Noname manuscript No.
(will be inserted by the editor)

Estimation of a Nonvisible Field-of-View Mobile Target Incoorperating Optical and Acoustic Sensors

Kuya Takami · Tomonari Furukawa · Makoto Kumon · Daisuke Kimoto · Gamini Dissanayake

Received: date / Accepted: date

Abstract This paper presents a Nonvisible Field-Of-View (NFOV) target estimation approach that incorporates optical and acoustic sensors. An optical sensor can accurately localize a target in its Field-Of-View (FOV) whereas the acoustic sensor could estimate the target location over a much larger space but only with limited accuracy. A recursive Bayesian estimation framework where observations of the optical and acoustic sensors are probabilistically treated and fused is proposed in this paper. A technique to construct the observation likelihood when two microphones are used as the acoustic sensor is also described. The proposed technique derives and stores the interaural level difference of observations from the two microphones for different target positions in advance and constructs the likelihood through correlation. A parametric study of the proposed acoustic sensing technique in a controlled test environment, and experiments with a NFOV target in an actual indoor environment are presented to demonstrate the capability of the proposed technique.

Keywords nonvisible field-of-view target estimation · recursive Bayesian estimation · interaural level difference · acoustic localization

K. Takami and T. Furukawa
Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA, USA
E-mail: {kuya, furukawa}@vt.edu

M. Kumon and D. Kimoto
Department of Mechanical System Engineering, Kumamoto University, Kumamoto, Japan
E-mail: kumon@gpo.kumamoto-u.ac.jp

G. Dissanayake
Center for Autonomous Systems, University of Technology, Sydney, NSW, Australia
E-mail: gamini.dissanayake@uts.edu.au

1 Introduction

Mobile target estimation, or more strictly target localization and tracking, 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. Indoor mobile target estimation, which is the focus of this paper, sees its significance in applications such as home security, home health care and urban search-and-rescue whilst its difficulty is primarily governed by the complexity of indoor structures [1,13]. Complex indoor structures make estimation problems challenging as they could introduce large unobservable regions when an optical sensor such as a camera is deployed [18]. This is because optical sensor's Field-Of-View (FOV) is determined by the range of the optical sensor and the Line-Of-Sight (LOS) from the optical sensor, which could be small in highly constrained environments. This gives rise to need for Nonvisible FOV (NFOV) mobile target estimation.

The coverage of a larger area can be achieved in four different ways. First is the most straightforward approach, which is to expand the optical FOV. It can be done by either implementing a single optical sensor with a larger FOV, such as an omnidirectional camera [16, 24], or multiple optical sensors each covering a different FOV, such as a spherical camera [11]. While such implementations are suitable for some applications, these approaches may still need to cope with a large unobservable area if environments are cluttered and complex since the FOV is determined by the LOS from the optical sensor(s). This approach, thus, is not essentially an adequate solution for the NFOV mobile target estimation.

The second approach deploys radio-frequency (RF) transmitters and receivers where an RF transmitter is

mounted on a target of concern. In one arrangement, RF receivers form a wireless sensor network (WSN), and numerical techniques are used localize a NFOV target by processing information of received signals such as signal intensity [2,5,25,35,19,9,10]. For example, Wang *et al.* [34] developed and applied a Bayesian filter using the chirp-spread-spectrum ranging. The achievement of sub-meter accuracy has been reported, but the accuracy significantly depends on the settings of the networking infrastructure including the location and the number of wireless sensor nodes [12,8]. A more accurate arrangement with minimal infrastructure uses "fingerprints" [1, 17]. In a static environment, there is a unique fingerprint at each location. A target can thus be localized by feature-matching the fingerprints. Whilst this arrangement could achieve high accuracy, the critical problem inherent in the RF based approach is its applicability to only targets with an RF transmitter [4,27,31,13].

In the third, acoustic sensors are used for target estimation. If the target is cooperative, it is possible to communicate with the target and estimate its location via sound. The approach most commonly utilizes the Time-Of-Arrival (TOA) information of acoustic signals. Chan *et al.* [3] and Riba and Urruela [28] localized a target by positioning three or more LOS acoustic sensors; furthermore, Nakadai *et al.* [23] and Sasaki *et. al.* [29] demonstrated LOS mobile robot localization using microphone array. Mak and Furukawa [21] considered the diffraction characteristics of low-frequency sound and estimated target location. The former is an acoustic version of WSN, but the reflective nature of sound waves confines acoustic sensor locations to LOS, which is not possible in many occasions. The latter enables NFOV target estimation, but the time of sound generation, which is often unknown, must be informed beforehand. The acoustic approach can be advantageously used for any target that can make sound unlike the RF based approach. The existing acoustic techniques, however, have not tackled the truly NFOV target estimation to the best of our knowledge. 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 [33,32,20]. The last approach enhances the NFOV target estimation with a sensor having a limited FOV such as an optical sensor by using a numerical technique. Mauler [22] stated the NFOV estimation problem mathematically, and Furukawa, et al. [6,7] developed the so-called Bayesian Search and Tracking (SAT) technique as 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 that is dynamically maintained by the recursive Bayesian estimation (RBE). While search with no detection is made possible in addition to tracking with detection, search has been found to fail unless the target is re-discovered within a short period after being lost. The unreliability of the belief significantly grows when there is no detection. It is, therefore, preferable to use any information on the target if available rather than purely relying on the information contained in no detection events.

This paper presents a NFOV target estimation approach, which incorporates an acoustic sensor in addition to an optical sensor. Whilst the optical sensor accurately localizes a target when the target is in the FOV, the acoustic sensor is used to estimate the target location even if the target is not in the optical FOV. The estimation is performed within the RBE framework where observations of the optical and acoustic sensors are each converted into an observation likelihood and making it possible to compute a joint observation likelihood via sensor fusion. Although the acoustic observation likelihood could be multi-modal with high uncertainty, the target belief updated in the past with sharply unimodal optical likelihoods when the target was in the optical FOV effectively acts as strong prior knowledge and allows accurate NFOV target estimation. A technique to construct an observation likelihood using an acoustic sensor composed of two microphones is also proposed. The proposed technique derives the interaural level difference (ILD) of observations from the two microphones for different target positions and stores the ILDs as fingerprints, or acoustic cues, *a priori*. Given a new acoustic observation, an acoustic observation likelihood is computed by quantifying the correlation of the ILD of the new observation to the stored ILDs. While the acoustic cues must be created in advance, the technique achieves the truly NFOV target estimation.

The paper is organized as follows. The following section reviews the conventional RBE that uses an optical sensor as well as the grid-based method. Section 3 presents the proposed target estimation approach incorporating an acoustic sensor. Section 4 demonstrates the efficacy of the proposed target estimation through experimental analysis, and conclusions are summarized in the final section.

2 Optical Recursive Bayesian Estimation

2.1 Target Motion Model and Optical Sensor Model

Consider a target t of concern, the motion of which is given by

$$\mathbf{x}_{k+1}^t = \mathbf{f}^t\left(\mathbf{x}_k^t, \mathbf{u}_k^t, \mathbf{w}_k^t\right) \tag{1}$$

where $\mathbf{x}_k^t \in \mathcal{X}^t$ is the state of the target at time step k, $\mathbf{u}_k^t \in \mathcal{U}^t$ is the set of control inputs of the target, and $\mathbf{w}_k^t \in \mathcal{W}^t$ is the "system noise" of the target. For simplicity, the target state describes the two-dimensional position.

In order for the formulation of the NFOV target estimation problem, this moving target is observed by a sensor platform s. To focus on the estimation of a mobile target, let the sensor platform be stationary and its global state be accurately known as $\tilde{\mathbf{x}}^s \in \mathcal{X}^s$. Note that $\tilde{()}$ is an instance of $()$. The sensor platform carries an optical sensor to observe the target. The FOV, or more precisely the "observable region", of the optical sensor s_c can be expressed with the probability of detecting the target $P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right)$ as

$$^{s_c}\mathcal{X}_o^t = \left\{\mathbf{x}_k^t | 0 < P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right) \leq 1\right\}.$$

Accordingly, the target position observed from the optical sensor, $^{s_c}\mathbf{z}_k^t \in \mathcal{X}^t$, is given by

$$^{s_c}\mathbf{z}_k^t = \begin{cases} ^{s_c}\mathbf{h}^t\left(\mathbf{x}_k^t, \tilde{\mathbf{x}}^s, {}^{s_c}\mathbf{v}_k^t\right) & \mathbf{x}_k^t \in {}^{s_c}\mathcal{X}_o^t \\ \emptyset & \mathbf{x}_k^t \notin {}^{s_c}\mathcal{X}_o^t \end{cases} \tag{2}$$

where $^{s_c}\mathbf{v}_k^t$ represents the observation noise, and $\emptyset$ represents an "empty element", indicating that the optical observation contains no information on the target or that the target is unobservable.

2.2 Recursive Bayesian Estimation

The RBE updates belief on a dynamical system, given by a probability density, in both time and observation. Let a sequence of observations of a moving target t by a stationary sensor platform s from time step 1 to time step k be $^s\tilde{\mathbf{z}}_{1:k}^t \equiv \{^s\tilde{\mathbf{z}}_\kappa^t | \forall\kappa \in \{1, ..., k\}\}$. Given the initial belief $p\left(\mathbf{x}_0^t\right)$, the sensor platform state $\tilde{\mathbf{x}}^s$ and a sequence of observations $^s\tilde{\mathbf{z}}_{1:k}^t$, the belief on the target at any time step k, $p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right)$, can be estimated recursively through the two stage equations; update and prediction.

In the prediction process, the target belief $p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right)$ is updated from that in the previous time step $p\left(\mathbf{x}_{k-1}^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right)$ by Chapman-Kolmogorov equation as

$$\begin{aligned} &p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right) \\ &= \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t\right) p\left(\mathbf{x}_{k-1}^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right) d\mathbf{x}_{k-1}^t, \end{aligned} \tag{3}$$

where $p\left(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t\right)$ is a probabilistic form of the motion model (1). Note that $p\left(\mathbf{x}_{k-1}^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right) = p\left(\mathbf{x}_0^t\right)$ when $k = 1$. The correction process, on the other hand, updates the belief using information available in the observations. The target belief $p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right)$ is corrected from the corresponding state estimated with the observations up to the previous time step $p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right)$ and a new observation $^s\tilde{\mathbf{z}}_k^t$ as

$$p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right) = \frac{q\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right)}{\int_{\mathcal{X}^t} q\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right) d\mathbf{x}_{k-1}^t}, \tag{4}$$

where

$$q(\cdot) = l\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}^s\right) p\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right), \tag{5}$$

and $l\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}^s\right)$ represents the likelihood of $\mathbf{x}_k^t$ given $^s\tilde{\mathbf{z}}_k^t$ and $\tilde{\mathbf{x}}^s$, which is a probabilistic version of the sensor model; i.e., Equation (2) if the sensor is optical. It is to be noted that the likelihood does not need to be a probability density since the normalization in Equation (4) makes the output belief be a probability density regardless of the formulation of the likelihood.

2.3 Modeling of Optical Observation Likelihood

The optical observation likelihood is modeled by first defining the "detectable region". Due to the existence of uncertainty, the observation of a no-empty element does not necessarily indicate that the target has been reliably detected. The detectable region of the optical sensor s_c that describes the region within which the sensor confidently detects the target is thus defined as:

$$^{s_c}\mathcal{X}_d^t = \left\{\mathbf{x}_k^t | \epsilon^t < P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right) \leq 1\right\} \subset {}^{s_c}\mathcal{X}_o^t,$$

where ϵ^t is a positive threshold value which judges the detection of the target. Given the observation $^s\tilde{\mathbf{z}}_k^t$, the optical observation likelihood is resultantly stated as

$$l^c\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \begin{cases} p\left(^s\tilde{\mathbf{z}}_k^t|\mathbf{x}_k^t, \tilde{\mathbf{x}}^s\right) & \exists^s\tilde{\mathbf{z}}_k^t \in {}^{s_c}\mathcal{X}_d^t \\ 1 - P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right) & \nexists^s\tilde{\mathbf{z}}_k^t \in {}^{s_c}\mathcal{X}_d^t, \end{cases} \tag{6}$$

where, depending on whether there exists a target within the detectable region, the upper and lower formulas return likelihoods with detection and no-detection events, respectively.

Figure 1 illustrates the configuration of the optical observation likelihood when a sensor platform is in a one-dimensional target space. When a target is not detected without having it in the detectable region, the likelihood tells where the target is unlikely to be and is thus represented as a heavily non-Gaussian distribution. When the target is detected, the likelihood becomes near-Gaussian with its peak located at the observed location. The closer the target to the sensor platform, the more accurate the estimation.

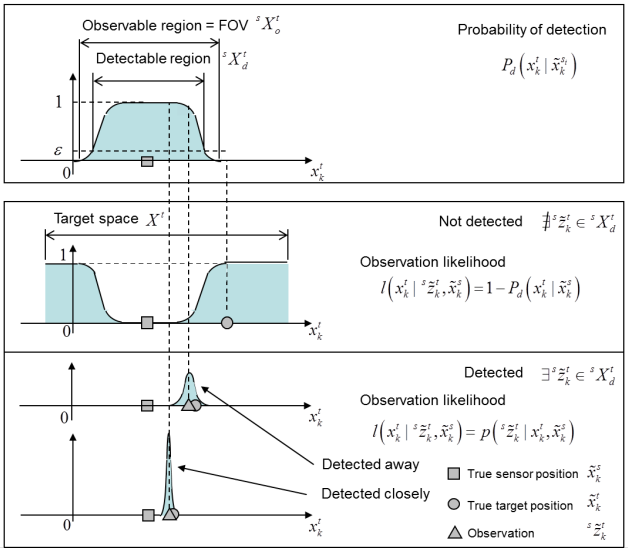


Fig. 1 Optical observation likelihood

2.4 Grid-based Method

Handling the heavily non-Gaussian no-detection likelihood necessitates the grid-based method for RBE. As the grid-based method represents the belief space in terms of regularly aligned grid cells, let the cell of concern be positioned at lth and mth partitions in x and y directions. At the grid cell $[l, m]$, the prediction and the correction are processed independently. The prediction requires the numerical evaluation of the Chapman-Kolmogorov equation in Equation (3) at each grid cell. Given the belief $p^{l,m}_{\mathbf{x}^t_{k-1}}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right)$ at time step k as well as the motion model $p^{l,m}_{\mathbf{x}^t_k|\mathbf{x}^t_{k-1}}$ constructed in the matrix form as the convolution kernel, the target belief at the grid cell $[l, m]$ can be predicted as

$$p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right) = \sum_{\alpha=0}^{I^t_x}\sum_{\beta=0}^{I^t_y} p^{\alpha,\beta}_{\mathbf{x}^t_k|\mathbf{x}^t_{k-1}} p^{l-\alpha,m-\beta}_{\mathbf{x}^t_{k-1}}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right). \quad (7)$$

where $\otimes$ indicates the convolution of the last belief with the motion model.

The correction requires the computation of Equation (4) at each grid cell. Given the predicted belief $p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right)$ and the observation likelihood $l^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_k\right)$, the target belief at the grid cell $[l, m]$ can be corrected as

$$p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k}\right) = \frac{q^{l,m}_{\mathbf{x}^t_k}(\cdot)}{\Delta x_r \Delta y_r \sum_\alpha \sum_\beta q^{\alpha,\beta}_{\mathbf{x}^t_k}(\cdot)}, \quad (8)$$

where

$$q^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k}\right) = l^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_k\right) p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right). \quad (9)$$

and $[\Delta x_r, \Delta y_r]$ is the size of the grid.

Whilst the generalized optical observation likelihood allows belief update and maintenance regardless of whether the target has been detected, the RBE with the optical observation likelihood does not update and maintain the belief effectively. Figure 2 illustratively depicts this problem where the FOV and the NFOV are given by the light blue and the white colors respectively. When the configuration of the target space is constrained complicatedly, the FOV becomes significantly limited compared to the target space. This makes the belief dominantly updated by the observation likelihood with no detection. If no detection continuously takes place, the belief keeps spread out with predictions and becomes highly uncertain and unreliable. The next section will describe the proposed target estimation incorporating an acoustic sensor to solve this problem.

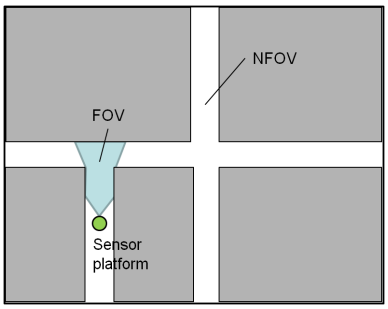


Fig. 2 Optical observation likelihood

3 Target Estimation Incorporating Optical and Acoustic Sensors

3.1 Acoustic Sensor Model and Observation Likelihood

The acoustic sensor incorporated in the proposed approach can observe a target on the non-line-of-sight (NLOS) or even in the NFOV though accuracy is limited due to the complex behavior of sound signals including reflection, refraction and diffraction. Because of its broad range, the observable region of the acoustic sensor could be considered unlimited when compared to that of the optical sensor. The acoustic sensor model s_a can be therefore constructed without defining an observable region unlike the optical sensor model:

$${}^{s_a}\mathbf{z}^t_k = {}^{s_a}\mathbf{h}^t\left(\mathbf{x}^t_k, \tilde{\mathbf{x}}^s_k, {}^{s_a}\mathbf{v}^t_k\right), \quad (10)$$

which is probabilistically equivalent to the likelihood given by

$$l^a\left(\mathbf{x}^t_k | {}^{s_a}\tilde{\mathbf{z}}^t_k, \tilde{\mathbf{x}}^s_k\right) = p\left({}^{s_a}\tilde{\mathbf{z}}^t_k | \mathbf{x}^t_k, \tilde{\mathbf{x}}^s_k\right). \quad (11)$$

Figure 3 illustrates the observation likelihood of the acoustic sensor in comparison to that of the optical sensor in Figure 1. The observation likelihood could be heavily non-Gaussian with multiple peaks if the target is on the NLOS though it is highly likely that one of the peaks is found near the location of the target as shown in the figure. The likelihood with a target on the LOS could still be multi-modal if there are adjacent structures that create reflective, refractive and/or diffractive sound signals, but it captures the target location more confidently with a sharper peak. The likelihood with a target on the LOS without adjacent structures will be a sharp near-Gaussian distribution since the direct sound dominates the observation.

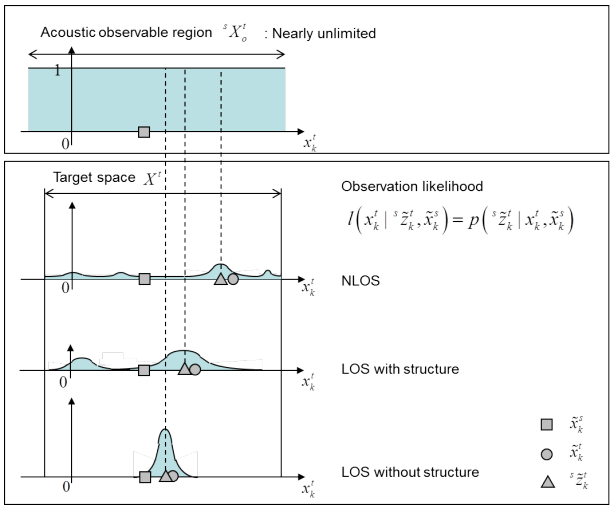


Fig. 3 Acoustic observation likelihood

3.2 RBE Using Joint Optical/Acoustic Observation Likelihood

Given the observation likelihood of the optical sensor $l^c\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)$ and that of the acoustic sensor $l^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)$, the proposed approach derives a joint likelihood by multiplying the two likelihoods:

$$l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right) = l^c\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right) l^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right) \tag{12}$$

in accordance to the canonical data fusion formula. The joint optical/acoustic likelihood may not be a probability density similarly to the optical and acoustic observation likelihoods.

Figure 4 illustratively shows the resulting joint optical/acoustic observation likelihood when the target is in the NFOV. The possible locations of the target are narrowed down since the optical likelihood with no detection clears out likelihood in the detectable region and dropped some peak(s) as shown in the figure. However, the joint likelihood could still remain heavily non-Gaussian with multiple peaks and thus may not solely make a good estimation about where the target is.

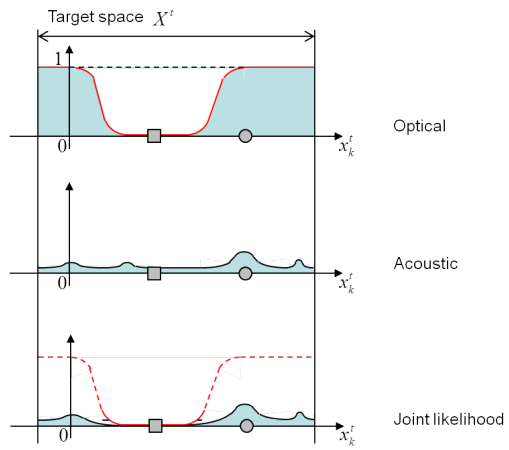


Fig. 4 Joint optical/acoustic observation likelihood

Figure 5 shows the further ability of the proposed approach in reliable target estimation. The proposed approach performs RBE by using the joint likelihood in correction (substituting Equation (12) into Equation (5)) within the standard RBE framework. Because sharpest and most Gaussian is the optical observation likelihood with detection, the prior belief is most determined by the last optical observation and remains a sharp Gaussian distribution. The posterior belief with the joint observation likelihood inherits this characteristics since the joint likelihood most likely captures the target location with a peak and magnifies the confidence of the prior belief with the joint likelihood.

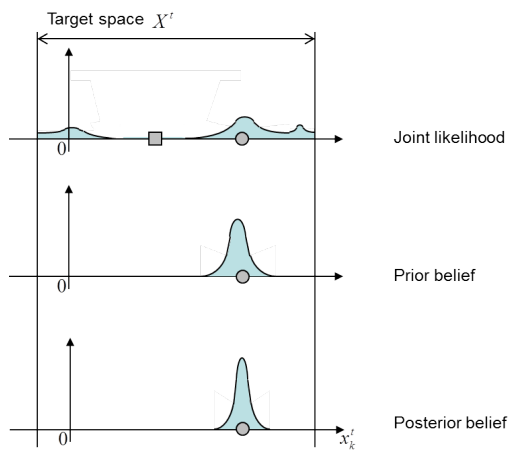


Fig. 5 RBE incorporating an acoustic sensor

3.3 Modeling of Acoustic Observation Likelihood

The technique proposed in this paper to model an acoustic observation likelihood uses two microphones as an acoustic sensor and constructs acoustic cues of the target in the environment of concern *a priori*. This is because of the potential of the proposed technique for NLOS target estimation through the preliminary investigations of the authors [15,14,26] and inability of the aforementioned existing techniques. Figure 6 shows a schematic diagram of the proposed technique to model an acoustic observation likelihood. We assume that the target emits sound with white noise, and we indeed use it to create the acoustic observation likelihood. A white noise sound emitted at a specific position by the target for a certain time period is first recorded by two microphones. After applying fast Fourier transform (FFT), the difference between the frequency-domain amplitude responses, known as the ILD, is then derived and further sampled to form an ILD vector within the frequency range of interest. The ILD vector is created with various target positions and each saved as an acoustic cue. The acoustic observation likelihood modeling essentially corresponds to creating the set of ILD vectors. When a target emitted a white noise sound, the ILD vector of the sound observation is compared to all the acoustic cues. The degree of similarity is then used to develop a correlation map indicating where the target is likely to be. The correlation map is the acoustic likelihood of the particular sound observation.

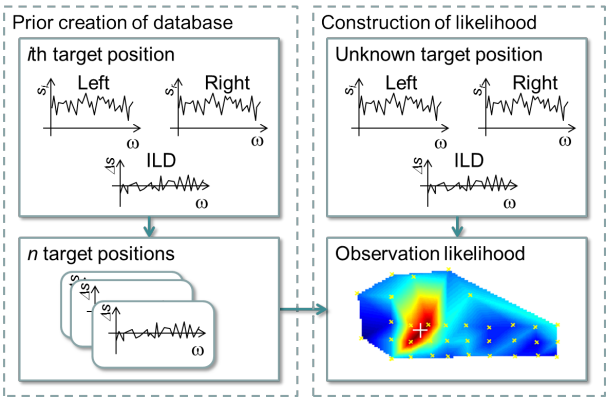


Fig. 6 Acoustic sensor

Mathematically, let the frequency-domain sound level of the target at the ith position $(\mathbf{x}_k^t)_i$, which is observed by the left and right microphones, be $s_l\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)$ and $s_r\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)$ where ω is a frequency of sound. The ILD for the ith position $(\mathbf{x}_k^t)_i$, $\Delta S\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)$, is then given by

$$\Delta S\left(\omega|\left(\mathbf{x}_k^t\right)_i\right) = 20\log\left|s_l\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)\right| - 20\log\left|s_r\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)\right|. \quad (13)$$

If the ILD is sampled at N frequencies $\mathbf{\Omega} = [\omega_1, \ldots, \omega_N]^\top$, the ILD vector can be described as

$$\mathbf{S}\left(\mathbf{\Omega}|\left(\mathbf{x}_k^t\right)_i\right) = \left[a_1\Delta S\left(\omega_1|\left(\mathbf{x}_k^t\right)_i\right), \ldots, a_N\Delta S\left(\omega_N|\left(\mathbf{x}_k^t\right)_i\right)\right]^\top, \quad (14)$$

where

$$a_j = \left\langle \min\left\{\left|s_l\left(\omega_j|\left(\mathbf{x}_k^t\right)_i\right)\right|, \left|s_r\left(\omega_j|\left(\mathbf{x}_k^t\right)_i\right)\right|\right\} - \epsilon\right\rangle. \quad (15)$$

In the equation, $\langle\cdot\rangle$ is Macaulay brackets, and $\min\{\cdot,\cdot\}$ returns the smaller value of the two entities. When the ILD vector is created with a set of frequencies $\tilde{\mathbf{\Omega}}$ at n known target positions, i.e., $(\tilde{\mathbf{x}}_k^t)_i, \forall i \in \{1, ..., n\}$, the acoustic cues to be prepared in advance and used to create the acoustic observation likelihood become $\mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right), \forall i \in \{1, ..., n\}$.

Given the ILD vector $\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)$ with observation ${}^s\tilde{\mathbf{z}}_k^t$ at unknown target position $\mathbf{x}_k^t$, the proposed technique quantifies the degree of correlation of the ith ILD vector to that of the unknown target position as

$$X\left(\left(\tilde{\mathbf{x}}_k^t\right)_i|{}^s\tilde{\mathbf{z}}_k^t\right) = \frac{1}{2}\left\{\frac{\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)^\top \mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)}{\left|\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)\right|\left|\mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|} - 1\right\}. \quad (16)$$

where $0 \leq X(\cdot) \leq 1$. The acoustic observation likelihood with the unknown target position $\mathbf{x}_k^t$ can be finally calculated as

$$l^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \sum_{i=1}^{n}\mu_i\left(\boldsymbol{\xi}_k^t\right)X\left(\left(\tilde{\mathbf{x}}_k^t\right)_i|{}^s\tilde{\mathbf{z}}_k^t\right), \quad (17)$$

where $\mu_i\left(\boldsymbol{\xi}_k^t\right)$ is a basis function of natural coordinates $\boldsymbol{\xi}_k^t$, which are transformed from $\mathbf{x}_k^t$. Based on the characteristics of the likelihood, the proposed technique uses the T-spline basis function such that $\mu_i\left(\boldsymbol{\xi}_k^t\right)$ in a T-mesh, for the two-dimensional parameter space case $\boldsymbol{\xi}_k^t = [\xi_k^t, \eta_k^t]^\top$, can be represented as

$$\mu_i\left(\boldsymbol{\xi}_k^t\right) = \mu_i^1\left(\xi_k^t\right)\mu_i^2\left(\eta_k^t\right) \quad (18)$$

where $\mu_i^{(\cdot)}$ is a cubic B-spline basis function. Further detailed formulations are found in [30]. It is to be noted here that other basis functions are also possible while the proposed technique uses T-spline basis functions.

4 NUMERICAL AND EXPERIMENTAL ANALYSIS

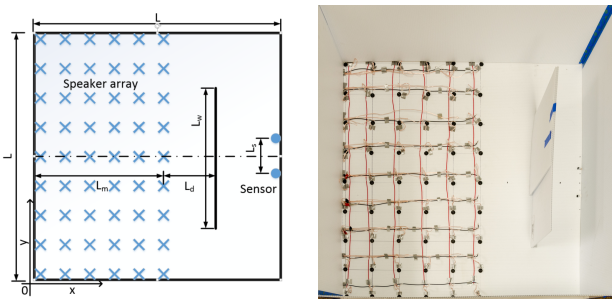
The efficacy of the proposed approach was examined experimentally in two steps. The first step was aimed at studying the capability and limitation of the proposed acoustic sensing technique by parametrically changing the complexity of the environment where the experimental system with a speaker array and a movable/replaceable wall was developed specifically for this study. After verifying the feasibility of the acoustic sensing for NLOS target localization, the applicability of the proposed approach to the estimation of a NFOV target in a complex practical environment was investigated. The investigation looked into the performance of both the joint optical/acoustic observation likelihood and the RBE with the joint likelihood.

4.1 Acoustic Observation of NLOS Target

Figure 7(a) shows the design of the experimental system that changes the complexity of the environment for the evaluation of the proposed acoustic sensing technique. An acoustic sensor consisting of two microphones is fixed next to an outer wall and faces open space where a speaker array and movable/replaceable wall(s) are placed. The complexity of the environment can be changed by varying the two parameters of the movable/replaceable wall: the distance of the wall to the edge of speaker array L_d and the length of the wall L_w. The longer the distance and/or the larger the length, the more complex the environment since sound from speakers result in more reflections.

Shown in the figure as blue crosses are speaker locations. A microcontroller controls speakers so that each speaker sequentially emits white noise sound for a programmed period. A set of ILDs for a wall setting can be thus collected automatically. Once the ILDs are collected, the ability of the proposed acoustic sensing technique is evaluated by emitting sound from a speaker at some location within the area of the speaker array and identifying the location in the form of observation likelihood. The location is not where one of the speakers of the speaker array is located to demonstrate the ability of the proposed technique in identifying the target at an arbitrary position.

Figure 7(b) shows the developed experimental system whereas the dimensions and other parameters used in the experiments are listed in Table 1. Sound was sampled and represented at 8,192 frequency bins within the audible range to capture its behavior accurately. 54 speakers were aligned to cover the open space. The distance and the length of the wall were varied to introduce both lightly NLOS and heavily NLOS environments. The case of two walls ($n_w = 2$) were tested in addition to the single wall case to make the environment more complex. The distance of only the wall closer to the acoustic sensor was varied.



(a) Schematic design (b) Developed system

Fig. 7 Experimental system for investigating influence of environmental complexity

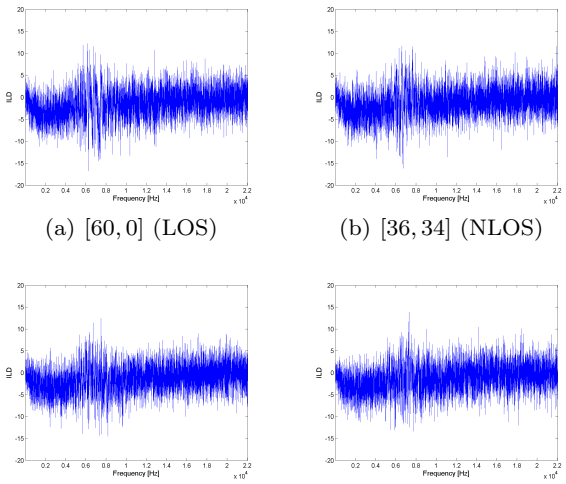
Figure 8 shows the six ILDs each observed with a target at one of the 54 positions when $[L_d, L_w, n_w] = [50, 40, 1]$. Two positions are on the LOS, and four are on the NLOS. It is first seen that the configuration of the ILD varies depending on the target position. The configuration is different even when the target is on the NLOS. This indicates that the ILD contains information on the target position no matter whether the target is on the LOS.

Figure 9 shows four resulting acoustic observation likelihoods each with a different condition. The former two cases are with a single wall having different distances and the same length whereas the latter two cases are with two walls having different lengths and the same distance. The result first indicates that the target location is well estimated when the distance is short (Figure 9(a)) or when the length is small (Figure 9(c)). The target is closer to LOS in these conditions since sound reaches the acoustic sensor with a small number of reflections. The identification of the target location in the remaining two cases (Figures 9(b) and 9(d)) is hard due to a number of sound reflections. The identification with two walls (Figures 9(c) and 9(d)) is seen to be harder than that with a single wall (Figures 9(a) and 9(b)) for the same reason. While the acoustic observation likelihood is heavily multi-modal in these cases, the target location is still captured by the highest peak (Figures 9(a)-9(c)) or at least by one of the peaks (Figure 9(d)). This demonstrates the ability of the proposed acoustic sensing technique for identifying the location of the NFOV target though with limited accuracy.

Figure 10(a) and 10(b) show the mean error of the acoustic observation likelihood when the distance and the length were varied for single and double wall cases. The mean error is a distance of the nearest peak of the acoustic observation likelihood to the true target location. The result of the mean error shows that the proposed technique could locate the target within 2 cm error in most (11) of the 12 cases for single wall case. The estimation is particularly good when the wall length was small. The mean error with two walls, meanwhile, is over 20 cm in four of nine cases. This indicates that the proposed acoustic sensing technique is not sufficient enough

Table 1 Dimensions and other parameters in the experiments

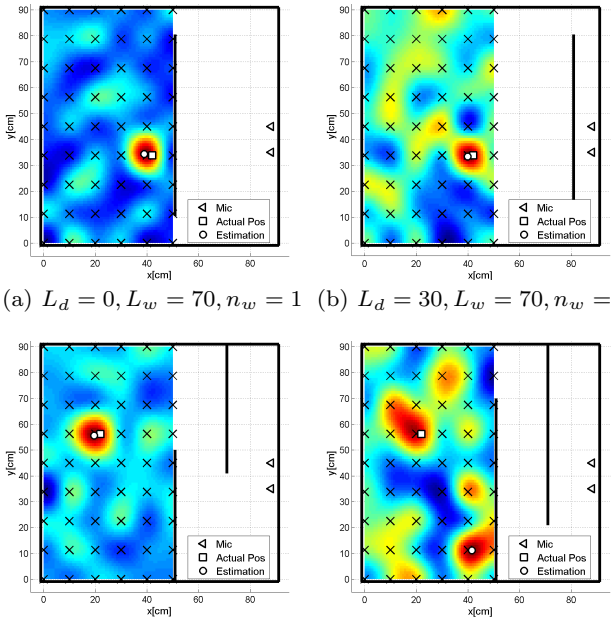
Parameter	Value	Parameter	Value
$\tilde{\mathbf{x}}^t$	$[42, 34]$ cm	L	90 cm
ω_1	0 Hz	L_m	50 cm
ω_N	22 kHz	L_s	10 cm
N	8,192	L_d	$\{0, 10, 20, 30\}$ cm
ϵ	0.01	L_w	$\{50, 60, 70\}$ cm
n	54	n_w	$\{1, 2\}$



(a) $[60, 0]$ (LOS) (b) $[36, 34]$ (NLOS)

(c) $[24, 56]$ (NLOS) (d) $[12, 90]$ (LOS)

Fig. 8 ILDs at six of the 54 positions

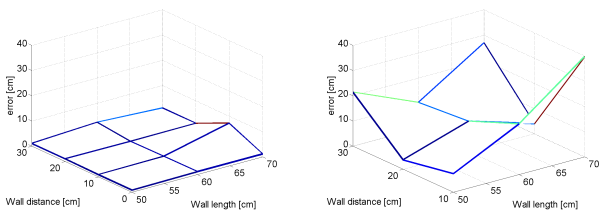


(a) $L_d = 0, L_w = 70, n_w = 1$ (b) $L_d = 30, L_w = 70, n_w = 1$

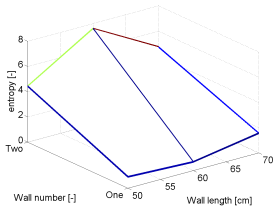
(c) $L_d = 20, L_w = 50, n_w = 2$ (d) $L_d = 20, L_w = 70, n_w = 2$

Fig. 9 Acoustic observation likelihoods for different environmental complexity

for NLOS localization. Figure 10(c) shows the variation of the differential entropy with respect to the number of walls and the wall length. It is seen that the addition of a wall dominantly increases uncertainty. While the proposed acoustic sensing technique is effective enough in relatively simple NLOS environments, enhancement is necessary when target estimation in more complex environments is pursued.



(a) Single wall mean error (b) Double wall mean error



(c) Diferential entropy for single and double wall

Fig. 10 Mean error and differential entropy of the acoustic observation likelihood with a single and double wall

4.2 NFOV Target Estimation in Complex Practical Environments

This subsection investigates the applicability of the proposed approach to NFOV target estimation in complex practical environments by first enhancing the acoustic sensing with the optical sensing and then executing the RBE with the joint optical/acoustic observation likelihood. Figure 11 shows the map of the indoor environment used to demonstrate the practical applicability of the proposed approach as well as the details of the demonstration. A sensor platform with a camera and two microphones for optical and acoustic observations was located in a corridor and faced with the open space the target could move around. The environment is complex with the existence of walls and structures, so that the FOV of the camera is significantly limited compared to the target space. Shown in the figure as yellow crosses are the target locations at which sound was emitted to collect ILDs. After the collection, the target was then moved along the lines indicated in the figure and emitted sound. The observation and estimation was examined at the four positions marked by red circles.

Figure 12 shows the target and the sensor platform. The sensor platform is with a camera and two microphones as aforementioned whereas the target is a wireless speaker. The same speaker was used to construct the acoustic observation likelihood. White noise was emitted from the speaker, and the parameters used to construct the acoustic observation likelihood are listed in Table 2.

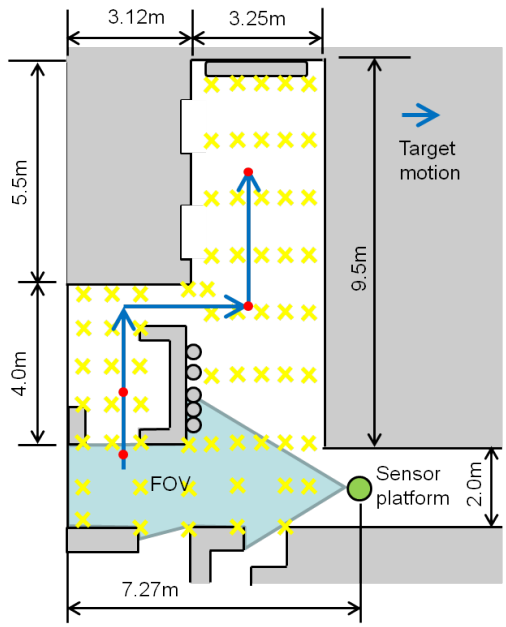


Fig. 11 Map of the test environment and demonstration

4.2.1 Joint Optical/Acoustic Observation of NFOV Target

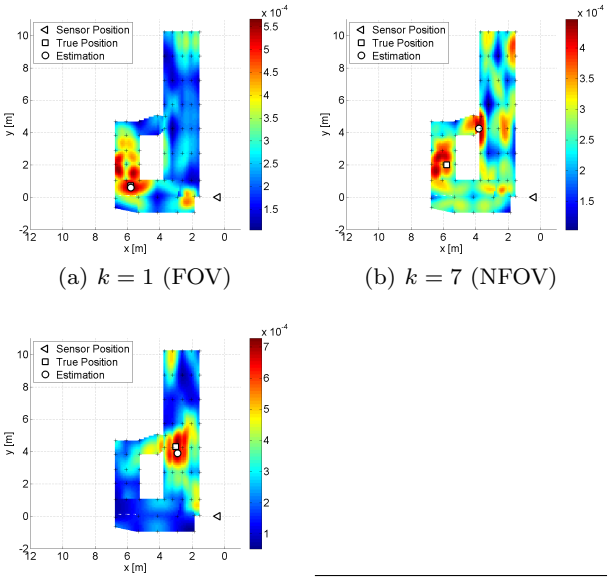
Figure 13 shows the acoustic observation likelihoods when the target moved and emitted sound at the four marked target positions, which are at the 1st, 7th, 35th and 51th steps. The

Fig. 12 Target and sensor platform

Table 2 Parameters for acoustic observation likelihood

Parameter	Value
ω_1	0 [Hz]
ω_N	386 [Hz]
N	100
ϵ	0.01
n	65

target is in the FOV only at the 1st step. The acoustic observation likelihood is seen to be multi-modal due to sound reflection even when the target is within the FOV and thus on the LOS. The proposed acoustic sensing technique has, however, been able to accurately identify the true target position at one of the peaks and successfully detect it except for the 7th step. Failure in the 7th step is a result of the limitation of acoustic sensing shown and concluded in the last subsection, but it is to be importantly noted that the target position is captured near the second highest peak.

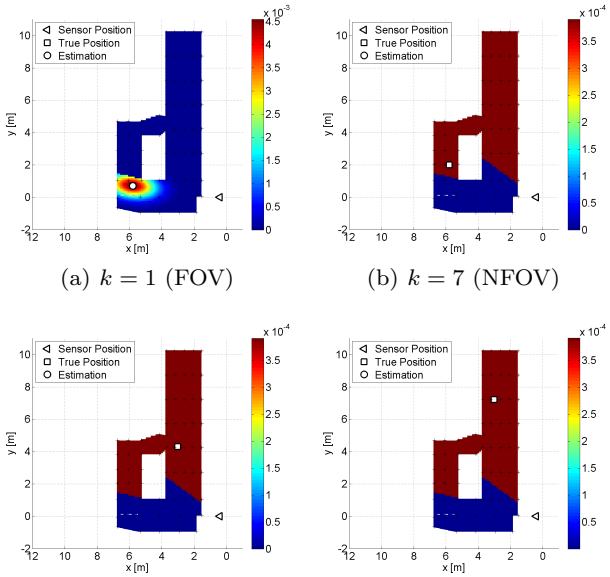


(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 13 Acoustic observation likelihoods

The effectiveness of the proposed acoustic sensing technique is further understood comparatively by seeing optical observation likelihoods in Figure 14. The optical sensor can identify the target accurately when it is within the FOV. The observation likelihood with the target outside the FOV can however provide no positive information on the target. Finally, Figure 15 shows the joint optical/acoustic observation likelihoods. It is seen that the joint observation likelihoods most narrow down the possible target locations by detecting the target dominantly with the optical observation likelihood when the target is within the FOV and with the acoustic observation likelihood when the target is outside the FOV. The wrong computation at the 7th step, however, remains and necessitates RBE for target estimation.



(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

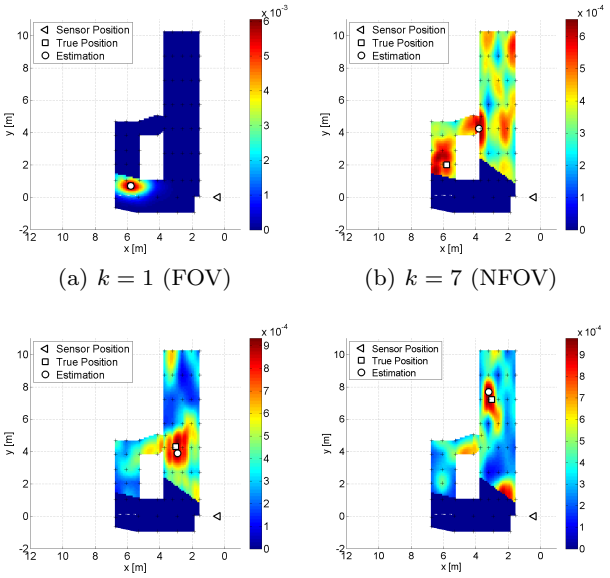
(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 14 Optical observation likelihoods

4.2.2 RBE with Joint Optical/Acoustic Observation Likelihoods

Having understood the limitation of the target detection with observations only in the last section, the effectiveness of the proposed RBE with the joint optical/acoustic observation likelihoods was investigated using the same test data. Without knowing the target motion well, the target motion model was given by a random walk model assuming that the target is a human who could move to any direction with equal probability.

Figure 16 shows the target belief estimated via RBE with the joint optical/acoustic observation likelihoods. The result shows that the target position is well estimated with all the time steps including the 7th step where the joint observation likelihood did not detect the target with the highest peak. Because the target was initially observed by the optical sensor, strong and accurate prior belief was constructed. Since the prior knowledge is updated by prediction with the random walk model and correction with the joint observation likeli-

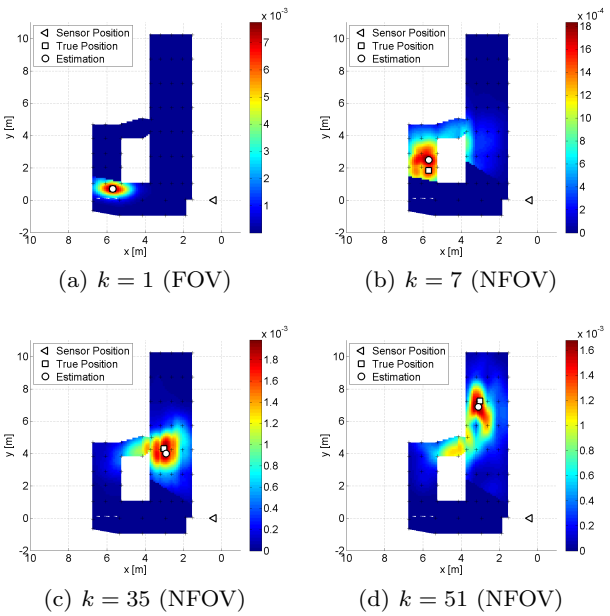


(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 15 Joint optical/acoustic observation likelihoods

hood, the proposed approach can eliminate wrong detections and estimate the target position near the true position.

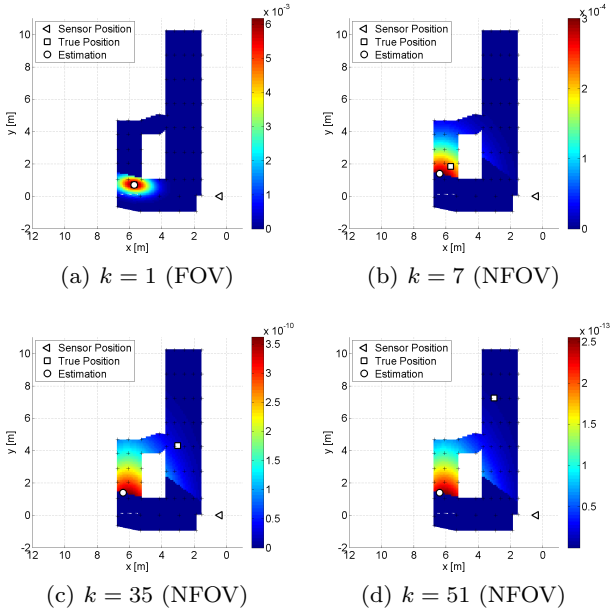


(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 16 Proposed optical/acoustic target estimation

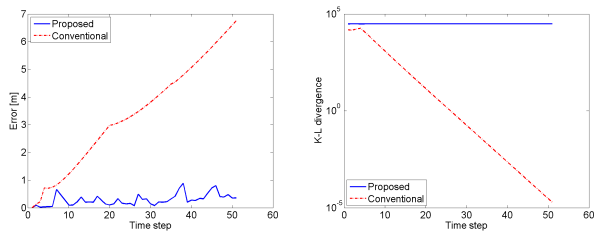
Figure 17 shows the target belief estimated conventionally with the optical observation likelihoods only to comparatively verify the effectiveness of the proposed approach. The result with the optical observation likelihoods is seen to estimate the target position wrongly when the target is in the NFOV. Because an inaccurate random-walk motion model is used, the target is estimated continuously at the location where it was lost. Finally, Figure 18 shows the results quantitatively evaluating the performance of the proposed approach. Figure 18(a) shows the transition of the mean error of the estimated target position from the true position whereas the transition of the Kullback-Leibler (KL) divergence is exhibited in Figure 18(b). The error transition indicates that the proposed approach maintains error within 1 m even when the target has not been lost from the FOV for some time whilst the conventional RBE with optical observation likelihoods increases the error with high gradient. The KL divergence transition also shows this behavior, indicating that the proposed approach maintains target information with the use of the acoustic sensor.



(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 17 Conventional optical target estimation



(a) Error vs time step (b) KL divergence vs time step

Fig. 18 Quantitative analysis

5 Conclusions

This paper has presented a NFOV target estimation approach which incorporates optical and acoustic sensors. The proposed approach performs RBE with joint optical/acoustic ob-

servation likelihoods. Although the acoustic observation likelihood could be multi-modal with high uncertainty, the target belief updated in the past with sharply unimodal optical likelihoods effectively acts as strong prior knowledge and enables accurate NFOV target estimation. A technique to construct an observation likelihood using an acoustic sensor composed of two microphones has also been proposed. The acoustic observation likelihood is created by correlating the ILD of the new acoustic observation with the ILDs collected in advance.

The first experiment studied the capability and limitation of the proposed acoustic sensing technique by parametrically changing the complexity of the environment. It has been found that the proposed technique can identify the location of a NLOS target but with limited accuracy particularly when the complexity of the environment is severe. The applicability of the proposed RBE approach with joint optical/acoustic observation likelihoods to the estimation of a NFOV target in a complex practical environment was the investigated as the second experiment. The result shows that the target position is well estimated at all the time steps even when the joint observation likelihood does not identify the target location well due to the use of prediction with a motion model and strong prior belief constructed with the past optical observation.

The paper has demonstrated the new concept for NFOV target estimation, and many challenges are still open for future study. One of the improvements to the approach is the enhancement of acoustic sensing by incorporating the interaural time difference (ITD) and the interaural phase difference (IPD) as well as the use of non-white noise sound so that the approach could be used for various applications. Other ongoing work includes the implementations of the proposed approach to the mobile sensor platform and to the infrastructure. The proposed approach will can be used for various practical applications including home security, home heath care and urban search-and-rescue by the extension.

References

1. 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*, volume 2, pages 775–784. Ieee, 2000. 1, 2
2. M. Bertinato, G. Ortolan, F. Maran, R. Marcon, A. Marcassa, F. Zanella, M. Zambotto, L. Schenato, and A. Cenedese. Rf localization and tracking of mobile nodes in wireless sensors networks: Architectures, algorithms and experiments. 2008. 2
3. Y. Chan, W. Tsui, H. So, and P. Ching. Time-of-arrival based localization under nlos conditions. *Vehicular Technology, IEEE Transactions on*, 55(1):17–24, 2006. 2
4. P. Chen. A non-line-of-sight error mitigation algorithm in location estimation. In *Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE*, pages 316–320. IEEE, 1999. 2
5. H. Dai, Z. Zhu, and X. Gu. Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network. *Journal of Network and Computer Applications*, 2012. 2
6. T. Furukawa, F. Bourgault, B. Lavis, and H. 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*, pages 2521–2526. IEEE, 2006. 2
7. T. Furukawa, L. C. Mak, H. DurrantWhyte, and R. Madhavan. Autonomous bayesian search and tracking, and its experimental validation. *Advanced Robotics*, 26(5-6):461–485, 2012. 2
8. P. Gao, W. Shi, W. Zhou, H. Li, and X. Wang. A location predicting method for indoor mobile target localization in wireless sensor networks. *International Journal of Distributed Sensor Networks*, 2013, 2013. 2
9. S. Gezici. A survey on wireless position estimation. *Wireless Personal Communications*, 44(3):263–282, 2008. 2
10. I. Guvenc and C. Chong. A survey on toa based wireless localization and nlos mitigation techniques. *Communications Surveys & Tutorials, IEEE*, 11(3):107–124, 2009. 2
11. N. D. Jankovic and M. D. Naish. Developing a modular active spherical vision system. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1234–1239. IEEE, 2005. 1
12. J. Jung and H. Myung. Indoor localization using particle filter and map-based nlos ranging model. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 5185–5190, 2011. 2
13. H. M. Khoury and V. R. Kamat. Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, 18(4):444–457, 2009. 1, 2
14. D. Kimoto and M. Kumon. On sound direction estimation by binaural auditory robots with pinnae. 35th Meeting of Special Interest Group on AI Challenges, 2011. 6
15. D. Kimoto and M. Kumon. Optimization of the ear canal position for sound localization using interaural level difference. 36th Meeting of Special Interest Group on AI Challenges, 2012. 6
16. M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia. People tracking and following with mobile robot using an omnidirectional camera and a laser. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 557–562. IEEE, 2006. 1
17. 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*, 20(3):555–559, 2004. 2
18. L. Ledwich and S. Williams. Reduced sift features for image retrieval and indoor localisation. Citeseer. 1
19. H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, 2007. 2
20. Y. Lu and M. Cooke. Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1793–1805, 2010. 2
21. 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*, pages 475–480. IEEE, 2009. 2
22. R. Mauler. *Recent Developments in Cooperative Control and Optimizatio*, chapter Objective Functions for Bayesian Control-Theoretic Sensor Management, II: MHC-Like Approximation, pages 273–316. Kluwer Academic Publishers, Norwell, MA, 2003. 2
23. K. Nakadai, H. Nakajima, M. Murase, S. Kaijiri, K. Yamada, T. Nakamura, Y. Hasegawa, H. G. Okuno, and

H. Tsujino. Robust tracking of multiple sound sources by spatial integration of room and robot microphone arrays. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 4, pages IV–IV. IEEE, 2006. 2

24. S. K. Nayar. Catadioptric omnidirectional camera. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 482–488. IEEE, 1997. 1
25. L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004. 2
26. Y. Noda and M. Kumon. Sound source direction estimation in the median plane by two active pinnae. 13th SICE System Integration Division Annual Conference, 2012. 6
27. E. A. Prigge. *A positioning system with no line-of-sight restrictions for cluttered environments*. PhD thesis, Stanford University, 2004. 2
28. J. Riba and A. Urruela. A non-line-of-sight mitigation technique based on ml-detection. In *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, volume 2, pages 153–156, May 2005. 2
29. Y. Sasaki, S. Kagami, and H. Mizoguchi. Online short-term multiple sound source mapping for a mobile robot by robust motion triangulation. *Advanced Robotics*, 23(1-2):145–164, 2009. 2
30. T. W. Sederberg, J. Zheng, A. Bakenov, and A. Nasri. T-splines and t-nurccs. In *ACM transactions on graphics (TOG)*, volume 22, pages 477–484. ACM, 2003. 6
31. C. K. Seow and S. Y. Tan. Non-line-of-sight localization in multipath environments. *Mobile Computing, IEEE Transactions on*, 7(5):647–660, 2008. 2
32. 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*, pages 1024–1028. IEEE, 2012. 2
33. J. Valin, F. Michaud, J. Rouat, and D. Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1228–1233. IEEE, 2003. 2
34. J. Wang, Q. Gao, Y. Yu, H. Wang, and M. Jin. Toward robust indoor localization based on bayesian filter using chirp-spread-spectrum ranging. *Industrial Electronics, IEEE Transactions on*, 59(3):1622–1629, 2012. 2
35. D. Zhang, Y. Yang, D. Cheng, S. Liu, and L. M. Ni. Cocktail: an rf-based hybrid approach for indoor localization. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010. 2