18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Labeled Multi-Bernoulli Track-Before-Detect for Multi-Target Tracking in Video

Tharindu Rathnayake Amirali Khodadadian Gostar Reza Hoseinnezhad* Alireza Bab-Hadiashar
RMIT University, Victoria 3083, Australia
Emails: s3471339@student.rmit.edu.au, {amirali.khodadadian,rezah,abh}@rmit.edu.au
*Corresponding Author, Tel: +61-9925-6135, reza.hoseinnezhad.com

***Abstract*—This paper presents a labeled multi-Bernoulli filter for track-before-detect with a special focus on visual tracking of multiple targets in video. We show that labeled multi-Bernoulli distribution is a conjugate prior for an image likelihood function with a specific separable form. Following a previously formulated likelihood function (with the desirable separable form) using background subtraction, we apply our proposed labeled multi-Bernoulli filter. Our simulation results show that the proposed solution can successfully track multiple targets in a public visual tracking dataset. Comparative results show superior tracking performance compared with recent competing methods.**

I. Introduction

In multi-target tracking, the data is often preprocessed into point measurements which are commonly known as detections. While this will reduce the computational and memory requirements, some valuable information can be lost during the detection process, specially under low signal-to-noise ratio (SNR). This may drastically reduce the tracker accuracy. A potential remedy is to formulate Bayesian filters that directly use all the information embedded in the *image observations* to update the prior. This approach has led to a relatively large family of multi-object filtering solutions called *track-before-detect* (TBD). In such solutions, the key point is to formulate a likelihood function (for image observations) for which a particular multi-object distribution is a conjugate prior.

Many particle filter-based TBD solutions have been developed for typical multi-target tracking in highly noisy images [1], [2], and an excellent survey of such solutions can be found in [3]. The track-before-detect approach has been also taken by many researchers to devise solutions for specific applications such as tracking with airborne radars [4] and STAP radars [5], acoustic source localization [6], distributed sensor networks [7] and visual tracking [8], [9].

Following the development of finite set statistics (FISST) by Mahler [10]–[12] and its extension to devise random set filtering solutions for particular multi-target tracking applications [13]–[20], various track-before-detect solutions in that framework were also formulated for tracking from noisy radar images [21] and visual tracking with training datasets [22], [23] and without training using background subtraction [8], [24], [25]. These techniques can successfully track multiple targets, without explicitly solving the data association problem (hence, fast computation due to relaxing from the burden of exponential explosion caused by data association in presence of numerous targets). However, the target labels are not managed by the filters and a label management strategy is usually employed after each iteration of the filter to propagate the target labels.

Recently, Vo and Vo [26] introduced two families of labeled random finite set distributions, namely the labeled multi-Bernoulli (LMB) and the generalized LMB (GLMB) distributions and devised the Vo-Vo filter for propagating those distributions in multi-target tracking applications with point measurements [27], [28]. While the LMB distribution is not a conjugate prior for point measurements, Reuter *et al.* [29] showed how the update step can be approximated based on the conjugacy of the GLMB distribution. Soon after, a track-before-detect solution for visual tracking devised for GLMB filters, showing that for image likelihood functions with a particular separable form, the GLMB prior is conjugate [30].

This paper presents a track-before-detect LMB filter with particular application for visual tracking. We prove that for a specific family of likelihood functions for image observations, the LMB prior is conjugate, and show how this can be exploited to implement as a multi-target visual tracking technique in which the labels are automatically managed and propagated within the TBD-LMB filter. Simulation results demonstrate that the proposed algorithm can successfully track multiple targets in a public visual tracking dataset.

II. Background

A random finite set is a spatial point pattern on the space of interest. Intuitively, it is a set with a random number of elements where the elements are also random variables. In this paper we use FISST density notion to develop the labeled RFS and for simplicity, the difference between FISST density and probability density is disregarded.

A. Notation

In this paper we use lowercase letters to represent single-object states (e.g. $\boldsymbol{x}$ and x), uppercase letters to represent multi-object states (e.g. $\boldsymbol{X}$ and X), blackboard bold letters to represent the spaces (e.g. $\mathbb{N}, \mathbb{X}$ and $\mathbb{L}$) and bold letters (e.g. $\boldsymbol{x}$ and $\boldsymbol{X}$) are used to denote labeled entities, so that they are distinguishable from the unlabeled entities [26].

B. Labeled RFS

In order to integrate a unique label to each target, each state $x \in \mathbb{X}$ is coupled with a unique label $\ell = (\ell_t, \ell_b) \in \mathbb{L} =$

$\{\alpha_i : i \in \mathbb{N}\}$, where $\mathbb{N}$ denotes the set of positive integers and all the α_i's are distinct [26]. Each label has two elements by convention: ℓ_t is the time stamp at which the object was born, and ℓ_b is an index to count the objects born at the same time stamp, which is used to distinguish targets born at the same time.

A labeled RFS $\mathbf{X}$ with state space $\mathbb{X}$ and discrete label space $\mathbb{L}$ is a subset of $\mathbb{X} \times \mathbb{L}$ with distinct labels, if $\mathbf{X}$ and its labels $\mathcal{L}(\mathbf{X}) = \{\mathcal{L}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$ have the same cardinality, which can be mathematically denoted as $|\mathcal{L}(\mathbf{X})| = |\mathbf{X}|$ or $\Delta(\mathbf{X}) \triangleq \delta_{|\mathbf{X}|}\mathcal{L}(\mathbf{x}) = 1$.

The density of a labeled RFS X is a function

$$\boldsymbol{\pi} : \mathcal{F}(\mathbb{X} \times \mathbb{L}) \rightarrow \mathbb{R}^{+} \cup \{0\}$$

with unit integration over the labeled multi-object state space, i.e. $\int_{\mathbb{X}\times\mathbb{L}} \boldsymbol{\pi}(\boldsymbol{X})\delta \boldsymbol{X} = 1$ where the set integral is defined for any function $\boldsymbol{h} : \mathcal{F}(\mathbb{X} \times \mathbb{L}) \rightarrow \mathbb{R}$ by [26]:

$$\int h(\boldsymbol{X})\delta(\boldsymbol{X}) = \sum_{i=0}^{\infty} \frac{1}{i!} \int \boldsymbol{h}(\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i\}) d(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i). \tag{1}$$

C. Labeled Multi-Bernoulli Distribution

A labeled multi-Bernoulli RFS $\mathbf{X}$ with state space $\mathbb{X}$, label space $\mathbb{L}$ and finite parameter set $\{(r^{(\varsigma)}, p^{(\varsigma)})\} : \varsigma \in \Psi$, is a multi-Bernoulli RFS on $\mathbb{X}$, augmented with labels corresponding to the successful non-empty Bernoulli components. If the Bernoulli component $(r^{(\varsigma)}, p^{(\varsigma)})$ yields a non-empty set, then the label of the corresponding state is given by $\alpha(\varsigma)$, where $\alpha : \Psi \rightarrow \mathbb{L}$ is a 1-1 mapping [28]. The LMB density with the above mentioned parameters is given by [28]:

$$\boldsymbol{\pi}(\boldsymbol{X}) = \Delta(\mathbf{X}) 1_{\alpha(\Psi)}(\mathcal{L}(\boldsymbol{X}))[\Phi(\boldsymbol{X}; \cdot)]^{\Psi}. \tag{2}$$

where

$$\Phi(\boldsymbol{X}; \varsigma) = \begin{cases} 1 - r^{(\varsigma)} & \text{if } \alpha(\varsigma) \notin \mathcal{L}(\boldsymbol{X}), \\ r^{(\varsigma)} p^{(\varsigma)}(x) & \text{if } (x, \alpha(\varsigma)) \in \mathcal{L}(\boldsymbol{X}) \end{cases}$$

Under the assumption that α mapping is an identity mapping, the above multi-target density can be compactly written as:

$$\boldsymbol{\pi}(\boldsymbol{X}) = \triangle(\boldsymbol{X}) w(\mathcal{L}(\mathcal{X})) p^{\boldsymbol{X}}, \tag{3}$$

where

$$w(L) = \prod_{i\in\mathbb{L}} \left(1 - r^{(i)}\right) \prod_{\ell\in L} \frac{1_{\mathbb{L}}(\ell)\ r^{(\ell)}}{1 - r^{(\ell)}} \tag{4}$$

$$p(x, \ell) = p^{(\ell)}(x). \tag{5}$$

For the sake of simplicity in notation, we denote the above density by $\boldsymbol{\pi} = \{r^{(\ell)}, p^{(\ell)}\}_{\ell\in\mathbb{L}}$.

D. LMB Propagation

Let $\mathbb{L}_k = \{k\} \times \mathbb{N}$ denote the label space for the targets born at time k, and $\boldsymbol{x} \in \mathbb{X} \times \mathbb{L}_k$ is the state of a target born at time k. The label space for all the targets at time k, including all the label spaces at previous label spaces is denoted by $\mathbb{L}_{0:k}$ and is recursively constructed by $\mathbb{L}_{0:k} = \mathbb{L}_{0:k-1} \cup \mathbb{L}_k$.

Let y_k denotes the image observation at time k, and $y_{1:k}$ denotes the ensemble of all observations acquired up to time k. Denote the density of the multi-object state at time k by $\boldsymbol{\pi}_k(\boldsymbol{X}|y_{1:k})$. In a Bayesian paradigm, the labeled density is recursively predicted and updated as follows [28]:

$$\boldsymbol{\pi}_{k+1|k}(\boldsymbol{X}|y_{1:k}) = \int \boldsymbol{f}_{k+1|k}(\boldsymbol{X}|\boldsymbol{X}_k) \boldsymbol{\pi}_k(\boldsymbol{X}_k|y_{1:k}) \delta \boldsymbol{X}_k \tag{6}$$

$$\boldsymbol{\pi}_{k+1}(\boldsymbol{X}|y_{1:(k+1)}) = \frac{g_{k+1}(y_{k+1}|\boldsymbol{X}) \boldsymbol{\pi}_{k+1|k}(\boldsymbol{X})}{\int g_{k+1}(y_{k+1}|\boldsymbol{X}_k) \boldsymbol{\pi}_{k+1|k}(\boldsymbol{X}_k) \delta \boldsymbol{X}_k} \tag{7}$$

where $\boldsymbol{f}_{k+1|k}(\cdot|\cdot)$ is the multi-object transition density from time k to $k+1$, $g_{k+1}(\cdot|y_{k+1})$ is the multi-object likelihood function at time k for the given image observation y_{k+1}, and the integrals are set integrals as defined in (1). Henceforward, for the sake of brevity in notations, we will drop the "given observation parts" ($|y_{1:k}$ and $|y_{1:(k+1)}$) of the density arguments, as the dependence of evolved densities on the past and current observations is obvious.

According to Reuter *et al.* (see [29], proposition 2), if the multi-object density is LMB with state space $\mathbb{X}$ and label space $\mathbb{L}$ and its density parametrized by $\boldsymbol{\pi} = \{r^{(\ell)}, p^{(\ell)}\}_{\ell\in\mathbb{L}}$ and provided that the multi-object birth model is an LMB with the same state space $\mathbb{X}$, a label space $\mathbb{B}$ that is disjoint from $\mathbb{L}$ (i.e. $\mathbb{X} \cap \mathbb{B} = \emptyset$) and parametrized density $\boldsymbol{\pi}_B = \{r_{B}^{(\ell)}, p_{B}^{(\ell)}\}_{\ell\in\mathbb{B}}$, then the predicted multi-object density is also an LMB with state space $\mathbb{X}$ and label space $\mathbb{L}_+ = \mathbb{L} \cup \mathbb{B}$. Furthermore, the parameters of the predicted LMB density are $\boldsymbol{\pi}_+ = \{r_{+,S}^{(\ell)}, p_{+,S}^{(\ell)}\}_{\ell\in\mathbb{L}} \bigcup \{r_{B}^{(\ell)}, p_{B}^{(\ell)}\}_{\ell\in\mathbb{B}}$, where

$$r_{+,S}^{(\ell)} = \eta_{S}^{(\ell)}\ r^{(\ell)} \tag{8}$$

$$p_{+,S}^{(\ell)}(x) = \langle p_{S}(\cdot, \ell) f(x|\cdot, \ell), p^{(\ell)}(\cdot)\rangle / \eta_{S}^{(\ell)} \tag{9}$$

in which $p_{S}(x, \ell)$ is the state-dependent *probability of survival* for an existing Bernoulli component with label ℓ, $f(x_{k+1}|x_k, \ell)$ denotes the single-object transition density, and $\eta_{S}(\ell) = \langle p_{S}(\cdot, \ell), p^{(\ell)}(\cdot)\rangle$. This is simply equivalent to predicting the existing unlabeled Bernoulli components according to equations of multi-Bernoulli filter and retaining the labels of the predicted components, then unifying them with the birth Bernoulli components that come with new labels.

III. LMB Update with Image Observations

Let us assume that the multi-object likelihood function for a given image observation y has the following separable form:

$$g(y|\boldsymbol{X}) = f(y) \prod_{(x,\ell)\in\boldsymbol{X}} g_y(x, \ell). \tag{10}$$

In a visual tracking application, the colour image sequence can be processed via background subtraction and the resulting grey-scale image can be used as the image observation for which a separable likelihood function in the form of (10) can be formulated (more details on this formulation can be found in [24], [30]). The update formula for a generalized labeled multi-Bernoulli (GLMB) density with the above separable likelihood has been derived in [30]–see equations (21)-(24). Treating the LMB density as a GLMB with only one term, we can use those results to derive the posterior density for an LMB prior with the likelihood function in the form of (10). Suppose that the LMB prior is given by equations (3)-(5). The posterior density will then be given by:

$$\boldsymbol{\pi}(\boldsymbol{X}|y) \propto \Delta(\boldsymbol{X}) w_y(\mathcal{L}(\boldsymbol{X}))\,[p(\cdot|y)]^{\boldsymbol{X}} \tag{11}$$

where

$$w_y(L) = [\eta_y]^L\ w(L) \tag{12}$$

$$p(x,\ell|y) = \frac{p^{(\ell)}(x) g_y(x,\ell)}{\eta_y(\ell)} \tag{13}$$

$$\eta_y(\ell) = \langle p^{(\ell)}(\cdot), g_y(\cdot,\ell)\rangle. \tag{14}$$

Proposition 1. The posterior density introduced by equations (11)-(14) represents an LMB density.

Proof. Substituting $w(L)$ from (4) in (12) and replacing the resulting $w_y(\mathcal{L}(\boldsymbol{X}))$ term in the update equation (11) leads to:

$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{X}|y) &\propto \Delta(\boldsymbol{X})[\eta_y]^{\mathcal{L}(\boldsymbol{X})}\ w(\mathcal{L}(\boldsymbol{X}))\,[p(\cdot|y)]^{\boldsymbol{X}} \\
&\propto \Delta(\boldsymbol{X})\prod_{i\in\mathbb{L}}(1-r^{(i)})\prod_{\ell\in\mathcal{L}(\boldsymbol{X})}\frac{1_{\mathbb{L}}(\ell) r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}}[p(\cdot|y)]^{\boldsymbol{X}} \\
&\propto \Delta(\boldsymbol{X})\prod_{i\in\mathbb{L}}(1-r^{(i)})\prod_{\ell\in\mathcal{L}(\boldsymbol{X})}\frac{1_{\mathbb{L}}(\ell)\frac{r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}}{\frac{1-r^{(\ell)}}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}} \\
&\quad [p(\cdot|y)]^{\boldsymbol{X}} \\
&\propto \Delta(\boldsymbol{X})\prod_{i\in\mathbb{L}}(1-r^{(i)})\prod_{\ell\in\mathcal{L}(\boldsymbol{X})}\frac{1_{\mathbb{L}}(\ell)\frac{r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}}{1-\frac{r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}} \\
&\quad [p(\cdot|y)]^{\boldsymbol{X}} \\
&= \Delta(\boldsymbol{X})\prod_{i\in\mathbb{L}}\left(1-\frac{r^{(i)}\eta_y(i)}{1-r^{(i)}+r^{(i)}\eta_y(i)}\right) \\
&\quad \prod_{\ell\in\mathcal{L}(\boldsymbol{X})}\frac{1_{\mathbb{L}}(\ell)\frac{r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}}{1-\frac{r^{(\ell)}\eta_y(\ell)}{1-r^{(\ell)}+r^{(\ell)}\eta_y(\ell)}}\,[p(\cdot|y)]^{\boldsymbol{X}}.
\end{aligned}$$

We note that in the last step, the proportionality turns into equality as we find the normalizing term to be $\prod_{i\in\mathbb{L}}\left(1-r^{(i)}+r^{(i)}\eta_y(i)\right)^{-1}$. The validity of this derivation stems from the fact that the resulting density has the same form of the LMB density (3) and integrates to 1[1]. Thus, the posterior density is LMB. □

[1]Note that besides finding the arrangement of probabilities of existence in the product terms being similar to (3), the $p(\cdot|y)$ terms each integrate to 1

Remark 1. The posterior LMB can be parametrized by $\boldsymbol{\pi}_{\text{updated}} = \{r^{(\ell)}_{\text{updated}}, p^{(\ell)}_{\text{updated}}\}_{\ell\in L}$ where

$$r^{(\ell)}_{\text{updated}} = \frac{r^{(\ell)}\ \langle p^{(\ell)}(\cdot), g_y(\cdot,\ell)\rangle}{1-r^{(\ell)}+r^{(\ell)}\ \langle p^{(\ell)}(\cdot), g_y(\cdot,\ell)\rangle} \tag{15}$$

$$p^{(\ell)}_{\text{updated}}(x) = \frac{p^{(\ell)}(x)\ g_y(x,\ell)}{\langle p^{(\ell)}(\cdot), g_y(\cdot,\ell)\rangle}. \tag{16}$$

IV. IMPLEMENTATION

Assume that at time step k the multi-Bernoulli posterior multi-target density $\boldsymbol{\pi}_k = \{r^{(\ell)}, p^{(\ell)}\}_{\ell\in\mathbb{L}}$ is given. In a sequential Monte Carlo (SMC) implementation, each density $p^{(\ell)}$ is represented by a set of weighted particles $\left\{(w^{(\ell,j)}, x^{(\ell,j)})\right\}$, and its particle approximation is given by:

$$p^{(\ell)} \approxeq \sum_j w^{(\ell,j)}\delta_{x^{(\ell,j)}}(x). \tag{17}$$

If the proposal densities $q^{(\ell)}_+$ and $b^{(\ell)}_+$ are given, then the predicted labeled multi-Bernoulli multi-target density $\boldsymbol{\pi}_{k+1|k} = \left\{\left(r^{(\ell)}_{+,S}, p^{(\ell)}_{+,S}\right)\right\}_{\ell\in\mathbb{L}} \cup \left\{\left(r^{(\ell)}_B, p^{(\ell)}_B\right)\right\}_{\ell\in\mathbb{B}}$ can be computed as

$$r^{(\ell)}_{+,S} = r^{(\ell)}\sum_j w^{(\ell,j)} p_{S,k}(x^{(\ell,j)}), \tag{18}$$

$$p^{(\ell)}_{+,S} = \sum_j \bar{w}^{(\ell,j)}_{P,+}\delta_{x^{(\ell,j)}_+}(x), \tag{19}$$

$$r^{(\ell)}_B = \text{parameter given by the birth model} \tag{20}$$

$$p^{(\ell)}_B = \sum_j \bar{w}^{(\ell,j)}_B\delta_{x^{(\ell,j)}_B}(x), \tag{21}$$

where

$$x^{(\ell,j)}_+ \sim q^{(\ell)}_+(\cdot|x^{(\ell,j)}_B, y) \tag{22}$$

$$w^{(\ell,j)}_{P,+} = \frac{w^{(\ell,j)} f(x^{(\ell,j)}_+|x^{(\ell,j)}) p_{S,k}(x^{(\ell,j)})}{q^{(\ell)}(x^{(\ell,j)}_+|x^{(\ell,j)}, y)}, \tag{23}$$

$$\bar{w}^{(\ell,j)}_{P,+} = w^{(\ell,j)}_{P,+} \setminus \sum_j w^{(\ell,j)}_{P,+}, \tag{24}$$

$$x^{(\ell,j)}_B \sim b^{(\ell)}_+(\cdot|y), \tag{25}$$

$$w^{(\ell,j)}_B = \frac{p_B(x^{(\ell,j)}_B)}{b^{(\ell)}_+(x^{(\ell,j)}_B|y)}, \tag{26}$$

$$\bar{w}^{(\ell,j)}_B = w^{(\ell,j)}_B \setminus \sum_j w^{(\ell,j)}_B. \tag{27}$$

Suppose that the predicted labeled multi-Bernoulli multi-object density $\boldsymbol{\pi}_+ = \left\{\left(r^{(\ell)}_+, p^{(\ell)}_+\right)\right\}_{\ell\in\mathbb{L}_+}$ is given with its density components , $p^{(\ell)}_+$ being represented by a set of weighted particles,

$$p^{(i)}_+ \approxeq \sum_j w^{(\ell,j)}_+\delta_{x^{(\ell,j)}_+}(x). \tag{28}$$

Then the updated labeled multi-Bernoulli multi-object parameters $\boldsymbol{\pi}(\cdot|y) = \left\{\left(r^{(\ell)}, p^{(\ell)}\right)\right\}_{\ell\in\mathbb{L}_+}$ is given by

$$r^{(\ell)} = \frac{r_+^{(\ell)} \varrho_+^{(\ell)}}{1 - r_+^{(\ell)} + r_+^{(\ell)} \varrho_+^{(\ell)}}, \tag{29}$$

$$p^{(\ell)} = \frac{1}{\varrho_+^{(\ell)}} \sum_j w_+^{(\ell,j)} g_y(x_+^{(\ell,j)}) \delta_{x_+^{(\ell,j)}}(x), \tag{30}$$

where (31)

$$\varrho_+^{(\ell)} = \sum_j w_+^{(\ell,j)} g_y(x_+^{(\ell,j)}). \tag{32}$$

A. Techniques to Guarantee Computational Tractability

The updated particles are resampled so that each labeled Bernoulli component retains a certain number of particles which is proportional to its existence probability $r^{(\ell)}$. Further, the number of particles is constrained between a minimum of $L_{\min}$ and a maximum of $L_{\max}$, following [21], [23], [25], [31]. During the pruning step, we discard all the labeled Bernoulli components which have a probability of existence $r^{(\ell)}$ less than a threshold denoted by r_{th}. This enables us to keep track of the growing number of labeled Bernoulli components which has an exponential growth with the number of targets, ensuring that only the Bernoulli components with a high probability of representing an existing target are retained and all the other components are discarded.

At the merging step, labeled Bernoulli components with an overlapping ratio of more than a threshold are merged together. In our visual tracking experiments, rectangular target blobs were used to represent targets. We calculate the overlapping ratio as the ratio of the area of intersection between two components to the area of the smaller rectangular target. In our experiments, targets with an overlapping ratio of more than 60% were merged. During the merging procedure, the existence probability of the merged labeled Bernoulli component is set to be the minimum between 0.999 and the addition of the existence probabilities of the two merged labeled Bernoulli components. All the particles used to represent the two merged labeled Bernoulli components are present in the resultant labeled Bernoulli component. The particle weights are scaled according to the probabilities of existence, and they sum up to 1.

When selecting a threshold for the overlapping ratio, we must consider the targets which evolve very close to each other. If a low overlapping ratio is used, the targets which are very close to each other may be tracked as a single target and this will reduce the accuracy of our tracker. On the other hand, if we use a very high overlapping ratio, two labeled Bernoulli components which represent the same target may not be merged and hence, the particular single target will be tracked as two different targets, resulting in a degradation of accuracy.

B. State Extraction

The most widely used state extraction methodology of extracting the targets with an existence probability higher than a certain threshold is used in our work. The threshold is application specific and a target $\bar{\boldsymbol{X}}$ is extracted

$$\bar{\boldsymbol{X}} = \Big\{ (\bar{x}, \ell) : r^{(\ell)} > \varepsilon \Big\}, \tag{33}$$

where ε is the threshold and $\bar{x} = \arg_x \max p^{(\ell)}(x)$. Using a high ε will prune the clutter tracks while delaying the inclusion of new tracks, whereas a low ε will include new tracks immediately at the expense of including more clutter tracks.

V. Simulation Results

We elaborate the results of our tracker using video sequences available in CAVIAR benchmark dataset. In our MATLAB implementation, targets are assumed to be rectangular blobs with a variable width and height, though they are constrained. The variability of the width and the height enables us to represent the targets moving towards and away from the camera–see Figure 1. The target state is a 4D vector comprising of the location of the top left corner of the blob (x,y), the width w and the height h of the blob.

We use a random walk model $x(k+1) = x(k) + e(k)$ where $e(k)$ is the Gaussian noise with zero mean and a covariance of $\varSigma = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_w^2, \sigma_H^2)$. All the targets are assumed to follow the aforementioned system model and the rationale behind using this model is that it is general enough to be compatible with almost all the human targets. Human targets are free to move in any direction and are mainly constrained in the amount of distant that they can travel in a single time step. Other common motion models such as constant velocity model are designed for objects like airplanes and cars that are limited in their maneuvers. However with much higher computational requirements we can use social behavioral models such as the one suggested in [32].

To detect the targets which enter into the camera field of view and to re-detect the targets which have been missed, we use a birth model. We assume that one target may appear in each of the four quarters of the image in each time step with a constant existence probability of 0.02 throughout the simulation. The location of the target is assumed to be uniformly distributed within the particular quarter. When additional information is available, such as positions of the gate entrances and elevator access points, we can use other complex birth models with different probability densities, such as the one described in [33], in which the birth process is modeled as a non-uniform probability distribution.

In our implementation, we do not use any information on the appearance of the target (such as the color), but size constraints on the targets are imposed. Upon applying background substation, we utilize morphological closing operation (See [34, p. 136]) with a square structuring element of size 3 pixels to remove all the small targets that appear due to noise after the thresholding step. The maximum and minimum number of particles per target are set to $L_{\max} = 1000$ and $L_{\min} = 100$, and probability of survival is assumed to be constant at $P_S = 0.99$.

To evaluate the accuracy of our tracker, we use CLEAR MOT metrics, which were first introduced in [35], Multiple

Table I
TRACKING PERFORMANCE COMPARISON FOR CAVIAR DATASET

Method	MOTA %	MOTP %
GM-PHD-STS	33.93	65.78
GM-PHD-DD	67.91	66.96
TBD-LMB (Ours)	76.63	85.37

Object Tracking Accuracy and Precision (MOTA and MOTP). MOTP metric is a measure of the error in consistency of tracked trajectories with the ground truth and MOTA is a measure on the number of false positives, detections, and identity switches throughout the tracking period. The MOTP metric is defined by:

$$MOTP = \frac{\sum_{i,k} d_k^i}{\sum_k c_k} \tag{34}$$

where c_k is the number of matches found for time k and d_i is the distance between the target x_i and its corresponding hypothesis h_i. The MOTA metric is defined by:

$$MOTA = 1 - \frac{\sum_k (m_k + fp_k + mme_k)}{\sum_k g_k} \tag{35}$$

where m_k, fp_k, mme_k and g_k are the number of misses, false positives, mismatches and number of ground truth objects at time k, respectively. These metrics have been widely utilized by the visual tracking community [35], [36] and for fair comparisons, we have computed them in our studies.

The test sequence is from CAVIAR dataset[2]. It shows four persons entering a lobby, from the far left corner and walking as a group to exit from an exit near the camera. The persons are entering from a far corner and hence the size of the targets are initially very small. Moreover, they exit from the seen from a point near to the camera and the target sizes are much larger. It is important to note that there is a small occlusion in this scenario. Although, our tracker is not natively designed to handle occlusions, it performs generally better than recently established random set-based visual tracking methods, namely the standard GM-PHD filter based Tracking System (GM-PHD-STS) [9], and the GM-PHD visual tracker with Data Driven importance sampling function (GM-PHD-DD) [37]. In a number of other visual tracking methods, the same dataset has been used for comparison purposes. The aforementioned methods were chosen since they share the same approach as our proposed method, that is, they all use random set-based multi-object filters as the underlying mathematical framework. The tracking results are given in Table I. They demonstrate that our tracker outperforms both trackers in terms of MOTA and MOTP values.

VI. CONCLUSION

A track-before-detect solution was proposed with labeled multi-Bernoulli assumption for the multi-object distribution. For a family of multi-object likelihood functions for image observations, which have a particular separable form, the LMB prior was proven to be conjugate, and the multi-Bernoulli parameters of the posterior were formulated. The resulting method was implemented using the Sequential Monte Carlo technique and applied to track a number of people in a public visual tracking dataset (CAVIAR). The results are promising, showing that our proposed TBD-LMB filter outperforms the competing random set-based tracking methods in terms of the MOTP and MOTA metrics.

[2]http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

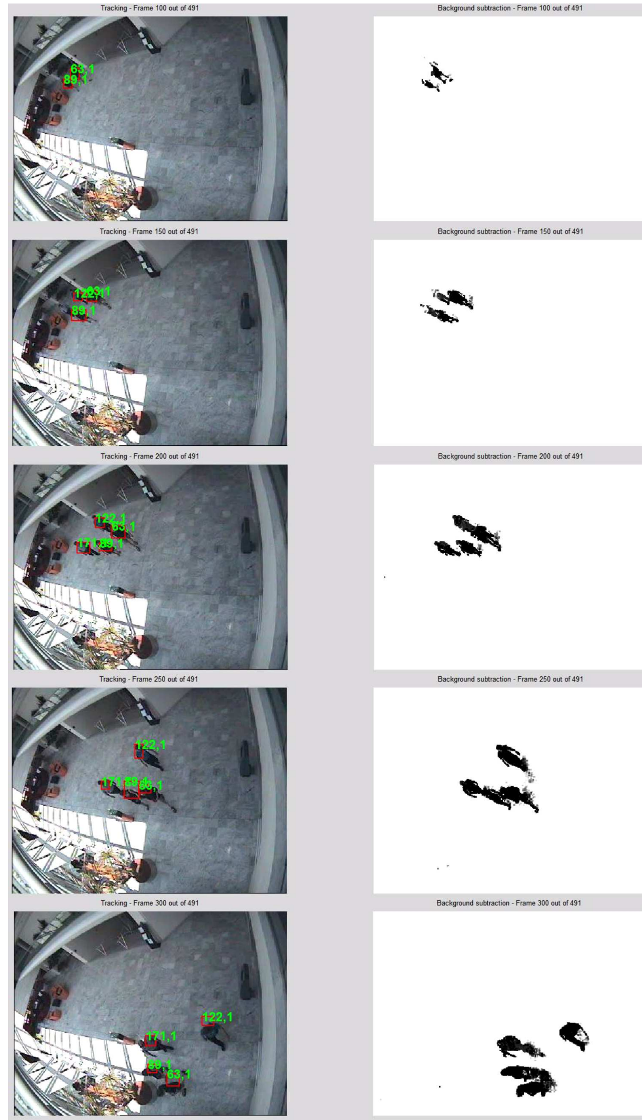

Figure 1. Screen shots of CAVIAR dataset tracking at t = 100, 150, 200, 250 and 300

ACKNOWLEDGMENT

This work was supported by ARC Discovery Projects grant DP130104404, and ARC Linkage Projects grant LP130100521.

REFERENCES

[1] D. Salmond and H. Birch, "A particle filter for track-before-detect (i)," in *Proceedings of the American control conference*, vol. 5, 2001, pp. 3755–3760.

[2] Y. Boers and H. DRIESSEN, "Particle filter track-before-detect application using inequality constraints," *IEEE transactions on aerospace and electronic systems*, vol. 41, no. 4, pp. 1481–1487, 2005.

[3] S. J. Davey, M. G. Rutten, and B. Cheung, "A comparison of detection performance for several track-before-detect algorithms," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 41, 2008.

[4] S. Buzzi, M. Lops, and L. Venturino, "Track-before-detect procedures for early detection of moving target from airborne radars," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 3, pp. 937–954, 2005.

[5] D. Orlando, L. Venturino, M. Lops, and G. Ricci, "Track-before-detect strategies for stap radars," *Signal Processing, IEEE Transactions on*, vol. 58, no. 2, pp. 933–938, 2010.

[6] M. F. Fallon and S. Godsill, "Acoustic source localization and tracking using track before detect," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1228–1242, 2010.

[7] T. A. Wettergren, "Performance of search via track-before-detect for distributed sensor networks," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 1, pp. 314–325, 2008.

[8] J. Wong, B. T. Vo, B. N. Vo, and R. Hoseinnezhad, "Multi-bernoulli based track-before-detect with road constraints," in *Information Fusion (FUSION), 2012 15th International Conference on*. IEEE, 2012, pp. 840–846.

[9] X. Zhou, Y. Li, B. He, and T. Bai, "Multi-target visual tracking with game theory-based mutual occlusion handling," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 4201–4206.

[10] R. P. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 39, no. 4, pp. 1152–1178, 2003.

[11] R. Mahler, "Phd filters of higher order in target number," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 4, pp. 1523–1543, 2007.

[12] R. P. Mahler, *Statistical multisource-multitarget information fusion*. Artech House, Inc., 2007.

[13] B.-N. Vo, S. Singh, and A. Doucet, "Sequential monte carlo methods for multitarget filtering with random finite sets," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 4, pp. 1224–1245, 2005.

[14] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3553–3567, 2007.

[15] ——, "The cardinality balanced multi-target multi-bernoulli filter and its implementations," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 409–423, 2009.

[16] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation in multi-object filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.

[17] W.-K. Ma, B.-N. Vo, S. S. Singh, and A. Baddeley, "Tracking an unknown time-varying number of speakers using tdoa measurements: A random finite set approach," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3291–3304, September 2006.

[18] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the Cardinalized Probability Hypothesis Density filter," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3553–3567, 2007.

[19] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Proc.*, vol. 54, no. 11, pp. 4091 – 104, 2006.

[20] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Tran. AES*, vol. 41, no. 4, pp. 1224–1245, 2005.

[21] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5129–5141, 2010.

[22] R. Hoseinnezhad, B.-N. Vo, D. Suter, and B.-T. Vo, "Multi-object filtering from image sequence without detection," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1154–1157.

[23] R. Hoseinnezhad, B.-N. Vo, B.-T. Vo, and D. Suter, "Visual tracking of numerous targets via multi-bernoulli filtering of image data," *Pattern Recognition*, vol. 45, no. 10, pp. 3625–3635, 2012.

[24] R. Hoseinnezhad, B.-N. Vo, and T. N. Vu, "Visual tracking of multiple targets by multi-Bernoulli filtering of background subtracted image data," in *International conference on Swarm Intelligence (ICSI'2011), Lecture Notes in Computer Science (LNCS)*, ser. 2, Y. Tan, Ed., vol. 6729, Chongqing, China, June 2011, pp. 509–518.

[25] R. Hoseinnezhad, B.-N. Vo, and B.-T. Vo, "Visual tracking in background subtracted image sequences via multi-bernoulli filtering," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 392–397, 2013.

[26] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3460–3475, 2013.

[27] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, 2014.

[28] ——, "Labeled random finite sets and the bayes multi-target tracking filter," 2013.

[29] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-bernoulli filter," *Signal Processing, IEEE Transactions on*, vol. 62, no. 12, pp. 3246–3260, 2014.

[30] F. P. Yuthika Punchihewa and R. Hoseinnezhad, "Multiple target tracking in video data using labeled random finite set," in *Control, Automation and Information Sciences (ICCAIS), 2013 International Conference on*, 2014, pp. 13–18.

[31] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on bernoulli filters: theory, implementation and applications," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3406–3430, 2013.

[32] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 261–268.

[33] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking," *Information Fusion*, vol. 8, no. 1, pp. 2–15, 2007.

[34] A. C. Bovik, *Handbook of image and video processing*. Academic press, 2010.

[35] B. Keni and S. Rainer, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, 2008.

[36] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan, "The clear 2006 evaluation," in *Multimodal Technologies for Perception of Humans*. Springer, 2007, pp. 1–44.

[37] Y.-D. Wang, J.-K. Wu, A. A. Kassim, and W. Huang, "Data-driven probability hypothesis density filter for visual tracking," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 8, pp. 1085–1095, 2008.