

Non-Field-Of-View Indoor Sound Source Localization based on Reflection and Diffraction

Kuya Takami^{1*}, Tomonari Furukawa¹ and Makoto Kumon²

Abstract—This paper presents a new approach which acoustically localizes a mobile target outside the **Field-of-View (FOV)**, or the **Non-Field-of-View (NFOV)** of an optical sensor, based on reflection and diffraction signals. In this approach, sound source reflection and diffraction signal **Time-of-Arrival (TOA)** and frequency dependent property construct two distinct observation likelihoods from single sound. The fusion of these likelihoods, a joint acoustic observation likelihood, is derived to estimate the sound source. The location of the **NFOV** target is finally estimated within the recursive Bayesian estimation framework. After the experimental parametric studies, the potential of the proposed approach for practical implementation has been demonstrated by the successful localization.

I. INTRODUCTION

Target localization and tracking, or mobile target estimation, in indoor environments has been a research challenge over several decades due to the existence of a variety of applications in addition to the significance and the difficulty of each application. It has variety of applications such as home security, home health care and urban search-and-rescue but its usefulness is limited by the complexity of indoor structures [1]–[3]. Indoor structures make estimation problems challenging as they can introduce large unobservable regions when an optical sensor such as a camera is deployed. This is because optical sensors’ **Field-of-View (FOV)** is determined by the **Line-of-Sight (LOS)** and range of the optical sensor, which could be small in highly constrained environments. However, if a target person is interactive, humans have capability of searching and finding the target person who is not in the **FOV** by communicating with the person and estimating the location. In such a case, target estimation is mostly performed not through vision but audition.

Recent work for **Non-Field-of-View (NFOV)** mobile target estimation has been tackled in three different ways. The first approach deploys target mounted **radio-frequency (RF)** transmitters and fixed receivers in the environment. In one arrangement, **RF** receivers form a **wireless sensor network (WSN)**, and numerical techniques are used to localize a **NFOV** target by processing information of received signals such as signal intensity [4], [5]. An improved arrangement with minimal infrastructure uses “fingerprints” [6], [7]. There

is a unique fingerprint at each location in a static environment. A target can thus be localized by feature-matching the fingerprints. Whilst this arrangement can achieve higher accuracy, the critical problem inherent in the **RF** based approach is its applicability only to near-**NFOV** target estimation [1], [8].

In the second approach, acoustic sensors are used for target estimation. Since sound signals are reflected by structures, it is possible to localize a **NFOV** target unlike the **RF** based approach provided that the sound signals contain information on the target location. The most common approach utilizes the **Time-of-Arrival (TOA)/Time-Difference-of-Arrival (TDOA)** information of acoustic signals [9]–[11]. Recently, Narang *et al.* [12] effectively used reflected **NFOV** sound source signal to assists robot navigation. The former **TOA/TDOA** based acoustic techniques, however, have not achieved true **NFOV** target estimation, and latter approach does not fully utilizes first arrival diffraction signal information. Furthermore, the majority of sound localization challenges have been focused on the direction of sound rather than its position due to complexity of sound wave propagation [13], [14].

The final approach enhances **NFOV** target estimation by including a sensor with a limited **FOV**, such as an optical sensor, by applying a numerical technique. Mauler [15] stated the **NFOV** estimation problem mathematically, and Furukawa, *et al.* [16], [17] developed a generalized numerical solution. In this technique, the event of “no detection” is converted into an observation likelihood and utilized to positively update probabilistic belief on the target. This belief is dynamically maintained by the **recursive Bayesian estimation (RBE)**. The technique, however, has been found to fail in target estimation unless the target is re-discovered within a short period after being lost. Kumon, *et al.* [18] incorporated an acoustic sensor to maintain belief with no optical detection more reliably. Nevertheless, the technique performed poorly unless the target re-entered the optical **FOV** since the acoustic sensing is only conducted in an assistive capacity. Extending Kumon’s approach, Takami, *et al.* [19] focused more towards complex indoor environment using **Interaural Level Difference (ILD)** *a priori* fixed microphone array knowledge. However, these approach require prior data collection and can not be applied to the dynamic sensor movements.

This paper presents a new acoustic approach to estimate a **NFOV** mobile target using sound wave physical properties, reflection and diffraction. In the approach, sound source reflection and diffraction signal **TOA** and frequency dependent

*Corresponding author

¹Kuya Takami, Tomonari Furukawa are with Mechanical Engineering, Virginia Polytechnic Institute and State University, USA {kuya, tomonari}@vt.edu

²Makoto Kumon is with Department of Mechanical System Engineering, Kumamoto University, Kumamoto, Japan makoto@gpo.kumamoto-u.ac.jp

property construct two distinct observation likelihoods from target sound. The fusion of these likelihoods, a joint acoustic observation likelihood, is derived to perform the target estimation. The location of the **NFOV** target is finally estimated within the recursive Bayesian estimation framework.

II. NFOV

A. Auditory Recursive Bayesian Estimation

The proposed approach is mathematically described as follows. Let the state of the robot s at time step k be $\bar{\mathbf{x}}_k^s \in \mathcal{X}^s$. Consider a target t , the state is given by $\mathbf{x}_k^t \in \mathcal{X}^t$, and a sequence of observations of the target t by the robot s from time step 1 to time step k given by ${}^s\tilde{\mathbf{z}}_{1:k}^t \equiv \{{}^s\tilde{\mathbf{z}}_\kappa^t | \forall \kappa \in \{1, \dots, k\}\}$. The **RBE** represents belief on the target in the form of a probability density function and iteratively updates the belief in time and observation. Let the belief given a sequence of observations and the robot state at time step $k-1$ be $p(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s)$. Chapman-Kolmogorov equation updates the prior belief in time, or predicts the belief at time step k , by the probabilistic motion model $p(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s)$:

$$p(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s) = \int_{\mathcal{X}^t} p(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s) p(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s) d\mathbf{x}_{k-1}^t. \quad (1)$$

Note that the motion model is $p(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t)$ if the target is not reactive to the robot. The observation update, or the correction process, is performed using the Bayes theorem. The target belief is corrected using the new observation ${}^s\tilde{\mathbf{z}}_k^t$ as

$$p(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_k^s) = \frac{q(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s)}{\int_{\mathcal{X}^t} q(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s) d\mathbf{x}_k^t}, \quad (2)$$

where $q(\cdot) = l(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t) p(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s)$ and $l(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s)$ represents the observation likelihood of \mathbf{x}_k^t given ${}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s$.

$$l(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s) = \prod_j l_j^a(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s) \quad (3)$$

where $l_j^a(\cdot)$ are the likelihoods of j th acoustic sensor.

B. Construction of Auditory NFOV Target Observation Likelihood

Figure 1 shows the overview of the approach proposed for constructing a **NFOV** target observation likelihood using an auditory sensor. The proposed approach extracts the first-arrival diffraction and reflection signals by taking the wave propagation physics into account. The proposed approach begins with obtaining a time-domain signal of a relatively impulsive sound at each microphone. In each curve, notable peaks are then extracted as candidate first-arrival diffraction and reflection signals. When each candidate signal is described in the frequency domain, the first-arrival diffraction and reflection signals can be identified since they are the

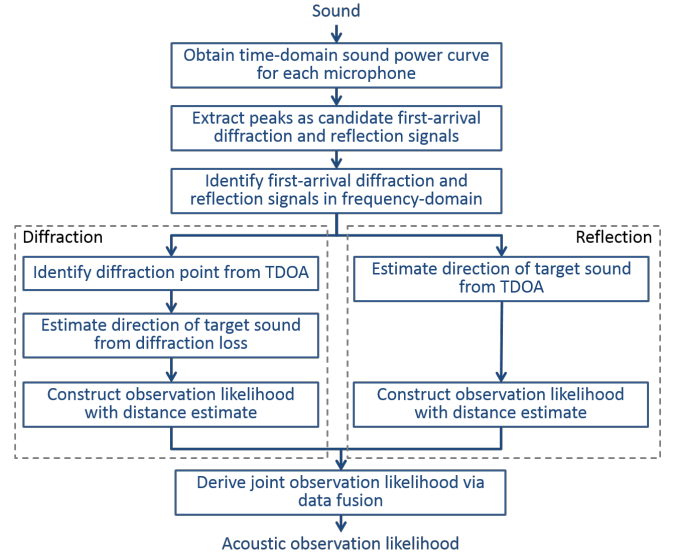


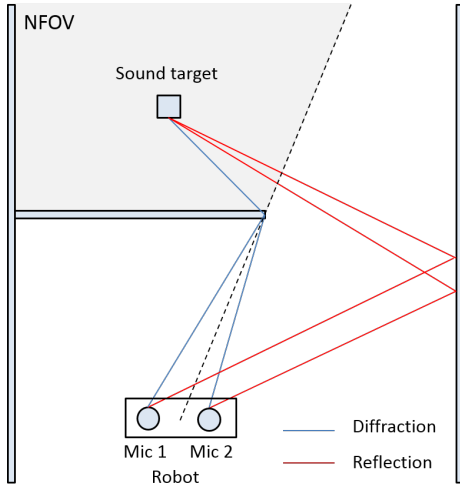
Fig. 1: Construction of auditory NFOV target likelihood

first signals that are correlated in low frequency range. The diffraction signal is then used to identify the so-called diffraction point by deriving the **TDOA** for each pair of microphones and further estimate the direction of target sound beyond the diffraction point from the loss of sound energy through diffraction, or the diffraction loss. An observation likelihood is eventually constructed by additionally estimating the distance from the sound magnitude and characteristics. The reflection signal estimates the target direction directly from the **TDOA** by mirroring and creating a virtual target. It also creates an observation likelihood with distance estimate by considering sound magnitude and characteristics and environmental properties. A joint observation likelihood is finally created by the fusion of the diffraction and reflection observation likelihoods.

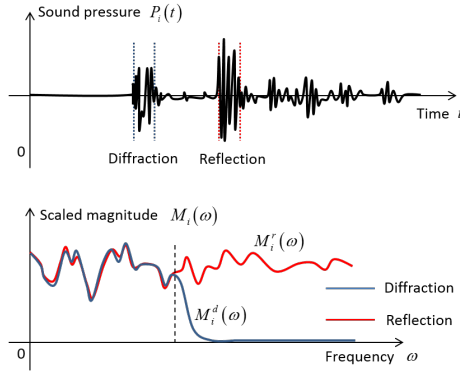
The proposed approach infers the location of the sound target using both the first-arrival diffracted and reflected sound signals. The next subsection describes the extraction of the first-arrival diffraction and reflection signals, followed by the target estimation using the diffracted and reflected signals in the subsequent two subsections. The final goal of this project is to develop a probabilistic **RBE** based framework, but the preliminary study has succeeded in the proof-of-concept in deterministic formulations. The two subsections will present the deterministic **NFOV** target estimation using diffraction and reflection sound waves. The final subsection derives the joint observation likelihood as a result of data fusion.

III. EXTRACTION OF FIRST-ARRIVAL DIFFRACTION AND REFLECTION SIGNALS

Figure 2 shows the extraction process of the diffraction and reflection signals illustratively in one of the simplest scenarios where a robot carrying two microphones receives sound emitted by a target in the **NFOV** in a two-dimensional



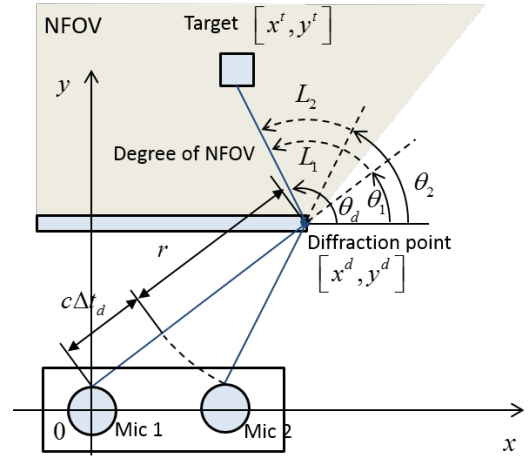
(a) Acoustic signals from NFOV target



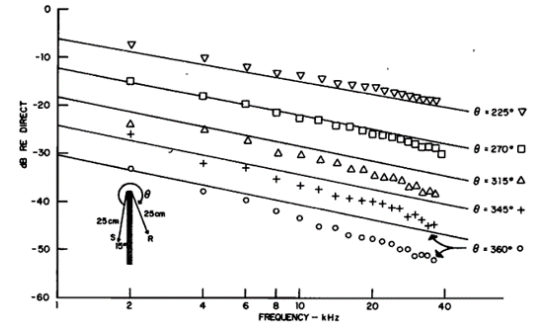
(b) Diffracted and reflected signals

Fig. 2: Auditory NFOV target observation

indoor environment with three walls (Figure 2(a)). As shown in the figure, sound waves emitted from the target reach the robot first through diffraction and second through reflection and, if the sound is relatively impulsive, the first-arrival diffraction and reflection signals can be extracted clearly. Figure 2(b) shows not only the sound pressure of sound in the time domain, $P_i(t)$, but also the magnitude of the resulting first-arrival diffraction and reflection signals in the frequency domain, $M_i^d(\omega)$ and $M_i^r(\omega)$, where $i \in \{1, 2\}$ is the index of microphone. Note that the magnitude is scaled to examine correlation. Signals are considered from the same sound source if they share the same characteristics in low frequency because low-frequency signals reflect and diffract. The proposed approach thus select the first set of signals that have the same low-frequency characteristics but are dissimilar in high frequency as the first-arrival diffraction and reflection signals of all the candidate signals. Diffraction signals have little high-frequency components whilst reflection signals see components in all frequencies. Needless to say, each of Microphones 1 and 2 constructs a different data set.



(a) Proposed approach



(b) Magnitude with different orientation angles [20]

Fig. 3: Estimation of sound direction from diffraction signals

A. Estimation of Sound Direction from Diffraction Signals

Figure 3(a) shows the notations used for target estimation from diffraction signals in the scenario introduced in the last subsection. Since the diffraction sound Microphones 1 and 2 receive is originated from the LOS location at which the sound diffracts, the proposed approach starts target estimation from diffraction signals with the selection of diffraction point from all candidates, which are corners of all structures. The measured quantity used for the selection is the **TD OA**, $\Delta t_d = t_{d2} - t_{d1}$, where t_{d1} and t_{d2} are the **TOAs** at Microphones 1 and 2 respectively. The diffraction point can be easily found from candidates as it satisfies the following equation:

$$(x^d)^2 + (y^d)^2 = (c\Delta t_d + r)^2. \quad (4)$$

where $[x^d, y^d]$ is the location of a candidate diffraction point, c is the speed of sound and r is a shorter distance between a microphone and the candidate diffraction point. With the diffraction point identified, the proposed approach further identifies the direction of the sound target from the diffraction point by analyzing the magnitudes of diffraction and reflection sounds $M_i^d(\omega)$ and $M_i^r(\omega)$. The loss of high-frequency signal components is assumed to be less with a microphone closer to **LOS** (Microphone 2 in this case) as there is no loss with a microphone on the **LOS** to the sound

target. This assumption, in fact, has been found to be valid by the work of Medwin a quarter-century ago [20] shown in Figure 3(b). The magnitude of diffraction sound drops when the “degree of **Non-Line-of-Sight (NLOS)**” represented by the orientation angle is increased. This makes the proposed approach define the diffraction loss as

$$L_i = \int [M_i^r(\omega) - M_i^d(\omega)] d\omega \geq 0, \forall i \in \{1, 2\} \quad (5)$$

and associate it with the degree of **NLOS**. The work of Medwin also shows that the diffraction loss is approximately proportional to the degree of **NLOS**. The sound direction from the diffraction point is given by

$$\theta_d = \theta_1 + \frac{\theta_2 - \theta_1}{L_1 - L_2} L_1. \quad (6)$$

B. Estimation of Sound Direction from Reflection Signals

Figure 4 shows the proposed approach for estimation of sound direction from reflection signals. Reflection makes the sound propagation and the subsequent target estimation complicated, but if the wall is smooth and yields specular reflection, the sound direction can be estimated easily by introducing a virtual target [21], which is located symmetrically to the real target relative to the wall of reflection. Let the position of the virtual target be $[\hat{x}^t, \hat{y}^t]$. The measured **TDOA** can be associated with the position of the virtual target as

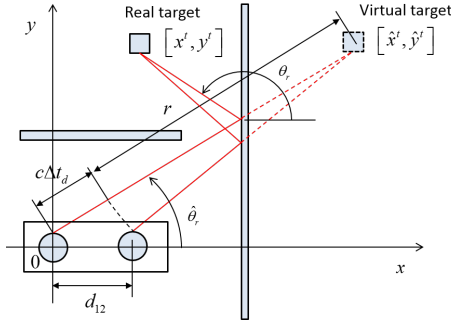


Fig. 4: Estimation of sound direction from reflection signals by proposed approach

$$\begin{cases} (\hat{x}^t)^2 + (\hat{y}^t)^2 = (r + c\Delta t_d)^2 \\ [(\hat{x}^t)^2 - d_{12}^2] + (\hat{y}^t)^2 = r^2 \end{cases}, \quad (7)$$

where d_{12} is a distance between Microphones 1 and 2. Since r is unknown unlike the diffraction, the two equations with three unknowns, \hat{x}^t , \hat{y}^t and r , introduce the relationship between \hat{x}^t and \hat{y}^t through the elimination of r . Derivation attempted as a preliminary study for this project yields the relationship as

$$(\hat{y}^t)^2 = \left(\frac{d_{12}^2}{c^2 \Delta t^2} - 1 \right) (\hat{x}^t)^2 - d_{12} \left(\frac{d_{12}^2}{c^2 \Delta t^2} - 1 \right) \hat{x}^t + \left[\frac{(d_{12}^2 + c^2 \Delta t^2)}{4c^2 \Delta t^2} - d_{12}^2 \right]. \quad (8)$$

The further mathematical manipulation shows that this equation asymptotically yields the sound direction as

$$\theta_r = \pi - \hat{\theta}_r = \lim_{r \rightarrow \infty} \tan^{-1} \frac{\hat{y}^t}{\hat{x}^t} = \cos^{-1} \frac{c\Delta t_d}{d_{12}}. \quad (9)$$

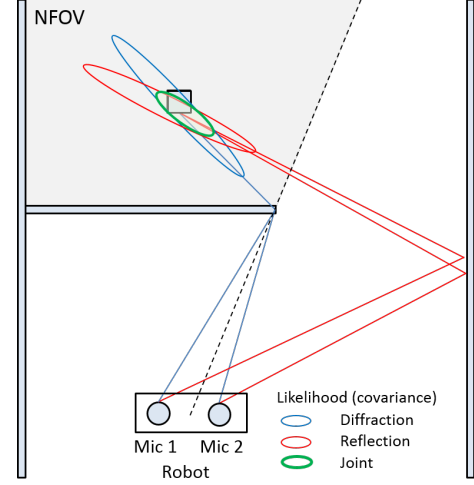


Fig. 5: Construction of joint observation likelihood through data fusion

C. Construction of Joint Observation Likelihood through Data Fusion

While the sound can be better identified in direction rather than distance, it is also possible to make an estimate on how far the sound target is. The proposed approach makes the estimate by utilizing any available information including the magnitude, sound patterns stored in a database, or sound characteristics in a knowledge base and constructs an observation likelihood for each of the diffraction and reflection signals by modeling uncertainties. For the j -th pair of microphones, the diffraction and reflection likelihoods are then combined to create an auditory joint observation likelihood via the canonical data fusion formula:

$$l_j^a(\mathbf{x}_k^t | \mathbf{z}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k) = l_j^d(\mathbf{x}_k^t | \mathbf{z}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k) l_j^r(\mathbf{x}_k^t | \mathbf{z}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k) \quad (10)$$

where $l_j^d(\cdot)$ and $l_j^r(\cdot)$ are the diffraction and reflection observation likelihood. Figure 6 illustrates the diffraction and reflection observation likelihoods as well as the joint observation likelihood where the observation likelihood is represented by an ellipsoid indicating a probability distribution with a covariance. The diffraction and reflection likelihoods are shown to have high eccentricity due to more accuracy in direction than in distance. Since the difference of the diffraction and reflection likelihoods in orientation may not be significant, the resulting auditory joint likelihood could also given by an ellipsoid with high eccentricity, but the proposed approach, utilizing the diffraction and reflection physics of sound, could estimate the location of the sound target.

IV. NUMERICAL ANALYSIS

The preliminary experimental study has demonstrated validity of the proposed approach through simulation.

TABLE I: Dimensions and other parameters for the experiments

Parameter	Value
x_s	
x_t	
L_w	
$[x_d y_d]$	

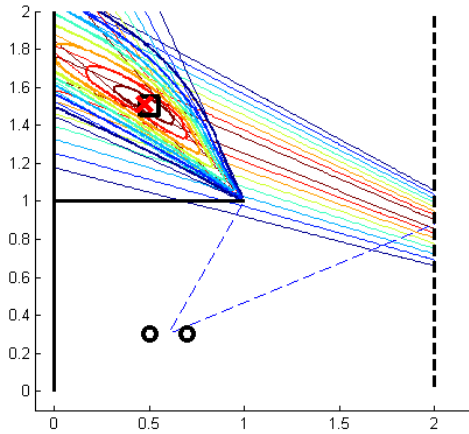


Fig. 6: Target estimation

V. CONCLUSIONS

SLAM - map of the environment and test in the scale of actual indoor environment

REFERENCES

- [1] N. B. Priyantha, H. Balakrishnan, E. D. Demaine, and S. Teller, "Mobile-assisted localization in wireless sensor networks," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1. IEEE, 2005, pp. 172–183.
- [2] H. M. Khoury and V. R. Kamat, "Evaluation of position tracking technologies for user localization in indoor construction environments," *Automation in Construction*, vol. 18, no. 4, pp. 444–457, 2009.
- [3] S. Argenti, P. Danes, P. Soueres *et al.*, "A survey on sound source localization in robotics: from binaural to array processing methods," 2014.
- [4] H. Dai, Z.-M. Zhu, and X.-F. Gu, "Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network," *Journal of Network and Computer Applications*, 2012.
- [5] I. Guvenc and C.-C. Chong, "A survey on toa based wireless localization and nlos mitigation techniques," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 3, pp. 107–124, 2009.
- [6] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2. Ieee, 2000, pp. 775–784.
- [7] A. M. Ladd, K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki, "On the feasibility of using wireless ethernet for indoor localization," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 555–559, 2004.

- [8] C. K. Seow and S. Y. Tan, "Non-line-of-sight localization in multipath environments," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 5, pp. 647–660, 2008.
- [9] 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.
- [10] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 826–836, 2003.
- [11] L. C. Mak and T. Furukawa, "Non-line-of-sight localization of a controlled sound source," in *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*. IEEE, 2009, pp. 475–480.
- [12] G. Narang, K. Nakamura, and K. Nakadai, "Auditory-aware navigation for mobile robots based on reflection-robust sound source localization and visual slam," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4021–4026.
- [13] Y. Tamai, Y. Sasaki, S. Kagami, and H. Mizoguchi, "Three ring microphone array for 3d sound localization and separation for mobile robot audition," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 2005, pp. 4172–4177.
- [14] P. Svaizer, A. Brutti, and M. Omologo, "Environment aware estimation of the orientation of acoustic sources using a line array," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 1024–1028.
- [15] R. Mauler, *Recent Developments in Cooperative Control and Optimization*. Kluwer Academic Publishers, Norwell, MA, 2003, ch. Objective Functions for Bayesian Control-Theoretic Sensor Management, II: MHC-Like Approximation, pp. 273–316.
- [16] T. Furukawa, F. Bourgault, B. Lavis, and H. F. Durrant-Whyte, "Recursive bayesian search-and-tracking using coordinated uavs for lost targets," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 2521–2526.
- [17] T. Furukawa, L. C. Mak, H. Durrant-Whyte, and R. Madhavan, "Autonomous bayesian search and tracking, and its experimental validation," *Advanced Robotics*, vol. 26, no. 5-6, pp. 461–485, 2012.
- [18] M. Kumon, D. Kimoto, K. Takami, and T. Furukawa, "Bayesian non-field-of-view target estimation incorporating an acoustic sensor," in *Proceedings of 2013 IEEE/RSJ International Conference of Intelligent Robots and Systems*, November 1-3, 2013 2013.
- [19] K. Takami, T. Furukawa, M. Kumon, and G. Dissanayake, "Non-field-of-view acoustic target estimation in complex indoor environment," in *Proc. of the Int. Conf. on Field and Service Robotics*, 2015.
- [20] H. Medwin, "Shadowing by finite noise barriers," *The Journal of the Acoustical Society of America*, vol. 69, no. 4, pp. 1060–1064, 1981.
- [21] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

ACKNOWLEDGMENT

Hanxing Liu