www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

BAYESIAN MULTIPLE PERSON TRACKING USING PROBABILITY HYPOTHESIS DENSITY SMOOTHING

S. Hernandez
Laboratorio de Procesamiento de Informacion Geoespacial
Universidad Catolica del Maule, Talca.
Chile
Email: shernandez@ucm.cl
M. Frean
School Of Engineering and Computer Science
Victoria University of Wellington.
New Zealand
Email: marcus.frean@ecs.vuw.ac.nz

Submitted: ***Accepted:*** ***Published:***

Abstract- We present a PHD filtering approach to estimate the state of an unknown number of persons in a video sequence. Persons are represented by moving blobs, which are tracked across different frames using a first-order moment approximation to the posterior density. The PHD filter is a good alternative to standard multi-target tracking algorithms, since overrides making explicit associations

between measurements and persons locations. The recursive method has linear complexity in the number of targets, so it also has the potential benefit of scaling well with a large number of persons being tracked. The PHD filter achieves interesting results for the multiple persons tracking problem, albeit discarding useful information from higher order interactions. Nevertheless, a backward state-space representation using PHD smoothing can be used to refine the filtered estimates. In this paper, we present two smoothing strategies for improving PHD filter estimates in multiple persons tracking. Results from using PHD smoothing techniques in a video sequence shows a slight gain in the cardinality estimates (meaning the number of persons in a particular video frame), but good performance in the individual location estimates.

Index terms*:* Power system, fault current, current limiter, permanent magnet, saturable core, magnetic current limiter, high temperature superconducting fault current limiter.

I. INTRODUCTION

A Bayesian method for the unknown number of targets with unknown association hypotheses has been formulated using point processes and random finite sets theories, under the name of the Probability Hypothesis Density (PHD) filter [1]. The method solves the troublesome multi–target estimation problem by approximating the complete posterior distribution of the filtering density by the first–order moment of a Poisson process. Alike the Kalman filter recursion, the PHD filter approach uses all observations from the past in order to produce instantaneous estimates of the number of targets and their locations. Moreover the PHD recursion can be efficiently computed in closed form using a Gaussian mixture representation or by means of stochastic integration using sequential Monte Carlo methods, so it is suitable for visual tracking applications [2].

Pedestrian counting and tracking is a challenging computer vision task, with applications in surveillance and video monitoring. Analyzing the size of a crowd along with the dynamics of the group and its members has the potential benefit of providing real-time detection of anomalies or events of particular interest. However, because of the complexity of extracting meaningful information from single or multiple cameras, the scope and availability of multiple target tracking techniques for crowd analysis has been restricted to constrained environments and calibration conditions [3].

Traditional target tracking algorithms for pedestrian tracking relies on intra-frame and inter-frame association hypotheses, which relates image measurements to predicted person locations [4]. In order to compute association hypotheses one has to make assumptions which are usually hard to satisfy in real environments, and specially difficult in crowded scenarios. Furthermore, occlusion reasoning and persons merging and splitting into groups, leaves a full posterior distribution on the number of persons and the association hypotheses being intractable [5].

Although the filtering approach provides a fairly accurate way to calculate an instant estimate of the state of a dynamic system, we might expect an improvement if we incorporate more information in the production of the estimate. Rather than considering only the past and current observations, the accuracy of the filtered estimate can be improved by also taking into account future observations [6]. This procedure is widely known as smoothing, and recent research has been undertaken on producing smoothed estimates for the PHD filter [7, 8].

In this paper, we consider unsupervised top-down Bayesian detection and tracking of multiple persons in crowded environments using the PHD filter approach. Even though the method is well suited for tracking a large number of persons observed in clutter which might come from illumination changes, the PHD approximation only holds for tracking scenarios where the signal–to–noise ratio is sufficiently high that a target can be well represented by the observed features [9]. Unfortunately, most of the state the art image processing techniques for person detection would require supervised learning techniques that are not well suited for real time applications [10], relies on multiple cameras [11] or computationally expensive appearance models [12].

More specifically, we propose a PHD smoothing approach for the problem of person tracking in crowded environments. Firstly, background segmentation is used to a generate foreground mask. Secondly, a simple 2D segmentation technique using ground plane information is used to perform person detection. Thirdly, the PHD filter is used to recursively estimate the number of persons and their locations. Fourthly, PHD smoothing is used to refine the instantaneous estimates. A schematic diagram of the procedure for performing detection and tracking is shown in Figure 1.

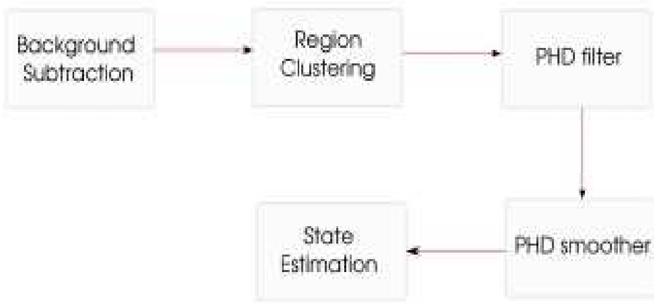


Figure 1. Schematic diagram of the tracking procedure

The contributions of this paper can be briefly summarized as:

1. A method to perform person tracking in crowded environments using a single static camera is proposed. Each person is assumed to move independently of each other, but no restrictions are made about its trajectory and velocity.
2. The PHD filter is described and the application to people tracking is also outlined. The method can deal with clutter originated from errors of the person detection technique and illumination changes.
3. We propose to use smoothing as a method to overcome some drawbacks of the PHD filter approach. Two different smoothing algorithms are presented and then tested using ground truth information. For that purpose, a suitable performance metric for multi–target tracking error estimation is also proposed.

Section II presents a summary of the application of the PHD filter to visual tracking. An introduction to the PHD filter is given in Section III and the sequential Monte Carlo implementation is also presented.

II. RELATED WORK

Detection and tracking of a moving person in a video can be achieved by means of comparing the difference between the current frames from a reference image. This technique is widely known as background subtraction, where the reference frame is usually termed the ‘background model’ [13]. The background model is a representation of the scene without moving parts, and the complexity of level of the model depends on the specific scenario. A basic background subtraction technique can use a unique image as the background model, however this technique easily fail when having small changes of luminance or in the geometry settings. The output of the background subtraction step is a set of connected regions of pixels belonging to the foreground, and is widely known as ‘blobs’. Each region has pixels that form an ellipse or a bounding box that can be tracked from frame to frame. Features of the connected regions are detections that can then be taken as noisy observations for a tracking system [14].

Tracking multiple humans is a challenging application because of the difficulty of generating a similarity function for a person using pixel information. Quantifying the information of a group of pixels using a person detection system can be potentially intractable, if we consider all possible orientations and occlusions. Early works for person detection considered vertical histograms where the head of the people can be distinguished, but this method is not robust in case of occlusion. More recent works have considered person detection using supervised learning by means of cascades of descriptors [15], requiring careful training and testing.

The application of the PHD filter to tracking multiple trajectories from features points in sequences of optical images was described in [16]. More recently, the sequential Monte Carlo (SMC) implementation of the PHD filter was applied to the problem of tracking multiple groups of persons in video [17]. Observations were taken from the moments of the blobs, and morphological operators were used to reduce the level of clutter in the system. The method was then compared with a Gaussian mixture implementation which explicitly accounts for birth, death and survival of targets [18]. The authors also provided a data–driven method for initializing the spatial density of birth and death in a scene.

The PHD filter was also used for tracking faces, people and vehicles using color based change detection in [19]. Since the PHD filter approach avoids computing associations between targets and estimated tracks, graph matching was proposed as a post-processing step for handling the data association problem. The authors reported improved accuracy of the algorithm in cluttered images. An extension to tracking 3D objects locations from multiple cameras have been proposed in [20]. The method is able to handle occlusions being present at a single camera, by fusing information from multiple cameras using the PHD filter. Further developments in the application of the PHD filter in visual tracking has been done by considering more data–driven approaches for designing birth and death proposals using scene information in [21] and [22].

III. PHD FILTER

The problem of performing joint detection and tracking of multiple objects has a natural interpretation under the theory of Poisson point processes [23]. In this case, a model-based approach for detection and tracking of multiple objects can be achieved by using the expectation of a random counting measure. Since a Poisson point process is invariant under transformations, such as thinning, superposition and random translations, the posterior distribution can be also approximated by a Poisson point process [24]. This property becomes extremely useful in visual tracking, where targets may randomly appear or disappear, leaving the number of targets to be modeled as a non-stationary discrete random variable.

A model for tracking multiple objects can perform filtering on a set-valued state X_k, given the history of set-valued observations $Z_{1:k}$. The approach is powerful enough for allowing a time-varying number of objects to appear and disappear, and because no particular order is required on the estimation procedure, the model avoids explicit data association. Furthermore, when using a Poisson spatial model of the new born targets and clutter, it is also possible to determine the expected number targets using the intensity measure of the resulting Poisson process [26,27].

The instances of the two RFS $X_k=\{x_1,x_2,\ldots,x_n\}$ and $Z_k=\{z_1,z_2,\ldots,z_m\}$ represents a set of targets and observations respectively. Bayesian filtering equations are constructed in a similar fashion as their single target filtering counterparts. In this case the RFS filtering and update equations can be written as follows:

$$p(X_k|Z_{1:k-1})=\int p(X_k|X_{k-1})\,p(X_{k-1}|Z_{k-1})\,\delta X_{k-1} \quad (1)$$

$$p(X_k|Z_{1:k})=\frac{p(Z_k|X_k)\,p(X_k|Z_{1:k-1})}{p(Z_k|Z_{k-1})} \quad (2)$$

The probability hypothesis density (PHD) $D(\cdot)$ is defined as the first-order moment or intensity function of a point process with density $p(\{x_1,\ldots,x_n\}|Z_{1:k})=j_n(x_1,\ldots,x_n)$. The PHD repackages the family of Janossy densities into a single function that specifies the probability of having a target x in a neighborhood of $\{x_1,\ldots,x_n\}$, such that the joint density can be written as:

$$D(x) \quad = \sum_{n=0}^{\infty}\frac{1}{n!}\int j_n(x,x,x_1,\ldots,x_n)\,dx_1,\ldots,dx_n \qquad (3)$$

A recursive formula for the filtering densities is given by:

$$D_{k|k-1}(x_k) = \int\left[\pi_s(x_{k-1})p(x_k|x_{k-1})+\gamma_{k|k-1}(x_k|x_{k-1})\right]D_{k-1|k-1}(x_{k-1})dx_{k-1}+b_{k|k-1}(x_k) \quad (4)$$

$$D_{k|k}(x_k)=L_z(x_k)\,D_{k|k-1}(x_k) \quad (5)$$

Where:

$$L_z(x_k)=\left[1-\pi_d(x_k)+\sum_{z\in Z_k}\frac{\pi_d(x_k)\,p(z|x_k)}{\lambda_c c_k(z)+D_k(z)}\right]$$

$$D_k(z)=\int\pi_d(x_k)\,p(z|x_k)\,D_{k|k-1}(x_k)\,dx_k$$

And:

$b_{k|k-1}(x)$: Spontaneous birth density

$\gamma_{k|k-1}(x|x')$: Probability of targets spawning

$p(x|x')$: Single target Markov transition density

$p(z|x')$: Single target likelihood function

$\pi_s(x)$: Probability of target survival

$\pi_d(x)$: Probability of target detection

λ_c : Average number of Poisson false alarms

$c_k(z)$: Spatial distribution of false alarms

The number of targets is calculated as the integral of the PHD $D(\cdot)$ or intensity function of the dynamic point process:

$$N_{k|k} = \int D_{k|k}(x)dx$$

Algorithm 1 describes the Sequential Monte Carlo (SMC) approximation to the PHD recursion as given in [28].

In the SMC implementation of the PHD filter, Monte Carlo samples are used to represent the intensity function, so a larger number of particles are used in areas where targets are more likely to exist. Assuming that we have sample from the posterior PHD distribution, clustering methods can be used for estimating the targets states. K-means and the Expectation-Maximization (EM) algorithms are the main approaches for state estimation for the PHD filter [29]. The total number of targets corresponds to the total particle mass, so target states are computed by clustering particles and using the centroids of each cluster. Furthermore, the authors in [29] also incorporated track continuity in the particle PHD filter by using validation techniques in the state estimation.

Since the PHD filter assumes low observation noise, parametric estimation using EM can be difficult. All data points would potentially be tightly clustered around their centers, introducing numerical instability in the calculation of the variances [31]. Furthermore, having only access to a re-sampled particle approximation could also produce a mismatch between model complexity and the amount of available data. Maximum likelihood approaches for parametric estimation suffers from local minima and over-fitting, as well dependency on the starting point. Bayesian approaches such as the Gibbs sampler can be used instead in order to overcome the problems of deterministic estimation using limited data [32,33].

Algorithm 1 Particle PHD filter

Require: $k \geq 1 \wedge \{w_{k-1}^i, x_{k-1}^i\}_{i=1}^{L_{k-1}}$

Step 1: Prediction step
for $i = 1, .., L_{k-1}$ **do**
  Sample $\tilde{x}_k^i \sim q_k(\cdot|x_{k-1}^i, Z_k)$
  Compute predicted weights $\tilde{w}_{k|k-1}^i = \frac{\phi(\hat{x}_k^i, Z_k)}{q_k(\hat{x}_k^i|x_{k-1}^i, Z_k)} w_{k-1}^i$
end for
for $i = L_{k-1}+1, .., J_k$ **do**
  Sample $\tilde{x}_k^i \sim p_k(\cdot|Z_k)$
  Compute predicted weights for the new born particles
  $\tilde{w}_{k|k-1}^i = \frac{1}{J_k} \frac{\gamma(\hat{x}_k^i)}{p_k(\hat{x}_k^i|Z_k)}$
end for
Step 2: Update step
for all $z \in Z_k$ **do**
  Compute $C_k(z) = \sum_{j=1}^{L_{k-1}+J_k} \psi_{z,k}(\hat{x}_k^j) w_{k|k-1}^j$
end for
for $i = 1, .., L_{k-1} + J_k$, update weights **do**
  $\tilde{w}_k^i = \left[\nu(x) + \sum_{z \in Z_k} \frac{\psi_{z,k}(\hat{x}_k^i)}{\kappa_k(z) + C_k(z)}\right] \tilde{w}_{k|k-1}^i$
end for
Step 3: Resampling step
Compute the total mass $\hat{N}_{k|k-1} = \sum_{j=1}^{L_{k-1}+J_k} \tilde{w}_k^j$
Resample $\{\tilde{w}_k^i/\hat{N}_{k|k}, \tilde{x}_k^i\}_{i=1}^{L_{k-1}+J_k}$
return $\{w_k^i, x_k^i\}_{i=1}^{L_k}$

IV. PHD SMOOTHING

The PHD filter algorithm provides an approximation to the expectation or first-order moment of the intensity measure of a Poisson point process. The method has the property of being able to explicitly model the birth and deaths of targets, as well as clutter and miss-detections, which can also be subject to spawning or merging. This model-based approach can be appealing in multiple tracking systems where the data association step is non-trivial or cannot be optimally solved.

An alternative solution for improving the PHD filter instantaneous estimates is to perform smoothing or retrodiction. Filtered estimates of the individual target states and the posterior cardinality distribution can be considerably improved by considering a higher data frame than the history of observations. More specifically, PHD filtering can be extended to smoothing and is expected to correct the abrupt changes on the estimated number of targets and their states that originate from errors propagated by the filtered distributions.

Let $X_k=\{x_1,\ldots,x_{n_k}\}$ be a set target states in and $Z_{1:T}$ a collection of set-valued measurements collected up to time $T\geq k$. The smoothed PHD can be written as follows:

$$D_{k|T}(x)=\int p(\{x\}\cup X_k|Z_{1:T})\delta X_k \qquad (6)$$

Accordingly, the smoothed number of targets can then be written as:

$$N_{k|T}=\int D_{k|T}(x)dx \qquad (7)$$

As with the standard linear and non-linear smoothing equations, the PHD smoothing problem might be approached by means of fixed-interval smoothing, fixed-lag smoothing or fixed-point smoothing. The algorithms presented here are not dependent on the data interval size, so they can be implemented under each one of these schemes. Notice that, since the PHD is only available for non-ordered sets, the full PHD smoothing distribution $p(X_{1:k}|Z_{1:T})$ is not available, so only the marginal PHD smoothing $D_{k|T}(x)$ in Equation 6 can be approximated. Sections IV-A and IV-B describe two possible approximations.

a. FORWARD-BACKWARD PHD SMOOTHER

Nandakumaran *et.al*. developed a Forward-Backward PHD (FB-PHD) smoother [7] based on a physical-space approach [34].

$$p(X_k|Z_{1:T}) = \int p(X_k, X_{k+1}|Z_{1:T})\delta X_{k+1} \quad (8)$$

$$= \int p(X_{k+1}|Z_{1:T})p(X_k|X_{k+1}, Z_{1:T})\,\delta X_{k+1} \quad (9)$$

$$= p(X_k|Z_{1:k}) \int p(X_{k+1}|Z_{1:T})p(X_{k+1}|X_k)p(X_{k+1}|Z_{1:k})\delta X_{k+1} \quad (10)$$

A particle approximation to the smoothing multi-target density can be written as:

$$\int_B D_{k+1|T}(x)dx = E[\sum_{x_{k+1}\in B} \mathbf{1}_B(x_{k+1})] \quad (11)$$

$$\approx \sum_{i=1}^{L_{k+1}} \mathbf{1}_B(x^i_{k+1})w^i_{k+1|T} \quad (12)$$

Algorithm 2 describes a sequential Monte Carlo approximation to the FB-PHD smoother.

Algorithm 2 Forward-Backward PHD smoother

```
Forward pass.
for k = 0, .., T do
    Perform SMC to get particles and weights {x^i_k, w^i_k}_{1≤i≤L_k}.
end for
Choose w^i_{T|T} = w^i_T.
Backward pass
for k = T − 1, .., 0 do
    for all i ∈ {1, ..., L_k} do
        μ^j_{k+1|k} = b_{k+1|k}(x^j_{k+1}) + Σ_{l=1}^{J_k} w^l_k [π_s(x^l_k) p(x^j_{k+1}|x^l_k) + γ_{k+1|k}(x^j_{k+1}|x^l_k)]
        w^i_{k|T} = w^i_{k|k} [Σ_{j=1}^{L_{k+1}} w^j_{k|T} π_s(x^i_k) p(x^j_{k+1}|x^i_k) / μ^j_{k+1|k}]
    end for
    Compute smoothed estimated number of targets N̂_{k|T} = Σ_{i=1}^{L_k} w^i_{k|T}
    Normalize {w^i_{k|T}}_{1≤i≤L_k} to get {N̂_{k|T} / L_k}_{1≤i≤L_k}.
end for
```

b. TWO-FILTER PHD SMOOTHER

Another approach for PHD smoothing can be achieved by means of the two-filter formula [35]. In this case, the PHD filter has to be combined with the output of a backward information filter, which propagates the posterior distribution of the random counting measure $N_{K|T}$ from Equation 9 to be represented by the following factorization:

$$p(X_k|Z_{1:T}) = p(X_k|Z_{1:k-1}, Z_{k:T}) \tag{13}$$

$$= \frac{p(X_k|Z_{1:k-1})\, p(Z_{k:T}|X_k)}{p(Z_{k:T}|Z_{1:k-1})} \tag{14}$$

$$\propto p(X_k|Z_{1:k-1})\, p(Z_{k:T}|X_k) \tag{15}$$

Where the backward information $p(Z_{k:T}|X_k)$ filter can be written as:

$$p(Z_{k+1:T}|X_k) = \int p(Z_{k+1:T}, X_{k+1}|X_k)\, \delta X_{k+1} \tag{16}$$

$$= \int p(X_{k+1}|X_k) p(Z_{k+1:T}|X_{k+1})\, \delta X_{k+1} \tag{17}$$

The SMC approximation for the backward predicted smoother can then be written as Algorithm 3

Algorithm 3 Two-Filter PHD smoother

Forward pass.
for $k = 0, .., T$ **do**
  Perform SMC to get particles and weights $\{x_k^i, w_k^i\}_{1 \le i \le L_k}$.
end for
Choose $w_{T|T}^i = w_T^i$.
Backward pass
for $k = T-1, .., 0$ **do**
  for all $i \in \{1, \ldots, L_k\}$ **do**
    $\psi_k^l = \sum_{h=1}^{J_k} \pi_d(x_k^l)\, p(z|x_k^l)$
    $L_z(x_k^l) = 1 - \pi_d(x_k^l) + \sum_{z \in Z_k} \frac{\pi_d(x_k) p(z|x_k^l)}{\lambda_c\, C_k(z) + \psi_k^l}$
    $\alpha_{k+1|k}^j = \sum_{l=1}^{J_k} w_{k|k}^l p(x_{k+1}^j | x_k^l)$
    $w_{k|T}^i = L_z(x_k^i) \left[\sum_{j=1}^{L_{k+1}} \frac{w_{k|T}^j \pi_s(x_k^i) p(x_{k+1}^j | x_k^i)}{\alpha_{k+1|k}^j} + 1 - \pi_s(x_k^i) \right]$
  end for
  Compute smoothed estimated number of targets $\hat{N}_{k|T} = \sum_{i=1}^{L_k} w_{k|T}^i$
  Normalize $\{w_{k|T}^i\}_{1 \le i \le L_k}$ to get $\{\frac{\hat{N}_{k|T}}{L_k}\}_{1 \le i \le L_k}$.
end for

V. PHD FILTER AND SMOOTHER FOR PERSON TRACKING AND COUNTING

Instead of using an explicit person detection system, we use a PHD filter approach to estimate the locations of an unknown number of persons. A constant velocity model is used as a generative model for the movement of a single person. The forward model calculates the new position of a person using a velocity vector that remains nearly constant in magnitude and direction.

$x_k = [x_x, x_y]^T$ be the transpose of a 2-dimensional position of a person in the image plane and $\dot{x}_k = [\dot{x}_x, \dot{x}_y]$ its velocity.

In a state-space representation the state vector of a person is written as an augmented vector $\mathbf{x_k} = \left[x_k, \dot{x}_k \right]$. A linear mapping F is used to model the dynamic behavior of a person with Gaussian noise w_k. The position of a single person at the discrete time K can be written as:

$$\mathbf{x_k} = F\mathbf{x_{k-1}} + w_k$$
$$w_k \sim \mathcal{N}(0, \Sigma_{\mathbf{x_k}})$$

where F is a linear transformation matrix in which dt represents the sampling time:

$$F = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The observations $\mathbf{y}_k = \left[y_x, y_y \right]$ only contain information about the position of a person, so velocity has to be estimated from previous measurements [40]. The velocity is related to the object position as $\dot{x}_k = (x_k - x_{k-1})/dt$ for each sampling interval dt. However, since the PHD filter does not perform inter-frame person association, velocity is sampled from a zero-mean Gaussian prior distribution $\mathcal{N}(0, \Sigma_{\dot{x}})$ with diagonal covariance.

The observations are related to the state of a person state by means of a linear transportation matrix G plus Gaussian observation noise v_k:

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

a. Indoor Tracking With Occlusions

In the first experiment, the indoor tracking video 2 sequence from the VISOR dataset (http://imagelab.ing.unimore.it/visor/) is used to illustrate the proposed technique for tracking with occlusions. A temporal Gaussian background model using the parameters specified in Table 1 was used for generating the foreground blobs.

Parameter	Value
Frame buffer (frames)	30
Learning rate	0.75

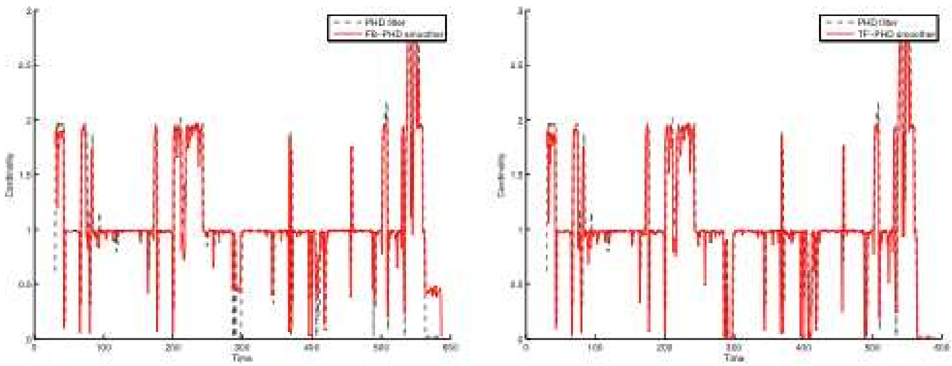
Table 1. Parameter settings for the background subtraction model

The SMC implementations of the PHD filter and the FB-PHD and TF-PHD smoothers are used to recursively estimate the number of persons and their locations. Parameters for the filter are shown in Table 2 and the cardinality estimates are shown in Figure 6. The PHD filter is not able to correctly estimate the number of persons in the presence of occlusions (frame 287 of the sequence). Because there are no detected persons (a.k.a. missed detections), the PHD filter estimate is strongly biased to the error, leaving all particles with negligible weights [41]. The FB-PHD smoother is able to alleviate this effect in a backward pass (see Figure 2(a)). However, this is not the case for the TF-PHD smoother which also uses the observations in order to compute the backward estimate.

Parameter	Value
Number of particles per target	1000
Poisson clutter rate (per unit value)	5e-5
Poisson birth rate (per unit value)	1e-5
uniform spatial clutter density	$U([1, 352] \times [1, 288])$
uniform spatial birth density	$U([1, 352] \times [1, 288])$
initial Poisson birth rate	3

target process noise	*diag*(15, 15, 3, 3)
target observation noise	*diag*(10, 50)
target survival rate	1
target detection rate	0.99

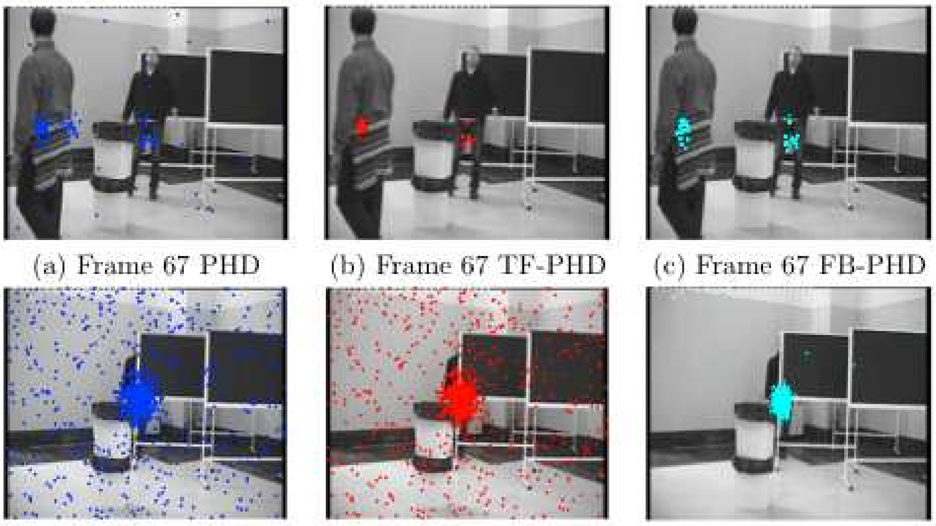
Table 2: Parameter settings for the PHD filter and smoother.



(a) FB-PHD smoother (b) TF-PHD smoother

Figure 2: Cardinality estimates for the PHD filter and smoother. The PHD filter (plotted in dashed lines) fails to estimate the number of persons in the presence of occlusion in frame 287. The FB-PHD smoother is able to recover from the error in a backward pass, but this is not the case for the TF-PHD smoother.

Figure 3 shows the Monte Carlo approximation to the PHD filter for the frame number 67 of the sequence. Location estimates are then obtained by using clustering techniques and the number of clusters corresponds to the PHD cardinality estimates. Both PHD smoothers are able to reduce uncertainty by means of removing spurious samples from the forward pass (see Figures 3(b) and 3(c)).



(a) Frame 67 PHD (b) Frame 67 TF-PHD (c) Frame 67 FB-PHD

(d) Frame 287 PHD (e) Frame 287 TF-PHD (f) Frame 287 FB-PHD

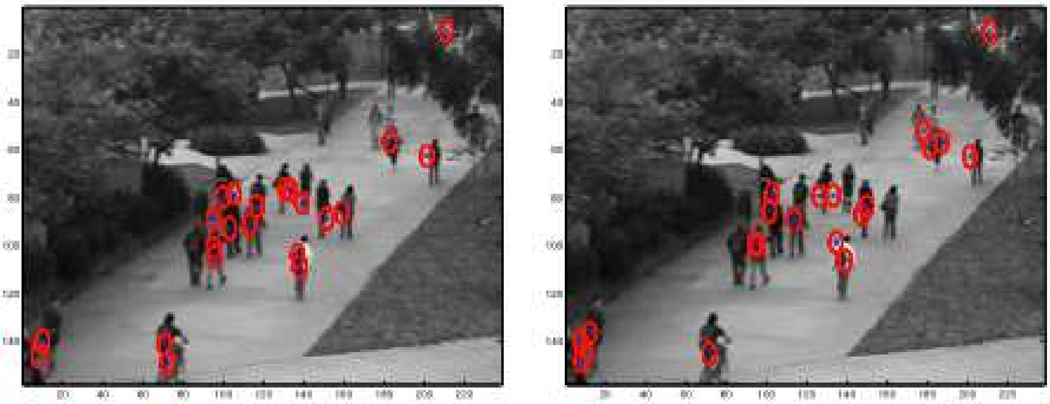
Figure 3: (a) Particle approximation the PHD filter in frame 67. (b) and (c) Reduced uncertainty in frame 67 using the TF-PHD smoother and FB-PHD smoother. (d) Particle approximation the PHD filter in frame 287 with occlusion. (e) the TF-PHD smoother suffers from the missed detection problem. (f) the FB-PHD smoother solves the occlusion problem.

b. People counting and tracking in crowded environments

In this worked example, the practical implications of using the PHD filtering in human tracking in real world surveillance scenarios are studied. For that purpose, a benchmark pedestrian database is used which is publicly available for testing new algorithms in crowd analysis. The UCSDPEDS (http://www.svcl.ucsd.edu/projects/peoplecnt/) dataset contains several videos of pedestrians taken from a stationary surveillance camera. The videos are 8-bit gray scale, with dimensions [238 × 158] at 10 frames per second. We focus on the persons counting and tracking task and the worked examples will show the PHD performance for this case. Figure 3 shows an example of a particular scene from the dataset.

Figure 3: Crowded scenario with multiple people walking in different directions. A single camera captures images at 10 frames per second and the goal is to track and count individual persons

Multiple observations from a single person caused by over-segmentation would cause problems in multi-target tracking methods. Moreover, incorrect person detections would worsen the SNR ratio, deteriorating the performance of the filter. In Figure 4 (frame 20 of the *vidf1 33 001.y* sequence of the dataset), the ellipses are used to enclose detected persons and due to the under-segmentation problem, a group of pedestrians is represented by a single target. Furthermore, because no person recognition has been performed, the estimates are not sensitive to the area occupied by a single person. Therefore, as a consequence of a poor SNR ratio, cardinality and state estimates becomes susceptible to under-segmentation and over-segmentation issues.



(a) PHD filter (b) TF-PHD smoother

Figure 4: Particle PHD filter and TF-PHD smoother estimates for frame 20.

Also, since the PHD filter does not perform any data association, the assessment of the error on individual person locations and velocities is not straightforward, requiring an additional step. Parameters for the PHD filter and smoothers are shown in Table 3.

Parameter	Value
Number of particles per target	150
Poisson clutter rate (per unit value)	1e-4
Poisson birth rate (per unit value)	1e-5
uniform spatial clutter density	U ([1, 238] × [1, 152])
uniform spatial birth density	U ([1, 238] × [1, 152])
initial Poisson birth rate	10
target process noise	*diag*(5, 5, .1, .1)
target observation noise	*diag*(8 4)
target survival rate	0.95
target detection rate	0.95

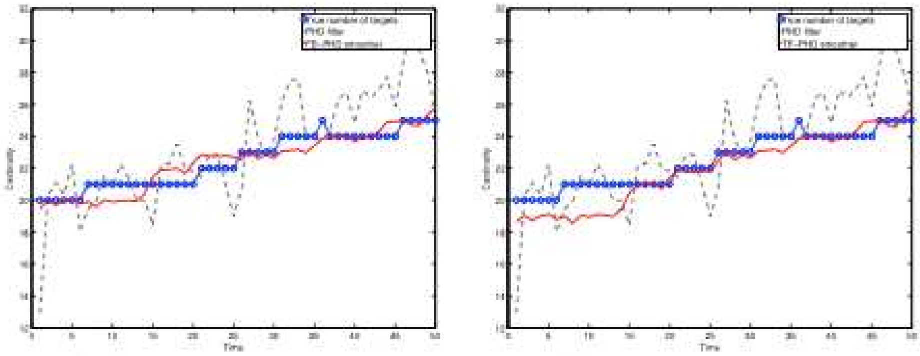
Table 3: Parameter settings for the PHD filter and smoother.

A person with a bicycle has larger area than the expected average, and as a result over-segmentation causes the PHD filter in Figure 4(a) to incorrectly estimate the number of targets in that area. Nevertheless, the TF-PHD smoother in Figure 4(b) is able to give an improved estimate in the region containing a single person.

The estimated number of targets in the backward step is less sensitive to fluctuations in the number of observations (see Table 4). Since estimates and ground truth might have different cardinalities, the OSPA error is used for comparison purposes [36,37]. Figure 5 shows the estimated number of persons for the PHD filter and both smoothers for the first 50 frames of the sequence.

Error	PHD	FB-PHD	TF-PHD
RMS	2.23	1.62	1.53
OSPA (EM)	1.61	1.61	1.60
OSPA (Gibbs sampler)	1.61	1.62	1.61

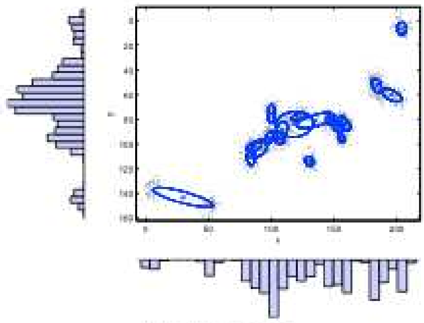
Table 4: OSPA error (with parameters p=2,c=2) for the PHD filter and fixed-interval smoothing for visual tracking.



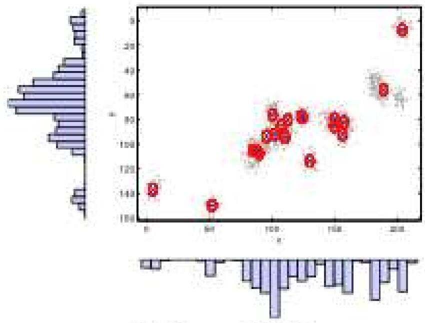
(a) FB-PHD smoother (b) TF-PHD smoother

Figure 4: Crowd counting estimates using the PHD filtering and smoothing. Both, the TF-PHD and the FB-PHD smoothers give an improved estimate of the number of targets.

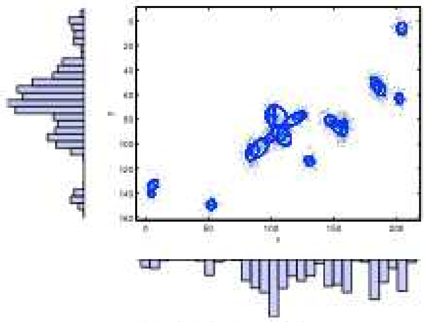
Person locations that are incorrectly addressed due to cardinality errors (wrongly estimated number of persons in the crowd) in the forward pass can be re-estimated in a backward pass. However, since re-sampling was performed in both steps, it is more challenging for the PHD smoothers to provide improved location estimates. Furthermore, since the PHD filter proposes individual samples for each person, location estimates are not sensitive to inter person distances. This issue is also inherited by particle PHD smoothers, so location estimates suffer from the same problem. Figure 5 shows the PHD filter, the FB-PHD and the TF-PHD smoothers using the EM algorithm and the Gibbs sampler in frame 14 of the dataset.



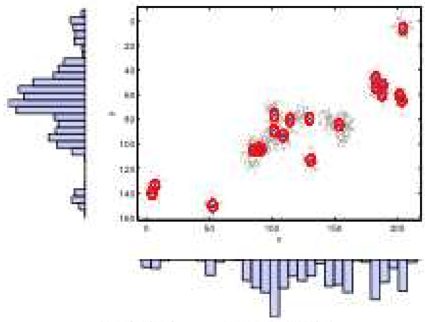
(a) EM PHD



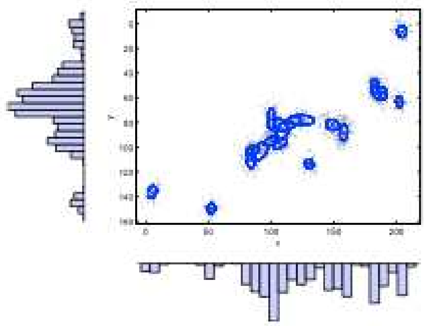
(b) Bayes PHD



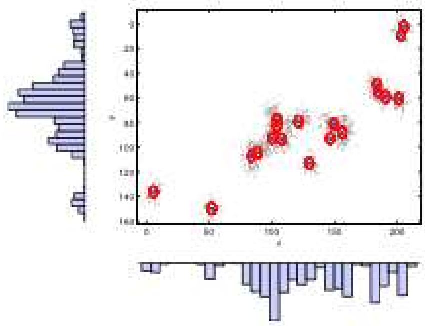
(c) EM FB-PHD



(d) Bayes FB-PHD



(e) EM TF-PHD



(f) Bayes TF-PHD

Figure 5: Particle approximations for frame 14 of the pedestrian tracking sequence. Location estimates from the PHD filter suffers from an incorrectly estimated number of persons. Since the Gibbs sampler is less sensitive to the initial conditions, it manages to allocate person locations more accurately and with less variance than the EM algorithm. Monte Carlo approximations by means of the FB-PHD and the TF-PHD smoothers provide improved estimates over the PHD filter alone.

Now the performance of the PHD smoothing approach on the sequence *vidf1 33 001.y* using fixed-lag implementations is analyzed. As opposed to fixed-interval, fixed-lag implementations can be implemented in real time using a small time lag. Four different time lags are considered and Table 5 shows the performance of the TF-PHD and the FB-PHD smoothers when the EM algorithm and the Gibbs sampler are used for state estimation. In this case we expected to have a large number of outliers in the estimated locations. Therefore, in order to measure the performance of smoothing over filtering, we choose the OSPA metric to be less sensitive to outliers.

Error	PHD	FB-PHD	TF-PHD
Fixed–lag (1 time step)			
RMS	2.62	2.11	2.11
OSPA(EM)	1.60	1.61	1.61
OSPA(Gibbs)	1.60	1.60	1.60
Fixed–lag (2 time steps)			
RMS	2.26	2.04	2.02
OSPA(EM)	1.60	1.59	1.60
OSPA(Gibbs)	1.62	1.59	1.60
Fixed–lag (3 time steps)			
RMS	2.26	1.88	1.86
OSPA(EM)	1.60	1.57	1.57
OSPA(Gibbs)	1.62	1.58	1.59
Fixed–lag (5 time steps)			
RMS	2.26	1.83	1.81
OSPA(EM)	1.60	1.57	1.57
OSPA(Gibbs)	1.62	1.57	1.57

Table 5: Cardinality and OSPA (c=2,p=2) error for the PHD filter and smoothers for visual tracking

Increasing the time-lag improves performance, but it can be seen that the OSPA error for both EM and Gibbs sampler estimation converges at time lag 5.

VI. CONCLUSIONS

An important remark on PHD filter for visual tracking can be discussed in terms of whether measurement-to-measurement and measurement-to-track associations are available or not. If a particular tracking scenario in consideration allows us to concatenate multiple single target filters, then standard multi-hypothesis approach will perform seamlessly without any distributional assumption (e.g. first-order moment approximations). However, if we cannot override clutter using gating techniques or we cannot distinguish between a new-born or an existing target, the algorithm would potentially end up having a combinatorial explosion in the number of association hypotheses.

The PHD filter was originally conceived in a somehow different scenario, where the expected value of the unknown number of targets is calculated by estimating the ratio of false measurements and the likelihood of a single target. Such modeling is useful in highly cluttered environments with targets having large signal-to-noise ratio.

This setup is not always well suited in visual tracking, where the first-order moment approximation has an adversarial effect in the estimation procedure which cannot always be alleviated in a backward pass. Nevertheless, we demonstrated the benefits of two PHD smoothing techniques for estimating person locations. Further work will consider integrating person detection schemes into the PHD filter. Using this approach, the likelihood of a single person would not only consider the false alarms ratio but also the geometry or the shape of each person being also defined by random parameters. Moreover, this stochastic model would also allow departing from the first-order moment approximation to the posterior, including persons interactions and larger occlusions.

REFERENCES

[1] R. Mahler, Multitarget Bayes filtering via first-order multitarget moments, IEEE Transactions on Aerospace and Electronic Systems 39 (4) (2003) 1152–1178.

[2] R. Mahler, Statistical Multisource-Multitarget Information Fusion, Artech House, Norwood, 2007.

[3] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, L. Q. Xu, Crowd analysis: a survey, Mach. Vision Appl. 19 (5-6) (2008) 345–357, ISSN 0932-8092.

[4] D. A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall Professional Technical Reference, ISBN 0130851981, 2002.

[5] T. Zhao, R. Nevatia, Bayesian human segmentation in crowded situations, in: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, vol. 2, ISSN 1063-6919, II–459–66 vol.2.

[6] M. Isard, A. Blake, A Smoothing Filter for CONDENSATION, in: Proceedings of the 5th European Conference on Computer Vision, vol. 1406, Springer-Verlag, 767–781, 1998.

[7] N. Nandakumaran, K. Punithakumar, T. Kirubarajan, Improved multi-target tracking using probability hypothesis density smoothing, in: Signal and Data Processing of Small Targets 2007, URL http://link.aip.org/link/?PSI/6699/66990M/1, 2007.

[8] S. Hernandez, Smoothing algorithms for the Probability Hypothesis Density filter (ECSTR10-13), Tech. Rep., Victoria University of Wellington, 2010.

[9] B.-N. Vo, B.-T. Vo, N. T. Pham, D. Suter, Bayesian multi-object estimation from image observations, in: Information Fusion, 2009. FUSION ’09. 12th International Conference on, 890–898, 2009.

[10] T. Zhao, R. Nevatia, B. Wu, Segmentation and Tracking of Multiple Humans in Crowded Environments, Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (7) (2008) 1198–1211, ISSN 0162-8828.

[11] S. Khan, M. Shah, Tracking multiple occluding people by localizing on multiple scene planes, IEEE Transactions of Pattern Analysis and Machine Intelligence 3 (2008) 505–519.

[12] D. Ramanan, D. A. Forsyth, A. Zisserman, Tracking people by learning their

appearance, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007) 65–81.

[13] C. Stauffer, W. Grimson, Learning patterns of activity using real-time tracking, Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (8) (2000) 747–757, ISSN 0162-8828.

[14] M. Isard, J. MacCormick, BraMBLe: a Bayesian multiple-blob tracker, Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on 2 (2001) 34–41.

[15] N. Dalal, B. Triggs, Histograms of Oriented Gradients for Human Detection, in: CVPR ’05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Volume 1, IEEE Computer Society, Washington, DC, USA, ISBN 0-7695-2372-2, 886–893, 2005.

[16] N. Ikoma, T. Uchino, H. Maeda, Tracking of feature points in image sequence by SMC implementation of PHD filter, in: SICE 2004 Annual Conference, 2004.

[17] Y. Wang, J. Wu, A. Kassim, W. Huang, Tracking a Variable Number of Human Groups in Video Using Probability Hypothesis Density, in: ICPR 2006. 18th International Conference on Pattern Recognition, 2006.

[18] Y.-D. Wang, J.-K. Wu, W. Huang, A. Kassim, Gaussian mixture probability hypothesis density for visual people racking, in: Information Fusion, 2007 10th International Conference on, 1–6, Dec. 2007.

[19] E. Maggio, E. Piccardo, C. Regazzoni, A. Cavallaro, Particle PHD Filtering for Multi-Target Visual Tracking, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007., 2007.

[20] N. Pham, W. Huang, S. Ong, Probability Hypothesis Density Approach for Multi-camera Multi-object Tracking, Computer Vision –ACCV 2007 (2007) 875–884.

[21] Y.-D. Wang, J.-K. Wu, A. Kassim, W. Huang, Data-Driven Probability Hypothesis Density Filter for Visual Tracking, Circuits and Systems for Video Technology, IEEE Transactions on 18 (8) (2008) 1085–1095, ISSN 1051-8215.

[22] E. Maggio, A. Cavallaro, Learning Scene Context for Multiple Object Tracking, Image Processing, IEEE Transactions on 18 (8) (2009) 1873–1884, ISSN 1057-

7149.

[23] R. L. Streit, Poisson Point Processes: Imaging, Tracking, and Sensing, Springer, 2010.

[24] J. F. C. Kingman, Poisson Processes, Oxford University Press, 1993.

[25] C. A. B, N. Vasconcelos, Bayesian Poisson regression for crowd counting, in: EEE 12th International Conference on Computer Vision, 2009.

[26] K. Gilholm, S. Godsill, S. Maskell, D. Salmond, Poisson models for extended target and group tracking, in: Signal and Data Processing of Small Targets 2005, SPIE, 2005.

[27] S. Hernandez, P. Teal, Multi-target Tracking with Poisson Processes Observations, in: Advances in Image and Video Technology, 2007.

[28] B. Vo, S. Singh, A. Doucet, Sequential Monte Carlo methods for multitarget filtering with random finite sets, IEEE Transactions on Aerospace and Electronic Systems 41 (4) (2005) 1224–1245.

[29] D. Clark, J. Bell, Multi-target state estimation and track continuity for the particle PHD filter, Aerospace and Electronic Systems, IEEE Transactions on 43 (4) (2007) 1441–1453, ISSN 0018-9251.

[30] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, ISBN 0387310738, 2006.

[31] M. Tobias, A. Lanterman, Techniques for birth-particle placement in the probability hypothesis density particle filter applied to passive radar, Radar, Sonar & Navigation, IET 2 (5) (2008) 351–365, ISSN 1751-8784.

[32] W. R. Gilks, S. Richardson, Markov Chain Monte Carlo in Practice, Chapman Hall, New York, 1996.

[33] C. P. Robert, G. Casella, Monte Carlo Statistical Methods (Springer Texts in Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, ISBN 0387212396, 2005.

[34] O. Erdinc, P. Willett, Y. Bar-Shalom, The Bin-Occupancy Filter and its Connection to the PHD Filters, Signal Processing, IEEE Transactions on .

[35] M. Briers, A. Doucet, S. Maskell, Smoothing algorithms for state–space models,

Annals of the Institute of Statistical Mathematics 62 (1) (2010) 61–89.

[36] O. E. Drummond, B. E. Fridling, Ambiguities in evaluating performance of multiple target tracking algorithms, in: O. E. Drummond (Ed.), Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 1698 of Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, 326–337, 1992.

[37] J. Hoffman, R. Mahler, Multitarget miss distance via optimal assignment, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 34 (3) (2004) 327–336.

[38] B. T. Vo, Random Finite Sets in Multi-Object Filtering, Ph.D. thesis, School of Electrical, Electronic and Computer Engineering. The University of Western Australia, 2009.

[39] D. Schuhmacher, B.-T. Vo, B.-N. Vo, A Consistent Metric for Performance Evaluation of Multi-Object Filters, Signal Processing, IEEE Transactions on 56 (8) (2008) 3447–3457, ISSN 1053-587X.

[40] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, Estimation with applications to tracking and navigation, Wiley Interscience, New York, 2001.

[41] M. Ulmke, D. Franken, M. Schmidt, Missed detection problems in the cardinalized probability hypothesis density filter, in: Information Fusion, 2008 11th International Conference on, 2008.