ACOUSTIC SIMULTANEOUS LOCALIZATION AND MAPPING (A-SLAM) OF A MOVING MICROPHONE ARRAY AND ITS SURROUNDING SPEAKERS

Christine Evers, Alastair H. Moore and Patrick A. Naylor

Imperial College London
Department of Electrical and Electronic Engineering
Exhibition Road, London, SW7 2AZ, United Kingdom
c.evers@imperial.ac.uk

ABSTRACT

Acoustic scene mapping creates a representation of positions of audio sources such as talkers within the surrounding environment of a microphone array. By allowing the array to move, the acoustic scene can be explored in order to improve the map. Furthermore, the spatial diversity of the kinematic array allows for estimation of the source-sensor distance in scenarios where source directions of arrival are measured. As sound source localization is performed relative to the array position, mapping of acoustic sources requires knowledge of the absolute position of the microphone array in the room. If the array is moving, its absolute position is unknown in practice. Hence, Simultaneous Localization and Mapping (SLAM) is required in order to localize the microphone array position and map the surrounding sound sources. In realistic environments, microphone arrays receive a convolutive mixture of direct-path speech signals, noise and reflections due to reverberation. A key challenge of Acoustic SLAM (a-SLAM) is robustness against reverberant clutter measurements and missing source detections. This paper proposes a novel bearing-only a-SLAM approach using a Single-Cluster Probability Hypothesis Density filter. Results demonstrate convergence to accurate estimates of the array trajectory and source positions.

Index Terms— Acoustic Simultaneous Localization and Mapping; Acoustic scene mapping; Moving microphone arrays.

1. INTRODUCTION

Audio signals often contain information about the acoustic environment that allows for the detection of events occluded for other sensors. The topic of Acoustic Scene Mapping (ASM) is hence becoming increasingly popular for applications such as home automation, teleconferencing, search-and-rescue robots, and Human-Robot Interaction (HRI). Acoustic scene maps represent the Cartesian positions and trajectories of sound sources in the surrounding environment. In order to obtain a scene map, instantaneous Directions-of-Arrival (DoAs) of sources are estimated using Sound Source Localization (SSL). Cartesian map feature positions are estimated over time from the DoAs by utilizing source tracking approaches.

In realistic environments, dominant sound reflections due to reverberation can cause SSL approaches to estimate spurious clutter DoAs as well as missing source detections, leading to estimation errors in acoustic maps. The adverse effects of reverberation become particularly problematic at large source-sensor distances. The accuracy of acoustic maps can be improved by allowing the microphone array to move towards key features of the map. This is particularly useful for applications utilizing ad-hoc arrays, as well as robot audition where a microphone array is installed in the head of a mobile robot. This paper focuses on the application of HRI, such that we consider a microphone array installed on a moving robot platform. Nonetheless, the proposed approach is applicable to any moving microphone array. Furthermore, sound sources and signals of principal interest are talkers and their speech signals respectively.

Allowing the microphone array, or robot, to roam freely within its environment facilitates exploration of the acoustic scene and hence improved mapping accuracy. However, many robots such as the humanoid NAO by Aldebaran Robotics are not equipped with sensors for localization of the robot position. The exact location of the robot within its environment is therefore unknown. Assuming accurate estimates of the Cartesian source positions, the robot location can be estimated by triangulation. However, DoAs are provided relative to the array, such that accurate knowledge of the robot location is required to estimate the source positions.

When using mobile platforms, ASM therefore results in a "chicken-and-egg" problem [1] of simultaneously *localizing* the position of a moving microphone array conditional on the source positions, whilst *mapping* the source positions conditional on the array location. This problem is also referred to as Simultaneous Localization and Mapping (SLAM) in the robotic community and has received extensive attention for visual sensors, see e.g., [2, 3, 4].

Recent work on Acoustic SLAM (a-SLAM) using speech signals is limited to only a few examples including [5, 6]. This is perhaps due to the very signficant challenges of SSL and tracking in realistic environments. Existing a-SLAM approaches suffer under high rates of clutter, resulting in the overestimation of the number of sources. Hence, false tracks are initialised, leading to increased estimation errors and making map management very difficult.

This paper proposes an approach for a-SLAM that is robust to clutter DoAs due to reverberation. We propose a novel bearing-only a-SLAM approach using the Single Cluster Probability Hypothesis Density (PHD) (SC-PHD) filter [7]. The robot position is predicted using a Rao-Blackwellised particle filter [8]. Relative to each robot particle, a bearing-only Gaussian Mixture PHD (GM-PHD) filter [9] is used to estimate the source positions over time. The proposed approach hence performs probabilistic a) triangulation of the sources, b) anchoring of the robot, and c) association of DoAs and sources.

In the following, the signal model is introduced in Section 2. Section 3 proposes the SLAM approach. Simulation results are provided in Section 4, and conclusions are drawn in Section 5.

The research leading to these results has received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465.

2. SLAM SYSTEM MODEL

2.1. State space

Let the unknown variables at each time $t = 1, \ldots, \infty$ consist of the time-varying robot position, $\mathbf{r}_t$, the number of sources, N_t, and the random finite set [10] of sources, $\boldsymbol{S}_t \triangleq \{\mathbf{s}_{t,n}\}_{n=1}^{N_t}$, containing the single-source states, $\mathbf{s}_{t,n}$. Defining $\boldsymbol{S}_t$ relative to $\mathbf{r}_t$, the unknown states, $\boldsymbol{X}_t \triangleq \{(\mathbf{r}_t, (N_t, \boldsymbol{S}_t))\}$, can be considered as a *single-cluster process* [7] with cluster center $\mathbf{r}_t$ and cluster points $\boldsymbol{S}_t$.

The multi-source state model accounts for source initialisation, survival between time steps, and termination, such that

$$\boldsymbol{S}_t = \left[\bigcup_{n=1}^{N_{t-1}} P(\mathbf{s}_{t-1,i})\right] \cup B_t, \tag{1}$$

where B_t is a birth process, and $P(\mathbf{s}_{t-1,i}) = \mathbf{s}_{t-1,j}$ if $\mathbf{s}_{t-1,i}$ persists between $t-1$ to t, and $P(\mathbf{s}_{t-1,i}) = \emptyset$ otherwise.

The single-source states are defined as $\mathbf{s}_{t,n} \triangleq \left[\hat{x}_{t,n}, \hat{y}_{t,n}, \hat{z}_{t,n}\right]^T$ with Cartesian source position, $(\hat{x}_{t,n}, \hat{y}_{t,n}, \hat{z}_{t,n})$, relative to the robot position. In this paper, the sources are assumed stationary, i.e.,

$$\mathbf{s}_{t,n} = \mathbf{s}_{t-1,n} + \mathbf{n}_{t,n}, \qquad \mathbf{n}_{t,n} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1},\, \boldsymbol{Q}\right) \tag{2}$$

for process noise $\mathbf{n}_{t,n}$ with covariance $\boldsymbol{Q} = \sigma_q^2\, \boldsymbol{I}_4$ and $\sigma_q^2 \ll 1$.

The robot position, $\mathbf{p}_t = \begin{bmatrix} x_{t,r} & y_{t,r} & \varsigma_{t,r} \end{bmatrix}^T$, containing the Cartesian position $(x_{t,r}, y_{t,r})$ and speed, $\varsigma_{t,r}$, is modelled as

$$\mathbf{p}_t = \boldsymbol{F}_{t,r}\, \mathbf{p}_{t-1} + \mathbf{v}_{t,\mathbf{p}}, \qquad \mathbf{v}_{t,\mathbf{p}} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1},\, \boldsymbol{\Sigma}_{t,\mathbf{v}}\right) \tag{3}$$

where $\mathbf{v}_{t,\mathbf{p}}$ is the process noise with covariance $\boldsymbol{\Sigma}_{t,\mathbf{v}}$ and the height, $z_{t,r}$, is constant and known. The dynamical model, $\boldsymbol{F}_t$, is general but expressed as a constant velocity model in this paper, such that

$$\boldsymbol{F}_{t,r} = \begin{bmatrix} 1 & 0 & \Delta_T \sin\gamma_{t,r} \\ 0 & 1 & \Delta_T \cos\gamma_{t,r} \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where Δ_T is the time delay between $t-1$ and t and $\gamma_{t,r}$ is the robot orientation, modelled as a random walk,

$$\gamma_{t,r} = \gamma_{t-1,r} + v_{t,\gamma}, \qquad v_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma_{v_{t,\gamma}}^2\right). \tag{5}$$

for process noise $v_{t,\gamma}$ with variance $\sigma_{v_{t,\gamma}}^2$. The unknown robot state is therefore defined as $\mathbf{r}_t \triangleq \begin{bmatrix} \mathbf{p}_t^T & \gamma_{t,r} \end{bmatrix}^T$.

2.2. Measurement space

Similar to the states, let the measurements, $\boldsymbol{Z}_t$ consist of the M_t source DoAs, $\boldsymbol{\Omega}_t \triangleq \{\boldsymbol{\omega}_{t,m}\}_{m=1}^{M_t}$ and the robot measurements, $\mathbf{y}_t \triangleq \begin{bmatrix} z_{t,v} & z_{t,\gamma} \end{bmatrix}^T$, containing the velocity and orientation instructions, $z_{t,v}$ and $z_{t,\gamma}$ respectively, supplied to the robot by its navigation system in order to follow a particular path [11]. As demonstrated in [11], $z_{t,v}$ and $z_{t,\gamma}$ diverge from $\varsigma_{t,r}$ and $\gamma_{t,r}$ due to physical imperfections. The robot measurements are hence modelled in this paper as

$$z_{t,v} = \mathbf{h}\, \mathbf{p}_t + w_{t,v} \qquad w_{t,v} \sim \mathcal{N}\left(0,\, \sigma_{w_{t,v}}^2\right) \tag{6a}$$

$$z_{t,\gamma} = \gamma_t + w_{t,\gamma} \qquad w_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma_{w_{t,\gamma}}^2\right) \tag{6b}$$

where $w_{t,v}$ and $w_{t,\gamma}$ are the speed and orientation measurement noise with variances $\sigma_{w_{t,v}}^2$ and $\sigma_{w_{t,\gamma}}^2$ respectively, and $\mathbf{h} \triangleq \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$.

The DoA, $\boldsymbol{\omega}_{t,m} \triangleq \left[\theta_{t,m}, \phi_{t,m}\right]^T$ due to source $\mathbf{s}_{t,n}$ for $m \in 1, \ldots, M_t$, contains the inclination $\theta = \cos^{-1}\left(z/\sqrt{x^2+y^2+z^2}\right)$ and azimuth $\phi = \arctan(y/x)$ and is modelled as

$$\boldsymbol{\omega}_{t,m} = g(\mathbf{s}_{t,n}) + \mathbf{m}_{t,m}, \qquad \mathbf{m}_{t,m} \sim \mathcal{N}\left(\mathbf{0},\, \boldsymbol{R}\right) \tag{7}$$

where $g(\mathbf{s}_{t,n})$ is the Cartesian-to-spherical transformation, and where $\mathbf{m}_{t,m}$ is the measurement noise with covariance, $\boldsymbol{R} = \sigma_r^2\, \boldsymbol{I}_2$.

The multi-source measurement process models source detections, missed detections and clutter due to reverberation, such that

$$\boldsymbol{\Omega}_t = \left[\bigcup_{n=1}^{N_t} D(\mathbf{s}_{t,n})\right] \cup C_t, \tag{8}$$

where C_t is the clutter process and $D(\mathbf{s}_{t,n}) = \boldsymbol{\omega}_{t,m}$ if $\mathbf{s}_{t,n}$ is detected and $D(\mathbf{s}_{t,n}) = \emptyset$ otherwise.

3. ACOUSTIC SLAM USING SC-PHD FILTERS

In order to fully describe the statistics of the unknown process, $\boldsymbol{X}_t$, its posterior Probability Density Function (pdf) should be estimated and propagated in time. For HRI applications, the interacting source of interest needs to be tracked whilst situational awareness of the environment is maintained, hence requiring tracking of multiple sources. For multi-source tracking, however, the pdf is numerically intractable. Rather than propagating the pdf, the posterior can be approximated by its first-order moment, the PHD [10]. For the single-cluster process, $\boldsymbol{X}_t$, the SC-PHD, $\lambda(\mathbf{x}_t|\boldsymbol{Z}_{1:t})$, can be factorised into the robot PHD, $\lambda(\mathbf{r}_t|\boldsymbol{Z}_{1:t})$, and the conditional source PHD, $\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t})$, such that [12]

$$\lambda(\mathbf{x}_t|\boldsymbol{Z}_{1:t}) = \lambda(\mathbf{r}_t|\boldsymbol{Z}_{1:t})\, \lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t}), \tag{9}$$

which can be estimated using the SC-PHD filter as proposed in [12]. In this paper we propose a novel bearing-only extension of the SC-PHD filter in order to estimate $\lambda(\mathbf{x}_t|\boldsymbol{Z}_{1:t})$ using source DoAs. Whilst the robot PHD is estimated using a Rao-Blackwellised particle filter (see Section 3.2), the source PHD is obtained using a GM-PHD filter (see Section 3.1). In order to estimate the Cartesian source positions from the DoAs, we propose to induce a range estimate at source birth, which is subsequently propagated in time by prediction of the Gaussian Mixture Model (GMM) components.

3.1. Bearing-only source PHD

The predicted source PHD, $\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1})$, is given by [10],

$$\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1}) = \lambda_b(\mathbf{s}_t|\mathbf{r}_t) + \lambda_s(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1}), \tag{10}$$

where $\lambda_b(\mathbf{s}_t|\mathbf{r}_t)$ is the birth PHD of sources initialized at time t, and $\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1})$ is the prediction of surviving sources from $t-1$.

Each DoA measurement can be either due to an existing source, a newborn source, or clutter. Newborn sources are therefore initialised in this paper from the measurements [13]. An estimate of the unmeasured range is also introduced at initialisation. The range estimate is propagated in time by probabilistic triangulation as previously proposed in [9]. For each DoA in $\{\boldsymbol{\omega}_{t,m}\}_{m=1}^{M_t}$, newborn source states are generated by drawing P random variates $\mathbf{m}_{b,0}^{(p)} \sim \mathcal{N}\left(\left[\boldsymbol{\omega}_{t,m}^T, r_0\right]^T, \boldsymbol{\Sigma}_{b,0}\right)$ for $p = 1, \ldots, P$ where r_0 is the prior range with variance $\sigma_{r_0}^2$, and the covariance is given by $\boldsymbol{\Sigma}_{b,0} \triangleq$

diag $\left[\boldsymbol{R}, \sigma_{r_0}^2\right]$, where $\boldsymbol{R}$ is the measurement noise covariance in (7). The predicted birth PHD is hence given by

$$\lambda_b(\mathbf{s}_t|\mathbf{r}_t) = \sum_{\ell=1}^{L} w_{b,0}^{(\ell)} \mathcal{N}\left(\mathbf{s}_t \,|\, \mathbf{m}_{b,0}^{(\ell)}, \boldsymbol{\Sigma}_{b,0}\right), \tag{11}$$

where $L = M\,P$, the Gaussian Mixture (GM) weights are given by $w_{b,0}^{(\ell)} = \frac{N_b}{L}$, and N_b is the expected number of births per time step.

Recalling (2), $\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1})$ is expressed as a GM [14], i.e.,

$$\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1}) = \sum_{j=1}^{J_{t-1}} w_s^{(j)} \mathcal{N}\left(\mathbf{s}_t|\mathbf{m}_s^{(j)}, \boldsymbol{\Sigma}_s^{(j)}\right), \tag{12}$$

where J_{t-1} is the number of GM components at $t-1$, the weights are $w_s^{(j)} = p_s\, w_{t-1}^{(j)}$ and the predicted mean, $\mathbf{m}_s^{(j)}$ and covariance, $\boldsymbol{\Sigma}_s^{(j)}$, are obtained from the Kalman Filter (KF) prediction equations [15]. To account for the time-varying robot position, the prior mean, $\hat{\mathbf{m}}_{t-1} \triangleq \mathbf{m}_{t-1} + \mathbf{p}_{t-1} - \mathbf{p}_t$, is used in the KF.

As both the newborn and surviving source components are modelled using GMs, the predicted PHD in (10) is equivalent to

$$\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t-1}) = \sum_{j=1}^{J_{t|t-1}} w_{t|t-1}^{(j)} \mathcal{N}\left(\mathbf{s}_t \,|\, \mathbf{m}_{t|t-1}^{(j)}, \boldsymbol{\Sigma}_{t|t-1}^{(j)}\right), \tag{13}$$

where $J_{t|t-1} = L + J_{t-1}$ such that $w_{t|t-1}^{(j)}$, $\mathbf{m}_{t|t-1}^{(j)}$ and $\boldsymbol{\Sigma}_{t|t-1}^{(j)}$ contain the surviving and birth components from (12) and (13). Knowledge is inferred from the measurements by updating the birth and surviving components and accounting for the probability of missed detections, such that the updated PHD, $\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t})$, is, [14]

$$\begin{aligned}\lambda(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{Z}_{1:t}) = &\sum_{j=1}^{J_{t-1}} (1-p_d) w_s^{(j)} \mathcal{N}\left(\mathbf{s}_t \,|\, \mathbf{m}_s^{(j)}, \boldsymbol{\Sigma}_s^{(j)}\right) \\ &+ \sum_{j=1}^{J_t} w_t^{(j,m)} \mathcal{N}\left(\mathbf{s}_t \,|\, \mathbf{m}_t^{(j,m)}, \boldsymbol{\Sigma}_t^{(j)}\right),\end{aligned} \tag{14}$$

for $J_t = M\,J_{t|t-1}$, where p_d is the constant probability of detection and the updated mean and covariance, $\mathbf{m}_t^{(j)}$ and $\boldsymbol{\Sigma}_t^{(j)}$ are given by the KF correction step. The updated weights, $w_t^{(j,m)}$, are given by

$$w_t^{(j,m)} = \frac{p_d\, w_{t|t-1}^{(j)}}{\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t)} \mathcal{N}\left(\boldsymbol{\omega}_{t,m} \,|\, g\left(\mathbf{m}_{t|t-1}^{(j)}\right), \boldsymbol{S}_t^{(j)}\right), \tag{15}$$

where $\boldsymbol{S}_t^{(j)}$ is the KF innovation covariance. The term $\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t)$, evalutes the single-source likelihood of measurement, $\boldsymbol{\omega}_{t,m}$, being due to either clutter, source birth or source survival, such that

$$\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t) = \kappa_{t,m} + \sum_{j=1}^{J_{t|t-1}} p_d\, w_{t|t-1}^{(j)}, \tag{16}$$

where $\kappa_{t,m} = \lambda_\kappa\, V\, \mathcal{U}(\boldsymbol{\omega}_{t,m})$ is the clutter PHD for room volume V with clutter rate, λ_κ. Clutter measurements are assumed uniformly distributed, such that $\mathcal{U}(\boldsymbol{\omega}_{t,m})$ denotes the uniform pdf of DoA $\boldsymbol{\omega}_{t,m}$ between $[0, 2\pi]$ in azimuth and $[0, \pi]$ in inclination.

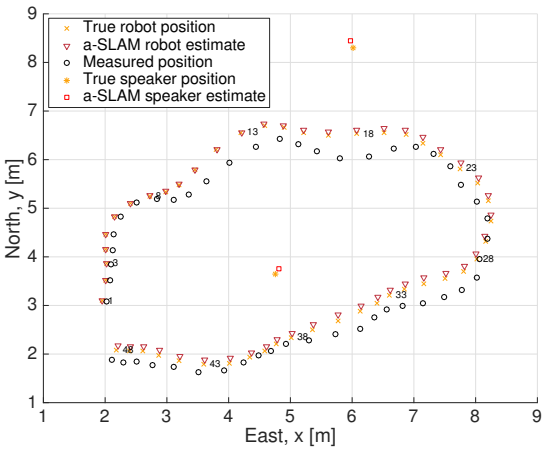

Fig. 1: Acoustic scene map showing a-SLAM robot (red triangles) and source (red squares; at $t = 50$) estimates; robot (orange crosses) and source (orange asterisks) ground truth; and robot trajectory from orientation measurements (black circles). Numbers: time stamps.

3.2. Robot PHD

Recalling the Gaussian state space model in (3), the optimal estimator of the robot position, $\mathbf{p}_t$, is given by the KF. However, the robot position is non-linearly dependent on the robot orientation, $\gamma_{t,r}$. The non-linearity can be tackled using a Rao-Blackwellised particle filter [16] to estimate $\mathbf{r}_t$. This approach samples I_t hypotheses, or particles, $\left\{\hat{\mathbf{r}}_t^{(i)}\right\}_{i=1}^{I_t}$ of the robot state. For each particle, one source GM-PHD (Section 3.1) is evaluated. As unlikely robot states result in low multi-source likelihood, resampling is used to ensure that only stochastically relevant particles propagate in time.

Assume that I_{t-1} particles, $\hat{\gamma}_{t-1}^{(i)}$ and $\hat{\mathbf{p}}_{t-1}^{(i)}$ of the robot orientation and position respectively, and their associated weights, $\alpha_{t-1}^{(i)}$, are available at time $t-1$ for $i = 1, \ldots, I_{t-1}$. At time t, P random variates, $\hat{\gamma}_t^{(i,p)}$, can be drawn from an importance function, $\pi(\gamma_{t,r}|\hat{\gamma}_{t-1}^{(i)}, z_{t,\gamma})$ [16]. For each of the resulting $I_t = P\,I_{t-1}$ particles, one KF realisation is evaluated to obtain a position particle, $\hat{\mathbf{p}}_t^{(i,p)}$, with covariance, $\boldsymbol{\Psi}_t^{(i,p)}$. The robot PHD hence is given by

$$\lambda(\mathbf{r}_t|\boldsymbol{Z}_{1:t}) = \sum_{i=1}^{I_{t-1}} \sum_{p=1}^{P} \alpha_t^{(i,p)} \delta_{\hat{\gamma}_t^{(i,p)}}(\gamma_{t,r}) \mathcal{N}(\mathbf{p}_t|\hat{\mathbf{p}}_t^{(i,p)}, \boldsymbol{\Psi}_t^{(i,p)}),$$

where $\delta_{\hat{\gamma}_t^{(i,p)}}(\gamma_{t,r})$ is the Dirac delta function of $\gamma_{t,r}$ evaluated at $\hat{\gamma}_t^{(i,p)}$, and $\alpha_t^{(i,p)}$ are the importance weights at t, given by

$$\alpha_t^{(i,p)} = \frac{\mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(i,p)})\, \hat{\alpha}_t^{(i,p)}}{\sum_{l=1}^{I_{t-1}} \sum_{m=1}^{P} \mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(l,m)})\, \hat{\alpha}_t^{(l,m)}}. \tag{17}$$

$\hat{\alpha}_t^{(i,p)}$ are the un-normalised importance weights given by

$$\hat{\alpha}_t^{(i,p)} = \alpha_{t|t-1}^{(i,p)}\, p(z_{t,\gamma}|\hat{\gamma}_t^{(i,p)})\, p(z_{t,v}|\hat{\mathbf{p}}_{t|t-1}^{(i,p)}), \tag{18}$$

where $p(z_{t,\gamma}|\hat{\gamma}_t^{(i,p)})$ and $p(z_{t,v}|\hat{\mathbf{p}}_{t|t-1}^{(i,p)})$ are the likelihood terms of the robot orientation and velocity obtained from the KF. The term

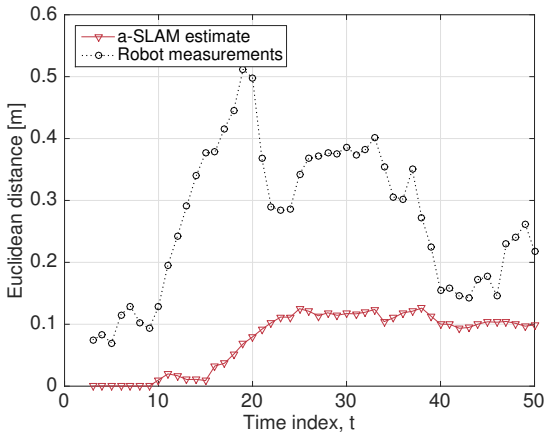

Fig. 2: Euclidean distance between ground truth and a-SLAM robot estimate compared to distance between truth and positions obtained from the orientation measurements.

$\mathcal{L}(\mathbf{\Omega}_t|\mathbf{r}_t^{(i,p)})$ in (17) is the multi-source pseudo-likelihood given by

$$\mathcal{L}(\mathbf{\Omega}_t|\mathbf{r}_t^{(i,p)}) \triangleq \prod_{m=1}^{M_t} \ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t^{(i,p)}), \tag{19}$$

for single-source likelihood terms, $\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t^{(i,p)})$, in (16).

4. RESULTS

The proposed a-SLAM approach is tested for simulated data. As SSL accuracy depends on the choice of algorithm and array design, the SLAM performance is evaluated independently by simulating DoAs with measurement error covariance $\sigma_w^2 = 4$ deg. Clutter is generated from a Poisson process with rate $\lambda_k V = 0.5$, corresponding to $0-2$ clutter DoAs per time step. The robot height and speed are 1.2 m and 0.2 m/s respectively for $T = 50$ time steps. The orientation measurement variance is $\sigma_{w_{t,\gamma}}^2 = 25$ deg. 2 sources are located at $(4.76, 3.65, 1.18)$ m and $(6.02, 8.30, 1.29)$ m in a $10.21 \times 10.33 \times 2.59$ m room.

The SC-PHD filter is run using 10 robot particles and $L = 150$ source births per measurement. Any source GM components outside of the room are reflected into the room along the estimated source direction. In order to avoid an explosion in the computational cost, systematic resampling [17] is applied to the robot particles. Point estimates of the robot state are extracted as a weighted particle mean. The source GM components are pruned to reduce the computational explosion as proposed in [14]; the truncation and merging thresholds are set to 10^{-9}m^2 and 2m^2 respectively, the maximum number of components after pruning is set to 300. Point estimates of the sources correspond to any components with weight ≥ 0.5 [14].

The estimated acoustic map is compared to the ground truth in Fig. 1. The same figure also shows the robot trajectory reconstructed from the orientation and velocity measurements using (3) without a-SLAM. The figure clearly indicates accurate estimation of the robot trajectory using a-SLAM. To verify this result, the Euclidean distance between the robot estimate and ground truth is shown in Fig. 2 and compared to the distance between the trajectory reconstructed from measurements and the ground truth. Results show an improvement of up to 40 cm distance error using a-SLAM.

Fig. 1 also shows that the source estimates converge towards the ground truth position. The azimuth estimates are plotted separately

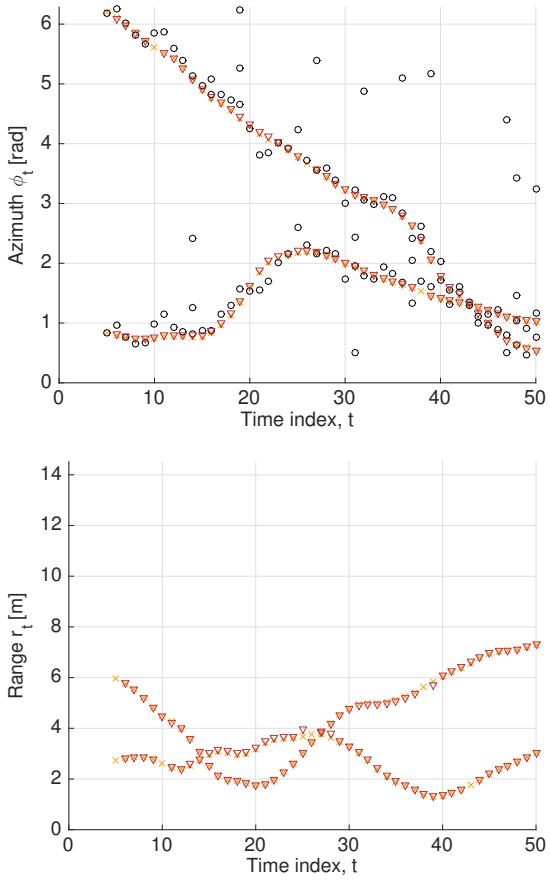

Fig. 3: a-SLAM performance of source estimates (red triangles) in (a) azimuth and (b) source-sensor range compared to ground truth (blue crosses) and DoA azimuth values (black circles).

over time against the ground truth and their DoA values in Fig. 3. Despite clutter and DoA error, no false tracks are initialized, hence avoiding an explosion in the number of map features. Missing track estimates are due to the method of state extraction, where the maximum weight of a source temporarily falls marginally below the extraction threshold. This effect occurs when the robot moves along a source DoA for a few steps, such that range uncertainty increases due to reduced spatial diversity.

Moreover, the results in Fig. 1 imply that the unmeasured range can be accurately inferred due to the spatial diversity of the moving robot platform. The a-SLAM range estimates are compared to the truth in Fig. 3, highlighting convergence of the range estimates.

5. CONCLUSION

This paper proposed an approach to the novel concept of a-SLAM. It was shown that localization of a moving microphone array and mapping of the surrounding sound sources are jointly dependent and can be simultaneously estimated using a bearing-only SC-PHD filter. Furthermore, due to spatial diversity of the moving array, the unmeasured distance between the sources and sensor can be inferred from the DoAs. Results verified accurate estimation of the source positions and range, and demonstrated accurate robot localization in terms of Euclidean distance from the ground truth.

6. REFERENCES

[1] J. Leonard and H. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Nov. 1991, pp. 1442–1447 vol.3.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.

[3] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping: Part II," *IEEE Robot. Autom. Mag.*, pp. 108–117, Sep. 2006.

[4] A. Nüchter, *3D Robotic Mapping*, ser. Springer Tracts in Advanced Robotics. Berlin-Heidelberg: Springer, 2009, vol. 52.

[5] J.-S. Hu, C.-Y. Chan, C.-K. Wang, and C.-C. Wang, "Simultaneous localization of mobile robot and multiple sound sources using microphone array," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2009, pp. 29–34.

[6] J.-S. Hu, C.-Y. Chan, C.-K. Wang, M.-T. Lee, and C.-Y. Kuo, "Simultaneous localization of a mobile robot and multiple sound sources using a microphone array," *Advanced Robotics*, vol. 25, no. 1-2, pp. 135–152, 2011.

[7] C.-S. Lee, D. E. Clark, and J. Salvi, "SLAM with dynamic target via single-cluster PHD filtering," *IEEE J. Sel. Topics Signal Process.*, no. 3, pp. 543–552, Jun. 2013.

[8] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer, 2001.

[9] C. Evers, J. Sheaffer, A. H. Moore, B. Rafaely, and P. A. Naylor, "Bearing-only acoustic tracking of moving speakers for robot audition," in *Proc. IEEE Intl. Conf. Digital Signal Processing (DSP)*, Singapore, Jul. 2015.

[10] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[11] L. George and A. Mazel, "Humanoid robot indoor navigation based on 2d bar codes: application to the nao robot," in *IEEE-RAS Intl. Conf. on Humanoid Robots (Humanoids)*, Oct 2013, pp. 329–335.

[12] C.-S. Lee, S. Nagappa, N. Palomeras, D. E. Clark, and J. Salvi, "SLAM with SC-PHD filters," *IEEE Robot. Autom. Mag.*, pp. 38–45, Jun. 2014.

[13] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1656–1668, Apr. 2012.

[14] B.-N. Vo and W.-K. Ma, "The Gaussian Mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[15] S. Gannot and A. Yeredor, "The Kalman filter," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2008, ch. 8, part B.

[16] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[17] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb 2002.