

柔軟索状レスキューロボットのための マイクロホン・加速度センサアレイを用いた3次元姿勢推定

3D Posture Estimation for a Hose-shaped Rescue Robot
using a Microphone and Accelerometer Array

○坂東 宜昭 糸山克寿 (京大) 昆陽雅司 田所諭 (東北大)
中臺一博 (東工大/HRI) 吉井和佳 (京大) 奥乃博 (早大)

Yoshiaki bando¹, Katsutoshi Itoyama¹, Masashi Konyo², Satoshi Tadokoro²,
Kazuhiro Nakadai³, Kazuyoshi Yoshii¹, and Hiroshi G. Okuno⁴

¹Kyoto University, ²Tohoku University,

³Tokyo Institute of Technology / Honda Research Institute Japan, and ⁴Waseda University

This paper presents an online method that estimates a 3D posture of a hose-shaped rescue robot using a microphone and accelerometer array. Posture (shape) estimation of a self-driving hose-shaped rescue robot is crucial for handling the robot body because the unseen robot posture deforms in narrow spaces under collapsed buildings. Conventional sound-based method that uses time-differences of arrivals (TDOAs) works only on a two-dimensional surface and is often hampered by the rubble around the robot. Our method eliminates the outliers of sound-based TDOA measurements, and compensates the lack of the posture information with the tilt information measured by accelerometers. Experimental results using a 3-m hose-shaped robot that was deployed in a simple 3D structure demonstrate that our method reduces the errors of initial states to about 20 cm in the 3D space.

Key Words: Rescue robot, Hose-shaped robot, Posture estimation, Sensor fusion

1 はじめに

柔軟索状レスキューロボット [1, 2] は細長い形状が特徴のレスキューロボットで、人間や動物が侵入できない瓦礫の間隙などを検索するために開発されている。リポート・オペレータは本ロボットに装着された小型カメラやマイクロホンを用いて被災者を検索する。例えば、小型車輪が装着された Active-Hose II [1] や、本体を覆う織毛の振動で駆動する Active Scope Camera (ASC) [2] などが報告されている。特に、ASC はアメリカの実際の災害現場での適用例も報告されている [3]。

柔軟索状レスキューロボットだけでなく、一般にレスキューロボットに搭載されるセンサ・システムは実際の災害現場では本来の性能を発揮できないことが多い [4]。例えば、ビデオカメラなどの画像センサは瓦礫内では視野が狭くなり、照明によってホワイトアウトやブラックアウトが生じ、ロボット周囲の撮影に支障をきたす [2]。また自己位置の絶対座標が取得できる GPS は、屋内や瓦礫内では GPS 衛星の信号が阻害されるため、精度が大きく低下する [4]。災害現場で頑健に動作するセンサ・システムの開発には、異なる種類のセンサを組み合わせ、互いの欠点を補完することが重要である [5]。

瓦礫が存在する複雑な環境で、柔軟索状レスキューロボットを効率的に操縦するためには、本ロボットの3次元姿勢(形状)推定が不可欠である。従来、様々なセンサを用いた姿勢推定法が提案されてきたが、それぞれ単体の手法を災害現場で使用するには課題が残されていた [6-8]。例えば、地磁気センサと加速度センサを用いる従来法では、地磁気が瓦礫に含まれる鉄片に干渉されるため、性能が劣化する [6]。また、ジャイロセンサを用いる従来法は、ジャイロセンサのドリフトにより、長時間の運用で誤差が累積する問題があった [7]。一方、音を用いた姿勢推定法が提案されており、ロボットに装着されたマイクロホンと小型スピーカの周囲の空間のみに依存して姿勢を推定できる [8]。また、到達時間差情報は、現在のマイクロホンと小型スピーカの位置関係

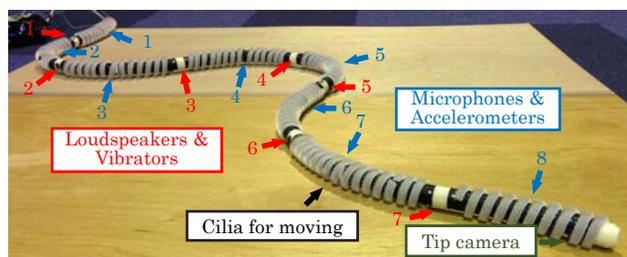


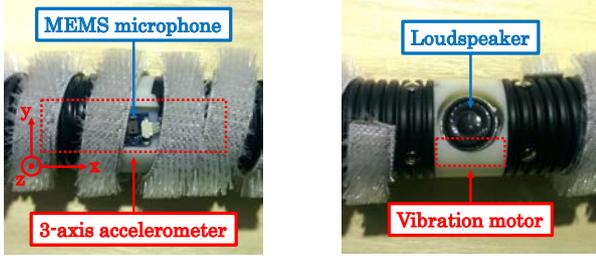
Fig.1 A hose-shaped rescue robot with an eight-channel microphone-accelerometer array.

を表しているため、累積誤差の問題を回避できる。しかし、本手法は周囲の瓦礫がマイクロホンとスピーカの間の音の伝搬を阻害する場合に性能が劣化する問題があった。

本研究では、柔軟索状レスキューロボットに複数のマイクロホン、スピーカ、および加速度センサを装着し、本ロボットの3次元姿勢推定を行う。ロボット上のマイクロホンと小型スピーカを用いて音の到達時間差を計測することで、これらの相対位置を推定できる [8-10]。また、加速度センサは重力加速度を計測することで各センサの傾き角度を外界に依存せず推定できる。到達時間差情報は、周囲の瓦礫によって性能が劣化することがあるが、到達時間差推定結果の外れ値を判定、除去し、これに伴う姿勢情報の欠如を傾き角度情報により補完する。

2 柔軟索状レスキューロボットの3次元姿勢推定

提案法は、一定時刻ごとに音の到達時間差と傾き角度を観測し、現在のロボットの姿勢を逐次的に推定する。音の到達時間差は信頼度推定を行い、信頼できるマイクロホンの結果のみを用いて姿勢を推定する。姿勢と信頼度の推定は、姿勢、観測および信



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

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

傾度の関係を表す状態空間モデルを定義し、switching Kalman filter [11, 12] によって逐次推論して行く。

2.1 柔軟索状レスキューロボット

図 1 に本稿で扱う柔軟索状レスキューロボットの写真を示す。本体は、直径 38 mm、全長 3 m のコルゲートチューブからなる。本ロボットは Fukuda ら [13] の Active Scope Camear と同じ駆動機構を搭載しており、ロボットの表面を覆う織毛を振動させて前進する。

本ロボットは、a) マイクロホンと加速度センサが装着されたモジュール (mic-acc モジュール、図 2-(a)) と、b) 小型スピーカと振動モータが装着されたモジュール (src-vib モジュール、図 2-(b)) の 2 つのモジュールを搭載する。図 3 に示すように、 $M = 8$ 個の mic-acc モジュールと $N = 7$ 個の src-vib モジュールが間隔 $l = 20$ cm で交互に配置されている。以下では、 k を観測の番号とし、 i 番目の mic-acc モジュールと src-vib モジュールの 3 次元座標をそれぞれ $\mathbf{u}_{ik} \in \mathbb{R}^3$ および $\mathbf{v}_{ik} \in \mathbb{R}^3$ と表す。

2.2 問題設定

本稿で扱う柔軟索状レスキューロボットの姿勢推定の問題設定を以下に示す。

入力	1) n_k 番目のスピーカから再生した試験音の各マイクロホンの到達時間差 $\tau_{n_k m_k}^k \in \mathbb{R}$ ($m = 1, \dots, M$)
	2) M 個の加速度センサで観測した傾き角度 $\psi_{m_k} \in \mathbb{R}$ ($m = 1, \dots, M$)
出力	ロボット上の mic-acc モジュールと src-vib モジュールの座標 $\mathbf{u}_{m_k} \in \mathbb{R}^3$ および $\mathbf{v}_{n_k} \in \mathbb{R}^3$

到達時間差 $\tau_{m_1 m_2}^k$ は、 n 番目のスピーカから再生した試験音の m_1 番目のマイクロホンと m_2 番目のマイクロホンの間での到達時間差を表す。試験音は、各観測ごとに順番にスピーカを選んで再生する。試験音には信号対雑音比を得やすい time-stretched pulse (TSP) [14] を使用し、到達時間差推定には残響に強い相互相関法である GCC-PHAT [15, 16] を用いる。傾き角度は、mic-acc モジュール内の 3 軸加速度センサを用いて次式で計測される。

$$\phi_{m_k} = \arctan \left(-\frac{a_{m_k}^x}{\sqrt{(a_{m_k}^y)^2 + (a_{m_k}^z)^2}} \right) \quad (1)$$

ここで、 $a_{m_k}^x$ 、 $a_{m_k}^y$ および $a_{m_k}^z$ はそれぞれ m 番目の加速度センサの x 、 y および z 軸の観測結果を表す。本稿では、傾き角度の計測中はロボットが静止していることを仮定する。

2.3 状態空間モデル

柔軟索状レスキューロボットの姿勢は、ロボット上の各モジュールを線分で結んだ折れ線モデルとして近似する。よって、 k 番目の観測を得たときの姿勢 ξ_k は、以下のように表す。

$$\xi_k = [\theta_{1k}, \dots, \theta_{(M+N-2)k}, \phi_{1k}, \dots, \phi_{(M+N-1)k}, l_{1k}, \dots, l_{(M+N-1)k}]^T \quad (2)$$

ここで、 θ_{ik} 、 ϕ_{ik} 、および l_{ik} はそれぞれ、各モジュール間の水平角度、垂直角度、および距離を表す。提案法で推定する潜在変

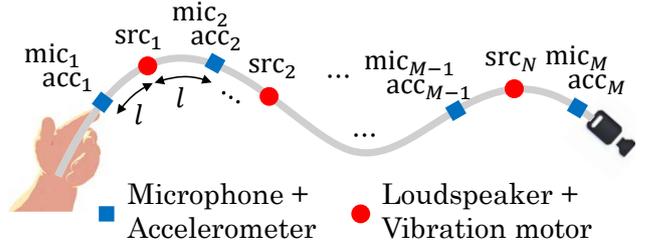


Fig.3 Arrangements of microphones, accelerometers, and loudspeakers on a hose-shaped robot.

数 \mathbf{z}_k は、ロボットの姿勢 ξ_k と、各マイクロホンの信頼度 $\mathbf{a}_k = [a_{1k}, \dots, a_{Mk}]^T \in \{0, 1\}^M$ とする。

$$\mathbf{z}_k = [\mathbf{z}_k, \mathbf{a}_k]^T \quad (3)$$

ロボット上の mic-acc モジュール及び src-vib モジュールの位置 \mathbf{u}_{mk} および \mathbf{v}_{nk} は、根本のモジュール位置 \mathbf{u}_{1k} とロボットの姿勢 ξ_k を用いて再帰的に計算できる。ここで、 \mathbf{x}_{ik}^* を $[\mathbf{u}_{1k}, \mathbf{v}_{1k}, \dots, \mathbf{v}_{N-1k}, \mathbf{u}_{M-1k}, \mathbf{v}_{Nk}, \mathbf{u}_{Mk}]$ の i 番目の要素とすると、各モジュールの相対位置は以下のように計算される。

$$\mathbf{x}_{ik}^* = \mathbf{x}_{i-1k}^* + l_{i-1k} \begin{bmatrix} \cos(\phi_{ik}^*) \cos(\theta_{ik}^*) \\ \cos(\phi_{ik}^*) \sin(\theta_{ik}^*) \\ \sin(\phi_{ik}^*) \end{bmatrix}, \quad (4)$$

$$\phi_{ik}^* = \sum_{j=1}^{i-1} \phi_{jk}, \quad \theta_{ik}^* = \sum_{j=1}^{i-2} \theta_{jk} \quad (5)$$

2.3.1 状態遷移モデル

本モデルでは、姿勢にはマルコフ性を仮定するが、マイクロホンの信頼度は各時刻で独立と仮定する。姿勢の状態遷移モデルは 1) 姿勢の運動モデルと 2) 姿勢の制約モデルの 2 つのモデルからなる。ロボットの運動モデル $q(\xi_k | \xi_{k-1})$ は直前の姿勢 ξ_{k-1} を用いて以下のように定義する。

$$q(\xi_k | \xi_{k-1}) = \mathcal{N}(\xi_k | \xi_{k-1}, \Sigma_k^\xi) \quad (6)$$

ここで、 $\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は、平均と分散パラメータを持つ多変量ガウス分布を表し、パラメータ Σ_k^ξ は、モデル誤差を表す共分散行列である。一方で、ロボットの姿勢の制約モデルはガウス分布を用いて以下のように表現する。

$$r(\xi_k) = \mathcal{N}(\xi_k | \xi, \Sigma^\xi) \quad (7)$$

ここで、 ξ および Σ^ξ はそれぞれ、姿勢の事前分布の平均および共分散行列を表す。これら 2 つのモデルは product of experts [17] に基づいて統合し、状態更新モデルを以下のように定義する。

$$p(\xi_k | \xi_{k-1}) \propto q(\xi_k | \xi_{k-1}) r(\xi_k) \quad (8)$$

マイクロホンの信頼度 \mathbf{a}_k には各時刻ごとに独立にベルヌーイ分布による事前分布を置く。

$$p(\mathbf{a}_k) = \prod_{m=1}^M \alpha^{a_{mk}} (1 - \alpha)^{1 - a_{mk}} \quad (9)$$

ここで、 α はベルヌーイ分布のハイパーパラメータである。

2.3.2 観測モデル

到達時間差 τ_k と傾き角度 ψ_k の観測モデル $p(\tau_k, \psi_k | \mathbf{z}_k)$ は、1) 到達時間差の観測モデル $p(\tau_k | \mathbf{z}_k)$ と 2) 傾き角度の観測モデル $p(\psi_k | \mathbf{z}_k)$ を用いて以下のように定義する。

$$p(\tau_k, \psi_k | \mathbf{z}_k) \propto p(\tau_k | \mathbf{z}_k) p(\psi_k | \mathbf{z}_k) \quad (10)$$

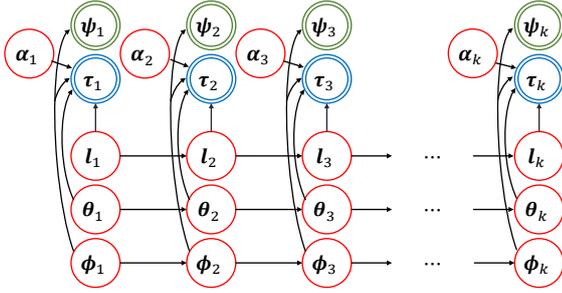


Fig.4 Graphical model of the proposed state-space model.

到達時間差の観測モデル $p(\tau_k|z_k)$ は、 k 番目の観測 τ_k に対して以下のように定義する。

$$p(\tau_k|z_k) \propto \prod_{m=1}^M \mathcal{N}(\tau_k | \tau_{n_k, m}(\xi_k), (\sigma^{\tau_1})^{a_{mk}} (\sigma^{\tau_2})^{1-a_{mk}}) \quad (11)$$

$$\tau_{n, m}(\xi_k) = \frac{|\mathbf{u}_{mk} - \mathbf{v}_{nk}| - |\mathbf{u}_{nk} - \mathbf{v}_{nk}|}{c} \quad (12)$$

ここで、 c は音速を表し、 \mathbf{u}_{mk} および \mathbf{v}_{nk} は式 4 で定義される mic-acc および src-vib モジュールの相対座標である。また、 σ^{τ_1} および σ^{τ_2} は到達時間差の観測ノイズの分散パラメータを表し、 σ^{τ_1} は対応するマイクロホンが信頼できるとき、 σ^{τ_2} は信頼できないときの観測誤差を表す。 σ^{τ_2} を大きく取る (e.g., 0.01s) ことで、到達時間差の外れ値を表現する。

傾き角度の観測モデル $p(\psi_k|z_k)$ は、以下のように定義する。

$$p(\psi_k|z_k) \propto \prod_{m=1}^M \mathcal{N}\left(\psi_k \middle| \sum_{i=1}^{2m-2} \phi_{ik} + \frac{1}{2}\phi_{(2m-1)k}, \sigma^\psi\right) \quad (13)$$

ここで、 σ^ψ は傾き角度の観測ノイズの分散パラメータを表す。

2.4 推論

提案法の状態空間モデル (図 4) は、switching Kalman filter [11] を用いて推論できる。ただし、現在の姿勢の予測分布 $p(\xi_k|\tau_{1:k-1}, \psi_{1:k-1})$ は、1 時刻前の推定結果 $p(\xi_{k-1}|\tau_{1:k-1}, \psi_{1:k-1}) = \mathcal{N}(\xi_{k-1}|\hat{\xi}_{k-1}, \hat{\Sigma}_{k-1}^\xi)$ を用いて以下のように計算する。

$$p(\xi_k|\tau_{1:k-1}, \psi_{1:k-1}) \propto \int q(\xi_k|\xi_{k-1})r(\xi_k)p(\xi_{k-1}|\tau_{1:k-1}, \psi_{1:k-1})d\xi_{k-1} \quad (14)$$

$$= \mathcal{N}(\xi_k|\bar{\xi}_k, \bar{\Sigma}_k^\xi) \quad (15)$$

$$\bar{\xi}_k = \bar{\Sigma}_k^\xi((\Sigma')^{-1}\mathbf{F}\hat{\xi}_{k-1} + (\Sigma^\xi)^{-1}\xi) \quad (16)$$

$$\bar{\Sigma}_k^\xi = ((\Sigma')^{-1} + (\Sigma^\xi)^{-1})^{-1} \quad (17)$$

$$\Sigma' = (\mathbf{F}^T(\hat{\Sigma}_{k-1}^\xi)^{-1}\mathbf{F} + \Sigma_k^\xi) \quad (18)$$

ここで、行列 \mathbf{F} は、運動モデルの遷移関数 (式 6) を行列に置き換えたものである。また、観測の予測分布 $p(\tau_k, \psi_k|\tau_{1:k-1}, \psi_{1:k-1}, \alpha_k)$ は、観測モデルに非線形変換を含むので、Unscented 変換 [12] を用いて近似計算を行う。

3 評価実験

提案法による 3 次元姿勢推定を、模擬瓦礫環境下で収録した実録音データを用いて評価した。

3.1 実験設定

図 5 に示すように、以下の 2 つの条件で姿勢推定を行った。

条件 1) 自由空間: 高さ 140 cm の脚立を用いて障害物のない自由空間にロボットを 3 次元的に配置した。

条件 2) 瓦礫環境: 6 本の木製の棒 (91 cm × 9 cm × 1.5 cm) を用いた模擬瓦礫環境内にロボットを配置した。

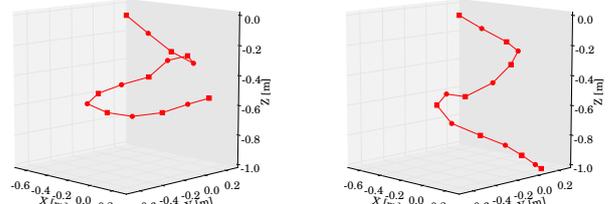


(a) Condition 1



(b) Condition 2

Fig.5 The conditions for experimental evaluation.



(a) Condition 1

(b) Condition 2

Fig.6 Ground-truth postures measured using a motion capture system.

本実験は、残響時間 (RT₆₀) が 800 ms の実験室にて行った。到達時間差推定の試験音には長さ 8,192 サンプルの TSP を用い、8 つのマイクロホンを同期して録音した。録音は同期 A/D 変換器 RASP-ZX (株式会社システムインフロンティア) を用いて 16 kHz, 24-bit サンプリングで行った。

Switching Kalman filter に与える姿勢の初期値 $\xi_0 = [\theta_{i0}, \dots, \phi_{i0}, \dots, l_{i0}, \dots]^T$ は以下のように決定した。各リンクの水平角度 θ_{i0} および垂直角度 ϕ_{i0} は平均が正解、標準偏差が 6° の正規分布からサンプリングした。また、各リンク長 l_{i0} は、ロボット上の mic-acc モジュールと src-vib モジュール間の距離である 0.2 m を与えた。本実験では、8 個の初期値に対して姿勢推定を行い、その平均と分散を評価した。観測の標準偏差 σ^{τ_1} , σ^{τ_2} , および σ^ψ はそれぞれ、 1.0×10^{-4} s, 1.0×10^{-2} s, および 20° に設定した。これらの値は実験的に決定した。

評価尺度には、平均位置誤差と先端位置誤差を用いた。平均位置誤差は全てのモジュールの位置誤差の平均を表す。また、先端位置誤差は、ロボットの根本の mic-acc モジュールの座標 \mathbf{u}_{1k} を正解に一致させた時の先端の mic-acc モジュールの座標 \mathbf{u}_{Mk} の位置誤差を表す。各モジュールの正解座標はモーションキャプチャシステム (Optitrack, NaturalPoint Inc.) を用いて計測した (図 6)。提案法 (TDOA+Tilt+SKF) は、1) 音の到達時間差のみを用いて姿勢推定する場合 (Only-TDOA) および、2) 到達時間差と傾き角度は用いるが、到達時間差の閾値処理で外れ値除去を行う場合 (TDOA+Tilt) と比較した。

3.2 実験結果

図 7-(a), 8-(a) に示すように、提案法 (TDOA+Tilt+SKF) は全ての条件で初期値の誤差を抑圧し平均位置誤差 0.2 m 以下で姿勢を推定できた。一方で、音の到達時間差のみを用いる手法 (Only-TDOA) や、閾値処理で到達時間差を棄却する手法 (TDOA+Tilt) は、障害物の存在する条件 2 で条件 1 より平均位置誤差が増加している。提案法はこれらの手法より高い精度で姿勢を推定できており、最大 57% 性能が向上している。特に、音の到達時間差のみで姿勢を推定した場合 (Only-TDOA)、誤差が 0.2 m を超えており、到達時間差と傾き角度情報を統合する提案法の有効性を確認できる。

積分型姿勢推定法であるジャイロセンサを用いる従来法は、35 秒程度の動作で 0.2 m 程度に先端位置誤差が増加する [7]。一方提案法は、図 7-(b), 8-(b) に示すように、初期値に誤差が含まれていても先端位置を 0.2 程度の精度で推定できており、十分な精度が得られた。また、図 7-(c), 8-(c) に提案法による推定結果の一例を示す。特に条件 2 において、ベースライン法は先端付近の推定誤差が大きいが、提案法は正しく推定できている。

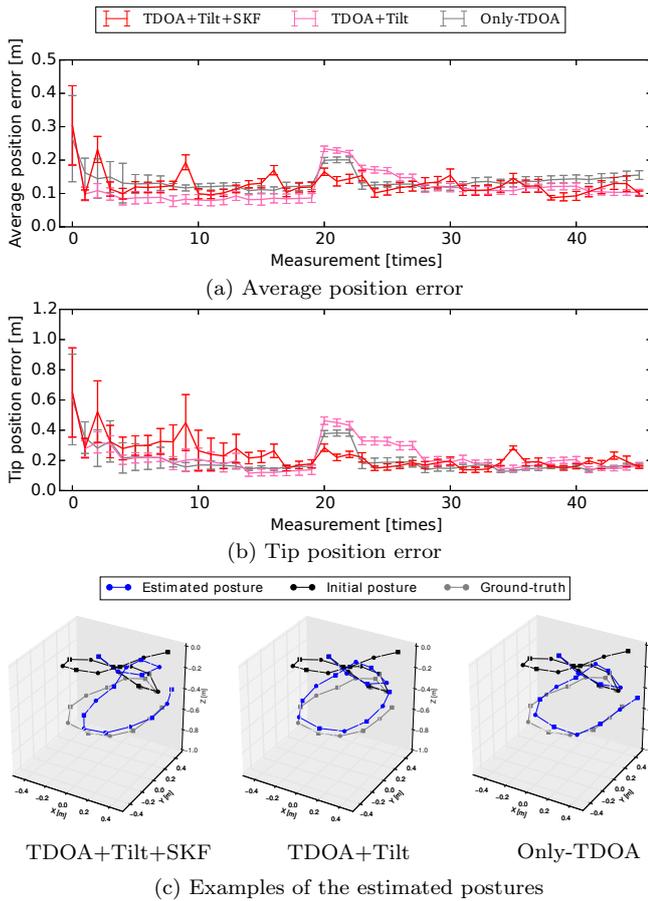


Fig.7 Estimation results of Condition 1.

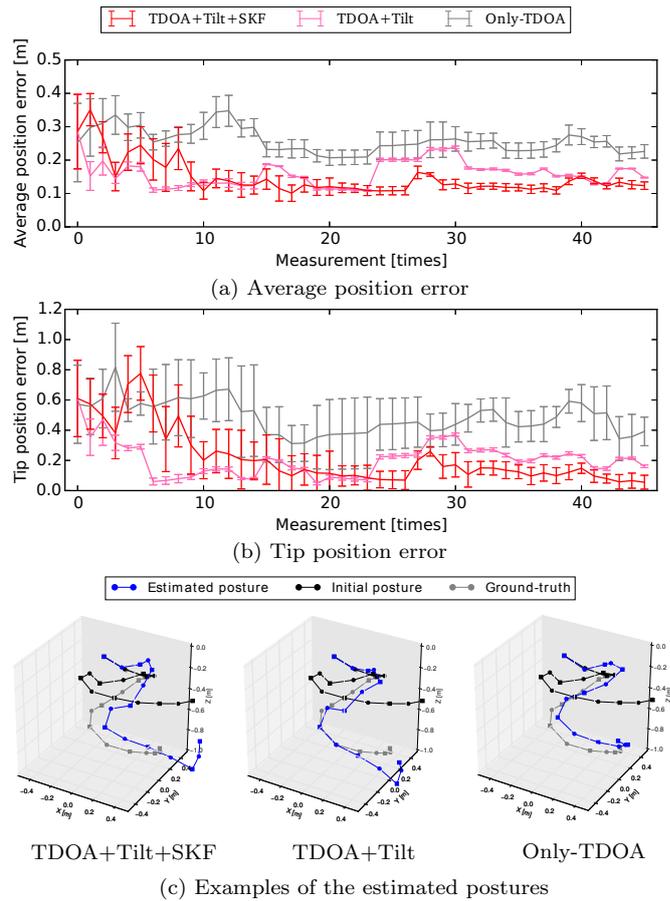


Fig.8 Estimation results of Condition 2.

4 おわりに

本稿では、柔軟索状レスキューロボットに複数のマイクロホン、スピーカ、および加速度センサを装着し、本ロボットの3次元姿勢推定を行った。提案法は、マイクロホンとスピーカから得られる音の到達時間差情報と、加速度センサから得られる各センサの傾き角度情報を統合して姿勢を推定する。音の到達時間差情報は、瓦礫など障害物によって性能が劣化する場合があるが、到達時間差推定結果の外れ値を判定、除去して対処する。さらに、不足した姿勢情報を、加速度センサから得られる傾き角度情報で補完して姿勢を推定する。外れ値判定、各センサ情報の統一と姿勢の推定は単一の状態空間モデルを switching Kalman filter を用いて推論を行う。3mの柔軟索状レスキューロボットを模擬瓦礫環境内に配置して収録した実録音信号による評価実験を行い、障害物存在下でも初期値の誤差を平均0.2m以下に抑圧できることを確認した。今後は、ジャイロセンサとの統合を行い、動的な姿勢の追跡を行う。音の到達時間差情報は、試験音を再生する時間が必要なので、時間分解能が高々数Hz程度である。動的な姿勢の変動を数百Hz以上の時間分解能で観測できるジャイロセンサと統合することで、駆動しているロボットの高精度な姿勢推定を目指す。

謝辞 本研究は、科研費基盤(S) No.24220006, 特別研究員奨励費 No. 15J08765, および ImPACT「タフ・ロボティクス・チャレンジ」の支援を受けた。

参考文献

- [1] A. Kitagawa *et al.* Development of small diameter Active Hose-II for search and life-prolongation of victims under debris. *JRM*, 15(5):474–481, 2003.
- [2] H. Namari *et al.* Tube-type Active Scope Camera with high mobility and practical functionality. In *IEEE/RSJ IROS*, 2012, 3679–3686.

- [3] S. Tadokoro *et al.* Application of Active Scope Camera to forensic investigation of construction accident. In *IEEE ARSO*, 2009, 47–50.
- [4] R. R. Murphy. *Disaster Robotics*. MIT Press, 2014.
- [5] R. R. Murphy. Navigational and Mission Usability in Rescue Robots. *JRSJ*, 28(2):142–146, 2010.
- [6] J. Lee *et al.* Non visual sensor based shape perception method for gait control of flexible colonoscopy robot. In *IEEE ICRB*, 2011, 577–582.
- [7] M. Ishikura *et al.* Shape estimation of flexible cable. In *IEEE/RSJ IROS*, 2012, 2539–2546.
- [8] Y. Bando *et al.* A sound-based online method for estimating the time-varying posture of a hose-shaped robot. In *IEEE SSRR*, 2014, 1–6.
- [9] H. Miura *et al.* SLAM-based online calibration for asynchronous microphone array. *AR*, 26(17):1941–1965, 2012.
- [10] N. Ono *et al.* Blind alignment of asynchronously recorded signals for distributed microphone array. In *IEEE WASPAA*, 2009, 161–164.
- [11] C.-J. Kim. Dynamic linear models with markov-switching. *JE*, 60(1-2):1–22, 1994.
- [12] E. A. Wan *et al.* The unscented kalman filter for nonlinear estimation. In *IEEE ASSPCCS*, 2000, 153–158.
- [13] J. Fukuda *et al.* Remote vertical exploration by Active Scope Camera into collapsed buildings. In *IEEE/RSJ IROS*, 2014, 1882–1888.
- [14] Y. Suzuki *et al.* An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *JASA*, 97(2):1119–1123, 1995.
- [15] C. Knapp *et al.* The generalized correlation method for estimation of time delay. *IEEE TASP*, 24(4):320–327, 1976.
- [16] C. Zhang *et al.* Why does PHAT work well in lownoise, reverberative environments? In *IEEE ICASSP*, 2008, 2565–2568.
- [17] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *NC*, 14(8):1771–1800, 2002.