Global Data Association for the Probability Hypothesis Density Filter Using Network Flows

Nicolai Wojke[1] and Dietrich Paulus[1]

***Abstract*— The Probability Hypothesis Density (PHD) filter is an efficient formulation of multi-target state estimation that circumvents the combinatorial explosion of the multi-target posterior by operating on single-target space without maintaining target identities. In this paper, we propose a multi-target tracker based on the PHD filter that provides instantaneous state estimation and delayed decision on data association. For this purpose, we reformulate the PHD recursion in terms of single-target track hypotheses and solve a min-cost flow network for trajectory estimation where measurement likelihoods and transition probabilities are based on multi-target state estimates. In this manner, the presented approach combines global data association with efficient multi-target filtering. We evaluate the approach on a publicly available pedestrian tracking dataset to present state estimation and data association capabilities.**

I. Introduction

The detection and tracking of multiple objects is a key challenge in many computer vision and robotics applications, such as driver assistance, autonomous driving, visual navigation, and object activity recognition. Traditionally, data association techniques such as the Joint Probabilistic Data Association Filter (JPDAF) [6] and Multiple Hypothesis Tracking (MHT) [18] have been applied to this topic. In the JPDAF, a single state hypothesis is generated by weighting individual measurements by their association likelihoods. In MHT, all possible hypotheses are tracked, but pruning schemes must be applied for computational tractability.

Due to recent progress in object detection and online learning, it has become increasingly popular to integrate appearance information into the problem formulation. Within the *tracking-by-detection* paradigm, multi-target tracking is typically formulated as global data association that requires discrete optimization. In particular, min-cost flow network formulations have become popular, because they allow for exact and efficient global optimization. A prominent formulation of min-cost flow tracking has originally been proposed by Zhang et al. [28]. They create a graph on the set of all measurements and find globally optimal trajectories using a min-cost flow algorithm. This formulation has been adopted by others in order to obtain better run-times when the number of objects is unknown [17] and to couple object detection and data association [25]. Dynamic programming [2], linear programming [7], and k-shortest path algorithms [3] have also been used to generate target trajectories from observations or fixed space-time locations.

[1]Active Vision Group, Institute for Computational Visualistics, University of Koblenz-Landau, 56070 Koblenz, Germany {nwojke, paulus}@uni-koblenz.de

Others apply graphical models to the data association problem: Yang and Nevatia [26], [27] formulate multi-target tracking as pair-wise conditional random field on low-level *tracklets*, i.e., reliable short term track fragments, to distinguish between multiple objects. Andriyenko et al. [1] apply a Markov random field formulation where they alternate between optimizing discrete data association and continuous state estimation. Milan et al. [12] formulate the discrete-continuous problem as conditional random field. Dehghan et al. [5] use structured learning for data association and propose a Lagrangian relaxation optimization.

An alternative approach to multi-target tracking has been developed within the signal processing community. Based on random finite set theory (e.g., [11]), an efficient approach that circumvents explicit data association has been proposed to estimate the state of an unknown, time varying number of targets. The Probability Hypothesis Density (PHD) filter [10] and its concrete Sequential Monte Carlo (SMC-PHD) [23] and Gaussian mixture (GM-PHD) [22] implementations have been successfully applied to this task. While the PHD filter does not provide track identities itself, there has been an effort to develop data association and track management schemes for the PHD filter. For example, Panta et al. [15] assign identities to individual particles of the SMC-PHD to obtain track-valued estimates and, similarly, in [14] the authors assign identities to individual mixture components of the GM-PHD. In [16], MHT is run on top of the output of a GM-PHD filter to provide target identities. More recently, Reuther et al. [19] proposed a multi-target tracker based on random finite set theory that propagates full multi-target densities, including their identifying label.

In this paper, we formulate multi-target state and trajectory estimation as separate, but related problems. We use the PHD filter as an online multi-target state estimator that provides (potentially safety-critical) information about dynamic objects. For delayed data association, we solve a min-cost flow network on the set of measurements. Multi-target state estimation and data association are coupled by the underlying model where measurement likelihoods and transition probabilities are based on multi-target estimates. The contributions of this paper are: (i) We provide a reformulation of the PHD recursion in terms of single-target track hypotheses, (ii) we show how to apply this reformulation to the min-cost flow tracking formulation of Zhang et al. [28]. To our knowledge, this method is the first to combine set-valued state estimation with global data association. The approach is general and can be applied to any tracking scenario where detections can be converted to point measurements.

The remainder of this paper is organized as follows: In Section II we give a brief introduction to random finite sets and the PHD filter. In Section III we describe the min-cost flow network formulation of the data association problem and in Section IV we give insights on how to implement the proposed approach in practice. In Section V we report on results from our experiments on a publicly available pedestrian tracking dataset. Finally, Section VI presents our conclusion.

II. Background

In this section we give a brief overview of random finite sets and multi-target Bayesian filtering. For a more complete introduction to methods described here, we refer the reader to [10] and [11].

A. Random Finite Sets

Finite set statistics (FISST) provides a set-theoretical foundation for information fusion that addresses many of the difficulties that arise in multi-target Bayesian filtering with unknown data association and unknown target appearance and disappearance. For this purpose, the theory provides a toolbox of mathematical procedures to systematically deal with set-valued random variables that have an unknown number of members, which are themselves random. The statistics of such a random finite set (RFS) can be described by two probability distributions: a discrete probability distribution for the cardinality of the set and a joint probability for the individual members of the set, given its cardinality.

Let X be a RFS that draws its instantiations from the hyperspace of all finite subsets $\mathcal{F}(\mathcal{X})$ of some space $\mathcal{X}$. The first-order moment of X is a non-negative function $v(\boldsymbol{x})$ defined on $\mathcal{X}$ which integrates to the expected number of elements that are also present in S for any closed subset $S \subseteq \mathcal{F}(\mathcal{X})$:

$$\int_S v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathrm{E}\left[|X \cap S|\right]. \tag{1}$$

This function is called the *probability hypothesis density* (PHD) or simply *intensity* of X. The PHD is of particular importance in our work for the following two reasons: (i) While the multi-target Bayes filter is computationally intractable just as its single-target counterpart, the PHD filter was developed to alleviate this computational intractability [22]. (ii) The PHD provides a useful connection between set-valued and vector-valued random variables. More precisely, intensity $v(\boldsymbol{x})$ of RFS X describes the zero-probability event $P(\boldsymbol{x} \in X)$ that $\boldsymbol{x}$ is contained in X [11].

Two important classes of RFSs are the Bernoulli RFS and the Poisson RFS. A Bernoulli RFS X is either empty with probability $1-r$ or contains exactly one member that is distributed according to a probability density $p(\boldsymbol{x})$. Further, the PHD of Bernoulli RFS X is $v(\boldsymbol{x}) = r \cdot p(\boldsymbol{x})$. Poisson RFS are uniquely characterized by their intensity. Let X be a Poisson RFS with intensity $v(\boldsymbol{x})$. Then, the expected number of elements in X is $\hat{N} = \int v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ and all of its members are independent and identically distributed (i.i.d.) according to $v(\boldsymbol{x})/\hat{N}$.

B. Multi-Object State Estimation

In multi-target Bayesian filtering, the set of all target states X_k and measurements Z_k at time k are reconceptualized as single set-valued random variables

$$X_k = \{\boldsymbol{x}_{k,1}, \ldots, \boldsymbol{x}_{k,N_k}\}, \tag{2}$$
$$Z_k = \{\boldsymbol{z}_{k,1}, \ldots, \boldsymbol{z}_{k,M_k}\}, \tag{3}$$

where no specific ordering on the respective collections of target states and measurements exists. Individual targets follow a single-target motion model $\boldsymbol{x}_k = f_{k|k-1}(\boldsymbol{x}_{k-1})$ and a single-target measurement model $\boldsymbol{z}_k = g_k(\boldsymbol{x}_k)$ describes the measurement generation process.

The RFS model for evolution of multi-target state X_k incorporates target motion, disappearance, and appearance:

$$X_k = \left[\bigcup_{\boldsymbol{x} \in X_{k-1}} S_{k|k-1}(\boldsymbol{x})\right] \cup B_k, \tag{4}$$

where $S_{k|k-1}(\boldsymbol{x})$ is a Bernoulli RFS that takes on either $\{f_{k|k-1}(\boldsymbol{x})\}$ if target $\boldsymbol{x}$ survives from time $k-1$ to k or $\emptyset$ otherwise and B_k is the RFS of spontaneous target appearance. According to the standard multi-target measurement model [11], measurements are either generated by a true target or clutter:

$$Z_k = \underbrace{\left[\bigcup_{\boldsymbol{x} \in X_k} \Upsilon_k(\boldsymbol{x})\right]}_{\Theta_k(X_k)} \cup K_k \tag{5}$$

where $\Upsilon_k(\boldsymbol{x})$ is a Bernoulli RFS that takes on $\{g_k(\boldsymbol{x})\}$ if $\boldsymbol{x}$ is detected and $\emptyset$ otherwise. The RFS K_k is the set of clutter measurements at time k.

While it is possible to derive update equations for propagating full multi-target densities using FISST, these are generally computationally intractable[1]. On the other hand, the PHD filter [10] is an approximation to the multi-target Bayes filter that propagates first-order moments of the multi-target state.

Let $v_{k-1}(\boldsymbol{x})$ be the intensity of multi-target state X_{k-1} and let $b_k(\boldsymbol{x})$ be the intensity of appearing targets B_k at time k. Further, let $p_{\mathrm{S}}(\boldsymbol{x})$ be the state-dependent probability that a target at $\boldsymbol{x}$ persists from time $k-1$ to time k and let $p_{k|k-1}(\boldsymbol{x} \mid \boldsymbol{x}')$ be the single-target motion model. Then, the intensity of predicted multi-target state $X_{k|k-1}$ is

$$v_{k|k-1}(\boldsymbol{x}) = b_k(\boldsymbol{x}) + \int p_{\mathrm{S}}(\boldsymbol{x}') p_{k|k-1}(\boldsymbol{x} \mid \boldsymbol{x}') v_{k-1}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'. \tag{6}$$

Given a sensor-specific probability of detection $p_{\mathrm{D}}(\boldsymbol{x})$ and the intensity $c_k(\boldsymbol{z})$ of clutter RFS K_k, the posterior intensity at time k is

$$v_k(\boldsymbol{x}) = v_{\mathrm{L},k}(\boldsymbol{x}) + \sum_{\boldsymbol{z}_{k,j} \in Z_k} v_{\mathrm{U},k}(\boldsymbol{z}_{k,j}, \boldsymbol{x}), \tag{7}$$

[1]However, approximations have been proposed where the multi-target density is approximated as a set of independent Bernoulli RFS [24].

where the first term accounts for missed detections

$$v_{\mathrm{L},k}(\boldsymbol{x}) = [1 - p_{\mathrm{D}}(\boldsymbol{x})]\, v_{k|k-1}(\boldsymbol{x}) \tag{8}$$

and the second term accounts for measurement-corrected updates

$$v_{\mathrm{U},k}(\boldsymbol{z}_{k,j}, \boldsymbol{x}) = \frac{p_{\mathrm{D}}(\boldsymbol{x}) p(\boldsymbol{z}_{k,j} \mid \boldsymbol{x}) v_{k|k-1}(\boldsymbol{x})}{c_k(\boldsymbol{z}_{k,j}) + \langle p_{\mathrm{D}}\, p(\boldsymbol{z}_{k,j} \mid \cdot), v_{k|k-1}\rangle}. \tag{9}$$

Above, we have written $p(\boldsymbol{z}_{k,j} \mid \boldsymbol{x})$ for the single-target measurement likelihood and

$$\langle p_{\mathrm{D}}\, p(\boldsymbol{z}_{k,j} \mid \cdot), v_{k|k-1}\rangle = \int p_{\mathrm{D}}(\boldsymbol{x}) p(\boldsymbol{z}_{k,j} \mid \boldsymbol{x}) v_{k|k-1}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{10}$$

for the likelihood that $\boldsymbol{z}_{k,j}$ has been generated by a target in $X_{k|k-1}$. We will use this inner product notation throughout the remainder of this paper.

III. Trajectory Estimation

The computational efficiency of the PHD filter stems from the fact that it operates on single-target space without maintaining target identities. While it is possible to estimate the number of targets and to recover their states as peaks in the intensity function [22], [23], the identity of the obtained state estimates remains unknown. In this section, we show how target trajectories can be recovered in a separate process. We formulate trajectory estimation as discrete optimization on the set of all measurements and use the intensity of the multi-target state for constructing entrance, exit, and transition probabilities.

A. Problem Formulation

Let $Z = Z_1 \cup \cdots \cup Z_M$ denote the set of measurements from time 1 to M and let $T_k = \{(t_{k_l}, i_{k_l})\}_{l=1}^{L_k}$ denote a single-target trajectory hypothesis, such that $T_k(l)$ is an index to the i_{k_l}-th measurement at time t_{k_l}, i.e., $\boldsymbol{z}_{T_k(l)} := \boldsymbol{z}_{t_{k_l}, i_{k_l}}$. Further, a multi-target association hypothesis is defined as the set of single-target trajectory hypotheses $T = \{T_k\}$. Then, the objective of trajectory estimation is the maximum a posteriori multi-target association hypothesis

$$T^* = \operatorname*{argmax}_T P(T \mid Z) \tag{11}$$

$$= \operatorname*{argmax}_T P(Z \mid T) P(T) \tag{12}$$

$$= \operatorname*{argmax}_T \prod_{\boldsymbol{z}_{t,i} \in Z} P(\boldsymbol{z}_{t,i} \mid T) \prod_{T_k \in T} P(T_k) \tag{13}$$

$$= \operatorname*{argmax}_T \sum_{\boldsymbol{z}_{t,i} \in Z} \log P(\boldsymbol{z}_{t,i} \mid T) + \sum_{T_k \in T} \log P(T_k). \tag{14}$$

The derivation above follows [28]. First, we use Bayes rule to factor the posterior into a measurement likelihood and a prior on target trajectories. Then, we assume measurements to be conditionally independent given the multi-target trajectory and that targets move independently of one another. Likewise, these assumptions are found in the derivation of the PHD recursion [10].

For modeling the prior on single-target trajectories we use a variable-length first-order Markov chain

$$P(T_k) = P(\{\boldsymbol{z}_{T_k(1)}, \ldots, \boldsymbol{z}_{T_k(L_k)}\}) \tag{15}$$

$$= P_{\mathrm{start}}(\boldsymbol{z}_{T_k(1)}) P_{\mathrm{link}}(\boldsymbol{z}_{T_k(2)} \mid \boldsymbol{z}_{T_k(1)}) \ldots P_{\mathrm{link}}(\boldsymbol{z}_{T_k(L_k)} \mid \boldsymbol{z}_{T_k(L_k-1)}) P_{\mathrm{end}}(\boldsymbol{z}_{T_k(L_k)}), \tag{16}$$

where P_{start} and P_{end} are measurement-dependent entry and exit probabilities and P_{link} is a transition probability between measurements. The measurement likelihood expresses the probability that measurement $\boldsymbol{z}_{t,i}$ has been generated by a true target and should be included in the multi-target association hypothesis, i.e.,

$$P(\boldsymbol{z}_{t,i} \mid T) = \begin{cases} P_{\mathrm{target}}(\boldsymbol{z}_{t,i}) & \exists\, T_k \in T : \boldsymbol{z}_{t,i} \in T_k, \\ 1 - P_{\mathrm{target}}(\boldsymbol{z}_{t,i}) & \text{otherwise.} \end{cases} \tag{17}$$

Substituting (17) into (14), the objective can be rewritten such that it depends only on measurements included in the multi-target association hypothesis:

$$T^* = \operatorname*{argmax}_T \sum_{\boldsymbol{z}_{t,i} \in T} \log l(\boldsymbol{z}_{t,i}) + \sum_{T_k \in T} \log P(T_k), \tag{18}$$

with

$$l(\boldsymbol{z}_{t,i}) = \frac{P_{\mathrm{target}}(\boldsymbol{z}_{t,i})}{1 - P_{\mathrm{target}}(\boldsymbol{z}_{t,i})}. \tag{19}$$

Further, if we reduce the search space by allowing each measurement to be associated to at most one target, such that $\forall k \neq l : T_k \cap T_l = \emptyset$, the objective of trajectory estimation can be mapped into a cost-flow network that allows for efficient and exact optimization [28]. For example, using a min-cost flow algorithm, the exact solution can be found with worst-case running time of $O(n^2 m \log n)$, where n is the number of measurements and m is the number of edges in the graph. Zhang et al. [28] report that, due to the structure of the problem, run times grow only linear with the number of measurements in practice. Greedy algorithms based on shortest path graph search and dynamic programming find near-optimal results in guaranteed linear time (e.g., [2], [17]).

For information on how to formulate the integer linear program that can be mapped into a cost-flow network we refer the reader to [17] and [28]. Instead, in the remainder of this section we show how to model the individual terms in the objective function using a PHD filter.

B. Track Hypothesis Generation

We rewrite the intensity of multi-target state X_k as a linear combination of individual track hypotheses. Let $Z_{1:k}$ denote the set of all measurements up to time k. Then, (7) can be rewritten as

$$v_k(\boldsymbol{x}) = u_k(\boldsymbol{x}) + \sum_{\boldsymbol{z}_{t,i} \in Z_{1:k}} q_{t,i}\, v_k^{(t,i)}(\boldsymbol{x}), \tag{20}$$

where $u_k(\boldsymbol{x})$ is the intensity of targets that remain undetected until time k, $v_k^{(t,i)}(\boldsymbol{x})$ is the intensity of a single track hypothesis for measurement $\boldsymbol{z}_{t,i}$, and $q_{t,i}$ is a scaling parameter that, as will be shown later in this section, can be interpreted as the probability that $\boldsymbol{z}_{t,i}$ has been generated by a target in $X_{k|k-1}$. Rewriting the intensity in this way allows us to construct first-order Markov transition probabilities between measurements.

Assume at time t we are given the intensity $v_{t|t-1}(\boldsymbol{x})$ of predicted multi-target state $X_{t|t-1}$ and the set of measurements Z_t. Then, we create a new track hypothesis for each measurement $\boldsymbol{z}_{t,i} \in Z_t$. Assuming a Bernoulli RFS, we initialize an intensity

$$v_t^{(t,i)}(\boldsymbol{x}) = r_t^{(t,i)} p_t^{(t,i)}(\boldsymbol{x}) \tag{21}$$

with probability of existence $r_t^{(t,i)} = 1$ and spatial density

$$p_t^{(t,i)}(\boldsymbol{x}) = \frac{p_{\mathrm{D}}(\boldsymbol{x}) p(\boldsymbol{z}_{t,i} \mid \boldsymbol{x}) v_{t|t-1}(\boldsymbol{x})}{\int p_{\mathrm{D}}(\boldsymbol{x}) p(\boldsymbol{z}_{t,i} \mid \boldsymbol{x}) v_{t|t-1}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}}. \tag{22}$$

The interpretation is as follows. One fundamental assumption in the derivation of the PHD filter is that $X_{t|t-1}$ is approximately Poisson [10]. Therefore, $\hat{N}_{t|t-1} = \int v_{t|t-1}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ i.i.d. targets are distributed according to $p_{t|t-1}(\boldsymbol{x}) = v_{t|t-1}(\boldsymbol{x}) / \hat{N}_{t|t-1}$. If measurement $\boldsymbol{z}_{t,i}$ has been generated by a true target, then its probability of existence is 1 and we can compute the posterior state distribution using a single-target Bayes update. Due to differences between single-target and multi-target likelihood modeling, the predicted intensity is reweighted by a state-dependent probability of detection. Using this track initialization scheme and scaling parameter

$$q_{t,i} = \frac{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \mid \cdot), v_{t|t-1} \rangle}{c_t(\boldsymbol{z}_{t,i}) + \langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \cdot), v_{t|t-1} \rangle} \tag{23}$$

the measurement-corrected terms in PHD update (9) can be written as $v_{\mathrm{U},t}(\boldsymbol{z}_{t,i}, \boldsymbol{x}) = q_{t,i} \, v_t^{(t,i)}(\boldsymbol{x})$.

In subsequent time steps, we propagate track hypotheses in a way that is aligned with the PHD filter. More specifically, we use the PHD predictor equation and the missed-detection case of the PHD update equation for propagating track hypotheses:

$$v_{k|k-1}^{(t,i)}(\boldsymbol{x}) = \int p_{\mathrm{S}}(\boldsymbol{x}') p_{k|k-1}(\boldsymbol{x} \mid \boldsymbol{x}') v_{k-1}^{(t,i)}(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}', \tag{24}$$

$$v_k^{(t,i)}(\boldsymbol{x}) = [1 - p_{\mathrm{D}}(\boldsymbol{x})] \, v_{k|k-1}^{(t,i)}(\boldsymbol{x}). \tag{25}$$

Finally, we propagate the intensity of undetected targets to account for target appearances. This intensity contains all targets that have entered the scene, but have not been detected until the most current time step:

$$u_{k|k-1}(\boldsymbol{x}) = b_k(\boldsymbol{x}) + \int p_{\mathrm{S}}(\boldsymbol{x}') p_{k|k-1}(\boldsymbol{x} \mid \boldsymbol{x}') u_{k-1} \, \mathrm{d}\boldsymbol{x}', \tag{26}$$

$$u_k(\boldsymbol{x}) = [1 - p_{\mathrm{D}}(\boldsymbol{x})] \, u_{k|k-1}(\boldsymbol{x}). \tag{27}$$

In (22)–(27), we have rewritten the PHD filter in terms of single-target track hypotheses. By doing so, we obtain a set of *measurement-induced* intensities at the current time step and a set of *legacy* intensities from previous time steps. The full multi-target intensity can be recovered from these individual track hypotheses as a weighted sum (c.f. Eq. 20).

In the following, we use the intensity associated with each measurement to derive measurement likelihoods and transition probabilities for trajectory estimation. It is worth noting that, formally, the PHD filter does not propagate single-target statistics. Instead, each track hypothesis must rather be thought of as a partition of the full multi-target PHD.

C. Measurement Likelihood

Taken that each measurement $\boldsymbol{z}_{t,i}$ must originate from either a true target in $X_{t|t-1}$ or clutter K_t, we compute the probability that $\boldsymbol{z}_{t,i}$ is indeed a true target by evaluating their respective intensities:

$$\begin{aligned} P_{\mathrm{target}}(\boldsymbol{z}_{t,i}) &= \frac{P(\boldsymbol{z}_{t,i} \in \Theta_t(X_{t|t-1}))}{P(\boldsymbol{z}_{t,i} \in K_t) + P(\boldsymbol{z}_{t,i} \in \Theta_t(X_{t|t-1}))} && (28) \\ &= \frac{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \mid \cdot), v_{t|t-1} \rangle}{c_t(\boldsymbol{z}_{t,i}) + \langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \mid \cdot), v_{t|t-1} \rangle} && (29) \\ &= q_{t,i}. && (30) \end{aligned}$$

Substituting (29) into (19), we get for the unnormalized measurement likelihood

$$\begin{aligned} l(\boldsymbol{z}_{t,i}) &= \frac{P_{\mathrm{target}}(\boldsymbol{z}_{t,i})}{1 - P_{\mathrm{target}}(\boldsymbol{z}_{t,i})} && (31) \\ &= \frac{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \mid \cdot), v_{t|t-1} \rangle}{c_t(\boldsymbol{z}_{t,i})}. && (32) \end{aligned}$$

This model faithfully characterizes sensor characteristics by taking into account the probability of detection and clutter statistics.

D. Transition Probability

For computing transition probabilities between measurements, we propagate each track hypothesis using the rewritten PHD recursion. Let $\boldsymbol{x}_{k|k-1}^{(t,i)}$ denote the target state that originates from measurement $\boldsymbol{z}_{t,i}$, propagated to time k. Then, we compute the probability that the target transitions into measurement $\boldsymbol{z}_{k,j}$ as the fraction of the multi-target measurement likelihood that is due to the partition of $v_{k|k-1}(\boldsymbol{x})$ corresponding to $\boldsymbol{x}_{k|k-1}^{(t,i)}$:

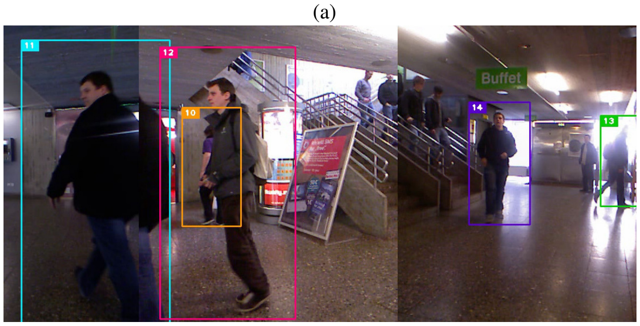
$$\begin{aligned} P_{\mathrm{link}}(\boldsymbol{z}_{k,j} \mid \boldsymbol{z}_{t,i}) &= \frac{P(\boldsymbol{z}_{k,j} \in \Upsilon_k(\boldsymbol{x}_{k|k-1}^{(t,i)}))}{P(\boldsymbol{z}_{k,j} \in \Theta_k(X_{k|k-1}))} && (33) \\ &= \frac{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{k,j} \mid \cdot), v_{k|k-1}^{(t,i)} \rangle}{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{k,j} \mid \cdot), v_{k|k-1} \rangle}. && (34) \end{aligned}$$

Using this term, the product of unnormalized measurement likelihood (32) and transition probability (34) equals the likelihood ratio between the target hypothesis and the null hypothesis, where the target hypothesis assumes that both measurements originate from the same target and the null hypothesis assumes $\boldsymbol{z}_{k,j}$ originates from clutter. This is in accordance with MHT where the same measure is used for scoring track hypotheses [8].

For computing entry probability P_{start} we proceed in a similar way, using the intensity of undetected targets in the nominator:

$$P_{\mathrm{start}}(\boldsymbol{z}_{t,i}) = \frac{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{t,i} \mid \cdot), u_{t|t-1} \rangle}{\langle p_{\mathrm{D}} \, p(\boldsymbol{z}_{k,j} \mid \cdot), v_{k|k-1} \rangle}. \tag{35}$$

(a)



(b)

Fig. 1: Selected example of tracking output on the RGB-D people dataset [9]. Track identities are colored coded.

The final term required for trajectory estimation is the probability that a single-target track ends at measurement $\boldsymbol{z}_{t,i}$. For this purpose we use the state-dependent probability of survival from the PHD recursion. More precisely, we average over the associated distribution of target state $\boldsymbol{x}_t^{(t,i)}$ to obtain the expected exit probability:

$$P_{\text{exit}}(\boldsymbol{z}_{t,i}) = \mathrm{E}_{p_t^{(t,i)}(\boldsymbol{x})}\left[1 - p_{\mathrm{S}}(\boldsymbol{x})\right] \quad (36)$$

$$= \langle 1 - p_{\mathrm{S}}(\boldsymbol{x}), p_t^{(t,i)} \rangle. \quad (37)$$

Note that this term is proportional to the intensity loss induced during PHD prediction (24).

IV. Implementation Details

In (22)–(27) we have rewritten the multi-target posterior intensity in terms of single track hypotheses. In practice, these modified update equations can be implemented using a bank of SMC-PHD filters [23], one for each track hypothesis. For estimating target states during online application, we calculate the expected number of targets from the full posterior and then select track hypotheses with highest intensity mass. Therefore, it is not necessary to apply additional partitioning or clustering algorithms in order to obtain state estimates from the posterior, as necessary in standard SMC-PHD implementations [23].

The min-cost flow network for trajectory estimation is incrementally built frame by frame. Optimization can be carried out either for the entire sequence at once or in smaller batches of a fixed number of frames, such that there is a maximum delay on data association.

Several pruning schemes can be implemented to manage the computational cost of the system. For example, we keep the size of the track hypothesis set manageable by removing tracks with intensity mass below a given threshold. Similarly, tracks may be discarded after a fixed number of missed detections. Further, when building the min-cost flow network, we add an edge to a potential succeeding measurement only when the transition probability is higher than a given threshold. This typically results in much more sparse graph structures.

TABLE I: Tracking Evaluation on RGB-D People Dataset [9]

	MOTP	**MOTA**	**FP**	**FN**	**ID**
Luber et al. [9]	N/A	**78.0%**	4.5%	**16.8%**	32
Munaro et al. [13]	73.7%	71.8%	7.7%	20.0%	19
SMC-PHD [23]	N/A	N/A	3.2%	34.0%	N/A
MCF-SMC-PHD	**74.6%**	75.1%	**1.5%**	23.3%	**7**

V. Experiments

For comparison with existing methods, we have applied our tracker to a publicly available dataset [9], [21] that contains a sequence of over 3000 RGB-D frames captured by three vertically mounted Microsoft Kinect sensors. The sensor configuration is placed in a busy university hall at approximately $1\,\mathrm{m}$ height. As input for our tracker we have used point measurements on the ground plane, generated by a people detector [13] that is available in current releases of the *Point Cloud Library* [20]. Target motion has been modeled using a constant velocity model. A uniform birth intensity and constant survival and detection probabilities have been used. In order to deal with continuously recurring false positive detections on distant walls, the clutter intensity has been configured based on a blurred version of an occupancy grid map, where clutter is more likely to be generated around static obstacles.

An illustrative example of tracking output is shown in Fig. 1. The results of our evaluation are summarized in Table I. The method of Luber et al. [9] is an implementation of MHT that integrates appearance information into the association likelihood. The method of Munaro et al. [13] uses Global Nearest Neighbor data association with appearance information integrated into the association likelihood. The SMC-PHD [23] denotes a plain sequential Monte Carlo implementation of the PHD filter that does not output target identities. The MCF-SMC-PHD is a sequential Monte Carlo implementation of the approach presented in this paper. Comparison with existing methods has been established through CLEAR MOT metrics [4]. The MOTP indicator is a measure of tracking precision in terms of position output. In accordance with [13], we use the bounding box overlap between ground truth tracks and tracker output for this purpose. The MOTA indicator is a measure of tracking errors in terms of false positives, false negatives, and mismatches. The validation threshold for counting correct matches has been set to at least 50% bounding box overlap. In addition, we report false positives (FP), false negatives (FN), and number of identity switches (ID).

In comparison to the method of Munaro et al. [13], who use the same detector as we do, our tracker produces similar MOTA and MOTP scores. The high number of false negatives in both approaches is mainly due to numerous annotated people that enter the scene through an elevated stairway where the detector does not produce reliable results. The fact that our tracker produces a lower number of false positives may be explained by the location-dependent clutter intensity. This also explains the higher number of false negatives, due to people walking close to walls where measurements are less likely generated by true targets. Note that our tracker produces considerably fewer ID switches than any of the methods that we have compared against.

The SMC-PHD does not output target identities. Therefore, we only report false positive and false negative rates. In comparison to our tracker, the plain SMC-PHD produces similar false positive rates. On the other hand, using the same parameter set, the number of false negatives is considerably higher for the SMC-PHD. This is due to people frequently passing by close to the camera, causing long-term occlusions. These are not well described by our parameter set where the probability of detection is assumed constant over the entire measurement space. In fact, inter-object occlusions are generally not easy to model in the PHD filter and the performance of our tracker is likewise affected in such situations. It is due to global optimization that the MCF-SMC-PHD more often picks up tracks after consecutive frames of missed detections.

All experiments have been conducted on a desktop computer with Intel(R) Core(TM) i7-2760QM CPU. Using 200 samples per track hypothesis, our single-threaded Python implementation of the PHD filter runs at approximately 16 fps. Batch trajectory estimation over the entire sequence (1133 frames) took about 2 s using an optimized min-cost flow solver written in C++. This accumulates to an overall run-time of approximately 15 fps with most time spent on state estimation.

VI. Conclusion

We have presented a novel formulation of min-cost flow data association for the Probability Hypothesis Density filter. The method provides instantaneous state estimates with delayed decision on target trajectories. Further, the method is general and can be applied to any tracking scenario that involves point measurements. In our experiments on a pedestrian tracking dataset, the proposed method outperforms existing approaches that integrate motion and appearance information into the association likelihood. Further, in contrast to many existing approaches following the tracking-by-detection paradigm, our method does not require learning of association probabilities. Instead, these probabilities are derived from multi-object statistics based on FISST. Long-term occlusions remain an open challenge: the performance of the tracker degrades with increasing number of inter-object occlusions. Therefore, we will investigate integration of explicit occlusion models in future work.

References

[1] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *CVPR*, pages 1926–1933, 2012.

[2] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In *CVPR*, pages 744–750, 2006.

[3] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(9):1806–1819, 2011.

[4] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP J. Image Video Process*, 2008, 2008.

[5] A. Dehghan, Y. Tian, P. H. Torr, and S. Mubarak. Target identity-aware network flow for online multiple target tracking. In *CVPR*, pages 1146–1154, 2015.

[6] T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE J. Ocean. Eng.*, 8(3):173–184, 1983.

[7] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In *CVPR*, pages 1–8, 2007.

[8] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg. Multiple hypothesis tracking revisited. In *ICCV*, pages 4696–4704, 2015.

[9] M. Luber, L. Spinello, and K. O. Arras. People tracking in RGB-D data with on-line boosted target models. In *IROS*, pages 3844–3849, 2011.

[10] R. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *IEEE Trans. Aerosp. Electron. Syst.*, 39(4):1152–1178, 2003.

[11] R. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, Norwood, MA, USA, 2007.

[12] A. Milan, K. Schindler, and S. Roth. Detection- and trajectory-level exclusion in multiple object tracking. In *CVPR*, pages 3682–3689, 2013.

[13] M. Munaro, F. Basso, and E. Menegatti. Tracking people within groups with RGB-D data. In *IROS*, pages 2101–2107, 2012.

[14] K. Panta, D. E. Clark, and Ba-Ngu Vo. Data Association and Track Management for the Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Trans. Aerosp. Electron. Syst.*, 45(3):1003–1016, July 2009.

[15] K. Panta, B.-N. Vo, and S. Singh. Improved probability hypothesis density (PHD) filter for multitarget tracking. In *ICISIP*, pages 213–218, 2005.

[16] K. Panta, B.-N. Vo, and S. Singh. Novel data association schemes for the probability hypothesis density filter. *IEEE Trans. Aerosp. Electron. Syst.*, 43(2):556–570, 2007.

[17] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *CVPR*, pages 1201–1208, 2011.

[18] D. B. Reid. An algorithm for tracking multiple targets. *IEEE Trans. Autom. Control*, 24(6):843–854, 1979.

[19] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer. The labeled multi-bernoulli filter. *IEEE Trans. Signal Process.*, 62(12):3246–3260, 2014.

[20] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *ICRA*, pages 1–4, 2011.

[21] L. Spinello and K. O. Arras. People detection in RGB-D data. In *IROS*, pages 3838–3843, 2011.

[22] B.-N. Vo and W.-K. Ma. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Process.*, 54(11):4091–4104, November 2006.

[23] B.-N. Vo, S. Sumeetpal, and A. Doucet. Sequential monte carlo methods for multitarget filtering with random finite sets. *IEEE Trans. Aerosp. Electron. Syst.*, 41(4):1224–1245, 2005.

[24] B.-T. Vo, B.-N. Vo, and A. Cantoni. The cardinality balanced multi-target multi-bernoulli filter and its implementations. *IEEE Trans. Signal Process.*, 57(2):409–423, 2009.

[25] Z. Wu, A. Thangali, S. Sclaroff, and M. Betke. Coupling detection and data association for multiple object tracking. In *CVPR*, pages 1948–1955, 2012.

[26] B. Yang and R. Nevatia. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. In *CVPR*, pages 1918–1925, 2012.

[27] B. Yang and R. Nevatia. An online learned CRF model for multi-target tracking. In *CVPR*, pages 2034–2041, 2012.

[28] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR*, pages 1–8, 2008.