Confidence-Aware Probability Hypothesis Density Filter for Visual Multi-Object Tracking

Nicolai Wojke and Dietrich Paulus
Active Vision Group, Institute for Computational Visualistics, University of Koblenz-Landau, 56070 Koblenz, Germany

Keywords: Visual Tracking, Multi-Object State Estimation.

Abstract: The Probability Hypothesis Density Filter (PHD) filter is an efficient recursive multi-object state estimator that systematically deals with data association uncertainty. In this paper, we apply the PHD filter in a tracking-by-detection framework. In order to mimic state-dependent false alarms, we introduce an adapted PHD recursion that defines clutter generators in state space. Further, we integrate detector confidence scores into the measurement likelihood. This extension is quite effective yet simple, which means that it requires few changes to the original PHD recursion, that it has the same computational complexity, and that there exist few parameters that must be adapted to the individual tracking scenario. Our evaluation on a popular pedestrian tracking dataset demonstrates results that are competitive with the state-of-the-art.

1 INTRODUCTION

Visual multi-object tracking is a key challenge in many computer vision applications. The problem is well studied and numerous approaches have been proposed. However, due to the combinatorial nature of data association, the problem remains challenging.

Within the last decade, it has become increasingly popular to formulate multi-object tracking as *tracking-by-detection*, where plausible object trajectories are found through global optimization. Zhang et al. (2008) provide a prominent formulation using a min-cost flow network. They create a graph on the set of all measurements and find globally optimal trajectories using a push-relabel algorithm. This formulation has been adopted by others in order to obtain better run-times: Pirsiavash et al. (2011) introduce a greedy path search based on dynamic programming, Berclaz et al. (2011) apply k-shortest path search. Others have extended the model to incorporate more structural information. For example, Deghan et al. (2015) integrate identity-specific association costs and propose a Lagrangian relaxation optimization.

Conventional multi-object tracking systems usually contain three components: state estimation, data association, and track handling. Therefore, these systems estimate the underlying object state, e.g., position and velocity, and perform association of measurements to objects on a frame-by-frame basis. Recently, a number of such conventional methods have been revisited and shown competitive performance. Notably, Kim et al. (2015) show that the classical multiple hypothesis tracking algorithm (Reid, 1979) can achieve state-of-the-art results when integrating online-learned appearance information into the association likelihood and Rezatofighi et al. (2015) have investigated an efficient solution to the joint probabilistic data association that, combined with a heuristic track handling scheme, achieves competitive results in dense tracking scenarios with substantial occlusions, false alarms, and missed detections. Relatedly, Segal and Reid (2013) use a novel parametrization of the classical data association problem to formulate a switching linear dynamical system that allows efficient inference in a message passing framework. Further, their formulation explicitly infers the number of objects and classifies detections into object and clutter categories. For this purpose, they use the detector confidence score as an additional observation. Integration of detector confidence scores has also been investigated Breitenstein et al. (2011) and Poiesi et al. (2013). Both integrate the detector confidence score as observations into a particle filtering framework.

The Probability Hypothesis Density (PHD) filter (Mahler, 2003) is a set-valued state estimator that is based on a relatively new, specialized theory for multi-object information fusion (Mahler, 2007). This theory provides comprehensive means of modeling

Wojke N. and Paulus D.
Confidence-Aware Probability Hypothesis Density Filter for Visual Multi-Object Tracking.
DOI: 10.5220/0006095801320139
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 132-139
ISBN: 978-989-758-227-1

multi-object phenomena and the PHD recursion itself deals with all notable sources of uncertainty involved in multi-object state estimation, including process and measurement noise as well as the uncertainty involved in data association. However, successful application of the PHD filter requires knowledge of the clutter process and the performance of the filter is known to degrade substantially if these parameters are chosen incorrectly. Therefore, a number of extensions have been proposed to learn the clutter process over time (Maggio and Cavallaro, 2009; Mahler et al., 2011). While the PHD filter does not provide track identities itself, it has recently been shown how these can be recovered in a network flow formulation (Wojke and Paulus, 2016).

In this paper, we explore the PHD filter in a tracking-by-detection framework. Therefore, our work builds upon the min-cost flow formulation of Wojke and Paulus (2016). Our contributions are as follows: First, we extend the standard PHD filter to mimic state-dependent false alarms. This is necessary, because in visual tracking scenarios clutter is dependent on the multi-object state. More specifically, due to localization inaccuracies, the object detector may fire false alarms in the surrounding of the true object location. In this paper, we present an adapted recursion that increases the accuracy of the cardinality estimate and reduces the number of false alarm tracks. Second, we provide a practical Sequential Monte-Carlo (SMC) implementation of a reformulated PHD recursion in terms of single-object track hypotheses (Wojke and Paulus, 2016). Our implementation is general, i.e., we make no *a priori* assumptions about the location of appearing objects and assume constant detection and survival probabilities. However, extension to more specific tracking scenarios is straight forward.

The remainder of this paper is organized as follows. In Section 2 we give a brief introduction to random finite sets and set-valued state estimation. In Section 3 we outline our adapted PHD recursion that accounts for confidence detector scores and describe a practical SMC implementation. In Section 4 we describe our experimental evaluation and we conclude in Section 5.

2 MULTI-OBJECT STATE ESTIMATION

In this section we give a brief overview of random finite sets and multi-object Bayesian filtering. For a more complete introduction to methods described here, we refer the reader to (Mahler, 2003, 2007).

Finite set statistics (FISST) provides a set-theoretical foundation for information fusion that addresses many of the difficulties that arise in multi-object Bayesian filtering with unknown data association and unknown object 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 $v(x)$ defined on $\mathcal{X}$ which integrates to the expected number of elements in X that are also present in S for any closed subset $S \subseteq \mathcal{F}(\mathcal{X})$:

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

This function is called the *probability hypothesis density* (PHD) or simply *intensity* of X. The PHD provides a useful connection between set-valued and vector-valued random variables: The intensity $v(x)$ of RFS X describes the zero-probability event $P(x \in X)$ that x is contained in X (Mahler, 2007).

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

$$X_k = \{x_{k,1}, \ldots, x_{k,N_k}\}, \tag{2}$$
$$Z_k = \{z_{k,1}, \ldots, z_{k,M_k}\}, \tag{3}$$

where no specific ordering on the respective collections of object states and measurements exists. Individual objects follow a single-object motion model $x_k = f_{k|k-1}(x_{k-1})$, and a single-object measurement model $z_k = g_k(x_k)$ describes the measurement generation process.

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

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

where $S_k(x)$ is a Bernoulli RFS that takes on either $\{f_{k|k-1}(x)\}$ if object x survives from time $k-1$ to k or $\emptyset$ otherwise, $T_k(x)$ is a RFS of targets that originate from x—this may be used to model, e.g., object splitting—and B_k is the RFS of spontaneous object appearances. According to the standard multi-object measurement model (Mahler, 2007), measurements are either generated by a true object or clutter:

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

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

Based on FISST, it is possible to derive an optimal multi-object Bayes filter that propagates multi-object densities. This Bayes filter is, however, generally computationally intractable (Mahler, 2007). In this work, we focus on the PHD filter (Mahler, 2003). The PHD filter is a computationally efficient alternative to the multi-object Bayes filter that propagates first-oder moments, instead.

3 CONFIDENCE-AWARE PHD FILTER

We present an adapted PHD recursion that mimics state-dependent false alarms through state-space clutter generators that survive for one time step only. The underlying idea is related to a recent extension of the PHD filter where the parameters of the clutter process are learned over time (Mahler et al., 2011). We follow this idea and present an alternative to the measurement model for state-dependent clutter proposed by Mahler (2014), which requires exhaustive summation over measurement partitions and is, therefore, computationally more demanding. Our model is much simpler, but requires detector confidence scores to guide the cardinality estimate of the PHD.

3.1 State-Space Clutter Generators

In what follows we use an augmented state space where each single-object state $w^{\mathrm{T}} = (x^{\mathrm{T}}, \beta)^{\mathrm{T}}$ contains a kinematic component x, e.g., position and velocity, and an object class identifier $\beta \in \{0, 1\}$ that is 0 for clutter and 1 for objects. The purpose of this augmentation is to mimic state-dependent false alarms using state-space clutter generators. Let $b_k(x, \beta)$ denote the intensity of appearing objects B_k, $\tau_k(x, \beta \mid x', \beta')$ the intensity of spawning objects $T_k(x', \beta')$, $v_{k-1}(x, \beta)$ the posterior intensity at time $k-1$, $p_{\mathrm{S}}(x, \beta)$ a state-dependent probability of survival, and $p_{k|k-1}(x \mid x')$ the single-object motion model that is independent of object class. Then, the predicted intensity at time k is

$$\begin{aligned} v_{k|k-1}(x,1) &= b_k(x,1) + \\ &\quad \langle p_{\mathrm{S}}(\cdot,1)\, p_{k|k-1}(x \mid \cdot), v_{k-1}(\cdot,1) \rangle, \end{aligned} \tag{6}$$

$$v_{k|k-1}(x,0) = N_{\mathrm{FA}}\, v_{k|k-1}(x,1), \tag{7}$$

where N_{FA} is the expected number of false alarms that are generated by the object detector for a given true object. Note that throughout the paper we use the inner product notation $\langle f, v \rangle = \int f(x) v(x)\, \mathrm{d}x$. Equation 6 is the standard PHD prediction (Mahler, 2003) for $\beta = 1$ without spawning objects, i.e., $\tau_k(x, 1 \mid x', \beta) = 0$. Equation 7 can be established as follows: For each new-born object and for every object that survives from previous times, create a Poisson clutter RFS with expected mean cardinality N_{FA} and set the probability of survival $p_{\mathrm{S}}(x, 0) = 0$, such that the RFS of surviving clutter is empty (i.e., clutter survives for one time step only). Then, the clutter birth intensity is a scaled version of the object birth intensity, the intensity of spawned clutter is a scaled version of the intensity of surviving objects, and the intensity of surviving clutter is zero. Note that, since the predicted clutter intensity (7) is a scaled version of the object intensity, it is not necessary to compute this term explicitly.

Now, let $z^{\mathrm{T}} = (y^{\mathrm{T}}, s)^{\mathrm{T}}$ denote a single-object measurement that contains a spatial component y and a detector confidence score s. Then, we assume the single-object measurement model factorizes into a spatial density conditional on object state and a probability density over the confidence score conditional on object class:

$$p(y, s \mid x, \beta) = p(y \mid x) p(s \mid \beta), \tag{8}$$

where, in the following, we abbreviate

$$p(s \mid \beta) = \begin{cases} P_{\mathrm{fg}}(s) & \beta = 1, \\ P_{\mathrm{bg}}(s) & \text{otherwise.} \end{cases} \tag{9}$$

Following Mahler et al. (2011), we can now compute the posterior for each object class separately. Let the detection probability be independent of object class $p_{\mathrm{D}}(x, \beta) = p_{\mathrm{D}}(x)$. Then, the posterior intensity of objects becomes:

$$\begin{aligned} v_k(x,1) &= [1 - p_{\mathrm{D}}(x)] v_{k|k-1}(x,1) + \\ &\quad \sum_{z_{k,j} \in Z_k} v_k{}^{(\mathrm{U})}(z_{k,j}, x, 1) \end{aligned} \tag{10}$$

with

$$v_k{}^{(\mathrm{U})}(z_{k,j}, x, 1) = \frac{P_{\mathrm{fg}}(s_{k,j}) p_{\mathrm{D}}(x) p_k(y_{k,j} \mid x) v_{k|k-1}(x,1)}{P_{\mathrm{bg}}(s_{k,j}) \left[c_k(y_{k,j}) + N_{\mathrm{FA}} \tau_k(y_{k,j}) \right] + P_{\mathrm{fg}}(s_{k,j}) \tau_k(y_{k,j})}, \tag{11}$$

where $\tau_k(y_{k,j}) = \langle p_D p_k(y_{k,j} \mid \cdot), v_{k|k-1}(\cdot, 1)\rangle$ is the intensity mass that accounts for the likelihood that $y_{k,j}$ has been generated by an object in $X_{k|k-1}$ and where $c_k(y_{k,j})$ is the intensity of state-independent clutter. Again, it is not necessary to write down the update equation for clutter objects, because they survive for one time step only.

Equations 6 and 10–11 represent our adapted PHD recursion. The derivation follows directly from our specific choice of clutter model. In the denominator of Equation 11, N_{FA} transfers intensity mass from the prior object intensity to the clutter intensity. For the cardinality estimate, this scaling factor controls how much emphasis should be put on the detector confidence score compared to the filtering process. Equation 10 collapses to the default PHD update for $P_{\text{fg}}(s_{k,j}) = P_{\text{bg}}(s_{k,j})$ with $N_{\text{FA}} = 0$.

3.2 Sequential Monte Carlo Implementation

We now describe a practical SMC implementation of the modified PHD recursion where we make use of two extensions that are complementary to the proposed clutter model: (i) We use an adapted sampling scheme for appearing objects that is more efficient when the birth intensity is uninformative (Ristic et al., 2012), (ii) we use a reformulation of the PHD recursion in terms of single-object hypotheses that can be mapped into a min-cost flow network to solve for target trajectories (Wojke and Paulus, 2016). Therefore, let $Z_{1:k}$ denote the set of measurements up to time k. Then, we partition the multi-object intensity into single-object track hypotheses

$$v_k(x) = \sum_{Z_{t,i} \in Z_{1:k}} q_{t,i} v_k^{(t,i)}(x), \tag{12}$$

where $v_k^{(t,i)}(x)$ is an intensity partition corresponding to the i-th measurement at time t and where $q_{t,i}$ is a scaling parameter.[1] Following Wojke and Paulus (2016), intensity $v_k^{(t,i)}(x)$ is proportional to the distribution over hypothetical object state $x_k^{(t,i)}$ that has generated measurement $z_{t,i}$ at time t and has since then not been detected. The scaling parameter $q_{t,i}$ accounts for the probability that $z_{t,i}$ has indeed been generated by an object in X_t, i.e., is not clutter. In our SMC implementation, we approximate each partition using a set of L_{T} samples and associated importance weights:

$$Q_k^{(t,i)} = \left\{ \left(w_k^{(t,i,n)}, x_k^{(t,i,n)} \right) \right\}_{n=1}^{L_{\text{T}}}, \tag{13}$$

$$v_k^{(t,i)}(x) \approx \sum_{n=1}^{L_{\text{T}}} w_k^{(t,i,n)} \delta(x - x_k^{(t,i,n)}). \tag{14}$$

From this particle approximation we can reconstruct the full multi-object intensity using (12). For notational brevity, we refer to this particle representation of the full multi-object intensity at time k as $Q_k = \{w_k^{(n)}, x_k^{(n)}\}_{n=1}^{L_k}$.

The following implementation consists of two steps: First, we propagate all *legacy* track hypotheses from the previous to the current time step. Then, we initialize a new *measurement-induced* track hypothesis for each newly arrived measurement. In terms of the PHD recursion, track propagation corresponds to prediction (6) and the missed detection case of update (10). Track initialization accounts for the measurement-corrected terms in update (10). At all times, the full multi-object intensity can be recovered from individual partitions using (12). Further, note that in the following implementation we use uninformed priors for spatial clutter and birth densities. In particular, we assume that state-independent clutter is Poisson with mean cardinality λ_c and uniform spatial density $p_{C_t}(y) = 1/V$, where V is the volume of the measurement space. Likewise, we assume no prior knowledge about the location of appearing objects. Therefore, we assume the birth intensity is Poisson with mean cardinality λ_b and place a uniform prior on appearing objects in measurement space $p_{B_k}(y) = 1/V$. It is, however, easy to adapt the presented algorithm to scene-specific layouts using more informed densities (e.g., higher birth probability at image borders).

Track Propagation. Assume at time k we are given particles $Q_{k-1}^{(t,i)}$ that approximate individual partitions of the posterior intensity at time $k-1$. Then, we propagate these *legacy* track hypotheses to time k as outlined in Listing 1. In lines 1–3 we multiply importance weights by the state-dependent probability of survival and sample from the single-object motion model to obtain a particle approximation $Q_{k|k-1}^{(t,i)}$ (c.f. Equation 6). In lines 4–6 we multiply importance weights by one minus the state-dependent probability of detection to account for the missed detection case of update (10) and obtain a particle approxmiation $Q_k^{(t,i)}$.

Track Initialization. Assume at time k we are given measurement set Z_k as well as particle set

[1]Note that in contrast to Wojke and Paulus (2016) we have no partition for the set of undetected targets. This is, because in the adapted sampling scheme of Ristic et al. (2012) it is assumed appearings objects are always detected.

Listing 1: Track propagation for a single legacy track $v_{k-1}^{(t,i)}(x)$.

for $n = 1, \ldots, L_\mathrm{T}$ **do**
    {Prediction}

$$w_{k|k-1}^{(t,i,n)} = p_\mathrm{S}(x_{k-1}^{(t,i,n)})w_{k-1}^{(t,i,n)}$$
$$x_{k|k-1}^{(t,i,n)} = p_{k|k-1}(\cdot \mid x_{k-1}^{(t,i,n)})$$

end for
for $n = 1, \ldots, L_\mathrm{T}$ **do**
    {Update}

$$w_k^{(t,i,n)} = \left[1 - p_\mathrm{D}(x_{k-1}^{(t,i,n)})\right] w_{k|k-1}^{(t,i,n)}$$
$$x_k^{(t,i,n)} = x_{k|k-1}^{(t,i,n)}$$

end for

$Q_{k|k-1}$ that approximates the predicted multi-object intensity. For each measurement $z_{k,j} \in Z_k$ we create a single-object track hypotheses as outlined in Listing 2. First, we update importance weights to account for the single-object measurement likelihood and state-dependent probability of detection (lines 1–3). Then, we draw samples from the birth intensity (line 4–6). Loosely following Ristic et al. (2012), we draw samples from

$$p_k(y_{k,j} \mid x)b_k(x) = p_k(y_{k,j} \mid x)p_{B_k}(x)\lambda_b, \tag{15}$$
$$= p_k(x \mid y_{k,j})p_{B_k}(y)\lambda_b, \tag{16}$$
$$= p_k(x \mid y_{k,j})\frac{\lambda_b}{V}, \tag{17}$$

where we assume the RFS of appearing objects is Poisson with expected number of objects λ_b and uniform spatial prior on measurement space $p_{B_k}(y_{k,j}) = 1/V$. Consequently, we draw samples from an *inverse measurement model* and set weights uniform such that they sum up to $b_k(z_{k,j}) = \lambda_b V^{-1}$. In practice, sampling from the inverse measurement model is more efficient when the birth intensity is uninformative, because birth samples are placed in areas where the measurement likelihood has high probability mass. In lines 7 and 8 we compute the probability that measurement $z_{k,j}$ has been generated by an object in X_k. Finally, in line 9 we resample to obtain L_T new particles with uniform weights.

Pruning and Data Association. Due to partitioning the intensity according to (12), the number of particles scales linearly with the number of measurements. However, only *few* track hypotheses contribute high intensity mass to the overall multi-object intensity. Therefore, at each time step, we prune track hypotheses with intensity mass below a given threshold.

Listing 2: Track initialization for measurement $z_{k,j} \in Z_k$.

1: **for** $n = 1, \ldots, L_{k-1}$ **do**
2:     Update weights of predicted intensity $v_{k|k-1}(x)$

$$w_k^{(k,j,n)} = p_\mathrm{D}(x_{k|k-1}^{(n)})p(y_{k,j} \mid x_{k|k-1}^{(k,j,n)})w_{k|k-1}^{(n)}$$
$$x_k^{(k,j,n)} = x_{k|k-1}^{(n)}$$

3: **end for**
4: **for** $n = 1, \ldots, L_\mathrm{T}$ **do**
5:     Draw birth samples from inverse measurement model

$$w_k^{(k,j,L_{k-1}+n)} = \frac{\lambda_b V^{-1}}{L_\mathrm{T}}$$
$$x_k^{(k,j,L_{k-1}+n)} \sim p(\cdot \mid y_{k,j})$$

6: **end for**
7: Compute

$$\tau_k(y_{k,j}) = \sum_{n=1}^{L_{k-1}+L_\mathrm{T}} w_k^{(k,j,n)}$$

8: Compute

$$q_{k,j} = \frac{P_\mathrm{fg}(s_{k,j})\tau_k(y_{k,j})}{P_\mathrm{bg}(s_{k,j})\left[\lambda_c V^{-1} + N_\mathrm{FA}\tau_k(y_{k,j})\right] + P_\mathrm{fg}(s)\tau_k(y_{k,j})}$$

9: Resample

$$\left\{w_k^{(k,j,n)}, x_k^{(k,j,n)}\right\}_{n=1}^{L_{k-1}+L_\mathrm{T}} \text{ to obtain } \left\{\frac{1}{L_\mathrm{T}}, x_k^{(k,j,n)}\right\}_{n=1}^{L_\mathrm{T}}$$

Further, to recover object trajectories, the adapted PHD recursion presented in this paper can be directly applied to the min-cost flow network of Wojke and Paulus (2016). The only parameter that is affected by our adaption is the probability of existence that is computed during track initialization (line 8). This term can be directly plugged into the original formulation. We refer the reader to the original publication (Wojke and Paulus, 2016) for further details on this part.

4 EXPERIMENTS

Evaluation has been carried out on the popular PETS'09 dataset (Ferryman and Shahrokni, 2009). For fair comparison, we used publicly available detections and ground truth provided by Andriyenko et al. (2012). Most of the sequences that we have evaluated on are medium or densly crowded scenarios with substantial occlusions, missed detections, and false alarms. Tracking was performed in 3D using a constant velocity motion model. Detections have been projected onto the ground plane using known camera calibration parameters. During all experiments, we used a single set of parameters. The motion model adds isotropic noise with standard deviation $\Delta t \cdot 0.5\,\mathrm{m}$ for the position and $\Delta t \cdot 1\,\mathrm{m/s}$ for the velocity, where

Table 1: Evaluation on PETS'09 dataset Ferryman and Shahrokni (2009): MT = Mostly Tracked, ML = Mostly Lost, ID = Number of ID switches.

Dataset	MOTA	MOTP	GT	MT	ML	ID	Rec.	Prec.
S1L1-2	61.8	**68.0**	44	**22**	12	**8**	64.3	**96.5**
Milan et al. (2013)	60.0	61.9	44	21	11	22	64.9	93.7
Rezatofighi et al. (2015)	**63.5**	64.5	44	17	**9**	13	**66.7**	95.8
S1L2-1	27.5	33.2	42	**7**	22	**34**	32.3	88.7
Milan et al. (2013)	29.6	58.8	42	2	21	42	30.9	**98.3**
Rezatofighi et al. (2015)	**32.8**	**57.6**	42	5	**15**	76	**38.6**	89.9
S2L1	86.2	**77.5**	19	**18**	**0**	**6**	96.0	90.8
Milan et al. (2013)	**90.1**	74.3	19	**18**	1	22	**96.8**	**94.1**
Rezatofighi et al. (2015)	–	–	–	–	–	–	–	–
S2L2	**58.6**	59.2	43	10	2	**139**	64.6	**93.3**
Milan et al. (2013)	58.1	**59.8**	43	**11**	1	167	65.1	92.4
Rezatofighi et al. (2015)	58.2	58.5	43	**11**	**0**	143	**69.8**	87.2
S2L3	44.5	**69.0**	44	11	20	**13**	46.2	**96.9**
Milan et al. (2013)	39.8	65.0	44	8	19	27	43.0	94.2
Rezatofighi et al. (2015)	**48.0**	62.3	44	**13**	**18**	23	**52.2**	93.4

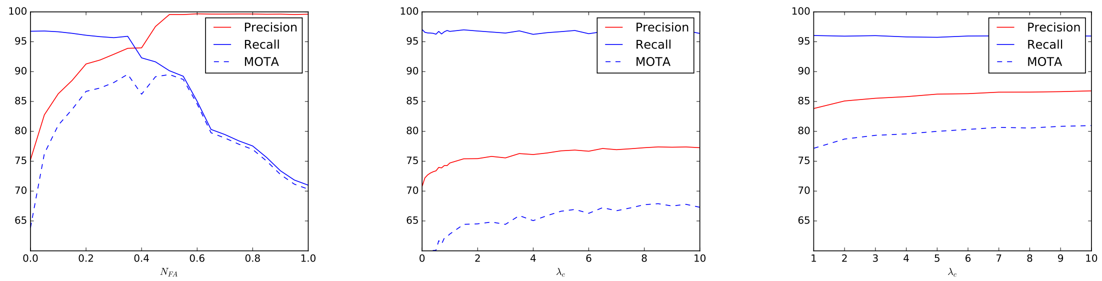
$\Delta t = 1/7$ is the time gap between consecutive frames. The measurement model adds isotropic noise with standard deviation $0.2\,\mathrm{m}$. When sampling from the inverse measurement model, the unobserved velocity was drawn from a normal distribution with standard deviation $1\,\mathrm{m/s}$. Further, we used $\lambda_c = 1.0$, $\lambda_b = 0.5$, $N_{\mathrm{FA}} = 0.2$, $p_{\mathrm{D}}(x) = 0.7$, and $p_{\mathrm{S}}(x) = 0.95$. The class-conditional likelihood of detector confidence scores $P_{\mathrm{fg}}(s)$ and $P_{\mathrm{bg}}(s)$ has been learned from data using Kernel Density Estimation with Gaussian kernel. For training, we used sequences S1L1-1 and S1L2-2 which have been excluded from evaluation.

We used the MOT challenge evaluation software (Leal-Taixé et al., 2015) to compute CLEAR MOT metrics (Bernardin and Stiefelhagen, 2008). All methods that we compare against use the same detections, ground truth, and evaluation criteria. Therefore, evaluation was carried out in 3D using a matching threshold of 1 m. The results of our evaluation are summarized in Table 1. Overall, the presented method performs well in terms of tracking precision, with consistently high ranked precision and MOTP scores. This underlines the state estimation capabilities of the PHD filter, even in dense tracking scenarios with substantial amount of false alarms. In terms of tracking accuracy, MOTA scores of our method are usually lower than those of Rezatofighi et al. (2015), but higher than those of Milan et al. (2013). At the same time, our method produces considerable fewer ID switches on all sequences. We observed that lower MOTA scores are mostly due to a larger number of false negatives. In crowded scenes with high number of detector failures, our tracker produced less, but stable tracks. Therefore the lower MOTA scores, but competitive performance in terms of ID switches and track statistics (MT/ML).

In a second experiment we have compared our confidence-aware PHD recursion against the original formulation of Wojke and Paulus (2016) to investigate our contribution on overall results. Using sequence S2L1 only, we exhaustively searched for optimal clutter parameters, while leaving all others parameters untouched. Plots of several tracking statistics against clutter parameters are shown in Figure 1. With a MOTA score of 89.5 we found the optimal value for state-dependent clutter at $N_{\mathrm{FA}} = 0.35$. Using $P_{\mathrm{fg}}(s) = P_{\mathrm{bg}}(s)$ and $N_{\mathrm{FA}} = 0$, i.e., applying the standard PHD recursion with uniform clutter, we found the optimal value for the expected number of false alarms at $\lambda_c = 10.0$ with a MOTA score of 67.3. Applying additional non maxima suppression, the MOTA score increases to 81.0. While the artificially high clutter rate alone suggests that the uniform distribution does not describe the false alarm process accurately, we see substantial improvement in tracking accuracy due to integration of detector confidence values.

5 CONCLUSIONS

The PHD filter provides a mathematically rigorous framework for multi-object state estimation that is relatively unexplored in the context of visual object



(a) N_{FA} (without non maxima suppression) (b) λ_c (without non maxima suppression) (c) λ_c (with non maxima suppression)

Figure 1: Clutter parameter analysis on S2L1: (a) with detector confidence scores, (b), (c) without detector confidence scores.

tracking. In this paper, we have presented an adapted PHD recursion that incorporates detector confidence scores to mimic state-dependent false alarms as well as a practical SMC implementation that can be integrated into the min-cost flow network formulation of Wojke and Paulus (2016). Our experiments revealed that integration of detector confidence scores has considerable impact on overall applicability of the PHD filter and, in general, our approach achieves results competitive with the current state of the art. FISST and the PHD filter may help to solve open multi-object tracking problems and there is ample opportunity for future work, e.g., integration of appearance information, application of more complex global data association formulations, and object group tracking.

REFERENCES

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

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

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

Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., and Van Gool, L. (2011). Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(9):1820–1833.

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

Ferryman, J. and Shahrokni, A. (2009). An overview of the PETS 2009 challenge. In *PETS*.

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

Leal-Taixé, L., Milan, A., Reid, I., Roth, S., and Schindler, K. (2015). MOTChallenge 2015: Towards a benchmark for multi-target tracking. *arXiv:1504.01942 [cs]*.

Maggio, E. and Cavallaro, A. (2009). Learning scene context for multiple object tracking. *IEEE Trans. Image Process.*, 18(8):1873–1884.

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

Mahler, R. (2007). *Statistical Multisource-Multitarget Information Fusion*. Artech House, Norwood, MA, USA.

Mahler, R. (2014). *Advances in statistical multisource-multitarget information fusion*. Artech House, Norwood, MA, USA.

Mahler, R., Vo, B.-T., and Vo, B.-N. (2011). CPHD filtering with unknown clutter rate and detection profile. *IEEE Trans. Signal Process.*, 59(8):3497–3513.

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

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

Poiesi, F., Mazzon, R., and Cavallaro, A. (2013). Multi-target tracking on confidence maps: An application to people tracking. *Comput. Vis. Image Und.*, 117(10):1257–1272.

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

Rezatofighi, S., Milan, A., Zhang, Z., Shi, Q., Dick, A., and Reid, I. (2015). Joint probabilistic data association revisited. In *ICCV*, pages 3047–3055.

Ristic, B., Clark, D., Vo, B.-N., and Vo, B.-T. (2012). Adaptive target birth intensity for PHD and CPHD filters. *IEEE Trans. Aerosp. Electron. Syst.*, 48(2):1656–1668.

Segal, A. and Reid, I. (2013). Latent data association: Bayesian model selection for multi-target tracking. In *ICCV*, pages 2904–2911.

Wojke, N. and Paulus, D. (2016). Global data association for the probability hypothesis density filter using network flows. In *ICRA*, pages 567–572.

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