

# A Random Finite Set Approach for Joint Detection and Tracking of Multiple Wideband Sources Using a Distributed Acoustic Vector Sensor Array

Xionghu Zhong and A. B. Premkumar

Centre for Multimedia and Network Technology, School of Computer Engineering,  
College of Engineering, Nanyang Technological University, Singapore, 639798.  
{xhzhong and asannamalai}@ntu.edu.sg

**Abstract**—This paper considers the problem of tracking multiple wideband acoustic sources in three dimensional (3-D) space using a distributed acoustic vector sensor (AVS) array. Least square approaches have been proposed to fuse the DOA measurements and estimate the 3-D position. However, the performance of position estimation can be seriously degraded by inaccurate DOA estimates, and also multiple source localization is impossible unless the DOA estimates can be associated to each source correctly. In this paper, A random finite set (RFS) approach is developed to jointly detect and track multiple wideband acoustic sources. An RFS is able to characterize the randomness of the state process (i.e., the dynamics of source motion and the number of active sources) as well as the measurement process (i.e., source detections, false alarms and miss detections). Since deriving a closed-form solution does not exist for the multi-source probability density, a particle filtering approach is employed to arrive at a computationally tractable approximation of the RFS densities. Simulations in different tracking scenarios demonstrate the ability of the proposed approaches in multiple acoustic source detection and tracking.

**Index Terms**—Acoustic vector sensor, direction of arrival, particle filtering, detection and tracking, random finite set.

## I. INTRODUCTION

Detection, localization and tracking of wideband acoustic sources in a three dimensional (3-D) space are important topics in signal processing and have many applications such as room speech enhancement, underwater target surveillance and sonar signal processing. The tasks are traditionally performed by using an array system either with large aperture or equipped with multiple hybrid arrays. In recent years, a new technology namely acoustic vector sensor (AVS) has been widely employed for acoustic source detection and localization, and different signal processing algorithms have been developed accordingly. AVS employs a co-located sensor structure and measures acoustic pressure as well as particle velocity at sensor position [1]. The manifold structure suggests that AVS has following advantages over traditional pressure sensors:

- 1) It produces both the azimuth and elevation information and enables 2-D DOA estimation with a single AVS.
- 2) Its elevation ranges between  $[-\pi/2, \pi/2]$  and allows elevation angle estimation unambiguously.
- 3) The manifold is independent of the source signal's frequency, which make AVS suitable for wideband source

signal or scenarios where the source signal's frequency is unknown *a priori*.

Due to these merits, both the theoretical aspects and the applications of AVS have been widely studied [1]–[4]. A full description of AVS in signal processing problems can be found in [1]. Traditional DOA estimation approaches such as Capon beamforming and subspace based approaches using AVSs have been investigated in [2], [3]. However, such investigations focus only on the DOA estimation, rather than the 3-D ( $x$ -,  $y$ -,  $z$ -) position estimation. Recently, advances in distributed sensor arrays in providing unprecedented capabilities for target detection and localization have motivated the deployment of distributed sensor arrays for acoustic source detection and localization [4], [5]. In [4], least square approaches have been developed for 3-D source position estimation. At each AVS, Capon beamforming is employed to estimate the DOA of the source. These DOA estimates are then employed to triangulate a 3-D location by using weighted least-square (WL) and re-weighted least-square (RWL) based approaches. However, such approaches assume that the source is static and relatively a large number of snapshots is required to achieve satisfactory performance. Also these approaches can be applied only for single source localization. For multiple sources, a data association method should be employed to associate the DOA estimates with each individual source.

In practice, existence of source is usually unknown and the number of sources may be time-varying in the tracking scene. Hence, more advanced approaches are needed to jointly detect and track a time-varying number of acoustic sources. In this paper, a random finite set (RFS) framework is introduced to deal with this problem by using a distributed AVS array. In essence, an RFS is a random process that is random in the cardinality as well as in the values of each element. It is thus able to naturally represent the state and measurement processes of multiple source detection and tracking problem. In the state space, each element of a RFS is a random vector which can be employed to describe the source motion dynamics, and the cardinality is a random variable which can be used to model the time-varying number of sources. Similar structure can also be constructed for the DOA measurements to describe the source detection, false alarms and missed

detection. Generally, RFS framework neglects the intrinsic data association between sources and measurements, and has been found promising for multi-object tracking problem [6]–[10]. For rigorous mathematical discipline of RFS framework and its application in multi-object tracking problem, the reader is referred to [6]–[10]. Particularly, RFS is employed for room acoustic source detection and tracking in [8].

The core contributions of this work is that RFS based approaches have been derived for multiple wideband acoustic source detection and tracking problem. The rest of this paper is organized as follows. In Section II, the AVS signal model is introduced. Section III presents the tracking algorithm developed for unknown and time-varying number of acoustic sources. Simulations are organized in Section IV and conclusions are drawn in Section V.

## II. SIGNAL MODEL AND INDIRECT METHOD

Assume that  $m_t$  wideband acoustic source signals  $s_m(t)$ , for  $m = 1, \dots, m_t$  impinge on an array equipped with  $N$  spatially distributed AVSs at discrete time  $t$ . The 2-D DOA of the  $m$ th source relative to the  $n$ th sensor ( $n = 1, \dots, N$ ) can be written as  $\theta_n^m(t) = [\phi_n^m(t), \psi_n^m(t)]^T$  where  $\phi_n^m(t) \in (-\pi, \pi]$  and  $\psi_n^m(t) \in [-\pi/2, \pi/2]$  represent the azimuth and elevation angles respectively, and superscript  $T$  denotes the transpose. AVS measures acoustic pressure as well as three component particle velocities. Let  $\mathbf{u}_n^m(t)$  be the unit direction vector pointing from the sensor toward to the source and given as

$$\mathbf{u}_n^m(t) = \begin{bmatrix} \cos \psi_n^m(t) \cos \phi_n^m(t) \\ \cos \psi_n^m(t) \sin \phi_n^m(t) \\ \sin \psi_n^m(t) \end{bmatrix}, \quad (1)$$

The received signal at the  $n$ th AVS can be modeled as [1]

$$\mathbf{y}_n(t) = \sum_{m=1}^{m_t} \mathbf{a}(\theta_n^m(t)) s_m(t) + \epsilon_n(t), \quad (2)$$

where  $\mathbf{a}(\theta_n^m(t)) = [1, (\mathbf{u}_n^m(t))^T]^T \in \mathbb{C}^{4 \times 1}$  is the steering vector, and  $\epsilon_n(t) \in \mathbb{C}^{4 \times 1}$  represent the channel noise including the pressure and velocity noise terms. Note that we have normalized the particle velocity terms by multiplying a constant term  $-\rho_0 c_0$ , where  $\rho_0$  and  $c_0$  represent the ambient density and the propagation speed of the acoustic wave in the medium respectively. For the measurement noise process, it is further assumed that  $\epsilon_n(t)$  is a sequence of complex-valued independent and identically distributed (i.i.d.) circular Gaussian random variables with zero mean and covariance matrix  $\mathbf{\Gamma}$ , given as  $\epsilon_n(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma})$ . The distances between the source and the sensors are assumed to be much larger than the maximum wavelength of acoustic signal. Each source signal has an i.i.d. random amplitude  $\varepsilon(t)$  and random phase  $\zeta(t)$ , i.e.,  $s(t) = \varepsilon(t)e^{j\zeta(t)}$ . This means that  $s(t)$  is a wide-band signal and is uncorrelated from one snapshot to the next.

Assume that  $T_0$  snapshots are considered at each time step  $k$ . When  $T_0$  is small, the source can be assumed stationary and  $\theta_n^m(k)$  is used to replace  $\theta_n^m(t)$  in the measurement frame.

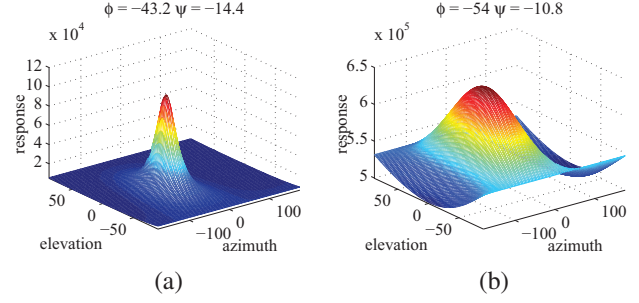


Fig. 1. Response of the Capon beamforming under the environments (a) SNR = 10dB; (b) SNR = -10dB. The source signal is located at  $(-44.4^\circ, -14.4^\circ)$ . The estimated DOA is labeled at the top of each figure.

Equation (2) can be written as

$$\mathbf{Y}_n(k) = \mathbf{A}(\theta_n(k))\mathbf{S}(k) + \epsilon_n(k), \quad (3)$$

where  $\mathbf{A}(\theta_n(k)) = [\mathbf{a}(\theta_n^1(k)), \dots, \mathbf{a}(\theta_n^{m_k}(k))]$  and  $\mathbf{S}(k) = [s_1(k), \dots, s_{m_k}(k)]^T$ . Capon beamforming response is [2]

$$\mathcal{P}_k^n(\theta) = \{\mathbf{A}^H(\theta)(\mathbf{R}_k^n)^{-1}\mathbf{A}(\theta)\}^{-1}, \quad (4)$$

where  $\mathbf{R}_k^n$  is the covariance matrix given as

$$\mathbf{R}_k^n = \mathbb{E}\{\mathbf{Y}_n(k)\mathbf{Y}_n(k)^H\} \approx \frac{1}{T_0}\mathbf{Y}_n(k)\mathbf{Y}_n(k)^H, \quad (5)$$

where  $\mathbb{E}$  is the expectation operation, and the superscript  $H$  denotes the conjugate transpose. The DOA estimation can easily be obtained by implementing a 2-D search over  $\theta$  which can maximize the output of Capon beamformer

$$\hat{\theta}_k = \arg \max_{\theta \in (-\pi, \pi] \times [-\pi/2, \pi/2]} |\mathcal{P}_k(\theta)|, \quad (6)$$

where  $|\cdot|$  denotes the amplitude of a complex value. Assume that the  $n$ th AVS is deployed at arbitrary locations  $\mathbf{x}_n^0 = [x_n^0, y_n^0, z_n^0]^T$  and the  $m$ th source is located at  $\mathbf{x}_{m,k} = [x_{m,k}, y_{m,k}, z_{m,k}]^T$ . According to the array geometry, the 2-D DOA  $\theta_n^m(k)$  is related to the source position by

$$\begin{aligned} \phi_n^m(k) &= \tan^{-1} \left( \frac{x_{m,k} - x_n^0}{y_{m,k} - y_n^0} \right); \\ \psi_n^m(k) &= \tan^{-1} \left( \frac{z_{m,k} - z_n^0}{\sqrt{(x_{m,k} - x_n^0)^2 + (y_{m,k} - y_n^0)^2}} \right); \end{aligned} \quad (7)$$

In [4], indirect approaches have been developed for 3-D localization. The DOA measurements at each AVS is estimated first by using (6). The DOAs are then regarded as measurements and employed to triangulate the 3-D source position by using weighted least-square (WL) and re-weighted WL (RWL) approaches. However, such approaches tend to be erroneous by the inaccurate DOA estimates. In noisy environments where the SNR is relatively high, Capon spectra is able to present the source DOA by a sharp peak as shown in Fig.1(a). However, when the SNR is low, the peak may be distorted and the estimated DOA may diverge from the ground truth, as shown in Fig. 1(b). Also, the approaches can be used only for single source localization. Although the authors suggest a possible

extension to multiple source localization, e.g., obtaining DOAs of multiple sources by enumerating multiple peaks. To achieve 3-D localization is not a trivial task since a sophisticated data association technique is required to identify the DOAs generated by each source.

### III. RANDOM FINITE SET APPROACH FOR JOINT DETECTION AND TRACKING

This section presents our solution towards to tracking an unknown and time-varying number of sources. Essentially, a RFS framework is formulated to characterize the randomness of source dynamics as well as measurement uncertainties. The RFS state process is introduced first.

#### A. RFS state model formulation

Assume that the  $m$ th source moves with a velocity  $\dot{\mathbf{x}}_{m,k} = [\dot{x}_{m,k}, \dot{y}_{m,k}, \dot{z}_{m,k}]^T$ . The source state  $\mathbf{x}_{m,k}$  can be constructed by cascading the position and velocity parts, i.e.,  $\mathbf{x}_{m,k} = [\mathbf{x}_{m,k}^T, \dot{\mathbf{x}}_{m,k}^T]^T$ . CV model [11] is employed here to model the source dynamics given as

$$\mathbf{x}_{m,k} = \mathbf{F}\mathbf{x}_{m,k-1} + \mathbf{G}\mathbf{v}_k, \quad (8)$$

where the coefficient matrix  $\mathbf{F}$  and  $\mathbf{G}$  are defined respectively by

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_3; \quad \mathbf{G} = \begin{bmatrix} \Delta T^2 \\ \Delta T \end{bmatrix} \otimes \mathbf{I}_3, \quad (9)$$

where  $\Delta T$  represents the time period in seconds between the previous and current time step, and  $\otimes$  denotes the Kronecker product, and  $\mathbf{v}(k)$  is a zero-mean real Gaussian process, i.e.,  $\mathbf{v}(k) \sim \mathcal{N}(\mathbf{0}, \Sigma_v)$ . For simultaneously detecting and tracking of unknown number of multiple acoustic sources, the parameters of interest will be the number of sources as well as the 3-D position of each source. The state of a single source at current time step  $k$  is stated as  $\mathbf{x}_{m,k}$ , for  $m = 1, \dots, m_k$ . All the parameters of interest can be characterized by using a single finite set, given as

$$\mathcal{X}_k = \{\mathbf{x}_{1,k}, \dots, \mathbf{x}_{m_k,k}\}, \quad (10)$$

where  $m_k = |\mathcal{X}_k|$  is the number of sources, with  $|\cdot|$  representing the cardinality. Given a realization  $\mathcal{X}_{k-1}$  of the RFS state at previous time step  $k-1$ , the source state  $\mathcal{X}_k$  at current step  $k$  is modeled by

$$\mathcal{X}_k = \mathcal{B}_k(\mathbf{b}_k) \cup \mathcal{S}_k(\mathbf{X}_{k-1}) \quad (11)$$

where  $\mathcal{B}_k(\mathbf{b}_k)$  is the state vector of sources born at time step  $k$ ,<sup>1</sup> and  $\mathcal{S}_k(\mathbf{X}_{k-1})$  denotes the RFS of states that have survived at time step  $k$ . For the birth process, we assume that:

- at most one source is born at a time step;

<sup>1</sup>Since the time-varying number of sources is considered, the source dynamics is not only the source motion itself, but also source birth and death process. In this paper, we use birth and death processes to describe the source appearance and disappearance in the tracking scene, and accordingly, the terminology 'born' is employed to represent that new source appears in the surveillance area, and 'die' refers to existing source which disappears from the surveillance area.

- $m_k < M_{\max}$  where  $M_{\max}$  is the maximum number of sources at each time step.

The first assumption is employed to simplify the problem. Also it is plausible to make such an assumption since the number of sources we have considered here is relatively small. In practice, it is possible that multiple sources turn up simultaneously. Further, the maximum number of sources in the surveillance area is bounded at  $M_{\max}$ , i.e.,  $m_k \leq M_{\max}$ . This means that when  $|\mathcal{X}_{k-1}| = M_{\max}$ , the new born state vector is an empty set, i.e.,  $\mathcal{B}_k = \emptyset$ . It is observed in [12] that for a single AVS, up to two sources can be uniquely identified. Hence,  $M_{\max} = 2$  is chosen in this work.<sup>2</sup> The source birth process can thus be formulated as

$$\mathcal{B}_k(\mathbf{b}_k) = \begin{cases} \emptyset, & \bar{h}_{\text{birth}}; \\ \{\mathbf{b}_k\}, & \bar{h}_{\text{birth}}; \\ \emptyset, & |\mathcal{X}_{k-1}| = M_{\max}. \end{cases} \quad (12)$$

where  $\bar{h}_{\text{birth}}$  and  $\bar{h}_{\text{birth}}$  are the hypotheses for birth process and non-birth processes respectively, and  $\mathbf{b}_k$  is the initial state vector under the birth hypothesis. For the initialization of the new source, the position part is assumed to be uniformly distributed over the possible position range, and the velocity part is assumed to be a Gaussian distribution around the ground truth velocity.  $\mathbf{b}_k$  is thus given as

$$\mathbf{b}_k = \mathbf{x}_0 \sim \mathcal{U}(\mathbf{x}) \times \mathcal{N}(\dot{\mathbf{x}}_0, \Sigma_0), \quad (13)$$

where  $\mathcal{U}(\mathbf{x})$  is a uniform distribution over the range of possible  $\mathbf{x}$  and  $\Sigma_0$  characterize the uncertainties in the velocity part. The survived state set  $\mathcal{S}_k(\mathcal{X}_{k-1})$  can be formulated by considering a death process. Assume that  $\bar{h}_{\text{death}}$  and  $\bar{h}_{\text{death}}$  are the hypotheses for death process and non-death processes respectively. When a death process happens, the corresponding state will be set as empty, and the remaining states will evolve following the motion dynamic equation (8).  $\mathcal{S}_k(\mathcal{X}_{k-1})$  can thus be given as

$$\mathcal{S}_k(\mathcal{X}_{k-1}) = \begin{cases} \mathcal{X}_{k-1} \setminus \{\mathbf{x}_{m,k}\}, & \bar{h}_{\text{death}} \text{ for} \\ & m\text{th source;} \\ \bigcup_{m=1}^{|\mathcal{X}_{k-1}|} \{\mathbf{F}\mathbf{x}_{m,k-1} + \mathbf{G}\mathbf{v}_k\}, & \bar{h}_{\text{death}}. \end{cases} \quad (14)$$

with  $\setminus$  denoting the set minus. The RFS state transition density can thus be expressed by the product of birth PDF and survival PDF, given as

$$p(\mathcal{X}_k | \mathcal{X}_{k-1}) = p(\mathcal{B}_k | \mathcal{X}_{k-1}) p(\mathcal{S}_k | \mathcal{X}_{k-1}) \quad (15)$$

Assume that the birth and death processes happen with prior probability  $P_{\text{birth}}$  and  $P_{\text{death}}$  respectively. The PDF of birth process can be formulated as

$$p(\mathcal{B}_k | \mathcal{X}_{k-1}) = \begin{cases} 1 - P_{\text{birth}}, & \mathcal{B}_k = \emptyset; \\ P_{\text{birth}} p(\mathbf{x}_0), & \mathcal{B}_k = \{\mathbf{x}_0\}; \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

<sup>2</sup>Note that when the number of sources is larger than two, an AVS array could always be employed.

To formulate the PDF of death process  $p(\mathcal{S}_k|\mathcal{X}_{k-1})$ , we firstly consider a single source death process. Source dies with following PDF

$$p(\mathcal{S}_k(\mathbf{x}_{m,k-1})|\mathcal{X}_{k-1}) = \begin{cases} P_{\text{death}}, & \mathcal{S}_k(\mathbf{x}_{m,k-1}) = \emptyset; \\ (1 - P_{\text{death}})p(\mathbf{x}_{m,k}|\mathbf{x}_{i,k-1}), & \mathcal{S}_k(\mathbf{x}_{m,k-1}) = \{\mathbf{x}_{m,k}\}; \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Note that for the death process, we also assume that at most one source can die at a time step. The total PDF of death process can be written as

$$p(\mathcal{S}_k|\mathcal{X}_{k-1}) = P_{\text{death}}^{m_k - m_{k-1}} (1 - P_{\text{death}})^{m_{k-1}} \sum_{1 \leq i_1 \neq i_m \leq m_{k-1}} \prod_{j=1}^{m_k} p(\mathbf{x}_{m,k}|\mathbf{x}_{i_j,k-1}) \quad (18)$$

where

$$\sum_{1 \leq i_1 \neq i_m \leq m_{k-1}} = \sum_{i_1=1}^n \sum_{i_2=1, i_2 \neq i_1}^n \cdots = \sum_{i_1=1}^n \sum_{i_m=1, i_m \neq i_{m-1} \neq \cdots \neq i_1}^n \quad (19)$$

Hence the RFS state transition PDF is formulated. In practice,  $P_{\text{birth}}$  and  $P_{\text{death}}$  are unknown and are obtained based on experimental studies. Usually, increasing  $P_{\text{birth}}$  or  $P_{\text{death}}$  will make the algorithm easier detect the source birth or death respectively. However, overly large  $P_{\text{birth}}$  or  $P_{\text{death}}$  will lead to an overestimation or underestimation on the number of sources.

### B. RFS model for 2-D DOA measurement

In [4], linear intersection methods have been developed to localize a single static source. It is desired that the largest peaks from Capon response (4) of each AVS corresponds to the source DOAs. However, due to interference of noise and interaction among source signals, spurious peaks may present and some of them may be even higher than the peaks corresponding to the real source. Hence, the peaks larger than a threshold (say 1/20 of the height of the largest peak) are collected to obtain the DOA estimates. This results in a finite set measurement model for the  $n$ th AVS at time step  $k$

$$\mathcal{Z}_k^n = \{\hat{\theta}_{1,k}^n, \dots, \hat{\theta}_{\kappa_k^n,k}^n\} \quad (20)$$

where  $\kappa_k^n = |\mathcal{Z}_k^n|$  denotes the number of measured DOAs. From this measurement set, following remarks can be made:

1. It is desirable that the DOAs generated by real sources are included in the measurement set. However, it is unknown *a priori* that which ones are generated by real sources and which one is due to a specific source  $\mathbf{x}_{m,k}$ ;
2. Due to interference from noise and multiple source signals themselves, spurious peaks may be collected and  $\mathcal{Z}_k^n$  may contain false DOA estimates;
3. Source may be undetected and no DOA measurement is related to the source. This case occurs when the two sources are closely spaced that the Capon beamforming cannot differentiate them due to a resolution problem.

Based on these observations, the measurement set can be divided into two separate sets which contain false DOA measurements generated by spurious peaks  $\mathcal{C}_k^n$  and measurements generated by sources  $\mathcal{G}_k^n$  respectively. Hence the measurement set is modeled by

$$\mathcal{Z}_k^n = \underbrace{\{\hat{\theta}_{1,k}^n, \dots, \hat{\theta}_{\bar{\kappa}_k^n,k}^n\}}_{\text{generated by sources } \mathcal{G}_k^n} \cup \underbrace{\{\hat{\theta}_{1,k}^n, \dots, \hat{\theta}_{\hat{\kappa}_k^n,k}^n\}}_{\text{false alarms } \mathcal{C}_k^n}, \quad (21)$$

where  $\bar{\kappa}_k^n$  and  $\hat{\kappa}_k^n$  are the number of source generated measurements and the number of false alarms respectively, i.e.,  $|\mathcal{G}_k^n| = \bar{\kappa}_k^n$  and  $|\mathcal{C}_k^n| = \hat{\kappa}_k^n$ , and  $\bar{\kappa}_k^n + \hat{\kappa}_k^n = \kappa_k^n$ . The source is detected with a probability of detection  $P_d$ . The number of false alarms is usually assumed to follow a Poisson distribution. Hence the probability of false alarms is

$$P_f(\hat{\kappa}_k^n) = \frac{e^{-\lambda_f} \lambda_f^{\hat{\kappa}_k^n}}{\hat{\kappa}_k^n!} \quad (22)$$

where  $\lambda_f$  is a Poisson factor which denotes the average rate of false alarms.

Given a DOA measurement which is generated by clutter, the likelihood is assumed to be a uniform distribution within the possible 2-D DOA range  $\theta$  given by,

$$p(\hat{\theta}_{i,k}^n|\mathcal{X}_k = \emptyset) = \mathcal{U}_\theta[(-\pi, \pi/2], [\pi, \pi/2]] = \frac{1}{2\pi^2}. \quad (23)$$

where  $\mathcal{U}_d[a, b]$  is a uniform distribution over the possible range  $[a, b]$  for variable  $d$ . If the measurement is generated by a real source, the likelihood is assumed to be the true DOA corrupted by an additive Gaussian noise. The likelihood can then be written as,

$$p(\hat{\theta}_{i,k}^n|\mathcal{X}_k = \{\mathbf{x}_{m,k}\}) = \mathcal{N}(\hat{\theta}_{i,k}^n; \mathbf{h}(\mathbf{x}_{m,k}), \Sigma_\theta). \quad (24)$$

where  $\mathbf{h}(\mathbf{x}_{m,k})$  is the measurement function (7) and  $\Sigma_\theta$  is the variance which describes DOA estimation error. Hence, when the source state is an empty set, the set likelihood is this

$$p(\mathcal{Z}_k^n|\mathcal{X}_k = \emptyset) = P_f(\kappa_k^n) \left( p(\hat{\theta}_{i,k}^n|\mathcal{X}_k = \emptyset) \right)^{\kappa_k^n} \quad (25)$$

The total likelihood is thus

$$p(\mathcal{Z}_k^n|\mathcal{X}_k) = \sum_{\tilde{\mathcal{Z}}_k^n \subseteq \mathcal{Z}_k^n} g(\tilde{\mathcal{Z}}_k^n|\mathcal{X}_k = \emptyset) P_d^{m_k} (1 - P_d)^{|\tilde{\mathcal{Z}}_k^n| - m_k} \sum_{1 \leq i_1 \neq i_m \leq |\tilde{\mathcal{Z}}_k^n|} \prod_{j=1}^{|\tilde{\mathcal{Z}}_k^n|} p(\hat{\theta}_{j,k}^n|\mathbf{x}_{i_j,k}) \quad (26)$$

Hence, the RFS representation of the time-varying number of sources and measurements processes is achieved. In the following section, the particle filtering approach will be introduced to obtain the source position as well as source number estimates.

### C. Sequential Monte Carlo tracking algorithm

Above description gives an RFS formulation for AVS signal based detection and tracking problem. The Bayesian recursive estimation of the posterior distribution of the RFS state  $p(\mathcal{X}_k|\mathcal{Z}_k)$  can be written as

$$p(\mathcal{X}_k|\mathcal{Z}_{1:k-1}) = \int_{\mathcal{F}} p(\mathcal{X}_k|\mathcal{X}_{k-1})p(\mathcal{X}_{k-1}|\mathcal{Z}_{1:k-1})\mu(d\mathcal{X}_{k-1}) \quad (27)$$

$$p(\mathcal{X}_k|\mathcal{Z}_{1:k}) \propto p(\mathcal{Z}_k|\mathcal{X}_k)p(\mathcal{X}_k|\mathcal{Z}_{1:k-1}) \quad (28)$$

where  $p(\mathcal{X}_k|\mathcal{X}_{k-1})$  characterizes the birth, death and survival processes of the state dynamics. The subscript  $\mathcal{F}$  is the collection of all finite subsets of the state space, and  $\mu(d\mathcal{X}_{k-1})$  is a measure of  $\mathcal{F}$ . Note that in the case that  $\mu$  is the Lebesgue measure,  $\mu(d\mathcal{X}_{k-1})$  is the same as  $d\mathcal{X}_{k-1}$ . Since this work is from an application point of view, mathematical concepts on the set integration and these PDF constructions are beyond the scope of this paper. Readers are referred to [7], [9] for a detailed RFS definition and derivation of these PDFs.

Since the source position is nonlinearly related to the measurements, closed-form solution for the PDF of DOA estimates is not available. In this work, a particle filtering approach is employed to approximate the PDFs. Assume that we have particles  $\mathcal{X}_{k-1}^{(\ell)}$  for  $\ell = 1, \dots, L$  at previous time step  $k-1$ , and the corresponding importance weight  $w_{k-1}^{(\ell)}$ . The particles at current time step  $k$  are generated according to the state dynamic process described in Section III-A, given as

$$\mathcal{X}_k^{(\ell)} \sim p(\mathcal{X}_k^{(\ell)}|\mathcal{X}_{k-1}^{(\ell)}) \quad (29)$$

Since a prior distribution is employed as the importance function, the particles are weighted by

$$w_k^{(\ell)} = w_{k-1}^{(\ell)}p(\mathcal{Z}_k|\mathcal{X}_k^{(\ell)}) \quad (30)$$

where

$$p(\mathcal{Z}_k|\mathcal{X}_k^{(\ell)}) = \prod_{n=1}^N p(\mathcal{Z}_k^n|\mathcal{X}_k^{(\ell)}) \quad (31)$$

After resampling, the posterior distribution is thus approximated by

$$p(\mathcal{X}_k^{(\ell)}|\mathcal{Z}_k) \approx \sum_{\ell=1}^L \tilde{w}_k^{(\ell)} \delta_{\mathcal{X}_k^{(\ell)}}(\mathcal{X}_k) \quad (32)$$

where  $\tilde{w}_k^{(\ell)}$  is the normalized weight.  $\delta_{\mathcal{X}}(\mathcal{Y})$  is a set-valued Dirac delta function. For brevity,  $\delta_{\mathcal{X}}(\mathcal{Y})$  is defined such that  $\delta_{\mathcal{X}}(\mathcal{Y}) = 1$  if  $\mathcal{X} \subseteq \mathcal{Y}$  and 0

Due to a RFS presentation, each particle may differ from others in dimension and the elements in it have an arbitrary order. Extracting the state estimation is thus not as straightforward as that in the known and fixed source number scenario. Based on the particles and the corresponding weights  $\{\mathcal{X}_k^{(\ell)}, w_k^{(\ell)}\}_{\ell=1}^L$ , the number of sources can be approximated by

$$\bar{m}_k \approx \sum_{\ell=1}^L w_k^{(\ell)}|\mathcal{X}_k^{(\ell)}| \quad (33)$$

Since the number of sources should be an integer, we obtain the estimation of source number by using a rounding operation, i.e.,  $\hat{m}_k = \lceil \bar{m}_k \rceil$ . A K-means algorithm is then employed to cluster all the RFS particles. The centroids of these clusters  $\{\hat{\mathbf{x}}_{m,k}\}_{m=1}^{\hat{m}_k}$  are taken as the final state estimates. It is worth mentioning that when setting the birth and death priors as zero, the RFS particle filtering approach is a general sequential importance sampling based particle filtering, which is widely employed for single source tracking problem. General sequential importance sampling based PF has been developed for DOA tracking of an acoustic source in [13], [14].

### D. Performance metric

The performance evaluation of multiple source tracking is fundamentally different from that of single source tracking in that the state estimation is a finite set in which the estimated source number may be different with the ground truth. Since the dimension of the state estimation can be different from the ground truth, the error should take both the state divergence and the dimension mismatch into account. Defining a metric which is able to characterize all these errors is a problem on its own. In [15], the authors proposed an optimal subpattern assignment (OSPA) metric for such a problem. A penalty value is employed in OSPA to transfer the cardinality error into the state error and then OSPA is able to present the performance on source number estimation as well as source position estimation. Assume that  $\hat{\mathcal{X}}_k = \{\hat{\mathbf{x}}_{1,k}, \dots, \hat{\mathbf{x}}_{\hat{m}_k,k}\}$  is an estimation of the ground truth state set  $\mathcal{X}_k = \{\mathbf{x}_{1,k}, \dots, \mathbf{x}_{m_k,k}\}$ . Note that here the state cardinality estimation  $\hat{m}_k$  may not equal to the ground truth  $m_k$ . The OSPA error metric is defined as [15]

$$e_{\text{OSPA}}(\hat{\mathcal{X}}_k, \mathcal{X}_k) = \min_{\sigma} \sqrt[p]{\frac{1}{m_k} \left( \sum_{i=1}^{\hat{m}_k} d^{(c)}(\hat{\mathbf{x}}_{i,k}, \mathbf{x}_{\sigma_i,k})^p + c^p(m_k - \hat{m}_k) \right)} \quad (34)$$

if  $\hat{m}_k \leq m_k$ , and  $e_{\text{OSPA}}(\hat{\mathcal{X}}_k, \mathcal{X}_k) = e_{\text{OSPA}}(\mathcal{X}_k, \hat{\mathcal{X}}_k)$  if  $\hat{m}_k > m_k$ . Here,  $c > 0$  is a penalty value which determines the relative weighting of the cardinality error component. The function  $d^{(c)}(\cdot)$  is defined as  $\min(c, d(\cdot))$  to guarantee that the distance error is cut off at  $c$ . In OSPA metric, an appropriate selection of  $c$  is able to transfer the cardinality error component into localization error as part of the total error, and  $p \in (0, \infty)$  is a real positive number. In [15], the authors suggest that  $p = 2$  is a more practical choice since it yields smooth distance curves. Also, such a choice make the localization error similar to the error metric in the single source scenario, i.e., root-mean-square error (RMSE). More straightforward, OSPA metric can be interpreted as two error parts contributed by state error and cardinality error, which can be respectively given as

$$e_{\text{loc}}^{(c)}(\hat{\mathcal{X}}_k, \mathcal{X}_k) = \min_{\sigma} \sqrt{\frac{1}{m_k} \sum_{i=1}^{\hat{m}_k} \|\mathbf{C}\hat{\mathbf{x}}_{i,k} - \mathbf{C}\mathbf{x}_{\sigma_i,k}\|^2} \quad (35)$$

and

$$e_{\text{card}}(\hat{\mathcal{X}}_k, \mathcal{X}_k) = \sqrt{\frac{c^2}{m_k} (\hat{m}_k - m_k)} \quad (36)$$

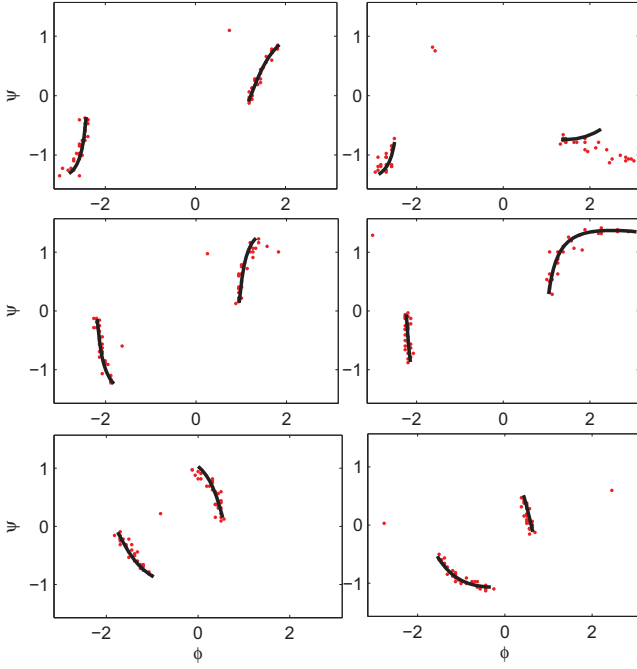


Fig. 2. Capon beamforming based DOA measurements from six sensors under SNR = 8dB and  $T_0 = 32$ . From top to bottom: #1, #3 and #5 (left), #2, #4 and #6 (right). The solid black lines are the ground truth DOAs and the red dots are the DOAs extracted from Capon beamformer.

where the minimization is taken over all possible permutation  $\sigma$ , and  $\mathbf{C} = [\mathbf{I}, \mathbf{0}]$  such that  $\mathbf{C}\mathbf{x}_k$  outputs the 3-D position part of the state. Generally, a large value of  $c$  means a significant penalty on cardinality error, and vice versa. The distance estimation error is cut off at  $c$  to make sure that the state error is smaller than the error made by cardinality mismatch. In next section, equation (34) will be used to evaluate the performance of time-varying number of sources tracking. For the parameters in the error metrics,  $p = 2$  and a moderate penalty value  $c = 100$  will be employed.

#### IV. SIMULATIONS

Six sensors are deployed to formulate a distributed AVS array. The sensor locations are: (30, -26, 40.39)m, (60, -21, 169.95)m, (0, 0, 0)m, (40, 38, -10.57)m, (-65, 40, -5.43) m, and (-100, -10, 51.80)m. Such a sensor deployment is exactly the same as that in [4]. a time-varying number of sources is considered: one (S1) is active from (-10, -40, -120)m to (-100, -140, -20)m, and the other (S2) from (10, 40, 120)m to (100, 140, 20)m with 30 time steps. S1 is active from time step 1 to 30, and S2 from time step 10 to 40. Such motion results in a velocity of  $\pm 3.5$ m/sec. roughly. The simulations are implemented under different noisy environments and with different numbers of snapshots. The background noise level is evaluated by SNR, and is simulated by adding the complex circular i.i.d. Gaussian noise into the received signal. The parameters for RFS approach

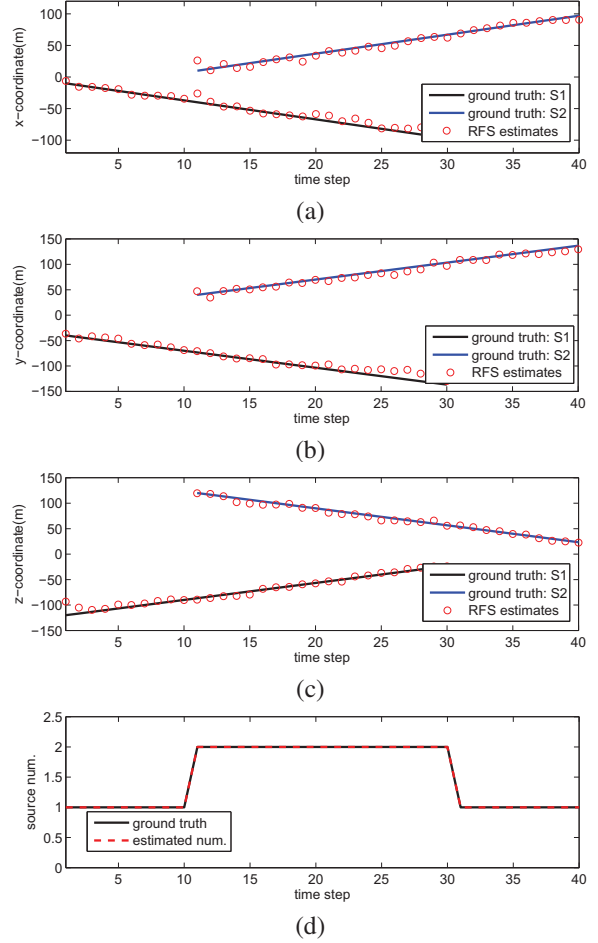


Fig. 3. Single source estimation results under SNR = 8dB and  $T_0 = 32$  for (a)  $x$ -; (b)  $y$ -; (c)  $z$ - coordinate; and (d) source number.

are set as:  $L = 1000$ ,  $P_{\text{birth}} = 0.8$ ,  $P_{\text{death}} = 0.15$ ,  $\lambda_f = 0.1$ ,  $\Sigma_\theta = 0.01\mathbf{I}_2$ , and  $\Sigma_0 = \Sigma_v = 0.01\mathbf{I}_{3M}$ . This parameter setup is found adequate for all following simulations. The source velocities are initialized around the ground truth. The initial positions are uniformly distributed over the possible source location area.

In the first simulation, the proposed approach is implemented under SNR = 0dB and  $T_0 = 32$ . Fig. 2 shows the DOA measurements from six sensors. It shows that good DOA estimation can be obtained based on Capon beamforming approach. Heavy false alarms and miss detections only happen in some time steps of sensor #2. The tracking results from a single implementation is shown in Fig. 3. The proposed tracking approach is able to track the number and the trajectories of the sources very accurately.

In the second simulation, implementation under SNR = -2dB and  $T_0 = 128$  is considered. Fig. 4 shows the DOA measurements from six sensors. It shows that the DOA estimation is not very accurate. For some sensors, there are heavy false alarms and miss detections in a number of time steps. The tracking results from a single implementation is shown

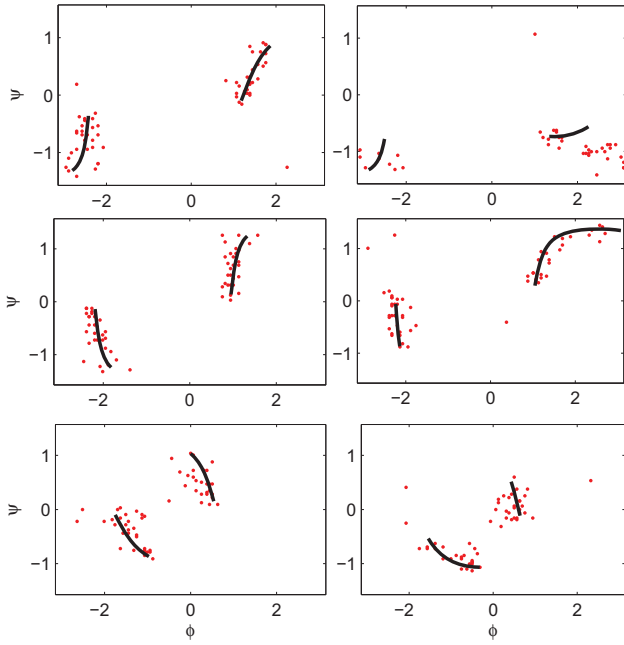


Fig. 4. Capon beamforming based DOA measurements from six sensors under  $\text{SNR} = -2\text{dB}$  and  $T_0 = 128$ . From top to bottom: #1, #3 and #5 (left), #2, #4 and #6 (right). The solid black lines are the ground truth DOAs and the red dots are the DOAs extracted from Capon beamformer.

in Fig. 5. The proposed tracking approach is able to track the number and the trajectories of the sources accurately. Even though when the DOA measurements are seriously distorted and large miss detection presents across some of the sensors.

To obtain the average detection and tracking performance, the proposed algorithm is also implemented under different noise environments and using different numbers of snapshots. The root-mean-square error (RMSE) over 100 Monte Carlo (MC) runs under  $\text{SNR} = 8\text{dB}$ ,  $T_0 = 32$ ;  $\text{SNR} = -2\text{dB}$ ,  $T_0 = 128$  and  $\text{SNR} = -6\text{dB}$ ,  $T_0 = 256$  is presented in Fig. 6. It shows that the proposed RFS tracking approach is able to provide good accuracy for 3-D location estimation as well as source number estimation. In most time steps, the sources can be detected and tracked. The degradation only happens at the source birth and death steps since the algorithm usually needs burn-in period to converge to the ground truth. However, such a burn-in period is very short in all simulations. Also, due to the interference between the sources, the position estimation performance is degraded when two sources are simultaneously active.

## V. CONCLUSIONS

A random finite set approach for 3-D wideband acoustic source tracking using a distributed AVS array is developed. Traditionally, least square approaches are proposed to fuse the DOA measurements and estimate the 3-D position. Such approaches can be rendered erroneous by inaccurate DOA

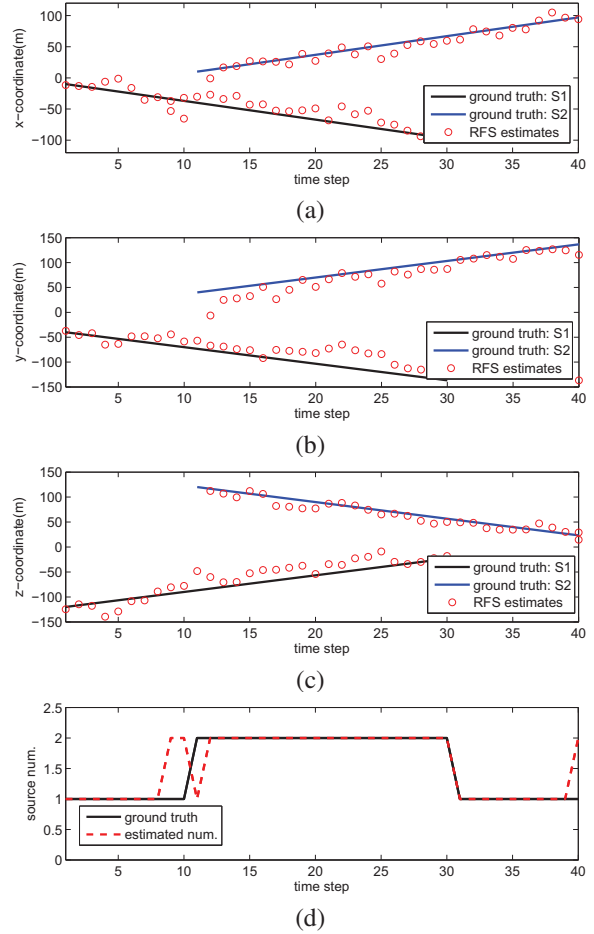


Fig. 5. Single source estimation results under  $\text{SNR} = -2\text{dB}$  and  $T_0 = 128$  for (a)  $x$ -; (b)  $y$ -; (c)  $z$ - coordinate; and (d) source number.

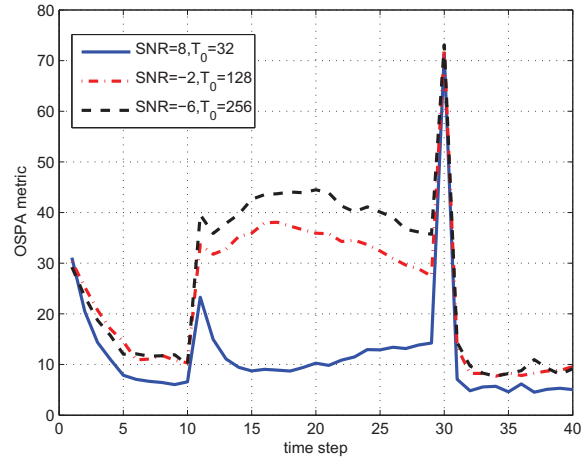


Fig. 6. Time-varying number of sources scenario: OSPA for 100 MC runs.

estimates, and also multiple source localization is impossible unless the DOA estimates can be associated to each individual

source correctly. In this work, an RFS framework is employed to formulate the source birth and death processes as well as false alarms and source detections. Hence, the time-varying nature of the source state process and DOA measurement process can be characterized. A particle filtering approach is also introduced to arrive at a computationally tractable approximation of the RFS densities. Simulations in different tracking scenarios demonstrate the ability of the proposed approaches in tracking a time-varying number of acoustic sources. In our future work, tracking a larger number of sources ( $M_{\max} > 2$ ) will be considered and the applications of the proposed approach in real acoustic environment will be investigated.

#### REFERENCES

- [1] A. Nehorai and E. Paldi, "Acoustic vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2481–2491, 1994.
- [2] M. Hawkes and A. Nehorai, "Acoustic vector-sensor beamforming and capon direction estimation," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2291–2304, 1998.
- [3] K. T. Wong and M. D. Zoltowski, "Self-initiating MUSIC-based direction finding in underwater acoustic particle velocity-field beamspace," *IEEE J. Ocean. Eng.*, vol. 25, no. 2, pp. 262–273, 2000.
- [4] M. Hawkes and A. Nehorai, "Wideband source localization using a distributed acoustic vector-sensor array," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1479–1491, 2003.
- [5] S. Zhou and P. Willett, "Submarine location estimation via a network of detection-only sensors," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3104–3115, 2007.
- [6] R. Mahler, "'statistics 101" for multisensor, multitarget data fusion," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 53–64, Jan 2004.
- [7] B.-N. Vo, S. Singh, and A. Doucet, "Sequential monte carlo methods for multitarget filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.
- [8] W.-K. Ma, B.-N. Vo, S. S. Singh, and A. Baddeley, "Tracking an unknown time-varying number of speakers using tdoa measurements: A random finite set approach," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3291–3304, 2006.
- [9] R. P. S. Mahler, *Statistical Multisource-Multitarget information fusion*. Artech House, 2007.
- [10] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic, "Bernoulli forward-backward smoothing for joint target detection and tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4473–4477, 2011.
- [11] X. Rong Li and V. P. Jilkov, "Survey of maneuvering target tracking. part I: Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
- [12] B. Hochwald and A. Nehorai, "Identifiability in array processing models with vector-sensor applications," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 83–95, 1996.
- [13] X. Zhong, A. B. Premkumar, and A. S. Madhukumar, "Particle filtering for 2-d direction of arrival tracking using an acoustic vector sensor," in *Proc. 19th European Signal Processing Conference*, 2011, pp. 245–249.
- [14] —, "Particle filtering and posterior Cramér-Rao bound for 2-D direction of arrival tracking using an acoustic vector sensor," *IEEE Sensors J.*, vol. 12, no. 2, pp. 363–377, Feb. 2012.
- [15] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.