State Extraction of Probability Hypothesis Density Filter Based on Dirichlet Distribution

Xiaoxi Yan
Institute of Integrated Automation
Xi'an Jiaotong University
Xi'an, China
yanxiaoxi1981@gmail.com

Chongzhao Han
Institute of Integrated Automation
Xi'an Jiaotong University
Xi'an, China
czhan@mail.xjtu.edu.cn

Jing Liu
Institute of Integrated Automation
Xi'an Jiaotong University
Xi'an, China
elelj20080730@mail.xjtu.edu.cn

Abstract - The Dirichlet distribution is applied to extract the multi-target states in the sequential Monte Carlo implementation of probability hypothesis density filter in this paper. The character of Dirichlet distribution with negative exponent parameters, on which this paper is mainly concentrated, is that it makes the components unstable and competitive with each other so that components with less evidence will be discarded during the iteration process. Maximum likelihood criterion is adopted for the proposed algorithm, while it is implemented by the expectation maximum algorithm. In order to reduce the time cost, the k-dimensional tree is applied to initialize the components of Dirichlet distribution. Simulation results prove that the proposed state extraction algorithm is superior to k-means algorithm and expectation maximum implementation of Gaussian mixture model.

Keywords: probability hypothesis density, state extraction, Dirichlet distribution, k-dimensional tree, expectation maximum.

1 Introduction

Multitarget tracking is to jointly estimate the number of targets and their states from a sequence of noisy and cluttered measurement sets. Multitarget tracking is a class of set-valued estimation problems in nature, in which both the number and the states of multiple targets are random. Although point process theory is recognized as the mathematical fundament for multitarget tracking [1], there have not been systematic and rigorous algorithms until probability hypothesis density (PHD) filter is proposed by Mahler in the tools of random finite sets statistics [2~3]. The probability hypothesis density, whose integral in any region of state space is the expected number of targets in that region, is a "single density" that collapses the multitarget posterior density into a single density on single-target state space. PHD filter incorporates target initiation, tracking, and termination into a whole procedure seamlessly. There are main two implementations of PHD filter, namely, sequential Monte Carlo (SMC) implementation and Gaussian-mixture implementation, which are suitable for nonlinear non-Gaussian dynamics and linear-Gaussian dynamics, respectively [4~5]. Analogous to the superiority of Kalman filter to alpha-beta filter, the performance of PHD filter can be obviously improved by including the second multitarget moment. Form this standpoint, the cardinalized PHD (CPHD) filter, which propagates not only PHD but also the distribution of number of targets, is developed in order to improve the performance of PHD filter [6]. However, CPHD filter is more intractable of computation in general, compared to PHD filter. Gaussian mixture implementation of CPHD filter has been developed under the linear Gaussian assumption about target dynamics and birth process [7]. As PHD and CPHD filters are promising and unified methodologies, they have been widely applied in many fields, such as distributed tracking, active-acoustic tracking, and sensor management and so on. The survey of their various applications can be referred to [8].

As far as the sequential Monte Carlo implementation of PHD filter is concerned, SMC method is applied in approximating the integrals in PHD filter based on the relation of random finite set statistics and probability density function [4]. The PHD is presented by a large set of weighted particles which are propagated over time using a generalizing importance sampling and resampling strategy. The summation of the weights of particles is the expected number of targets since the integral of PHD over the surveillance is the expected number of targets, while the summation is 1 in original particle filter. The output of SMC implementation of PHD filter is a weighted particle set approximating the PHD, not point estimates of individual target states, since the PHD filter's output is an intensity function estimate. As multitarget states are generally pursued in various applications, algorithms for computing point estimates from this weighted particle set become very important. The goal of this paper is to extract the multitarget states from the reasampled particle set based on Dirichlet distribution with negative exponent parameters. This work is meaningful for the consummation of SMC implementation of PHD filter.

The remaining parts are organized as following. The state extraction algorithm based on Dirichlet distribution with negative exponent parameters, are presented in section 2.

Section 3 is devoted to the simulation study. Conclusion is provided in section 4.

2 Algorithm

The problem on which this paper is concentrated, is the state extraction algorithm by which the estimated target state set $\hat{X}$ is achieved from the resampled particle set $\mathbf{X}=\left\{x^{(1)},\ldots,x^{(n)}\right\}$ (i.e., the output of SMC implementation of PHD filter). It is for convenience that the time indexes of particles are neglected here.

Dirichlet distribution, often denoted $Dir(\alpha)$, is a family of continuous multivariate probability distributions parameterized by the vector α of positive reals, which indicate the evidence of corresponding components [9]. It is the multivariate generalization of beta distribution. Let $\pi=(\pi_1,\ldots,\pi_k)\sim Dir(\alpha)$, where order $k\geq 2$ and vector $\alpha=(\alpha_1,\cdots,\alpha_k)$ with parameters $\alpha_1,\ldots,\alpha_k>0$. It has a probability density function with respect to Lebesgue measure on Euclidean space R^{k-1} given by

$$p(\pi_1,\ldots,\pi_{k-1};\alpha_1,\ldots,\alpha_k)=\frac{1}{B(\alpha)}\prod_{i=1}^{k}\pi_i^{\alpha_i-1}, \quad (1)$$

for all $\pi_1,\ldots,\pi_{k-1}>0$ satisfying $\pi_1+\ldots+\pi_{k-1}<1$, where π_k is an abbreviation for $1-\pi_1-\ldots-\pi_{k-1}$ (namely, $\pi_1+\ldots+\pi_{k-1}+\pi_k=1$). The normalizing constant is the multinomial beta function, which can be expressed in terms of the gamma function:

$$B(\alpha)=\frac{\prod_{i=1}^{k}\Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)},\ \alpha=(\alpha_1,\ldots,\alpha_k) \quad (2)$$

Define the $\alpha_0=\sum_{i=1}^{k}\alpha_i$. Then, the expectation and variance of π_i are

$$E[\pi_i]=\frac{\alpha_i}{\alpha_0} \quad (3)$$

and

$$Var[\pi_i]=\frac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(\alpha_0+1)}=\frac{E(\pi_i)(1-E(\pi_i))}{\alpha_0+1}, \quad (4)$$

respectively.

Furthermore, if $i\neq j$,

$$\operatorname{cov}(\pi_i,\pi_j)=\frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}, \quad (5)$$

where the covariance is the scalar since π_i and π_j are scalars. All the pairwise correlations are negative.

Dirichlet distribution is the conjugate prior of the categorical distribution and multinomial distribution in Bayesian statistics. A conjugate prior is an algebraic convenience, giving a closed-form for the posterior and showing intuitively how likelihood updates a distribution. From the standpoint of Bayes' theorem, the resampled particles are assumed to the observations of the estimated multitarget states in the SMC implementation of PHD filter in the proposed algorithm. Then, the likelihood function is the multinomial distribution, where each particle is assumed to be from one of estimated states. The resampled particles are assumed to be independent and identical distribution. Here the Dirichlet prior with negative exponents is adopted for the mixing probabilities of estimated multitarget states,

$$p(\pi_1,\ldots,\pi_k)\propto\exp\left\{c_m\sum_{m=1}^{k}\log\pi_m\right\}, \quad (6)$$

where $c_m=-N/2$ and N is the number of parameters specifying each component in the following part. This Dirichlet prior is the standard Jeffrey prior when c_m is equivalent to -1/2. The estimated state with larger mixing probability is with more evidence from the particle set. The nature of Dirichlet prior with negative exponents is unstable so that it can make components of estimated multitarget states unstable and promotes competition among these components. It is the character of Dirichlet distribution that we use in this paper to extract the estimated multitarget states by iteration procedure after larger initialization of estimated states from particle set (i.e., the initial k is larger.).

Let $\mathbf{X}=[X_1,\ldots,X_d]^T$ be a d-dimensional random variable of particle, with $x=[x_1,\ldots,x_d]^T$ representing one particular particle. As discussion in above that each particle is assumed to be from one of the multitarget estimated states, $\mathbf{X}$ can be modeled as a k-component Gaussian mixture distribution

$$p(x\mid\theta)=\sum_{m=1}^{k}\pi_m p_m(x\mid\theta_m), \quad (7)$$

where $\pi_1,\ldots,\pi_k$ the mixing probabilities, $\theta_m=\{\mu_m,C_m\}$ the set of mean and covariance defining the m th Gaussian component corresponding to one of the estimated states, and $\theta=\{\theta_1,\ldots,\theta_k,\alpha_1,\ldots,\alpha_k\}$ the whole set of parameters. The mixing probabilities must satisfy

$$\sum_{m=1}^{k}\pi_m=1. \quad (8)$$

The mixing probabilities of components in Gaussian mixture distribution are modeled by Dirichlet distribution (6). The log-likelihood of the resampled particle set is

$$\log p\left(\mathbf{X}/\theta\right)=\sum_{i=1}^{n}\log\sum_{m=1}^{k}\pi_m p\left(x^{(i)}/\theta_m\right). \quad (9)$$

In this paper, the maximum likelihood (ML) estimate is adopted

$$\hat{\theta}=\arg\max_{\theta}\left\{\log p\left(\mathbf{X}/\theta\right)\right\}. \quad (10)$$

The ML estimate is implemented by expectation maximum (EM) algorithm, in which the missing part is a set of n labels $Z=\left\{\mathbf{z}^{(1)},\ldots,\mathbf{z}^{(n)}\right\}$ associated with the n particles, indicating which state component each particle is from. Each label is a binary vector $\mathbf{z}^{(i)}=\left[z_1^{(i)},\ldots,z_k^{(i)}\right]$. Then, the complete log-likelihood is

$$\log p\left(\mathbf{X},Z/\theta\right)=\sum_{i=1}^{n}\sum_{m=1}^{k}z_m^{(i)}\log\left(\pi_m p\left(x^{(i)}/\theta_m\right)\right). \quad (11)$$

The estimated mixing parameters consist of the number of components, means and covariances of components during the iterative procedure. E-step an M-step of EM algorithm are alternatingly applied until the relative change of complete log-likelihood is less than the threshold $\varepsilon=10^{-4}$. It is obvious in (11) that the complete log-likelihood $\log p\left(\mathbf{X},Z/\theta\right)$ is linear with respect to the missing Z, since the elements of Z are binary. Thus, the computation of conditional expectation of the complete likelihood in the E-step, given $\mathbf{X}$ and the current estimate $\hat{\theta}\left(t-1\right)$, is reduced to the computation of conditional expectation of Z

$$W\equiv E\left[Z\mid\mathbf{X},\hat{\theta}\left(t-1\right)\right], \quad (12)$$

where t is the index of iteration in EM algorithm. Then, the Q-function is

$$Q\left(\theta,\hat{\theta}\left(t-1\right)\right)=\log p\left(\mathbf{X},W/\theta\right), \quad (13)$$

where the conditional expectation of the elements of Z are given by

$$w_m^{(i)}=\frac{\hat{\pi}_m\left(t-1\right)p\left(x^{(i)}\mid\hat{\theta}_m\left(t-1\right)\right)}{\sum_{j=1}^{k}\hat{\pi}_j\left(t-1\right)p\left(x^{(j)}\mid\hat{\theta}_j\left(t-1\right)\right)}. \quad (14)$$

In M-step, the component parameters (means and covariance matrices) are updated according to

$$\hat{\theta}\left(t\right)=\arg\max_{\theta}Q\left(\theta,\hat{\theta}\left(t-1\right)\right). \quad (15)$$

The update of mixing probabilities is more critical than that of means and covariance of components in M-step, since some components with less evidence from particle set will be discarded. The mixing probabilities are updated by

$$\hat{\pi}_m\left(t\right)=\frac{\max\left\{0,\left(\sum_{i=1}^{n}w_m^{(i)}-\frac{N}{2}\right)\right\}}{\sum_{j=1}^{k}\max\left\{0,\left(\sum_{i=1}^{n}w_j^{(i)}-\frac{N}{2}\right)\right\}}, \quad (16)$$

where the mixing probabilities of some Gaussian components of the estimated states with less evidence from the particle set are set to be zero to implement their removements from the mixture model. In this way, the Dirichlet distribution with negative exponents for weights drives the irrelevant components to extinction. The number of the components is reduced while the ML solution is being searched. Although there is no proof of optimality, it seems reasonable to discard the component m when its weight π_m becomes negative. It also confirms that the mixing weights stay nonnegative. The Dirichlet distribution has influence only on mixing weights while it has no direct influence on means and covariances of components. When one component's updated weight becomes zero, the weights should be normalized in order to redistribute its evidence to survival components. The mean and covariance of the component with zero updated weight are not updated in M-step. As the Gaussian components are adopted here, the mean and covariance of Gaussian component with non-zero mixing probabilities in (16) are updated by

$$\hat{\mu}_m\left(t\right)=\left(w_{sum}\right)^{-1}\sum_{i=1}^{n}x^{(i)}w_m^{(i)} \quad (17)$$

and

$$\hat{C}_m\left(t\right)=\left(w_{sum}\right)^{-1}\sum_{i=1}^{n}\left(x^{(i)}-\hat{\mu}_m\left(t\right)\right)\left(x^{(i)}-\hat{\mu}_m\left(t\right)\right)^{T}w_m^{(i)} \quad (18)$$

where $w_{sum}=\sum_{i=1}^{n}w_m^{(i)}$ are the summation of $w_m^{(i)}$ and they are similar with the references [10~11].

In the initialization of the proposed algorithm, the number of Gaussian components of estimated states is assumed to be very large so that the initial components can spread throughout the resampled particle set. Then, the work to be done after initialization is to remove the components with less evidence from the resampled particle set. Generally, the initial mean vectors are randomly chosen from particle set. However, there are thousands of particles in the particle set so that a much larger number of Gaussian components have to be initialized in order to cover the whole particle set. Thus, the random initialization of mean vectors leads to intractable computation in the whole algorithm. To reduce the computation, a k-dimensional (k-d) tree, which defines a recursive partitioning of the data space into disjoint subspaces, is applied to initialize the mean vectors in this paper [12]. This recursion of k-d tree terminates if a

prespecified number of buckets have been created. Then, the bucket centers are adopted as the initial means vectors of Gaussian components. The number of Gaussian components of k-d tree initialization is much smaller than that of random initialization, under the case that both they can achieve successful initializations (i.e., they can cover the whole particle set.). Therefore, the time cost of the algorithm with k-d tree initialization is much less. The initial covariance matrices are 1/10 of the mean global diagonal covariance matrix

$$\sigma^2 = \frac{1}{10d} trace\left(\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)}-\mathbf{m}\right)\left(x^{(i)}-\mathbf{m}\right)^T \right), \quad (19)$$

where $\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}$ is the global data mean of particle set. The initial weights of are the same as the diagonal of the initial larger number of components.

The free covariance is adopted for each component in the algorithm. Thus, the number of parameters specifying each component is $N = d + d\left(d+1\right)/2$. In the iterations of EM algorithm, covariance matrix of component may be close to singular occasionally. Analogous to the soft constraint on the covariance matrices [13], a much smaller diagonal matrix is added in the covariance matrix at each time step in proposed algorithm.

3 Simulation study

A two-dimensional scenario with unknown and time varying number of targets is considered here. The surveillance region is [-1000, 1000] × [-1000, 1000] (in meter). The state of each target consists of position and velocity, while measurement of target is the position. Each target moves according to the following dynamics,

$$x_k = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} T^2/2 & 0 \\ 0 & T^2/2 \\ T & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} v_{1,k} \\ v_{2,k} \end{bmatrix},$$

where $x_k = \left[x_{1,k}, x_{2,k}, x_{3,k}, x_{4,k}\right]^T$ the state, $\left[x_{1,k}, x_{2,k}\right]^T$ the position, $\left[x_{3,k}, x_{4,k}\right]^T$ the velocity at time k , and $T = 1$ the sampling period. The process noises $\left\{v_{1,k}\right\}$, $\left\{v_{2,k}\right\}$ are mutually independent zero-mean Gaussian white noise with respective standard deviations $\sigma_{v_1} = 5$ (m/s^2) and $\sigma_{v_1} = 5$ (m/s^2). There are four targets over 100 scans in total, and the number of targets is unknown and variable. Target 1 and target 2 remain existing during 100 scans. Target 3 appears at time step 1 and disappears at time step 95. Target 4 appears at time step 5 and disappears at time step 100. Each existing target has a constant probability of survival $e_{k/k-1} = 0.95$. There is no spawning during the tracking procedure for simplicity. New targets can appear spontaneously according to a Poisson point process with intensity function $\gamma_k = 0.2N\left(\cdot;\overline{x},Q\right)$, where

$$\overline{x} = \begin{bmatrix} -600 \\ -500 \\ 0 \\ 0 \end{bmatrix}, Q = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix}$$

and $N\left(\cdot;\overline{x},Q\right)$ denotes a normal density with mean $\overline{x}$ and covariance Q .

The target-originated measurements are given by

$$y_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix}$$

with $\left\{w_{1,k}\right\}$ and $\left\{w_{2,k}\right\}$ mutually independent zero-mean Gaussian white noise with standard deviations $\sigma_{w_1} = \sigma_{w_2} = 10$ m. The measurement noise is assumed independent of the process noise. The detection probability $P_D\left(x\right)$ of every target is assumed 1.

Clutter process is modeled as a Poisson random finite set, while each clutter measurement is assumed to be uniform over the surveillance region. Thus, clutter intensity is

$$\kappa_k\left(z_k\right) = \lambda \bullet c_k\left(z_k\right),$$

where λ is the average number of clutter measurements per scan, namely, the parameter of Poisson distribution, and $c\left(z\right)$ is the uniform distribution of clutter. In simulation, parameter λ is assumed 50.

One thousand particles per expected target are used in the PHD filter, thus, the number of particles varies throughout the simulation. The importance sampling densities are transition density of single target and spontaneous birth density. The proposed algorithm of state extraction is applied in the reampled particle set approximating the intensity to obtain the estimated states. The large number of Gaussian components in Dirichlet distribution at the initialization step is set to be 30. Thus, the desired number of buckets in k-d tree portioning the particle set for the initialization of Dirichlet components should be slightly larger than 30. The desired number of buckets in k-d tree is assumed to be 35 here. Then, the 30 initial means of components are selected from these 35 centers of buckets in k-d tree. The individual x and y coordinates of the true tracks of four targets and the estimates of SMC implementation of PHD filter with the proposed state extraction algorithm over the surveillance region at each time step are displayed in Fig.1 and Fig.2 after one Monte

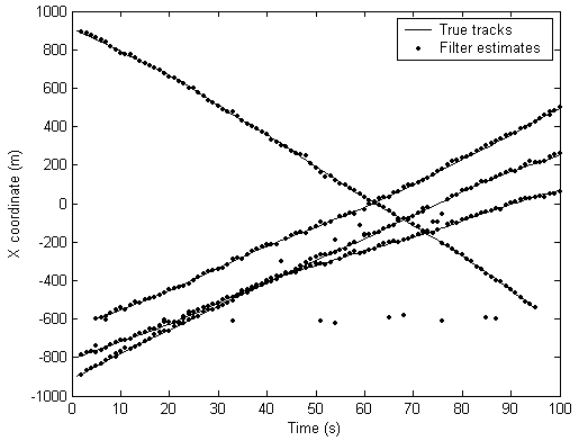

Figure 1: True tracks and estimates of X coordinates

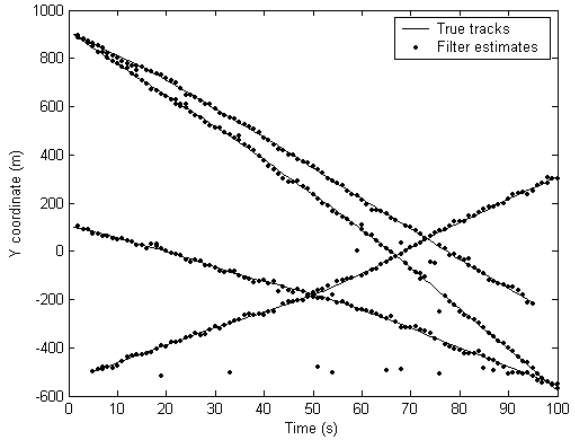

Figure 2: True tracks and estimates of Y coordinates

Carlo simulation. The plots indicate that the proposed state extraction algorithm is able to estimate the multitarget states when the estimates of the number of targets are correct.

As the cardinalities of true and estimated of multitarget states set may be not equal, the Wasserstein distance is used for performance evaluation of the SMC implementation of PHD filters with the proposed state extraction algorithm. The Wasserstein distance is defined by

$$d_p(\hat{X}, X) = \min_{C} \sqrt[p]{\sum_{i=1}^{|\hat{X}|}\sum_{i=1}^{|X|} C^{ij} \left\| \hat{x}^i - x^j \right\|^p}$$

$\hat{X}$ and X are the estimated and true sets of multitarget states [14]. The minimum is taken over the set of all transportation matrices C (a transportation matrix is one whose entries C^{ij} satisfy $C^{ij} \geq 0$, $\sum_{j=1}^{|X|} C^{ij} = 1/\left|\hat{X}\right|$). It is not defined if either X or $\hat{X}$ is empty. The typical

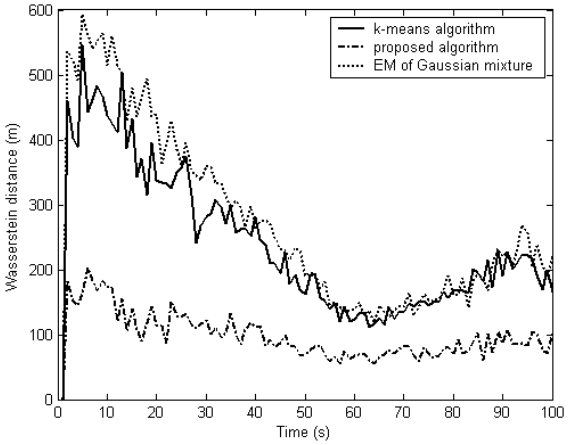

Figure 3: The averaged Wasserstein distances of 100 Mont Carlo simulations

k-means algorithm and EM implementation of Gaussian mixture model, whose detail can be referred to [15], are chosen as the comparison algorithms for the proposed algorithm of state extraction. Fig. 3 shows the averaged Wasserstein distances between the positions estimates of the multiple target states and ground truth at each time step. It is obvious that the proposed algorithm of state extraction is superior to other two algorithms, while the k-means algorithm is slight better than EM implementation of Gaussian mixture model. The reason is that the proposed algorithm based on Dirichlet distribution with negative exponent parameters extracts the means of components with large evidence from particle set as the estimated states.

4 Conclusion

A state extraction algorithm based on Dirichlet distribution with negative exponent parameters is proposed for the sequential Monte Carlo implementation of PHD filter. It is an important nature of Dirichlet distribution with negative exponent parameters that it is unstable. The proposed algorithm is implemented by EM algorithm with the ML criterion, under the assumption that each particle is from one of the estimated states. A k-d tree is also applied in initialization to reduce the time cost. The proposed algorithm is superior to the k-means algorithm and EM implementation of Gaussian mixture model for state extraction from the resampled particle set.

References

[1] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes (Second Edition)*, Springer, New York, 2002.

[2] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Transactions on Aerospace and Electronic Systems*, Vol 39, No. 4, pp. 1152–1178, October 2003.

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

[4] B.-N. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, Vol 41, No. 4, pp. 1224–1245, October 2005.

[5] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, Vol 54, No. 11, pp. 4091–4104, November 2006.

[6] R. Mahler, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, Vol 43, No. 4, pp. 1523-1543, October 2007.

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

[8] R. Mahler, "A survey of PHD filter and CPHD filter implementations," *Signal Processing, Senor Fusion, and Target Recognition* XVI 2007, Orlando, 9 April, Proc. 6967, pp. 65670O-1-65670O-12.

[9] K. Samuel, K. Balakrishman, and J. Norman, *Continous multivariate distributions*, Wiley, New York, 2000.

[10] M. Figueiredo and A.K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 24, No. 3, pp. 381-396, March 2002.

[11] Z. Zivkovic and F. Van der Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 26, No. 5, pp. 651–656, May 2004.

[12] R.F., Sproull, "Refinements to nearest-neighbor searching in k-dimensional trees," *Algorithmica*, Vol. 6, No. 4, pp. 579-589, April, 1991.

[13] M. Kloppenburg and P. Tavan, "Deterministic annealing for density estimation by multivariate normal mixtures," *Physical Review E*, Vol. 55, pp. R2089-R2092, March, 1997.

[14] J. Hoffman, and R. Mahler, "Multitarget miss distance via optimal assignment," *IEEE Transactions on Systems, Man, and Cybernetics–Part A*, Vol 34, No. 3, pp. 327-336, May 2004.

[15] D.E. Clark, and J. Bell, "Multi-target state estimation and track continuity for the particle PHD filter," *IEEE Transactions on Aerospace and Electronic Systems*, Vol 43, No. 4, pp.1441-1453, October 2007.