr et al. called ACF detector [29]. The detector used in the experiment integrates a set of image channels (normalised gradient magnitude, a histogram of oriented gradients, and LUV colour channels) to extract various types of features in order to discriminate objects from the background.

Assuming that the object state \( x_k = [p_x,k \dot{p}_x,k, p_y,k \dot{p}_y,k]^T \) (\( x \)- and \( y \)-positions and velocities) is observed with additive Gaussian noise, the measurement likelihood function is given by

\[
\begin{align*}
& g_k(z_k | x_k) = \mathcal{N}(z_k; H x_k, \Sigma), \tag{21}
\end{align*}
\]

where \( \mathcal{N}(\cdot; m, P) \) denotes a normal distribution with mean \( m \) and covariance \( P \), \( z_k \) is the response from the designated detector; \( H = [1000; 0010] \), i.e. \( x \)- and \( y \)-positions are observed by the detector, \( \Sigma \) is the covariance matrix of the observation noise.

Sample detection results in Fig. 5 contain false positive detections from other types of object with similar shapes as pedestrians. Based on our experiences ACF detector is robust to partial occlusions, however, there are more false positive detections than other single-model based detectors [30]. Thus, it is relatively difficult to remove false positive detections by hard thresholding when it is used for visual scenes with time-varying imaging conditions or moving cameras. In particular, in a visual scene from autonomous vehicles, the average number of clutters (i.e. clutter rate) is varying with respect to the change in the field of view due to the vehicle pose change. The basic assumption behind the existing visual multi-object tracking is that the offline-designed detector, e.g. histogram of oriented gradient (HOG) detector [29, 30], gives reasonably clean detections. Thus, direct data association algorithms such as [8, 9] show reasonable performance with a minor number of clutter measurements. However, in practice, there are false positive detections which make data association results in accurate and computationally intensive.
In this paper, we propose a new multi-object tracking algorithm for unknown clutter rate based on two interconnected random finite set filters. The unknown clutter rate is estimated using the one-step RMB filter. Then, trajectories of objects are estimated using the boosted GLMB filter with the estimated clutter rate online. Two filters are sharing tracking parameters so that there is no need for tuning of clutter parameters. Comparison results via a synthesised non-linear multi-object tracking and visual tracking datasets (video surveillance and biomedical) with state-of-the-art online trackers illustrate that the proposed multi-object tracker shows reliable performance. An interesting future research direction would be the extension of the tracking algorithm to adaptive survival probability, multiple image feature and handling of missed-detections for further improvement.

6 References

[1] Mahler, R.: 'Statistical multisource-multitarget information Fusion' (Artech House, Norwood, MA, 2007)
[2] Bar-Shalom, Y., Fortmann, T.E.: 'Tracking and data Association' (Academic Press, San Diego, 1998)
[3] Mahler, R., Vo, B.-T., Vo, B.-N.: 'CPHD filtering with unknown clutter rate and detection profile', IEEE Trans. Signal Process., 2011, 59, (8), pp. 3497–3513
[4] Reid, D.: 'An algorithm for tracking multiple targets', IEEE Trans. Autom. Control, 1979, 24, (6), pp. 843–854
[5] Breitenstein, M.D., Reichlin, F., Leibe, B., et al.: 'Online multiperson tracking-by-detection from a single, uncalibrated camera', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (9), pp. 1820–1833
[6] Laet, T.D., Bruyninckx, H., Schutter, J.D.: 'Shape-based online multitarget tracking and detection for targets causing multiple measurements: variational Bayesian clustering and lossless data association', IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33, (12), pp. 2477–2491
[7] Hoseinnezhad, R., Vo, B.-N., Vo, B.-T.: 'Visual tracking in background subtracted image sequences via multi-Bernoulli filtering', IEEE Trans. Signal Process., 2013, 61, (2), pp. 392–397
[8] Zhang, L., Li, Y., Nevatia, R.: 'Global data association for multi-object tracking using network flows'. Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, USA, 2008
[9] Milan, A., Roth, S., Schneider, K.: 'Continuous energy minimization for multitarget tracking', IEEE Trans. Pattern Anal. Mach. Intell., 2014, 36, (1), pp. 58–72
[10] Mahler, R.: 'Multi-target Bayes filtering via first-order multi-target moments', IEEE Trans. Aerosp. Electron. Syst., 2003, 39, (4), pp. 1152–1178
[11] Mahler, R.: 'PHD filters of higher order in target number', IEEE Trans. Aerosp. Electron. Syst., 2007, 43, (3), pp. 1523–1543
[12] Rezaei Fighi, S.H., Gould, S., Vo, B.-T., et al.: 'Multi-target tracking with time-varying clutter rate and detection profile: application to time-lapse cell images'. IET Comput. Vis., 2018, Vol. 12 Iss. 5, pp. 728-734

© The Institution of Engineering and Technology 2018
microscopy sequences’, *IEEE Trans. Med. Imaging*, 2015, 34, (6), pp. 1336–1348.

[13] Beard, M., Vo, B.-T., Vo, B.-N.: ‘Multi-target filtering with unknown clutter density using a bootstrap GM-CPHD filter’, *IEEE Signal Process. Lett.*, 2013, 20, (4), pp. 323–326.

[14] Vo, B.-T., Vo, B.-N., Cantoni, A.: ‘Analytic implementations of the cardinalized probability hypothesis density filter’, *IEEE Trans. Signal Process.*, 2007, 55, (7), pp. 3553–3567.

[15] Vo, B.-T., Vo, B.-N., Hoseinnezhad, R., et al.: ‘Robust multi-Bernoulli filtering’, *IEEE J. Select. Top. Signal Process.*, 2013, 7, (3), pp. 399–409.

[16] Kim, D.Y., Jeon, M.: ‘Robust multi-Bernoulli filtering for visual tracking’. Proc. IEEE Conf Control Automation and Information Sciences, Gwangju, South Korea, 2014.

[17] Vo, B.-T., Vo, B.-N., Cantoni, A.: ‘The cardinality balanced multi-target multi-Bernoulli filter and its implementations’, *IEEE Trans. Signal Process.*, 2009, 57, (2), pp. 409–423.

[18] Hoseinnezhad, R., Vo, B.-N., Vo, B.-T., et al.: ‘Visual tracking of numerous targets via multi-Bernoulli filtering of image data’, *Pattern Recognit.*, 2012, 45, (10), pp. 3625–3635.

[19] Vo, B.-T., Vo, B.-N., Phung, D.: ‘Labeled random finite sets and multi-object conjugate priors’, *IEEE Trans. Signal Process.*, 2013, 61, (13), pp. 3460–3475.

[20] Schuhmacher, D., Vo, B.-T., Phung, D.: ‘Labeled random finite sets and the Bayes multi-target tracking filter’, *IEEE Trans. Signal Process.*, 2014, 62, (24), pp. 6554–6567.

[21] Schuhmacher, D., Vo, B.-T., Vo, B.-N.: ‘A consistent metric for performance evaluation of multi-object filters’, *IEEE Trans. Signal Process.*, 2008, 56, (8), pp. 3447–3457.

[22] Yoon, J.H., Yang, M.-H., Lim, J., et al.: ‘Bayesian multi-object tracking using motion context from multiple objects’. Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV), Hawaii, USA, 2015.

[23] Kuo, C.-H., Nevatia, R.: ‘How does person identity recognition help multi-person tracking?’. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Colorado Springs, Colorado, USA, 2011.

[24] Andriluka, M., Roth, S., Schiele, B.: ‘People-tracking-by-detection and people-detection-by-tracking’. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, 2008.

[25] Ess, A., Leibe, B., Gool, L.V.: ‘Depth and appearance for mobile scene analysis’. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Minneapolis, Minnesota, USA, 2007.

[26] Ferryman, J., Shahrokni, A.: ‘PETR09: Dataset and challenge’. IEEE Workshop Performance Evaluation of Tracking and Surveillance, Snowbird, Utah, USA, 2009.

[27] Yoon, J.H., Lee, C.-R., Yang, M.-H., et al.: ‘Online multi-object tracking via structural constraint event aggregation’. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016.

[28] Chenouard, N., Bloch, I., Olivo-Marin, J.-C.: ‘Multiple hypothesis tracking for cluttered biological image sequences’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, 35, (11), pp. 2736–2750.

[29] Dollár, P., Tu, Z., Perona, P., et al.: ‘Integral channel features’. BMVC, London, UK, 2009.

[30] Dalal, N., Triggs, B.: ‘Histogram of oriented gradients for human detection’. Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), San Diego, California, USA, 2005.

[31] Henriques, J.F., Caseiro, R., Martins, P., et al.: ‘High-Speed tracking with k-means and kmedoids correlation filters’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, 37, (3), pp. 583–596.

[32] Yuan, Y., Liu, Y., Wang, Q.: ‘Tracking as a whole: multi-target tracking by modeling group behavior with sequential detection’, *IEEE Trans. Intell. Transp. Syst.*, 2017, 18, (2), pp. 3339–3349.

[33] Wang, Q., Fang, J., Yuan, Y.: ‘Multi-cue based tracking’, *Neurocomputing*, 2014, 131, pp. 227–236.