2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Congress Center Hamburg
Sept 28 - Oct 2, 2015. Hamburg, Germany

Microphone-Accelerometer Based 3D Posture Estimation for a Hose-shaped Rescue Robot

Yoshiaki Bando[1], Katsutoshi Itoyama[1], Masashi Konyo[2], Satoshi Tadokoro[2],
Kazuhiro Nakadai[3], Kazuyoshi Yoshii[1], and Hiroshi G. Okuno[4]

***Abstract*— 3D posture estimation for a hose-shaped robot is critical in rescue activities due to complex physical environments. Conventional sound-based posture estimation assumes rather flat physical environments and focuses only on 2D, resulting in poor performance in real world environments with rubble. This paper presents novel 3D posture estimation by exploiting microphones and accelerometers. The idea of our method is to compensate the lack of posture information obtained by sound-based time-difference-of arrival (TDOA) with the tilt information obtained from accelerometers. This compensation is formulated as a nonlinear state-space model and solved by the unscented Kalman filter. Experiments are conducted by using a 3 m hose-shaped robot with eight units of a microphone and an accelerometer and seven units of a loudspeaker and a vibration motor deployed in a simple 3D structure. Experimental results demonstrate that our method reduces the errors of initial states to about 20 cm in the 3D space. If the initial errors of initial states are less than 20 %, our method can estimate the correct 3D posture in real-time.**

I. Introduction

Hose-shaped rescue robots have been developed for probing spaces under collapsed buildings that humans or animals cannot go into [1]–[3]. They are characterized by a thin, long and flexible body, and have self-locomotion mechanisms for penetrating into narrow spaces. The Active Hose-II robot [1], for example, has small powered wheels that enable it to move forward, and the Active Scope Camera robot [2], [3] can move forward by vibrating cilia covering its body. The second robot was used for an actual search-and-rescue mission in Jacksonville, Florida, USA in 2008 [4].

Sensor systems on rescue robots including the hose-shaped robots typically do not work well in the extreme environments where such robots are intended to be used [5]–[8]. The accuracy of the GPS, for example, is degraded because the rubble in collapsed buildings blocks signals from the satellites [5], and a video camera inserted into narrow gaps often fails to capture the views there because the lighting causes whiteout or blackout conditions [6]. To develop robust sensor systems, it is thus essential to integrate various modalities compensating each other's weaknesses [9]–[12].

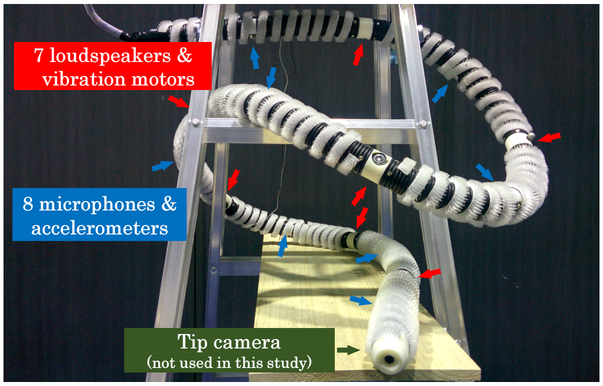


Fig. 1. We aim to estimate the posture of a hose-shaped robot using microphones, loudspeakers, and accelerometers installed on the robot.

To control the flexible body of a hose-shaped robot in an unseen complex environment, it is crucial to robustly estimate its 3D posture (shape). Although there are many posture estimation methods using various types of sensors [13]–[15], these methods face some problems in the disaster environments. The performances of magnetometer-accelerometer based method, for example, are degraded in the disaster environments because magnetic fields are easily affected by the steel frames of collapsed buildings [13]. Also proposed was a sound-based posture estimation method that can be used in a closed space allowing sound propagation among microphones and loudspeakers installed on the robot [14]. Nevertheless, a sound-based method often fails to estimate the robot posture accurately because the obstacles around the robot block sound propagation.

In this paper, we present a microphone-accelerometer based 3D posture estimation method for a hose-shaped robot equipped with a set of microphones, loudspeakers, and accelerometers (Fig. 1). The microphones and loudspeakers allow to estimate their relative positions using the time differences of arrival (TDOAs) of a reference signal emitted from the loudspeakers, and the accelerometers are used for estimating their tilts by measuring the acceleration of gravity. Since the TDOA-based method is degraded in a rubble-containing environment, we exclude TDOAs distorted by rubble and fill up the lack of posture information with the tilt information. To do this, we detect TDOAs of direct sound by excluding outliers, and estimate the robot posture based on a nonlinear state-space model integrating TDOA and tilt information by using the unscented Kalman filter (UKF) [16].

[1] Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan {yoshiaki, itoyama, yoshii}@kuis.kyoto-u.ac.jp

[2] Graduate School of Information Science, Tohoku University, Miyagi, 980-8579, Japan {konyo, tadokoro}@rm.is.tohoku.ac.jp

[3] Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Saitama, 351-0114, Japan / Honda Research Institute Japan Co., Ltd. nakadai@jp.honda-ri.com

[4] Graduate Program for Embodiment Informatics, Waseda University, Tokyo 169-0072, Japan okuno@aoni.waseda.jp

*Demo page: http://winnie.kuis.kyoto-u.ac.jp/members/yoshiaki/demo/iros2015/

978-1-4799-9993-4/15/$31.00 ©2015 IEEE

II. Related Work

This section reviews related work on posture estimation. We first introduce existing methods for estimating the shape of a flexible cable and then review methods for simultaneous localization of microphones and sound sources.

A. Shape Estimation for Flexible Cables

Lee *et al.* [13] estimated the shape of a flexible sensor tube by using a sensor network system based on several electronic compass units including a 3-axis accelerometer and 3-axis magnetometer. The robot shape is modeled as a kinematic chain and is estimated by using orientation information obtained by the compass units. Since the orientation of each unit is determined in part from magnetometer information, this method fails to estimate the posture when the magnetic field is distorted.

Ishikura *et al.* [15] estimated the posture of an Active Scope Camera robot, which is one kind of hose-shaped robot, by using gyrometers. They formulated a flexible dynamics model and estimated the posture using the UKF. Since this method uses only proprioceptive sensors, the estimation accuracy is not affected by disaster environments. The estimation performance, however, is often degraded by the vibration of the robot. They had also examined a vision-based localization method that estimates the movement of a tip camera by extracting the corresponding points in images [6]. This method, though, often fails to extract these points when a tip light causes over-exposures. Both of these integral-type methods suffer from the cumulative error problem.

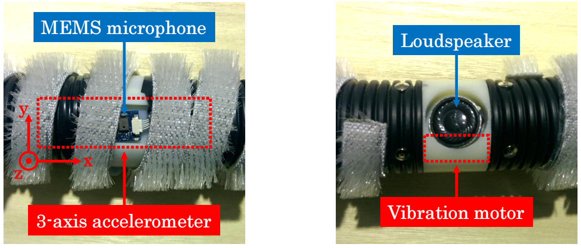
B. Localization of Microphones and Loudspeakers

There are many non-integral-type sound-based methods for estimating microphone and loudspeaker positions simultaneously [17]–[20]. Since recorded sound depends only on the current relative positions of the microphones and loudspeakers, the cumulative error problem can be avoided.

Ono *et al.* [17] defined a *blind alignment problem* in which the positions of an asynchronous microphone array and sound sources are estimated from the TDOAs between the microphone channels. They solved this problem by using an auxiliary function approach. Their method is designed for offline use and is based on the assumption that the positions of sound sources and microphones are stable.

Rather than using the TDOA-based approach, Chen *et al.* [18] developed a method that localizes microphones and sound sources simultaneously on the basis of the sound source energy, which depends on the distance between the sound source and the microphone. The microphone and sound source positions estimated by this method are based on a sound attenuation model. Since the method is based on the sound energy, though, its performance is severely degraded by external noise.

Miura *et al.* [19] developed an TDOA-based online simultaneous localization method using a moving sound source. This method localizes microphones and a moving sound source, and it detects the time delays between the channels by using the EKF-SLAM framework developed to enable a mobile robot to estimate its location and a map of the field simultaneously. The microphone positions are regarded as the map and the position of the moving sound source is modeled as the self-location. This method estimates the positions of microphones on a robot while a moving person is clapping his hands.



(a) Microphone and accelerometer (b) Loudspeaker and vibrator

Fig. 2. Modules with a microphone and accelerometer or a loudspeaker and vibrator.

A TDOA-based online method was proposed for estimating the 2D posture of a hose-shaped robot that has a set of microphones and loudspeakers installed on its body [14], [20]. This method is based on a nonlinear state-space model representing the dynamics of the robot posture and can estimate the positions of the moving loudspeakers and microphones, which represent the time-varying posture, by using the UKF [20]. This method tackles the obstacle problem (TDOAs distorted by rubble) by detecting outliers among the estimated TDOAs and excluding them [14]. Since the number of TDOAs of direct sounds decreases in narrow spaces, this method fails to estimate the robot posture accurately in a rubble-containing environment.

III. Microphone-Accelerometer Based 3D Posture Estimation

In the proposed method of microphone-accelerometer based 3D posture estimation, the posture of a hose-shaped robot is estimated by repeating the following four steps: 1) generate a reference signal from each loudspeaker, one by one, 2) estimate the reference signal's TDOAs at the microphones, 3) estimate the tilts at 3-axis accelerometers, and 4) estimate the robot posture from the estimated TDOAs and tilts by using the UKF.

A. Prototype Hose-shaped Robot

Fig. 1 shows a prototype hose-shaped robot used in this study. The body is a corrugated tube with a diameter of 38 mm and a total length of 3 m. This robot has a self-propelling mechanism the same as that of the hose-shaped robot called the Active Scope Camera [3]. The entire surface of the robot is covered by cilia and the robot moves forward by vibrating the cilia.

This robot has two types of modules, one with a microphone (mic) and 3-axis accelerometer (acc) (Fig. 2(a)) and the other with a small loudspeaker (src) and vibration motor (vib) (Fig. 2(b)). As shown in Fig. 3, $M = 8$ mic-acc modules and $N = 7$ src-vib modules are positioned on the robot at a regular interval $l = 20$ cm. The distance between the modules at the ends is 2.8 m.

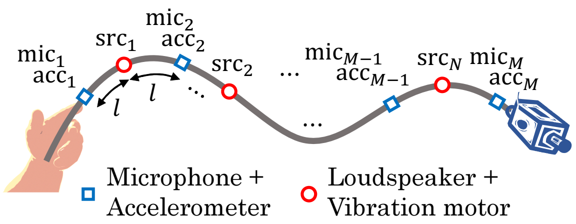


Fig. 3. Arrangements of microphones, accelerometers, and loudspeakers.

B. Problem Specification

We denote the microphones, accelerometers, and loudspeakers as mic_m, acc_m $(m = 1, \cdots, M)$, and src_n $(n = 1, \cdots, N)$, respectively, where $N = M - 1$. We define k as the measurement index and the mic-acc module and src-vib module positions as $\boldsymbol{u}_{m,k}$ and $\boldsymbol{v}_{n,k}$, respectively.

The problem of the microphone-accelerometer based posture estimation is defined as follows:

Input: 1) TDOAs $\tau^n_{m_1,m_2,k} \in \mathbb{R}$ $(m_1, m_2 \in \mathcal{M}_k)$ when src_n omits a reference signal, and 2) tilt angles at the accelerometers $\psi_{1,k}, \cdots, \psi_{M,k} \in \mathbb{R}$.
Output: The positions of each mic-acc module $\boldsymbol{u}_{m,k} \in \mathbb{R}^3$ and each src-vib module $\boldsymbol{v}_{n,k} \in \mathbb{R}^3$.

where $\tau^n_{m_1,m_2,k}$ represents a TDOA between mic_{m_1} and mic_{m_2} and $\mathcal{M}_k$ represents a set of indices for microphones that record the direct sound of the reference signal. The TDOAs $\tau^n_{m_1,m_2,k}$ and microphone indices $\mathcal{M}_k$ are estimated from synchronized M-channel audio signals $\boldsymbol{z}_k(t) \in \mathbb{R}^M$ obtained by recording a reference signal $s(t) \in \mathbb{R}$ (Sec. III-C.1). The tilts $\psi_{1,k}, \cdots, \psi_{M,k}$ are estimated from M-channel 3-axis accelerometer measurements $\boldsymbol{a}_{1,k}, \cdots, \boldsymbol{a}_{M,k} \in \mathbb{R}^3$ (Sec. III-C.2). We assume the robot is stable when the acceleration is measured because we estimate the tilts of the modules from gravitational acceleration.

C. Feature Extraction

We estimate the robot posture by using TDOAs and tilts at the mic-acc modules calculated from the M-ch audio signal $\boldsymbol{z}_k(t)$ and the accelerometer measurements $\boldsymbol{a}_{1,k}, \cdots, \boldsymbol{a}_{M,k}$.

1) TDOA Estimation: For robustness in rubble-containing environments, we estimate the TDOAs and detect TDOAs of direct sound by excluding outliers. Since the TDOA between two adjacent microphones cannot be longer than the sound propagation time for the interval length on the robot ($2l$) in an open space, our method excludes the TDOA that does not satisfy this theorem. We formulate the set of indices $\mathcal{M}_k$ for microphones that record the direct sound of the reference signal as follows:

$$\mathcal{M}_k = \{m | m \text{ satisfies valid}(m)\} \tag{1}$$

$$\mathrm{valid}(m) = \begin{cases} \mathrm{valid}(m-1) \wedge |\tau^n_{m,m-1,k}| < \frac{2l}{c} & \text{if } m > n \\ |\tau^n_{n+1,n,k}| < \epsilon & \text{if } m = n \\ \mathrm{valid}(m+1) \wedge |\tau^n_{m+1,m,k}| < \frac{2l}{c} & \text{if } m < n \end{cases} \tag{2}$$

where c and ϵ represent the speed of sound in an open space and a threshold parameter for regarding the TDOA $\tau^n_{n+1,n,k}$ as small enough, respectively.

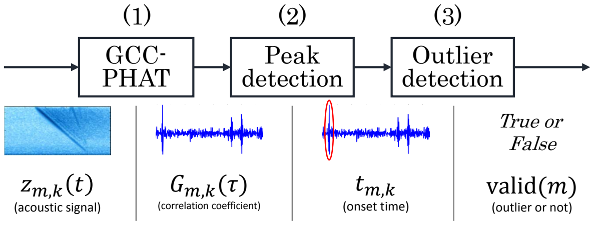


Fig. 4. Onset estimation for estimating TDOA.

A TDOA $\tau^n_{m_2,m_1,k}$ between mic_{m_1} and mic_{m_2} when src_n omits a reference signal is estimated from the difference of onset times at the microphones $t^n_{m_1,k}$ and $t^n_{m_2,k}$:

$$\tau^n_{m_2,m_1,k} = t^n_{m_2,k} - t^n_{m_1,k}. \tag{3}$$

The onset time of the m-th microphone $t^n_{m,k}$ is calculated by detecting the first peak of a correlation coefficient between the reference signal and recorded signal (Fig. 4). The correlation coefficient is calculated based on GCC-PHAT [21], which is robust against reverberation when the signal-to-noise ratio (SNR) is high [22]. To obtain a high-SNR reference signal from the environment noises, we use a time-stretched pulse (TSP) [23] as a reference signal. The TSP is a frequency-modulated sine wave that smoothly decreases the instantaneous frequency from the Nyquist frequency to zero Hz over time. As the frequency of a TSP is different every time, the SNR of a TSP becomes high. This feature also avoids spatial aliasing in the microphone array, and enables us to playback the signal with a small loudspeaker which has a limited frequency characteristics.

2) Tilt Estimation: We estimate the tilts at the accelerometers by measuring the direction of gravitational acceleration. The output of tilt estimation is a set of tilts $\psi_{m,k}(m = 1, \cdots, M)$ at the accelerometers acc_m. The tilt $\psi_{m,k}$ is estimated from the accelerometer measurements $\boldsymbol{a}_{m,k}$ as follows:

$$\psi_{m,k} = \arctan\left(-a^x_{m,k} \Big/ \sqrt{(a^y_{m,k})^2 + (a^z_{m,k})^2}\right) \tag{4}$$

where $a^x_{m,k}$, $a^y_{m,k}$, and $a^z_{m,k}$ represent the elements of the input acceleration $\boldsymbol{a}_{m,k}$, respectively.

D. State-Space Model of Robot Posture

Our method estimates the 3D posture of a hose-shaped robot by using TDOAs estimated using the microphones and tilts estimated from the accelerometers. More specifically, we formulate a state-space model that associates a state space representing the 3D robot posture with an observation space representing the TDOA and tilt of each mic-acc module (Fig. 5). The current posture is estimated by using the UKF.

As shown in Fig. 6, The robot posture is modeled as a serially-connected link model. A posture at the k-th measurement, $\boldsymbol{\xi}_k$, is defined as follows:

$$\boldsymbol{\xi}_k = [\theta_{1,k}, \cdots, \theta_{M+N-2,k}, \phi_{1,k}, \cdots, \phi_{M+N-1,k}, \\ l_{1,k}, \cdots, l_{M+N-1,k}]^{\mathrm{T}}, \tag{5}$$

where $\theta_{i,k}$, $\phi_{i,k}$, and $l_{i,k}$ are a horizontal link angle, a vertical link angle, and a link length, respectively.

The relative positions of the microphones and loudspeakers on the robot, $\boldsymbol{u}_{m,k}$ and $\boldsymbol{v}_{n,k}$, are calculated recursively from the first position $\boldsymbol{u}_{1,k}$. Suppose that $\boldsymbol{x}^*_{i,k}$ is the i-th

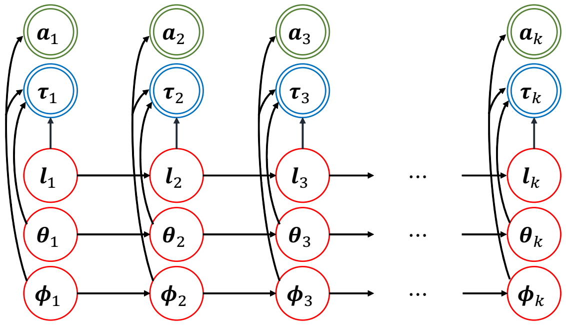


Fig. 5. Graphical representation of the proposed state-space model.

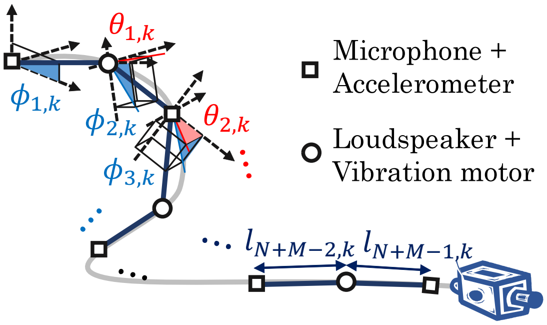


Fig. 6. 3D serially-connected link model of robot posture.

member of $[\boldsymbol{u}_{1,k}, \boldsymbol{v}_{1,k}, \cdots, \boldsymbol{u}_{M-1,k}, \boldsymbol{v}_{N,k}, \boldsymbol{u}_{M,k}]$. Then each position is given by

$$\boldsymbol{x}^*_{i,k} = \boldsymbol{x}^*_{i-1,k} + l_{i-1,k} \begin{bmatrix} \cos(\phi^*_{i,k})\cos(\theta^*_{i,k}) \\ \cos(\phi^*_{i,k})\sin(\theta^*_{i,k}) \\ \sin(\phi^*_{i,k}) \end{bmatrix}, \tag{6}$$

$$\phi^*_{i,k} = \sum_{j=1}^{i-1} \phi_{j,k}, \qquad \theta^*_{i,k} = \sum_{j=1}^{i-2} \theta_{j,k}. \tag{7}$$

1) State Update Model: The state update model $p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ is based on two concepts: a) posture dynamics and b) posture constraint. The posture dynamics $q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ is represented as random walk:

$$q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \mathcal{N}(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}, \boldsymbol{Q}_k), \tag{8}$$

where $\boldsymbol{Q}_k \in \mathbb{R}^{L\times L}$ is the covariance matrix of the process noise. The posture constraint $r(\boldsymbol{\xi}_k)$, on the other hand, is modeled as a Gaussian distribution:

$$r(\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{\xi}_k|\bar{\boldsymbol{\xi}}, \boldsymbol{P}), \tag{9}$$

where $\bar{\boldsymbol{\xi}} \in \mathbb{R}^L$ and $\boldsymbol{P} \in \mathbb{R}^{L\times L}$ are the mean and covariance matrix of the feasible posture.

These two distributions are integrated for the state update model $p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ on the basis of the product of experts [24]:

$$p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \frac{1}{A} q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) r(\boldsymbol{\xi}_k), \tag{10}$$

where $A = \int q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) r(\boldsymbol{\xi}_k) d\boldsymbol{\xi}_k$ is a normalization factor.

2) Measurement Model: The measurement model $p(\boldsymbol{\tau}_k, \boldsymbol{\psi}_k|\boldsymbol{\xi}_k)$ is formulated with two sub models: a) a TDOA measurement model $p(\boldsymbol{\tau}_k|\boldsymbol{\xi}_k)$ and b) a tilt measurement model $p(\boldsymbol{\psi}_k|\boldsymbol{\xi}_k)$ as follows:

$$p(\boldsymbol{\tau}_k, \boldsymbol{\psi}_k|\boldsymbol{\xi}_k) = q(\boldsymbol{\tau}_k|\boldsymbol{\xi}_k) r(\boldsymbol{\psi}_k|\boldsymbol{\xi}_k) \tag{11}$$

The TDOA measurement model $q(\boldsymbol{\tau}_k|\boldsymbol{\xi}_k)$ is defined using a set of TDOAs $\tau^{n_k}_{m_k,n_k,k}$ where the m_k is the one of the filtered microphone indices $\mathcal{M}_k$:

$$q(\boldsymbol{\tau}_k|\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{\tau}_k|[\tau^{n_k}_{m_1,n_k,k}(\boldsymbol{\xi}_k)|m_1 \in \mathcal{M}_k]^{\mathrm{T}}, \boldsymbol{R}^{\tau}_k), \tag{12}$$

where $\boldsymbol{R}^{\tau}_k$ represents the covariance matrix of the measurement noise and TDOA $\tau^n_{m_1,m_2,k}(\boldsymbol{\xi}_k)$ is calculated by using the distances between the two microphones and the loudspeaker as follows:

$$\tau^n_{m_1,m_2,k}(\boldsymbol{\xi}_k) = \frac{|\boldsymbol{u}_{m_2,k} - \boldsymbol{v}_{n,k}| - |\boldsymbol{u}_{m_1,k} - \boldsymbol{v}_{n,k}|}{c}, \tag{13}$$

where c represents the speed of sound.

The tilt measurement $\boldsymbol{\psi}_k$ is a set of tilts angles $\psi_{m,k}$ at mic-acc modules:

$$q(\boldsymbol{\psi}_k|\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{\psi}_k|[\psi_1(\boldsymbol{\xi}_k), \cdots, \psi_M(\boldsymbol{\xi}_k)]^{\mathrm{T}}, \boldsymbol{R}^{\psi}_k) \tag{14}$$

where $\boldsymbol{R}^{\psi}_k$ represents the covariance matrix of the measurement noise and tilt $\psi_m(\boldsymbol{\xi}_k)$ is calculated by accumulating the vertical link angles $\phi_{a,k}$ as follows:

$$\psi_m(\boldsymbol{\xi}_k) = \frac{1}{2}\sum_{i=1}^{2m-2} \phi_{i,k} + \frac{1}{2}\sum_{i=1}^{2m-1} \phi_{i,k} \tag{15}$$

IV. Evaluation

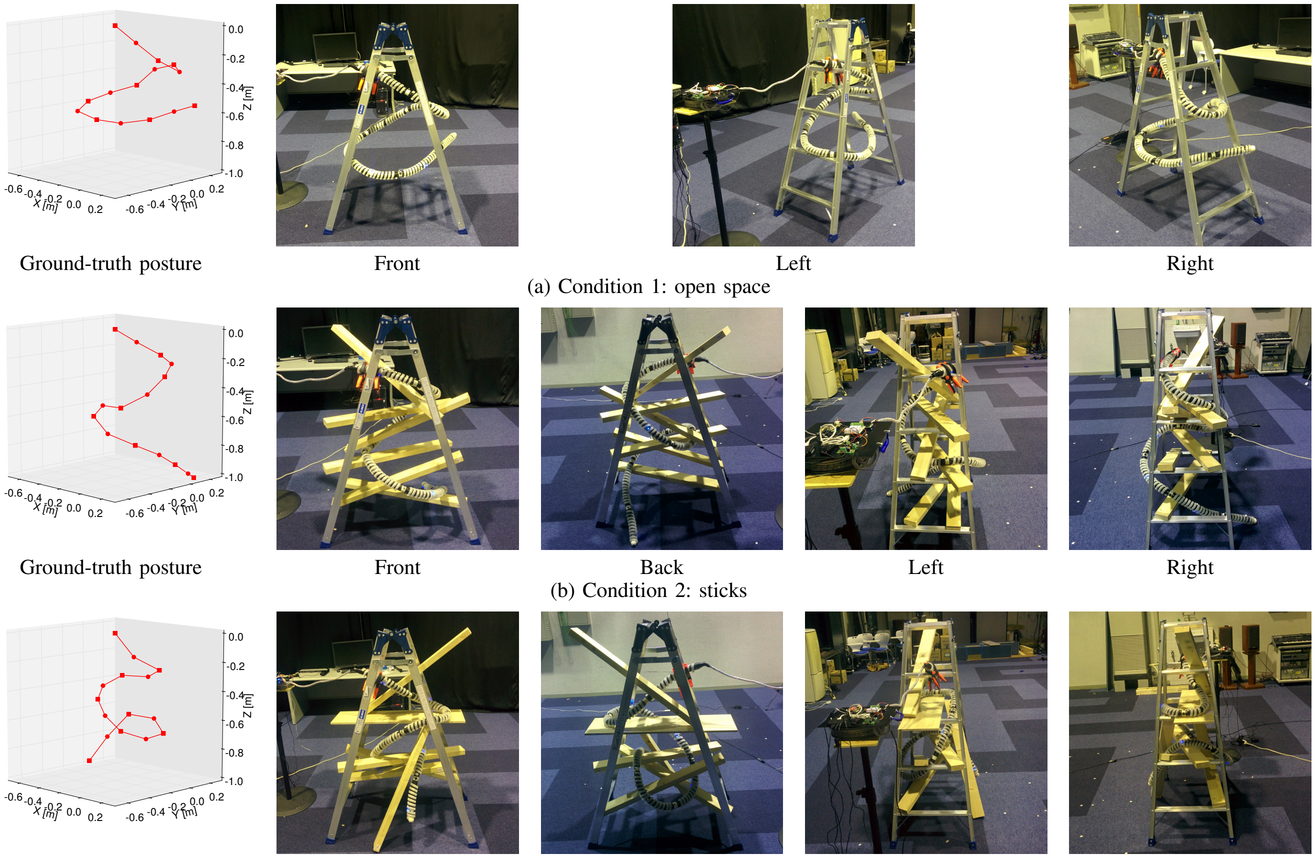
This section reports an experiment evaluating the proposed method of 3D posture estimation in rubble-containing environments.

A. Experimental Settings

We compared the proposed method integrating microphone and accelerometer information with a conventional method estimating the posture by using only microphone information. This experiment was conducted in an experimental room where the reverberation time RT$_{60}$ was 800 ms. As shown in Fig. 7, we estimated the robot postures in the following three conditions:

1) **Open space**: There was no rubble around the robot. The robot curved three-dimensionally on a stepladder 140 cm high.
2) **Sticks**: Six wooden sticks (91 cm × 9 cm × 4 cm) representing rubble were placed around the robot.
3) **Sticks and plate**: In another rubble-containing environment, the six wooden sticks and a wooden plate (91 cm × 25 cm × 1.5 cm) were placed around the robot.

We used a TSP reference signal that had a length of 8,192 samples (512 ms) at 16 kHz, and recorded with a synchronized A/D converter RASP-ZX (Systems In Frontier Corp.). The initial state $\boldsymbol{\xi}_0 = [\theta_{i,0}, \cdots, \phi_{i,0}, \cdots, l_{i,0}, \cdots]^{\mathrm{T}}$ of the UKF was determined in the following manner. The initial horizontal and vertical link angles $\theta_{i,0}$ and $\phi_{i,0}$ were sampled from a Gaussian distribution whose mean corresponded to the ground-truth posture and standard deviation was 6°. The link lengths $l_{i,0}$ were set to 0.2 m which was the distance between mic-acc and src-vib modules on the robot. The threshold of the TDOA estimation ϵ was set to $0.04/340$ sec. The other parameters were determined experimentally.



Ground-truth posture Front Left Right

(a) Condition 1: open space

Ground-truth posture Front Back Left Right

(b) Condition 2: sticks

Ground-truth posture Front Back Left Right

(c) Condition 3: sticks and plate

Fig. 7. Three conditions for experimental evaluation. Ground-truth postures were measured using a motion capture system.

We implemented the proposed method by using Python without multiprocessing. The estimation was conducted with a standard laptop computer with an Intel Core i7–3517U CPU (2-core, 1.9 GHz) and 4.0 GB of memory. The CPU time and elapsed time for the whole estimation algorithm with 50 measurements were 8.561 s and 9.129 s, respectively. These values were small enough compared with the whole signal length of the reference signals (25.6 s) that our method could work in real time.

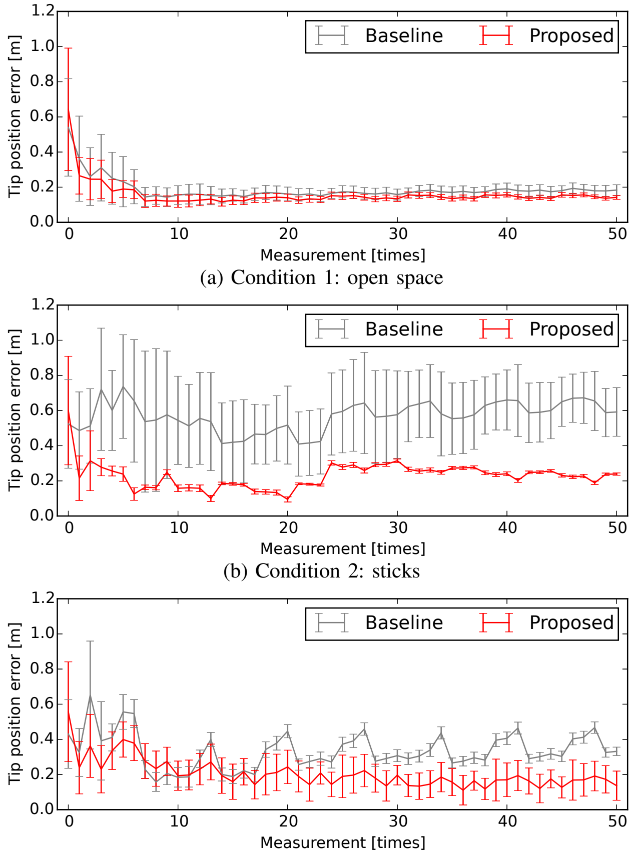
We evaluated the tip position error and average estimation error. The tip position error was the distance between the ground-truth and estimated positions of the tip module (8-th mic-acc module). The average estimation error was the average distance between the ground-truth and estimated positions of all the modules. The ground-truth position of each module was measured using a motion capture system (OptiTrack, NaturalPoint Inc.). The estimation errors were evaluated with 32 different initial states. Since the conventional sound-only method, which does not consider the tilt information, has rotation ambiguity at the x-axis of the 1-st mic-acc module, we rotated the estimated posture to make the average estimation error as small as possible.

B. Experimental Results

As shown in Figs. 8 and 9, in all conditions, the proposed method suppressed the tip position errors at the initial states to about 0.2 m and suppressed the average position errors there to less than 0.2 m. Moreover, when the robot was placed in rubble-containing environments (conditions 2 and 3), the baseline sound-based method failed to estimate the robot's posture. The proposed method, on the other hand, robustly suppressed the estimation errors

As shown in Fig. 10, in the all conditions, the postures estimated by the proposed method were close to the ground-truth posture, whereas when the robot was placed in condition 2 or 3, the first joint angle estimated by the conventional method was significantly different from the ground-truth posture. Both of the rubble-containing environments had a wooden stick in front of the joint place (2nd src-vib module) to prevent estimation of the robot posture. This shows that in the proposed method the lack of information at the joint was compensated by the information obtained from the accelerometers.

As shown in Fig. 11, when the errors of the initial state of the Kalman filter were larger than those in the other conditions, the estimation error became larger. In this condition the standard deviation of initial errors was set to 30 deg, whereas in the other conditions it was set to 6 deg. This shows that our method is sensitive to the initial state. This is because mirror symmetrical ambiguity could not solved even if both TDOA and tilt information were used. A promising solution to this problem is to predict

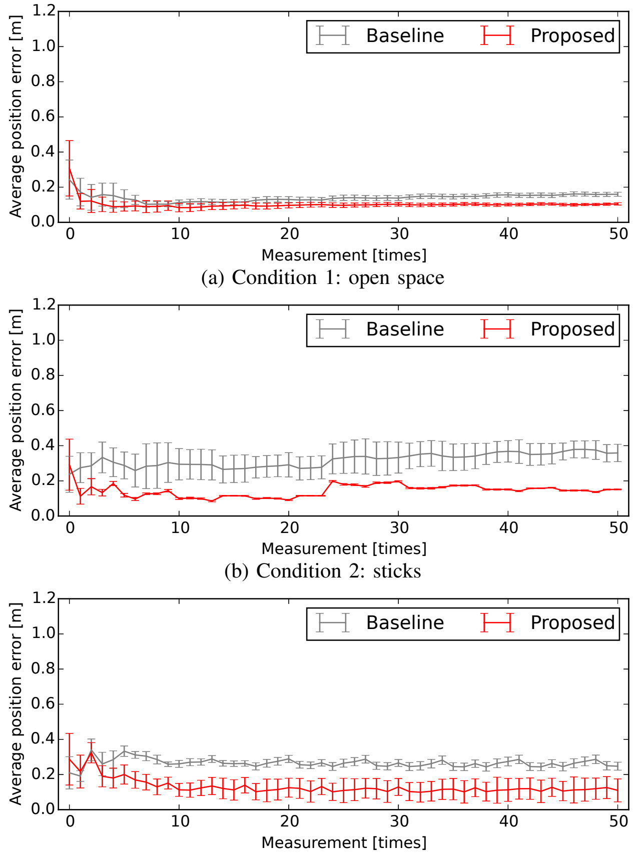


(a) Condition 1: open space

(b) Condition 2: sticks

(c) Condition 3: sticks and plate

Fig. 8. Tip position errors obtained by proposed and baseline methods in the three conditions. Polylines and error bars indicate the mean and standard deviation for 32 different initial states, respectively.



(a) Condition 1: open space

(b) Condition 2: sticks

(c) Condition 3: sticks and plate

Fig. 9. Average position errors obtained by proposed and baseline methods in three conditions. Polylines and error bars indicate the mean and standard deviation for 32 different initial states, respectively.

the time-varying posture of a moving robot in a dynamical manner. Since the posture at the initial insertion is given with a insertion-guide pipe [2], we can obtain the current posture by tracking the time-varying posture during the insertion. It was shown that a sound-based method can track the moving posture by considering the posture change rate [20]. Integration with sequential information obtained by accelerometers and gyrometers would also be beneficial for further improvement of 3D time-varying posture estimation.

V. Conclusion

This paper presented a 3D posture estimation method using microphones and accelerometers for a hose-shaped rescue robot. Since correct TDOAs are not always obtained at all microphones if a reference signal is blocked by some obstacles, our method incorporates tilt information obtained by the accelerometers for estimating a robot posture robustly in rubble-containing environments. We formulated a nonlinear state-space model that integrates TDOA and tilt information and used the unscented Kalman filter for posture estimation. Experiments using a 3 m hose-shaped robot with eight microphones and accelerometers and seven loudspeakers showed that our method successfully reduced the tip position errors of the initial states to about 0.2 m even when the robot was placed in rubble-containing environments.

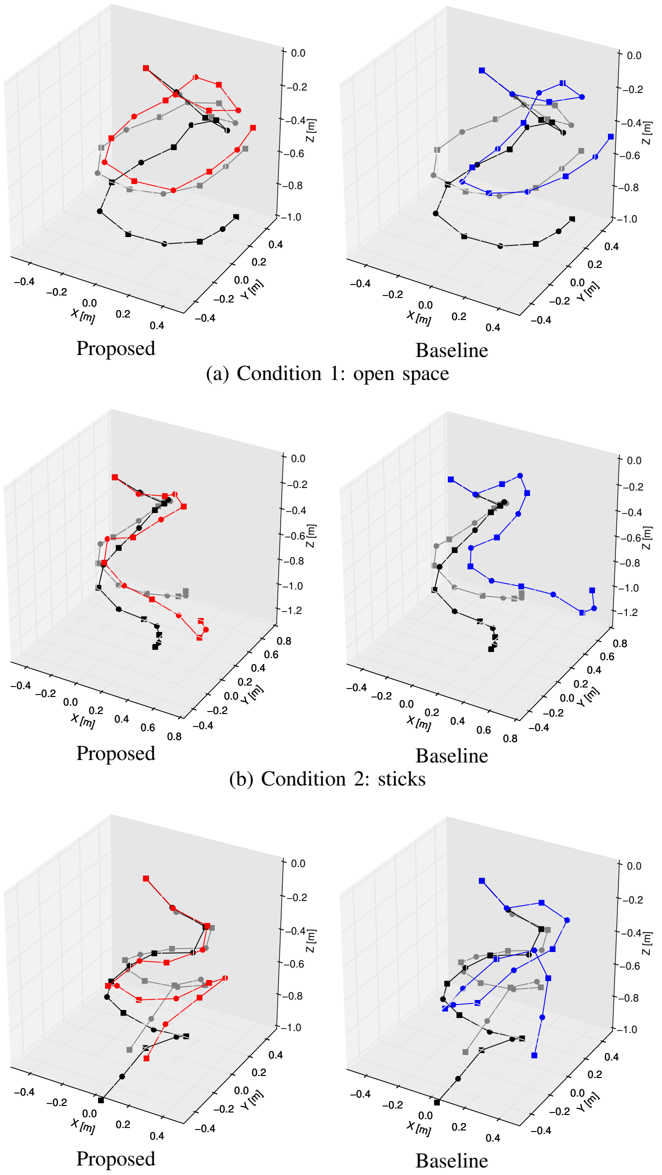
Although our method can work well to a certain extent in rubble-containing environments, further performance improvement would be feasible by adaptively changing a threshold that accepts only correct TDOAs. To estimate the 3D time-varying posture of a moving hose-shaped robot, we plan to integrate angular-velocity information obtained by gyrometers into a unified state-spate model. This approach could reduce the initial-state sensitivity of the method because the mirror symmetrical problem of posture estimation could be solved by focusing on dynamical change of the posture. Since the proposed state-space model was designed to maintain the dynamics of the posture, it is easy to integrate these sensors. The resulting method will be evaluated in a realistic disaster environment. A series of our studies will help the remote operator to freely manipulate a hose-shaped robot in an unseen rubble-containing environment.

Acknowledgements This study was partially supported by ImPACT Tough Robotics Challenge and by JSPS KAKENHI No. 24220006 and No. 15J08765.

References

[1] A. Kitagawa *et al.*, "Development of small diameter Active Hose-II for search and life-prolongation of victims under debris," *Journal of Robotics and Mechatronics*, vol. 15, no. 5, pp. 474–481, 2003.

[2] K. Hatazaki *et al.*, "Active Scope Camera for urban search and rescue," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 2596–2602.

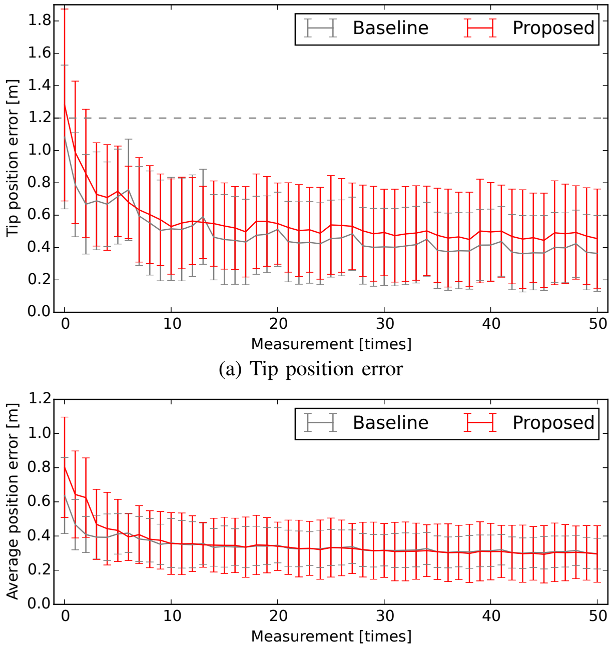


(a) Condition 1: open space

(b) Condition 2: sticks

(c) Condition 3: sticks and plate

Fig. 10. Examples of estimated postures at the 50-th measurement. Black and gray lines represent initial and ground-truth postures, respectively.



(a) Tip position error

(b) Average position error

Fig. 11. Tip and average position errors with larger errors of initial states, in condition 3. The standard deviation of initial errors was set to 30 deg (it was set to 6 deg in the other conditions).

[3] J. Fukuda *et al.*, "Remote vertical exploration by Active Scope Camera into collapsed buildings," in *IEEE/RSJ IROS*, 2014, pp. 1882–1888.

[4] S. Tadokoro *et al.*, "Application of Active Scope Camera to forensic investigation of construction accident," in *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2009, pp. 47–50.

[5] R. R. Murphy, *Disaster Robotics*. MIT Press, 2014.

[6] M. Ishikura *et al.*, "Vision-based localization using Active Scope Camera – accuracy evaluation for structure from motion in disaster environment," in *IEEE/SICE International Symposium on System Integration*, 2010, pp. 25–30.

[7] S. Weiss *et al.*, "Monocular-SLAM–based navigation for autonomous micro helicopters in GPS-denied environments," *Journal of Field Robotics*, vol. 28, no. 6, pp. 854–874, 2011.

[8] K. Schmid *et al.*, "Stereo vision based indoor/outdoor navigation for flying robots," in *IEEE/RSJ IROS*, 2013, pp. 3955–3962.

[9] S. Lynen *et al.*, "A robust and modular multi-sensor fusion approach applied to MAV navigation," in *IEEE/RSJ IROS*, 2013, pp. 3923–3929.

[10] M. Tailanian *et al.*, "Design and implementation of sensor data fusion for an autonomous quadrotor," in *IEEE International Instrumentation and Measurement Technology Conference*, 2014, pp. 1431–1436.

[11] J. M. Santos *et al.*, "A sensor fusion layer to cope with reduced visibility in SLAM," *Journal of Intelligent & Robotic Systems*, pp. 1–22, 2015.

[12] R. R. Murphy, "Navigational and Mission Usability in Rescue Robots," *Journal of the Robotics Society of Japan*, vol. 28, no. 2, pp. 142–146, 2010.

[13] J. Lee *et al.*, "Non visual sensor based shape perception method for gait control of flexible colonoscopy robot," in *IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 577–582.

[14] Y. Bando *et al.*, "Posture estimation of hose-shaped robot by using active microphone array," *Advanced Robotics*, vol. 29, no. 1, pp. 35–49, 2015.

[15] M. Ishikura *et al.*, "Shape estimation of flexible cable," in *IEEE/RSJ IROS*, 2012, pp. 2539–2546.

[16] E. A. Wan *et al.*, "The unscented kalman filter for nonlinear estimation," in *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[17] N. Ono *et al.*, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 161–164.

[18] M. Chen *et al.*, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *IEEE WASPAA*, 2007, pp. 22–25.

[19] H. Miura *et al.*, "SLAM-based online calibration for asynchronous microphone array," *Advanced Robotics*, vol. 26, no. 17, pp. 1941–1965, 2012.

[20] Y. Bando *et al.*, "A sound-based online method for estimating the time-varying posture of a hose-shaped robot," in *IEEE International Symposium on Safety, Security, and Rescue Robotics*, 2014, pp. 1–6.

[21] C. Knapp *et al.*, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[22] C. Zhang *et al.*, "Why does PHAT work well in lownoise, reverberative environments?" in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 2565–2568.

[23] Y. Suzuki *et al.*, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1119–1123, 1995.

[24] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.