2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Bayesian Non-Field-of-View Target Estimation Incorporating An Acoustic Sensor

Makoto Kumon, Daisuke Kimoto, Kuya Takami and Tomonari Furukawa

Abstract— **This paper presents non-field-of-view (NFOV) target estimation incorporating an acoustic sensor, which consists of two microphones. The proposed approach derives the interaural level difference (ILD) of observations from the two microphones for different target positions and stores the ILDs as database *a priori*. Given a new acoustic observation on a target, an acoustic observation likelihood is created by calculating the correlation of the ILD of the new observation to the stored ILDs. A joint observation likelihood is then developed by fusing the optical and acoustic observation likelihoods, and the recursive Bayesian estimation updates and maintains belief on the target using the joint observation likelihood. The proposed approach detects a target positively using an acoustic sensor even if it is outside the field of view of the optical sensor and localizes the target accurately by estimating it within the RBE. The efficacy of the proposed approach was first validated by experimental studies. Further numerical demonstrations then show the applicability of the proposed approach to the NFOV target estimation.**

I. Introduction

Optical sensors have been primary sensors for target tracking due to their high-precision localization capability. The optical tracking is however effective only if the target stays within the field-of-view (FOV), which is determined by the range of the optical sensor and the line-of-sight (LOS) from the optical sensor. The FOV is often significantly smaller than the area the target explores. Once it has been lost from the FOV and has not been immediately rediscovered, the target moves over the vast area and may not be any longer rediscovered for tracking.

Historically, the capability of tracking was improved by enhancing the accuracy of target estimation where the enhancement was achieved by probabilistically handling sensor observations and estimating the belief on the target. In order to handle different levels of uncertainty, various recursive Bayesian estimation (RBE) techniques have been developed and utilized accordingly. The most successful techniques in the early stage were Kalman filter (KF) and its variants [15]. The KF-based techniques represent the observation in terms of Gaussian probability density and update the target belief also in terms of the Gaussian probability density at the acquisition of every observation. While the KF-based techniques enhanced the accuracy by prioritizing the significance of observations based on their density, the enhancement is possible only when the target is within the FOV.

The problem of the KF-based techniques was tackled by two approaches. One was the replacement of the KF-based techniques by RBE techniques that can handle non-Gaussian distributions such as the sequential Monte Carlo (SMC) methods also known as the particle filter methods [12], the sequential Quasi-Monte Carlo (SQMC) methods [8] and their variants. The chance of losing the target becomes less since the target location is estimated more accurately. As the handling of non-Gaussian distributions does not solely enhance the accuracy when the target is lost, another approach attempted was the improved sensor modeling where negative observations of no-detection are treated as positive information. Mauler [13] described the observation likelihood without a detection event in terms of the negation of the probability of detection.

Furukawa, et al [2], [4], [5], [6] generalized the unified observation likelihood by defining the observable and detectable regions and enabled RBE using the grid-based method [1] due to the heavily non-Gaussian representation of the no-detection observation likelihood. The grid-based RBE further demonstrated successful real time estimation using a graphics processing unit (GPU). Despite the successful implementation, tracking with the unified observation likelihood, however, has been found to frequently fail in practice. This is because no-detection is not very positive information without positively estimating where the target is likely to be. Since the FOV of the optical sensor is significantly limited compared to the target space, the detection of a target is not easy when the target has been lost for some time.

This paper presents non-field-of-view (NFOV) target estimation incorporating an acoustic sensor, which consists of two microphones. If the target is cooperative, it is possible to communicate with the target and estimate its location via sound. The proposed approach derives the interaural level difference (ILD) of observations from the two microphones for different target positions and stores the ILDs as database *a priori*. Given a new acoustic observation, an acoustic observation likelihood is created by calculating the correlation of the ILD of the new observation to the stored ILDs. A joint observation likelihood is then developed by fusing the optical and acoustic observation likelihoods, and the RBE updates and maintains belief on the target using the joint observation

This work was primarily supported by the US Army International Technology Center (ITC) - Pacific with additional funding from the Office of Naval Research Global (ONRG), Deputy Assistant Secretary of the Army for Defense Exports and Cooperation (DASA-DEC), The U.S. Army Tank Automotive Research, Development and Engineering Center (TARDEC), Asian Office of Aerospace Research Development (AOARD). Partial support by National Science Foundation (NSF) is also acknowledged.

This work was supported by National Science Foundation and Japan Society for Promotion of Science

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

K. Takami and T. Furukawa are with Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA tomonari@vt.edu

likelihood. The proposed approach detects a target positively using an acoustic sensor even if it is outside the FOV of the optical sensor and localizes the target accurately by estimating it within the RBE.

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 III presents the proposed target estimation approach incorporating an acoustic sensor. Section IV demonstrates the efficacy of the proposed target estimation through experimental analysis, and conclusions are summarized in the final section.

II. Optical Recursive Bayesian Estimation

A. Target Motion Model and Optical Sensor Model

Consider a target t of concern, the motion of which is discretely 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 when it is not within the observable region.

B. 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 is updated in time. 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

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

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 in observation. 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

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

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.

C. 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 unified 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 represented as a heavily non-Gaussian distribution. When the target is detected, the likelihood becomes Gaussian or near-Gaussian with its peak located at the observed location by the detection. The closer the target to the sensor platform, the more accurate the estimation.

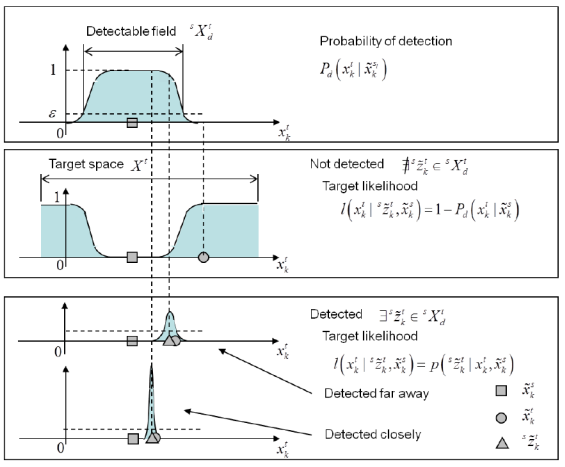


Fig. 1. Optical observation likelihood

D. Grid-based Method

Handling the heavily non-Gaussian no-detection likelihood necessitates the grid-based method as a RBE technique. 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 $\oslash$ indicates the convolution of the last belief with the motion model.

The correction requires the computation of Equations (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 unified observation likelihood allows belief update and maintenance regardless of an detection event, the RBE with the unified observation likelihood does not update and maintain the belief effectively. Figure 2 illustratively depicts this problem where the FOV and the NFOV is 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 and thus maintains the belief in an unreliable manner. The next section will describe the proposed target estimation incorporating an acoustic sensor to solve this problem.

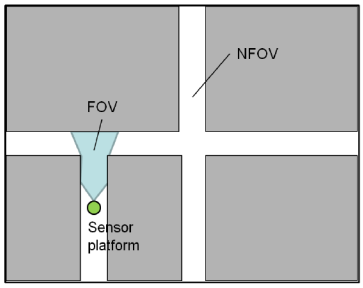


Fig. 2. Optical observation likelihood

III. Target Estimation Incorporating Acoustic Sensors

A. Acoustic Sensor Model and Observation Likelihood

While the optical sensor positively and thus accurately estimates the target position when the target is within its FOV and the detectable region, the target position cannot be estimated positively when the target is outside the detectable region. Unlike optical sensors, the acoustic sensor consisting of two or more microphones can observe a target outside the detectable region, even when the target is outside the observable region, the NFOV, or further not on the LOS, the non-line-of-sight (NLOS). The proposed approach thus utilizes an acoustic sensor in addition to the optical sensor. 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. Although the accuracy is not high, the acoustic sensor can positively estimate the target state even if the target is in the NFOV. At the same time, the observation likelihood could become heavily non-Gaussian with multiple peaks due to the influence of reflected sounds. It is possible that one of the peaks is found near the location of the sensor platform as shown in the figure. The likelihood with a target on the LOS is sharper due to the less influence of reflected sounds but could be still multi-modal. The likelihood with a target on

the LOS and in close proximity will be a sharp Gaussian or near-Gaussian distribution since the direct sound dominates the observation.

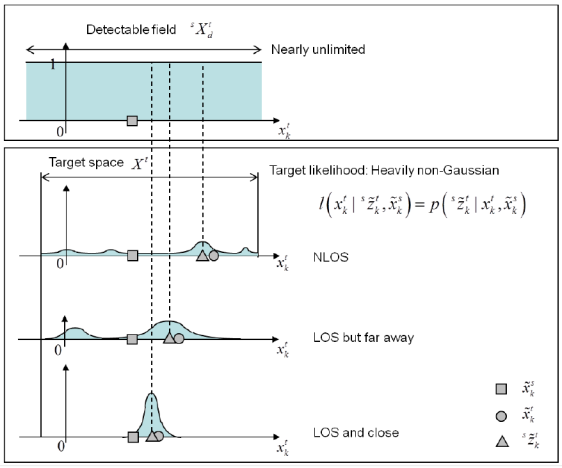


Fig. 3. Acoustic observation likelihood

B. RBE Using Joint 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:

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

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 acoustic/optical observation likelihood when the target is on the NLOS. Since the joint likelihood within the detectable region is cleared by the optical likelihood, possible locations of the target are significantly narrowed down. Some peak(s) could be dropped as shown in the figure. Both the optical observation likelihood and the acoustic observation likelihood could be represented by a heavily non-Gaussian distribution. As a result, the joint optical/acoustic observation likelihood could also become a heavily non-Gaussian distribution. The substitution of Equation (12) into Equation (5) allows the standard RBE and updates and maintains the target belief without any modification.

C. Modeling of Acoustic Observation Likelihood

The proposed approach models an acoustic observation likelihood from the observations of the sound target emitted at the two microphones based on the preliminary investigations of the authors [10], [11], [14]. Figure 5 shows a schematic diagram of the procedure to model an acoustic observation likelihood. We assume that the target emits sound with white noise and 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

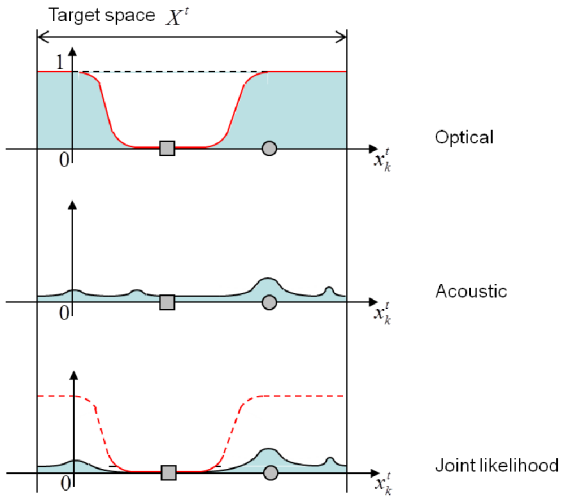


Fig. 4. Joint optical/acoustic observation likelihood

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 from various positions in the target space and saved as a reference response. 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 reference responses. The degree of similarity is then used to develop a similarity map. The similarity map is the acoustic likelihood of the particular sound observation.

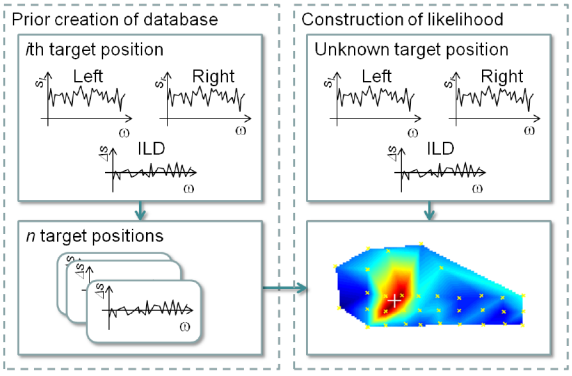


Fig. 5. Acoustic sensor

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

for the ith position $(\tilde{\mathbf{x}}_k^t)_i$, $\Delta S_i(\omega)$, is then given by

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

If the ILD is associated with the target location at the set of frequencies $\mathbf{\Omega} = [\omega_1, ..., \omega_N]^\top$, the ILD vector can be described as

$$\mathbf{S}_i(\mathbf{\Omega}) = [a_1\Delta S_i(\omega_1), ..., a_N\Delta S_i(\omega_N)]^\top, \quad (14)$$

where

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

In the equation, $\langle\cdot\rangle$ is Macaulay brackets, and $\min\{\cdot,\cdot\}$ returns the smaller value of the two entities. The acoustic observation likelihood modeling results in the ILD vectors for n positions within the target space, i.e., $\mathbf{S}_i^*(\mathbf{\Omega}), \forall i \in \{1, ..., n\}$.

Given the ILD vector $\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t)$ created from ${}^s\tilde{\mathbf{z}}_k^t$ with the unknown target position $\mathbf{x}_k^t$, the proposed approach quantifies its degree of correlation to the ith ILD vector as

$$X\left(\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t), \mathbf{S}_i^*(\mathbf{\Omega})\right) = \frac{\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t)^\top \mathbf{S}_i^*(\mathbf{\Omega})}{|\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t)|\,|\mathbf{S}_i^*(\mathbf{\Omega})|}. \quad (16)$$

The acoustic observation likelihood with the particular $\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t)$ can be finally calculated as

$$l^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \sum_i^N \mu\left(\mathbf{x}_k^t\right) X\left(\mathbf{S}(\mathbf{\Omega}|\mathbf{x}_k^t), \mathbf{S}_i^*(\mathbf{\Omega})\right), \quad (17)$$

where $\mu(\mathbf{x}_k^t)$ is a shape function developed by adjacent measurements. It is to be noted that we will not specify the best shape function to use in this paper since every shape function has advantages and disadvantages.

IV. NUMERICAL AND EXPERIMENTAL ANALYSIS

The efficacy of the proposed approach was investigated numerically and experimentally in two steps. The first step experimentally examined the effectiveness of the acoustic observation likelihood in NFOV target detection. After verification, the effectiveness of the RBE with the joint optical/acoustic observation likelihood in NFOV target estimation was demonstrated.

Figure 6 shows the indoor test environment used for both the studies with dimensions. A sensor platform is located in a corridor and faced with the open space the target can move around. The FOV 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 model the acoustic observation likelihood function. After the modeling of the acoustic observation likelihood, 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 7 shows the target and the sensor platform. The target is a wireless speaker whereas the sensor platform has two microphones as an acoustic sensor in addition to a camera as an optical sensor. White noise was emitted from the speaker, and the parameters used to construct the acoustic observation likelihood are listed in Table I.

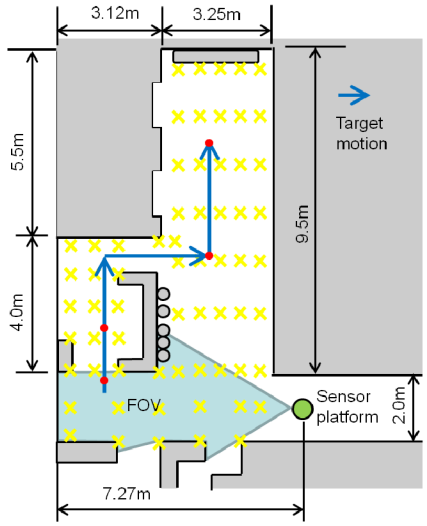


Fig. 6. Map of the test environment

Fig. 7. Target and sensor platform

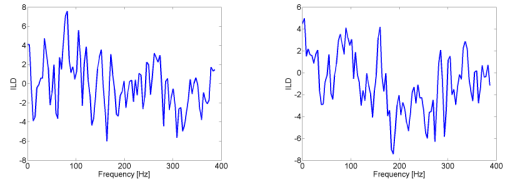
TABLE I
WHITE NOISE DATA

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

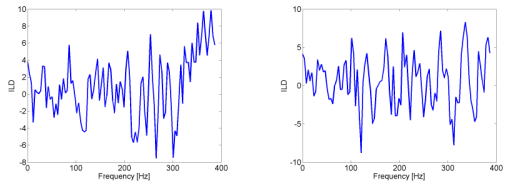
A. Effectiveness of Acoustic Observation Likelihood

Figure 8 shows the six ILDs each observed with a target at one of the 65 positions. Two positions are in the FOV, and four are in the NFOV. It is first seen that the configuration of the ILD varies depending on the target position. This indicates that the ILD contains information on the target

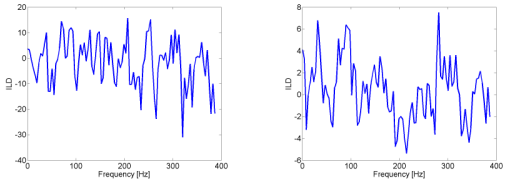
position. The ILD configuration is different even when the target is out of the LOS and thus in the NFOV. The capability of the proposed approach in detecting the target in the NFOV can be expected from this result.



(a) $[2.8, 9]$ (NFOV) (b) $[2.8, 5]$ (NFOV)



(c) $[2.8, 1]$ (FOV) (d) $[5, 1]$ (FOV)

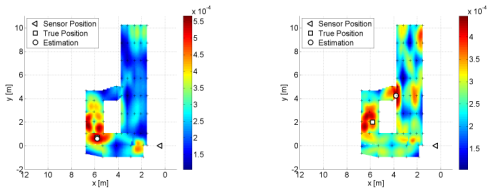


(e) $[6, 5]$ (NFOV) (f) $[5, 4.3]$ (NFOV)

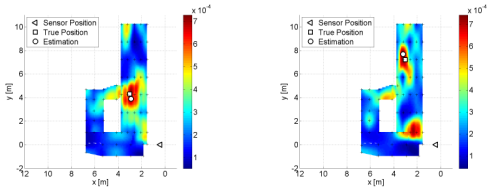
Fig. 8. ILDs at six positions of the 65 positions

Figure 9 shows the acoustic observation likelihoods when the target moved and emitted sound at the four target positions marked by red circles, which are at the 1st, 7th, 35th and 51th steps. The target is in the FOV only at the 1st step. The acoustic observation likelihood is seen to be multimodal due to the excitation of reflected sounds even when the target is within the FOV and thus on the LOS. The proposed approach has, however, been able to correctly capture the true target position at one of the peaks and successfully detect it except for the 7th step.

The effectiveness of the proposed approach is further understood comparatively by seeing Figure 10 with optical observation likelihoods. 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 11 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



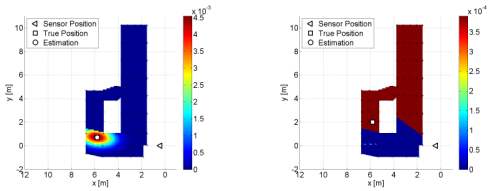
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)



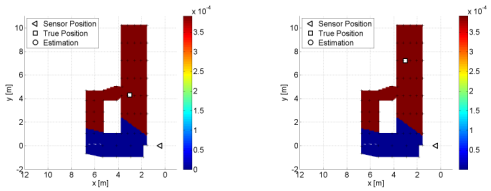
(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 9. Acoustic observation likelihoods

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)

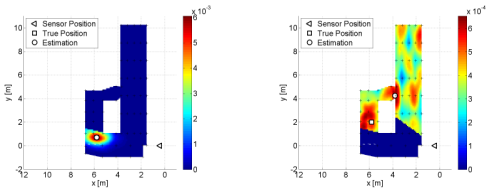


(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Fig. 10. Optical observation likelihoods

B. Effectiveness of the Proposed RBE in Target Estimation

Having understood the limitation of the target detection with observations only in the last section, the effectiveness of the proposed RBE approach was investigated by performing RBE with the same joint optical/acoustic observation likelihoods. Without knowing the target motion well, the target motion model was given by a random walk model assuming



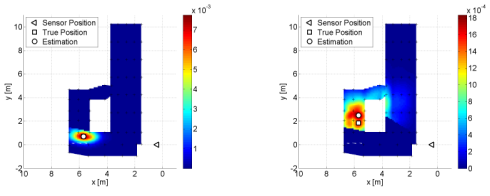
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

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

Fig. 11. Joint optical/acoustic observation likelihoods

that the target is a human who could move to any direction with equal probability.

Figure 12 shows the target belief estimated via RRE 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. Since the prediction supports the continuous movement of the target, the proposed approach was able to eliminate wrong detections and estimate the target position near the true position.

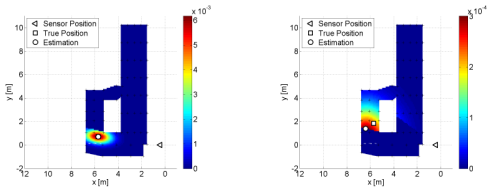


(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

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

Fig. 12. Proposed optical/acoustic target estimation

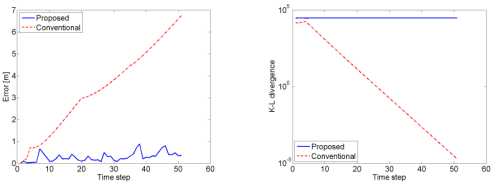
Figure 13 shows the target belief estimated conventionally with the optical observation likelihoods only to comparatively verify the effectiveness of the proposed approach. The result with the optical observation likelihoods is seen to estimate the target position wrongly when the target is in the NFOV. Because an inaccurate random-walk motion model is used, the target is estimated continuously at the location where it was lost. Finally, Figure 14 shows the results quantitatively evaluating the performance of the proposed approach. Figure 14(a) shows the transition of the error of the estimated target position from the true position whereas the transition of the Kullback-Leibler (KL) divergence is exhibited in Figure 14(b). The error transition indicates that the proposed approach maintains low error 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 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. 13. Conventional optical target estimation



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

Fig. 14. Quantitative analysis

V. CONCLUSIONS

This paper has presented target estimation incorporating an acoustic sensor. The proposed approach derives and stores the ILDs for different target positions *a priori*. Given a new acoustic observation, an acoustic observation likelihood is calculated and, further fused with the optical observation likelihood, creates a joint optical/acoustic observation likelihood. The RBE updates and maintains the target belief using the joint observation likelihood. The proposed approach not only guarantees the estimation performance of the optical RBE but also enables positive target estimation regardless of the observability of the target. The first set of experimental investigations has shown the effectiveness of the acoustic observation likelihood and the joint optical/acoustic observation likelihood in NFOV target detection. After verification, the superiority of the RBE with the joint optical/acoustic observation likelihood to the conventional optical RBE has been demonstrated in NFOV target estimation.

The paper has demonstrated the new concept of incorporating an acoustic sensor for NFOV target estimation, and many challenges are still open for future study. The approach was so far proposed for the stationary sensor platform. The ongoing research is aimed at generalizing the approach to handle dynamic sensor platforms. Other issues of immediate interest include the enhancement of acoustic sensing using 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.

REFERENCES

[1] Bergman, N. *Recursive Bayesian Estimation Navigation and Tracking Applications*, Ph.D Dissertation, Linkopings University, 1999.

[2] Bourgault, F., Goktogan, A., Furukawa, T. and Durrant-Whyte, H. F., "Coordinated Search of a Lost Target in a Bayesian World," *Journal of Advanced Robotics*, pp. 187-195, 2004.

[3] F. Bourgault, T. Furukawa and H. F. Durrant-Whyte, Optimal Search for a Lost Target in a Bayesian World, in *Proc. Int. Conf. on Field and Service Robotics*, Mt. Fuji, pp. 239-246 (2003).

[4] Bourgault, T., Furukawa, T. and Durrant-Whyte, H. F., "Optimal Search for a Lost Target in a Bayesian World," Eds. S. Yuta and H. Asama, Field and Service Robots IV, Springer Tracts in Advanced Robotics (STAR), Springer-Verlag, Vol. 24, pp. 209-222, 2006.

[5] Furukawa, T., Bourgault, F., Lavis, B. and Durrant-Whyte, H.F., "Bayesian Search-and-Tracking Using Coordinated UAVs for Lost Targets," *2006 IEEE International Conference on Robotics and Automation*, Orlando, May 14-18, 2006, pp. 2521-2526, 2006.

[6] Furukawa, T., Mak, L.D., Durrant-Whyte, H.F. and Madhevan, R., "Autonomous Bayesian Search and Tracking and its Experimental Validation," *Advanced Robotics*, Vol. 26, No. 5-6, pp. 461-485, 2012.

[7] B. Grochosky, *Information-theoretic Control of Multiple Sensor Platforms*, Ph.D thesis, University of Sydney (2002).

[8] Guo, D. and Wang, X., "Quasi-Monte Carlo Filtering in Nonlinear Dynamic Systems," *IEEE Transactions on Signal Processing*, 54(6), pp. 2087-2098, 2006.

[9] Hammersley, J. M. "Monte Carlo Methods for Solving Multivariable Problems," *Ann. New York Acad. Sci.*, 86, pp. 844-874, 1960.

[10] Kimoto, D. and Kumon, M., "Optimization of the Ear Canal Position for Sound Localization Using Interaural Level Difference," *Proceedings of the 36th Meeting of Special Interest Group on AI Challenges*, pp.14-18, 2012 (in Japanese).

[11] Kumon, M. and Kimoto, D., "On Sound Direction Estimation by Binaural Auditory Robots with Pinnae," *Proceedings of the 35th Meeting of Special Interest Group on AI Challenges*, pp.48-53, 2011 (in Japanese).

[12] Liu, J.S. and Chen, R., "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93(443), pp. 1032-1044, 1998.

[13] Mauler, R., Objective Functions for Bayesian Control-Theoretic Sensor Management, II: MHC-Like Approximation, in *Recent Developments in Cooperative Control and Optimization*, S. Butenko, R. Murphey and P. Pardalos (Eds.), pp. 273-316, Kluwer Academic Publishers, Norwell, MA (2003).

[14] Noda Y. and Kumon, M., "Sound Source Direction Estimation in the Median Plane by Two Active Pinnae," *Proceedings of the 13th SICE System Integration Division Annual Conference*, pp.1643-1646, 2012 (in Japanese).

[15] Sorenson, H.W. Ed., *Kalman Filtering: Theory and Application*, New York: IEEE, 1985.