一般社団法人 人工知能学会
Japanese Society for Artificial Intelligence

人工知能学会資料
JSAI Technical Report
SIG-Challenge-049-1 (11/25)

Non-Line-of-Sight Sound Localization in Unknown Indoor Environments

Tomonari FURUKAWA
Department of Mechanical Engineering
Virginia Tech
tomonari@vt.edu

Abstract

This paper presents a new approach that localizes Non-Line-of-Sight (NLOS) targets in unknown indoor environments. Sensors used in the approach are an auditory sensor, an visual sensor and sensors for Simultaneous Localization and Mapping (SLAM). The visual sensor localizes a target when the target is on its Line-of-Sight (LOS), but it is also used to recognize the environment and localize itself in conjunction with the SLAM sensors. The auditory sensor localizes the target even if the target is not on the LOS of the visual sensor. In order to localize a NLOS target in an unknown environment, the proposed approach extracts and analyzes the first-arrival diffraction signal and the first-arrival reflection signal. The estimation is performed within the Recursive Bayesian Estimation (RBE) framework where observations of the visual and auditory sensors are each converted into an observation likelihood. The ability of the proposed approach was experimentally validated in a controlled indoor environment.

1 Introduction

Indoor environments where humans stay and work are typically so complex with many structures or obstructions that the Line-of-Sight (LOS) region is significantly limited. Here is an example conversation that you could have when you are in such a Non-Line-of-Sight (NLOS) environment:

A: "Where are you?".
B: "I am here".
A: "I am getting close. Where are you?".
B: "I am here".
A: "I found you!".

If the target person is communicative, humans search for and find the target person who is not in the Field-of-View (FOV) by estimating the location of the sound. Audition is used not only as a means of communication but also as a sensor for target localization, and is as important as vision due to such multi-functional capabilities. Robotic audition, if the NLOS localization is made possible, becomes a useful tool for both the co-robots and people who are blind.

Past work on the localization of a NLOS target has been conducted in three different ways. The first approach, forming a Wireless Sensor Network (WSN), localizes the target by measuring the intensity of the transmitted signal at each wireless receiver and fusing the measurements of all the receivers under the LOS assumption [2, 4, 17, 27, 13, 7, 8]. Radio signals are commonly used since sound signals reflect excessively and do not create unique signal intensity. The approach is easy to install, but the accuracy depends on the validity of the LOS assumption [3, 18, 20, 10].

In the second approach, Time-of-Arrival (TOA) information of the received signal is used for target localization. The approach most commonly utilizes acoustic signals due to their slow speed compared to that of radio signals. The majority of sound localization challenges have been however focused on the direction of sound rather than its position due to complexity of sound wave propagation [25, 21]. For a NLOS target, Mak and Furukawa [14] located it by using the diffraction characteristics of low-frequency sound though the time of sound generation, which is often unknown, must be informed beforehand.

The last approach enhances the NLOS target localization numerically [15, 5, 6]. The approach localizes the target by updating and maintaining its probabilistic belief in the framework of Recursive Bayesian Estimation (RBE) and processing observations as likelihoods. In the use of an optical sensor, the event of "no detection" is converted into an observation likelihood describing no probability that the target exists. While the no-detection information is still useful, the belief, however, becomes highly unreliable unless the target is rediscovered within a short period after being lost. Takami et al. [24, 23] incorporated a stereo auditory sensor such that the NLOS target can be detected using positive information accordingly. Their approach however needs to collect acoustic cues of the environment in advance. It is not applicable if the environment is unknown.

This paper presents an extensive approach that localizes NLOS targets in unknown indoor environments. Sensors to be implemented in the proposed approach are an auditory sensor, an visual sensor and sensors for Simultaneous Localization and Mapping (SLAM). The visual sensor accurately localizes a target when the target is on its LOS, but it is also used to recognize the environment and localize itself in conjunction with the SLAM sensors. The auditory sensor localizes the target even if the target is not on the LOS of the visual sensor. In order to localize a NLOS target in an unknown environment, the first-arrival diffraction signal and the first-arrival reflection signal are extracted and analyzed. The estimation is performed within the RBE framework where observations of the visual and auditory sensors are each converted into an observation likelihood.

The paper is organized as follows. The following section presents the proposed approach that uses the visual and the auditory sensors and localizes NLOS targets in unknown indoor environments. Section 3 presents the proposed extraction of the first-arrival diffraction and reflection signals and the construction of the joint auditory observation likelihood. While its application as an assistive and training device for people who are blind or visually-impared is described in Section 4, Section 5 shows the results of the preliminary experimental study and demonstrates the efficacy of the proposed approach. Conclusions are summarized in the final section.

2 Hybrid Visual/Auditory RBE for Unknown Environments

2.1 Overview

Figure 1 shows a schematic diagram of the hybrid visual/auditory approach proposed in this paper. This is to localize a NLOS target in an unknown environment. Sensors used are an auditory sensor, a visual sensor and SLAM sensors. The auditory sensor is a microphone array whereas the visual sensor is a RGB-D sensor, which measures not only RGB information but also depth information and recognizes three-dimensional (3D) surroundings on the LOS. SLAM sensors are those to enable SLAM and include an Inertial Measurement Unit (IMU).

The proposed approach operates as follows. Similarly to the past work of the author, it deploys the framework of the grid-based RBE and estimates a sound target in terms of a non-Gaussian belief [5, 6, 11]. Since environments are assumed to be unknown, the proposed approach incorporates SLAM where the RGB-D sensor is utilized in addition to the SLAM sensors. The self-location is thus known, so only the target pose is estimated in the RBE. The RGB-D sensor detects and locates a sound target if it is on the LOS or in the FOV. Otherwise, a sound target is processed as "not detected", and the RGB-D image is used primarily to recognize 3D surroundings including geometry and material which in-

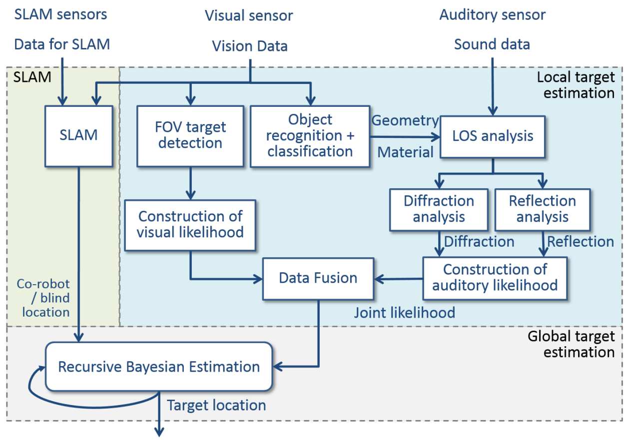


Figure 1: Schematic diagram of the proposed approach

fluence NLOS sound propagation. The microphone array is the sensor that locates a NLOS target in addition to a LOS target. Based on the visual information on the surroundings, a NLOS target is identified by extracting and analyzing the first-arrival diffraction signal and the first-arrival reflection signal. The proposed approach allows localization in unknown environments because it is sound physics based and does not thus require spatial information. Fusion of the visual and auditory observation likelihoods results in a joint likelihood, which updates belief in the glsrbe.

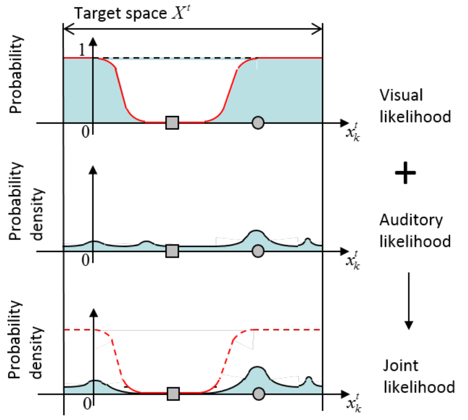
2.2 Mathematical Formulation

The mathematical framework of the hybrid visual/auditory RBE is as follows. Let the state of the robot s and the map updated by SLAM at time step $k-1$ be $\bar{\mathbf{x}}_{k-1}^s \in \mathcal{X}^s$ and $\bar{\mathbf{m}}_{k-1} \in \mathcal{M}$ respectively. Given a sequence of observations by the robot s from time step 1 to time step $k-1$, ${}^s\tilde{\mathbf{z}}_{1:k-1}^t \equiv \{{}^s\tilde{\mathbf{z}}_\kappa^t | \forall\kappa \in \{1, ..., k\}\}$, the RBE iteratively updates the belief on the state of a target t, $\mathbf{x}_k^t \in \mathcal{X}^t$, in time and observation. Let the belief given the sequence of observations and the robot state and the map at time step $k-1$ be $p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s, \bar{\mathbf{m}}_{k-1}\right)$. Chapman-Kolmogrov equation updates the prior belief in time, or predicts the belief at time step k, by the probabilistic motion model $p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s, \bar{\mathbf{m}}_{k-1}\right)$:

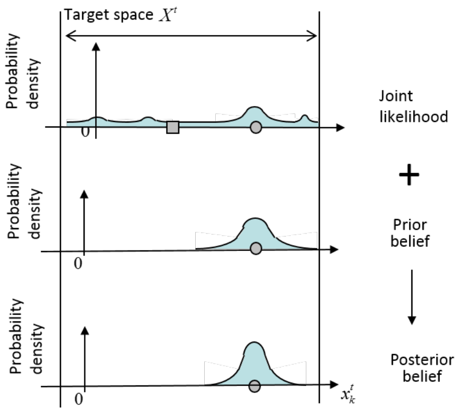
$$p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s, \bar{\mathbf{m}}_{k-1}\right) = \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s, \bar{\mathbf{m}}_{k-1}\right) p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t,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

The observation update, or the correction process, is performed using the Bayes theorem. The target belief is corrected using the new observation ${}^s\tilde{\mathbf{z}}_k^t$ as

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



(a) Joint visual/acoustic observation likelihood



(b) RBE incorporating an acoustic sensor

Figure 2: Hybrid visual/auditory target estimation

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

One of the core technologies proposed in this paper is the hybrid use of visual and auditory sensors. This is given by

$$l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) = l^c\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) l^a\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) \tag{3}$$

where $l^c(\cdot)$ and $l^a(\cdot)$ are the likelihoods of the visual sensor (RGB-D camera) and the auditory sensor (microphone array). In order to maximize information, the camera observation is used not only to detect a target if it is in the FOV but also to construct the no-detection likelihood if the target is outside the FOV:

$$l^c\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) = \begin{cases} p\left({}^s\tilde{\mathbf{z}}_k^t | \mathbf{x}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) & \exists^s\tilde{\mathbf{z}}_k^t \in {}^s\mathcal{X}_d^t \\ 1 - P_d\left(\mathbf{x}_k^t | \bar{\mathbf{x}}_k^s\right) & \nexists^s\tilde{\mathbf{z}}_k^t \in {}^s\mathcal{X}_d^t, \end{cases} \tag{4}$$

where ${}^s\mathcal{X}_d^t$ is the camera FOV or, more precisely, the detectable region. The effectiveness of Equation (4) is thoroughly investigated by the author in the context of autonomous search and tracking.

While the derivation of $l^a(\cdot)$ is most challenging and thus will be dealt with separately in the next section, the advantage of Equation (3) in RBE is illustratively shown in Figure 2. The possible locations of the target are narrowed down even though the no-detection likelihood is used in visual sensing since the likelihood clears out the joint likelihood in the FOV and dropped some peak(s) as shown in Figure 2(a). Because sharpest and most Gaussian is the visual observation likelihood with detection, the prior belief is most determined by the last visual observation and remains a sharp Gaussian distribution as shown in Figure 2(b). 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.

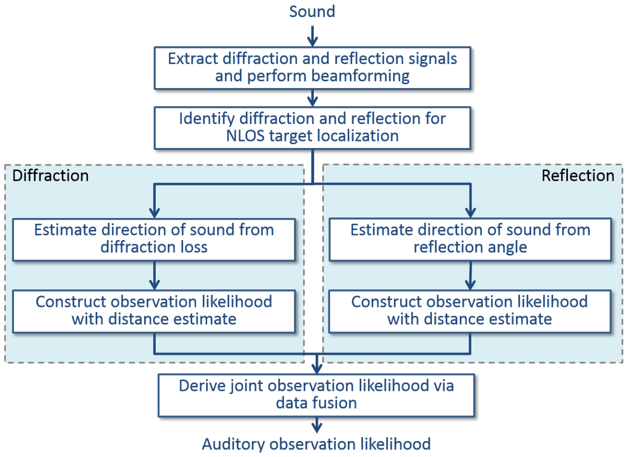
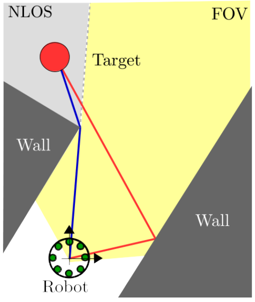


Figure 3: Construction of auditory NFOV target likelihood

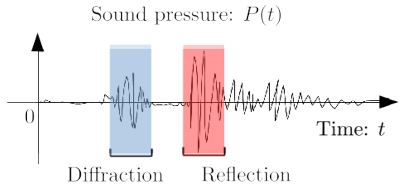
3 Physics based NLOS Auditory Observation Likelihood

3.1 Overview

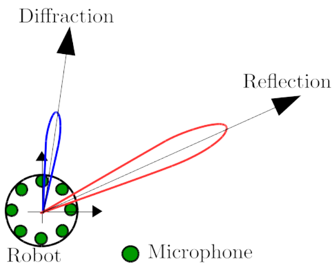
Figure 3 shows the overview of how to construct a NLOS auditory observation likelihood using the physics of sound wave propagation. Unlike radio signals, sound signals reflect significantly without penetrating into different media while they also diffract at low frequencies [1]. The proposed approach begins with obtaining a time-domain signal of a relatively impulsive sound at the microphone array. Notable signals are then extracted as candidate first-arrival diffraction and reflection sounds. The sound target is considered in a NLOS region if the diffraction and reflection signals behave as expected. An acoustic beamformer identifies the directions of the diffraction and reflection signals in the LOS region, and the diffraction and reflection points are identified to further localize the sound target in the NLOS region. The direction of the diffraction signal beyond the LOS is inferred by deriving the loss of sound energy through diffraction, or the diffraction loss. That of the reflection signal is identified by considering the orientation of the reflection wall. Diffraction and reflection observation likelihoods are eventually constructed by additionally incorporating knowledge on the distance from the sound magnitude and characteristics. An auditory observation likelihood is finally created by fusing the diffraction and reflection observation likelihoods.



(a) Acoustic signals from NFOV target

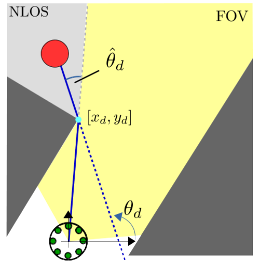


(b) Extraction of diffraction/reflection signals

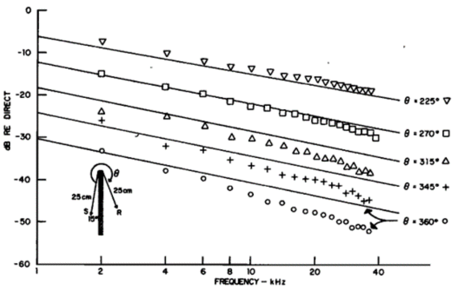


(c) Direction of signals identified though beamforming

Figure 4: Auditory NFOV target observation



(a) Diffraction sound estimation



(b) Magnitude with different orientation angles [16]

Figure 5: Diffraction sound estimation

3.2 Extraction of First-arrival Diffraction and Reflection Signals

Figure 4 shows the identification of a NLOS sound target and the extraction of the diffraction and reflection signals proposed in this paper illustratively in one of the simplest scenarios where a robot carrying a microphone array receives sound emitted by a target in the NLOS in a two-dimensional indoor environment with three walls (Figure 4(a)). Figure 4(b) shows the pressure of sound in the time domain, $P_i(t)$. As shown in the figures, sound waves emitted from a NLOS target reach the robot first through diffraction and second through reflection and, if the sound is relatively impulsive, the first-arrival diffraction and reflection signals can be extracted clearly. Because the sound energy loss from diffraction is larger than that from reflection, the sound target can be recognized NLOS if the absolute peak of the first signal is smaller than the second signal (The first-arrival signal is strongest with a LOS target as it is the LOS signal). Extraction becomes challenging for complex environments, but various existing techniques proposed to extract signals or select thresholds for extraction reportedly achieve successful extraction and identify candidate diffraction and reflection signals [12, 9, 26, 8]. The extraction results in the identification of directions of diffraction and reflection sounds through acoustic beamforming shown in Figure 4(c). Once it observes the sound directions, the RGB-D camera identifies the end of LOS and thus identifies the diffraction and reflection points as well as the orientation of the reflection wall and the reflection angle.

3.3 Estimation of Sound Direction from Diffraction Signals

Figure 5(a) shows the notations used for estimating sound direction from diffraction signals in the scenario introduced in the last subsection. The direction angle with respect to the robot frame to estimate is defined by θ_d whereas the NLOS angle is given by $\hat{\theta}_d$. The diffraction point is given by $[x_d, y_d]$. Of these, the diffraction point is known from the result of beamforming and the depth measurement, so the NLOS angle $\hat{\theta}_d$ must be further identified.

The proposed approach identifies the angle by analyzing the magnitudes of diffraction and reflection sounds, $M^d(\omega)$ and $M^r(\omega)$, which are extracted after representing the sound signals with respect to frequency ω. The loss of sound energy is assumed to be more if the NLOS angle is more. This assumption, in fact, has been found to be valid by the work of Medwin a quarter-century ago [16] shown in Figure 5(b). The magnitude of diffraction sound drops when the "level of NLOS" represented by the orientation angle is increased. Assuming that reflection is specular with negligible loss, this makes the proposed approach define the diffraction loss as

$$L_d = \int \left[M^r(\omega) - M^d(\omega) \right] \mathrm{d}\omega \geq 0 \tag{5}$$

and associates it with the level of NLOS. The work of Medwin also shows that the diffraction loss is approxi-

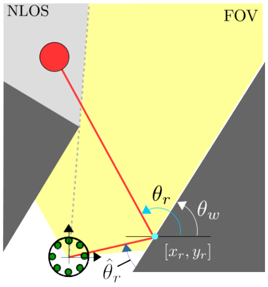


Figure 6: Reflection sound estimation

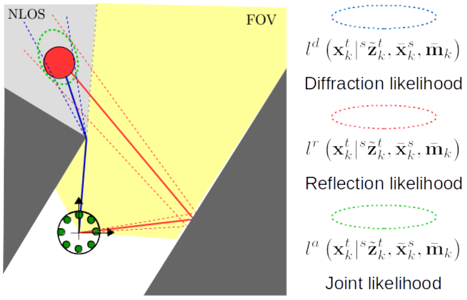


Figure 7: Joint acoustic observation likelihood

mately proportional to the level of NLOS. The nonlinear polynomial can be derived from the presented graph and enables more accurate identification of diffraction angle; $\hat{\theta}_d = f(L_d)$. Having the NLOS angle specified, the diffraction angle θ_d is derived as

$$\theta_d = \hat{\theta}_d + \tan^{-1}\frac{y_d}{x_d}. \tag{6}$$

3.4 Estimation of Sound Direction from Reflection Signals

Figure 6 shows the proposed approach for estimation of sound direction from reflection signals. Reflection makes the sound propagation and the subsequent target estimation complicated, but if the wall is smooth and yields specular reflection, the sound direction can be estimated easily [19]. Let the reflection angle with respect to the robot frame to derive be θ_r and the sound direction to the reflection wall and the reflection point be $\hat{\theta}_r$ and $[x_r, y_r]$ respectively. Since both $\hat{\theta}_r$ and $[x_r, y_r]$ are known from the preceding measurement, the orientation of the wall with respect to the robot frame is given by

$$\theta_w = \hat{\theta}_r + \tan^{-1}\frac{y_r}{x_r}. \tag{7}$$

The reflection angle is resultantly given by

$$\theta_r = \hat{\theta}_r + \theta_w = 2\hat{\theta}_r + \tan^{-1}\frac{y_r}{x_r}. \tag{8}$$

3.5 Construction of Joint Acoustic Observation Likelihood

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

$$\begin{aligned} & l^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) \\ & \quad = l_j^d\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) l_j^r\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) \end{aligned} \tag{9}$$

where $l^d(\cdot)$ and $l^r(\cdot)$ are the diffraction and reflection observation likelihood.

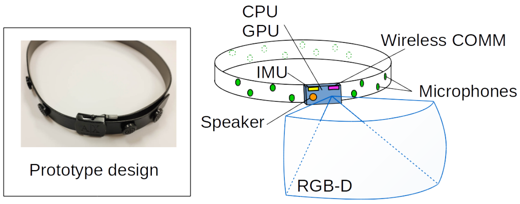
Figure 7 illustrates the diffraction and reflection observation likelihoods as well as the joint observation likelihood where the observation likelihood is represented by an ellipsoid indicating a probability distribution with a covariance. The diffraction and reflection likelihoods are shown to have high eccentricity due to more accuracy in direction than in distance. Since the difference of the diffraction and reflection likelihoods in orientation may not be significant, the resulting auditory joint likelihood could also be given by an ellipsoid with high eccentricity, but the proposed approach, utilizing the diffraction and reflection physics of sound, could estimate the location of the sound target.

4 Assistive/Training Devices for Blind and Visually Impaired People

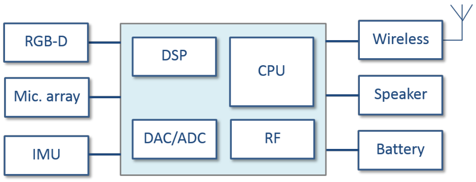
Figure 8 shows the schematic diagram of the wearable assistive/training device for blind and visually impaired people to be developed. The wearable device is chosen to be a belt partly because it is one of the most common wearings and partly because it can implement a ringed microphone array as well as other sensors/components with some weight including an RGB-D camera. The major components of the device are (1) a multi-story ringed microphone array, (2) an RGB-D camera, (3) an IMU, (4) a speaker, (5) a central unit and (6) a wireless module. Because of the multi-story design, the microphone array can form acoustic beams in not only the horizontal direction but also in the vertical direction. This allows the identification and removal of sound components coming from the floor and the ceiling and the extraction of the corresponding first-arrival diffraction and reflection signals. The RGB-D camera, which works based on the principle of time-of-flight, structured light or stereovision, is used not only to recognize LOS 3D environments for NLOS target estimation but also to perform SLAM together with an IMU. The speaker is used to provide feedback by sound for assistive and training use. The NLOS target estimation can be performed on the central unit since its computation can be made light in weight.

5 Preliminary Experimental Validation

This subsection presents a result of the preliminary proof-of-concept. Figure 9 shows the test environment developed and the actual test conducted in the anechoic

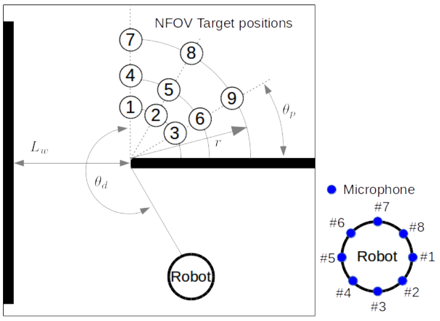


(a) Mechanical design



(b) Electronic design

Figure 8: Assistive/training devices for blind people



(a) Test environment

(b) NLOS test with Daniel Kish

Figure 9: Experimental proof-of-concept settings

chamber. The capability of the microphone array, as well as sighted and blind people including the pioneer of human echolocation, Daniel Kish, was tested in the environment. Only two walls were placed in the anechoic chamber to make the problem two-dimensional (2D) for the proof-of-concept purpose. A mechanical clicker created impulsive sound at positions labeled 1-9, and the robot with the microphone array and human testees were supposed to identify the correct sound direction and location [22].

Figure 10 shows the sound amplitude in the time domain collected at Microphones 6 and 8. The result shows that the diffraction signal and reflection signal are distinctly separable having enough time interval between the signals. More signal processing efforts will be necessary when more natural sounds are deployed, and the signal processing effort is ongoing while completing the proof-of-concept. Figure 11 shows the constructed joint

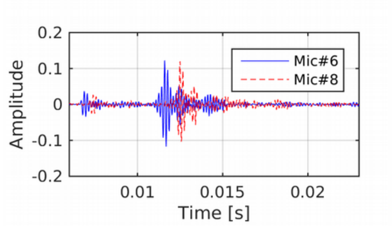
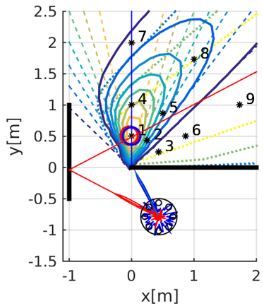
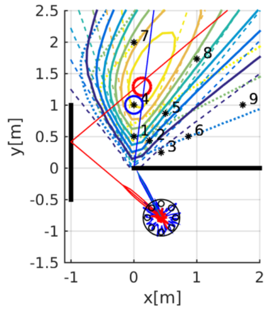


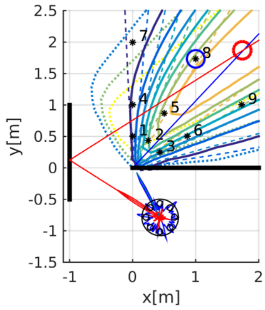
Figure 10: Time-domain signals



(a) Position 1



(b) Position 4



(c) Position 8

Figure 11: Result of proof-of-concept

observation likelihood together with the true target location colored blue when the sound source was located at Positions 1, 4 and 8. The maximum likelihood is shown by a red circle to compare to the true location. It can be first seen that the diffraction and reflection points are always identified well to indicate that the sound source is in the NLOS region. The accuracy drops particularly for targets in a severely NLOS region, but the proposed approach could still make a good estimate.

6 Conclusions

This paper has presented a new approach that localizes NLOS targets in unknown indoor environments. While the visual sensor localizes a target when the target is on its LOS, the auditory sensor stably localizes the target even if the target is not on the LOS of the visual sensor. In order for a NLOS target in an unknown environment, the proposed approach extracts and analyzes the first-arrival diffraction signal and the first-arrival reflection signal. The RBE localizes the NLOS most reliably and accurately. The ability of the proposed approach was experimentally validated in a controlled indoor environment.

The work presented in this paper is only the first step for the NLOS target localization in unknown indoor environments. Ongoing work includes the experimental validation in the uncontrolled indoor environment, the development of the assistive/training device for people who are blind or visually impaired and the use of human voice.

Acknowledgments

The work was primarily supported by National Science Foundation EAGER program (Award #1554961). The author also express his gratitude to Makoto Kumon, Kuya Takami and Hangxin Liu for their collaboration to the work.

References

[1] Giovanni Bellusci, Junlin Yan, Gerard JM Janssen, and Christian CJM Tiberius. An ultra-wideband positioning demonstrator using audio signals. In *Positioning, Navigation and Communication, 2007. WPNC'07. 4th Workshop on*, pages 71–76. IEEE, 2007.

[2] Marco Bertinato, Giulia Ortolan, Fabio Maran, Riccardo Marcon, Alessandro Marcassa, Filippo Zanella, Matrizio Zambotto, Luca Schenato, and Angelo Cenedese. Rf localization and tracking of mobile nodes in wireless sensors networks: Architectures, algorithms and experiments. 2008.

[3] Pi-Chun 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.

[4] Huan Dai, Zhao-Min Zhu, and Xiao-Feng Gu. Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network. *Journal of Network and Computer Applications*, 2012.

[5] Tomonari Furukawa, Frederic Bourgault, Benjamin Lavis, and Hugh F Durrant-Whyte. Recursive bayesian search-and-tracking using coordinated uavs for lost targets. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2521–2526. IEEE, 2006.

[6] Tomonari Furukawa, Lin Chi Mak, Hugh Durrant-Whyte, and Rajmohan Madhavan. Autonomous bayesian search and tracking, and its experimental validation. *Advanced Robotics*, 26(5-6):461–485, 2012.

[7] Sinan Gezici. A survey on wireless position estimation. *Wireless Personal Communications*, 44(3):263–282, 2008.

[8] Ismail Guvenc and Chia-Chin Chong. A survey on toa based wireless localization and nlos mitigation techniques. *Communications Surveys & Tutorials, IEEE*, 11(3):107–124, 2009.

[9] Zoubir Irahhauten, Giovanni Bellusci, Gerard JM Janssen, Homayoun Nikookar, and CCJM Tiberius. Investigation of uwb ranging in dense indoor multipath environments. In *Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on*, pages 1–5. IEEE, 2006.

[10] Hiam M Khoury and Vineet R Kamat. Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, 18(4):444–457, 2009.

[11] Makoto Kumon, Daisuke Kimoto, Kuya Takami, and Tomonari Furukawa. Bayesian non-field-of-view target estimation incorporating an acoustic sensor,. In *Proceedings of 2013 IEEE/RSJ International Conference of Intelligent Robots and Systems*, November 1-3, 2013 2013.

[12] Joon-Yong Lee and Sungyul Yoo. Large error performance of uwb ranging in multipath and multiuser environments. *Microwave Theory and Techniques, IEEE Transactions on*, 54(4):1887–1895, 2006.

[13] Hui Liu, H. Darabi, P. Banerjee, and Jing 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.

[14] Lin C Mak and Tomonari 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.

[15] 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.

[16] H Medwin. Shadowing by finite noise barriers. *The Journal of the Acoustical Society of America*, 69(4):1060–1064, 1981.

[17] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004.

[18] Eric A Prigge. *A positioning system with no line-of-sight restrictions for cluttered environments.* PhD thesis, Stanford University, 2004.

[19] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of Audio Engineering Society*, 45(6):456–466, June 1997.

[20] Chee Kiat Seow and Soon Yim Tan. Non-line-of-sight localization in multipath environments. *Mobile Computing, IEEE Transactions on*, 7(5):647–660, 2008.

[21] Piergiorgio Svaizer, Alessio Brutti, and Maurizio 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.

[22] Kuya Takami, Tomonari Furukawa, Makoto Kumon, and Lin Chi Mak. Non-field-of-view indoor sound source localization based on reflection and diffraction. In *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2015 IEEE International Conference on*, pages 59–64. IEEE, 2015.

[23] T; Kumon M; Dissanayake G. Takami, K.; Furukawa. Non-field-of-view acoustic target estimation in complex indoor environment. In *Field and Service Robotics*, June 24-26 2015.

[24] Tomonari; Kumon Makoto; Dissanayake Gamini Takami, Kuya; Furukawa. Estimation of a nonvisible field-of-view mobile target incoorperating optical and acoustic sensors. *Autonomous Robots*, 7(26):1–17, 2015.

[25] JeanMarc Valin, François Michaud, Jean Rouat, and Dominic 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.

[26] Chi Xu and Choi L Law. Delay-dependent threshold selection for uwb toa estimation. *IEEE communications letters*, 12(5):380–382, 2008.

[27] Dian Zhang, Yanyan Yang, Dachao Cheng, Siyuan Liu, and Lionel M Ni. Cocktail: an rf-based hybrid approach for indoor localization. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010.