

# マイクロホンアレイを用いた 駆動機構付ホース型ロボットの姿勢推定

○坂東 宜昭<sup>1</sup> 糸山 克寿<sup>1</sup> 昆陽雅司<sup>2</sup> 田所 諭<sup>2</sup> 中臺 一博<sup>3</sup> 吉井 和佳<sup>1</sup> 奥乃 博<sup>4</sup>  
<sup>1</sup> 京都大学 大学院情報学研究科 <sup>2</sup> 東北大学 大学院情報科学研究科  
<sup>3</sup> (株)ホンダ・リサーチ・インスティテュート・ジャパン <sup>4</sup> 早稲田大学 理工学術院

## 1. はじめに

ホース型ロボット [1-3] とは細長い形状が特徴のレスキューロボットで、瓦礫の隙間に挿入し先端のカメラを用いて被災者の捜索を行う。これまでに、タイミングベルトと小型車輪を用いた Active-Hose [1] や、繊維の振動を用いた Active Scope Camera [2, 3] などが開発されている。

柔軟に形状変化するホース型ロボットの姿勢制御では、リアルタイム姿勢推定が不可欠である。既存の境界センサによる姿勢推定法 [4] は加速度センサとジャイロセンサの値を積分して姿勢を推定する積分型計測法である。そのため、長時間の運用では累積誤差が蓄積するという問題があった。

本研究では稼働時間に依存しない非積分型計測法の姿勢推定手法を開発する。ホース型ロボットにマイクロホンアレイと小型スピーカを装着し、小型スピーカから再生する試験音のマイクロホン間の到達時間差を手がかりとした姿勢推定を行う。到達時間差は現在のマイクロホンと音源の位置だけに依存するので、累積誤差問題を回避できる。さらに、マイクロホンアレイを使用して音源定位・分離に応用できるので、被災者の位置推定や音声強調も可能となる。

これまで我々は、ロボットは静止しているものと仮定し、音を用いた姿勢推定法を開発してきた [5]。しかし、駆動機構を有するホース型ロボットでは、姿勢変動の追跡時に精度が低下するという問題があった。本稿ではこの問題を解決するため、姿勢の変化速度推定を行えるよう手法を拡張する。さらに、駆動機構を搭載したプロトタイプ・ホース型ロボットを用いて、実際にロボットを駆動させながら姿勢推定精度を評価する。

## 2. 姿勢推定の問題設定と概要

図2に示すように、本稿ではホース型ロボットに、マイクロホンと小型スピーカを交互に間隔  $l$  だけ離して装着する。各マイクロホンと小型スピーカはそれぞれ手元から順に  $\text{mic}_1, \text{src}_1, \dots, \text{src}_N, \text{mic}_M$  とする。ここで、 $M, N$  はそれぞれマイクロホンと小型スピーカの個数を表し、 $N = M - 1$  である。 $\text{mic}_m, \text{src}_n$  の各座標を、それぞれ  $\mathbf{u}_{m,k}, \mathbf{v}_{n,k} \in \mathbb{R}^2$  とする。 $k$  は観測のインデックスである。

本稿で扱う姿勢推定は次の手続きに沿って逐次実行される。まず 1) 各小型スピーカから順に試験音を再生し、2) 1回の試験音の録音から試験音の各マイクロホンへの到達時間差を推定する。さらに、3) 得られた到達時間差から姿勢を表すマイクロホンと小型スピーカの

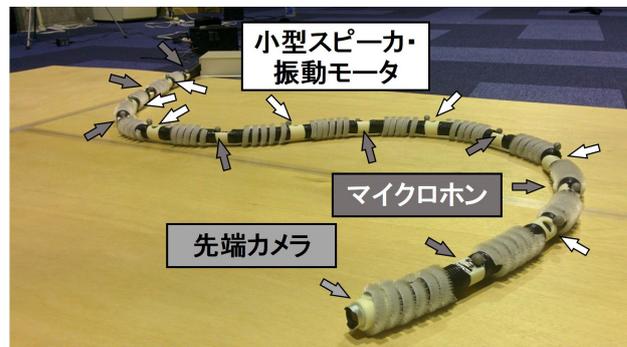


図1 駆動機構を有しマイクロホンと小型スピーカを搭載したプロトタイプ・ホース型ロボット

座標  $\mathbf{u}_{m,k}, \mathbf{v}_{n,k}$  を推定する。

以下に本稿で扱う問題設定を示す。

入力:  $\text{src}_n$  で再生した長さ  $G$  の試験音  $S(t) \in \mathbb{R}^G$  を録音した  $M$  チャンネル音響信号  $\mathbf{Z}_k(t) \in \mathbb{R}^{M \times G}$   
 出力: ロボット上のマイクロホンと小型スピーカの座標  $\mathbf{u}_{m,k}, \mathbf{v}_{n,k}$   
 仮定: 手元側のマイクロホンと小型スピーカの座標  $\mathbf{u}_{1,k}, \mathbf{v}_{1,k}$  は既知

ただし、 $t$  は時間を表し、試験音とは到達時間差推定のために小型スピーカで再生する元信号である。到達時間差はマイクロホンと小型スピーカの相対的な位置関係を表すので、出力はロボット上のマイクロホンと小型スピーカの相対位置  $\mathbf{u}_{m,k}, \mathbf{v}_{n,k}$  である。また、仮定として  $\mathbf{u}_{1,k}, \mathbf{v}_{1,k}$  を既知としても一般性を失わない。

マイクロホンと音源の位置を同時推定する関連研究に、補助関数法による手法 [6] と EKF-SLAM による手法 [7] がある。前者はオフラインで動作することを前提としておりロボットの姿勢推定には適さない。後者はオンライン手法だが運動モデルが既知の単一音源を仮定している。瓦礫中では音源の移動は困難であり、静止した 1 音源ではマイク位置推定できないため、本問題には適さない。提案法はホース上に配置された複数個の音源を仮定し、ロボットの姿勢を表す状態空間モデルを用いたオンライン推定を実現する。

## 3. 到達時間差を用いた姿勢推定

提案法は、得られた試験音の到達時間差から、姿勢モデルの事後確率をガウス分布で近似し Unscented Kalman Filter(UKF) [8] を用いて点推定する。

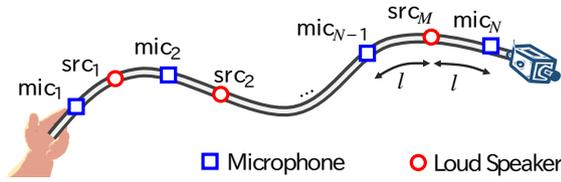


図2 ロボット上のマイクロホンと小型スピーカ配置

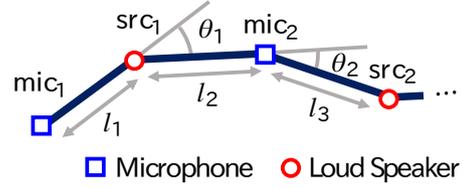


図3 姿勢のモデル

### 3.1 姿勢と観測のモデル

図3に示すように、ホース型ロボットの姿勢  $\zeta_k$  は、隣り合うマイクロホンと小型スピーカ間を線分で結んだリンクモデルで近似する。よって姿勢  $\zeta_k$  は、各ノード間の角度  $\theta_{i,k}$  ( $1 \leq i \leq N + M - 2$ ) とノード間距離  $l_{j,k}$  ( $1 \leq j \leq N + M - 1$ ) の  $2N + 3M - 3$  次元ベクトルで表現する。

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

推定する状態変数  $\xi_k$  は、ロボットの運動に対応するために姿勢  $\zeta_k$  とその速度ベクトル  $\dot{\zeta}_k$  を含む  $L = 4N + 4M - 6$  次元のベクトルとする。

$$\xi_k = [\zeta_k^T, \dot{\zeta}_k^T]^T \quad (2)$$

ただし  $\dot{\theta}_{i,k}$ ,  $\dot{l}_{j,k}$  はそれぞれ  $\theta_{i,k}$ ,  $l_{j,k}$  の変化速度を表す。各マイクロホンと小型スピーカの座標  $\mathbf{u}_{m,k}$ ,  $\mathbf{v}_{n,k}$  は  $\mathbf{x}_{i,k}$  を  $[\mathbf{u}_{1,k}, \mathbf{v}_{1,k}, \dots, \mathbf{u}_{M-1,k}, \mathbf{v}_{N,k}, \mathbf{u}_{M,k}]$  の  $i$  番目の要素として次のように計算される。

$$\mathbf{x}_{i,k} = \mathbf{x}_{i-1,k} + l_{i,k} \times [\cos(\sum_{i'=1}^{i-1} \theta_{i',k}), \sin(\sum_{i'=1}^{i-1} \theta_{i',k})] \quad (3)$$

**観測モデル** 観測モデル  $p(\mathbf{y}_k | \xi_k)$  は、小型スピーカ  $\text{src}_n$  が再生時のマイクロホン  $\text{mic}_{m_1}$ ,  $\text{mic}_{m_2}$  間の到達時間差  $\tau_{m_1 \rightarrow m_2}^n(\xi_k)$  を用いて次のように表現する。

$$p(\mathbf{y}_k | \xi_k) = \mathcal{N}(\mathbf{y}_k | \mathbf{T}(\xi_k), \mathbf{R}_k) \quad (4)$$

$$\mathbf{T}(\xi_k) = [\tau_{n \rightarrow n+1}^n(\xi_k), \dots, \tau_{n \rightarrow n-1}^n(\xi_k), \tau_{n \rightarrow n+1}^n(\xi_k), \dots, \tau_{n \rightarrow M}^n(\xi_k)]^T \quad (5)$$

ただし  $\mathbf{y}_k \in \mathbb{R}^{N-1}$ ,  $\mathbf{R}_k \in \mathbb{R}^{N-1 \times N-1}$  はそれぞれ、観測ベクトルと観測誤差を表す共分散行列である。

到達時間差  $\tau_{m_1 \rightarrow m_2}^n(\xi_k)$  はロボット上のマイクロホンと小型スピーカの座標から次のように定義する。

$$\tau_{m_1 \rightarrow m_2}^n(\xi_k) = \frac{D^{n,m_2}(\xi_k) - D^{n,m_1}(\xi_k)}{C} \quad (6)$$

ここで  $D^{n,m}(\xi_k)$  は  $\text{src}_n$  と  $\text{mic}_m$  間の距離を表し、 $C$  は音速を表す。

**状態更新モデル** 状態更新モデルは、1) ロボットの移動を表す遷移モデルと 2) 状態変数の制約を表す事前分布から構成する。遷移モデル  $q(\xi_k | \xi_{k-1})$  では、1時刻前の姿勢  $\zeta_{k-1}$  とその変化速度  $\dot{\zeta}_{k-1}$  から現在の姿勢を予測する：

$$q(\xi_k | \xi_{k-1}) = \mathcal{N}(\xi_k | [\zeta_{k-1} + \dot{\zeta}_{k-1}, \dot{\zeta}_{k-1}]^T, \mathbf{Q}_k) \quad (7)$$

ただし  $\mathbf{Q}_k \in \mathbb{R}^{L \times L}$  はモデル誤差を表す共分散行列である。状態変数への制約  $r(\xi_k)$  はガウス分布で表現する：

$$r(\xi_k) = \mathcal{N}(\xi_k | \xi, \mathbf{P}) \quad (8)$$

ここで、 $\xi \in \mathbb{R}^L$  と  $\mathbf{P} \in \mathbb{R}^{L \times L}$  はそれぞれ姿勢の制約を表すガウス分布の平均と共分散行列である。これらを Product of Experts (PoE) [9] として統合し、状態更新モデルを以下のように記述する：

$$p(\xi_k | \xi_{k-1}) = \frac{1}{A} q(\xi_k | \xi_{k-1}) r(\xi_k) \quad (9)$$

ただし、 $A$  は  $A = \int p(\xi_k | \xi_{k-1}) d\xi_k$  となる正規化パラメータである。

### 3.2 推論アルゴリズム

3.1章の姿勢モデルの推定には、状態変数の事後分布がガウス分布に従うと仮定して UKF を用いる。UKF では、 $k$  回目までの観測  $\mathbf{y}_{1:k}$  が得られた時の状態変数の事後分布  $p(\xi_k | \mathbf{y}_{1:k})$  を推定するために、状態変数の予測分布  $p(\xi_k | \mathbf{y}_{1:k-1}) = \int p(\xi_k | \xi_{k-1}) p(\xi_{k-1} | \mathbf{y}_{1:k-1}) d\xi_{k-1}$  を Unscented 変換 [8] により近似する。

提案法の状態更新モデルのうち、遷移モデルはガウス分布の線形変換で、制約を表す事前分布はガウス分布で定義される。よって、状態の予測分布  $p(\xi_k | \mathbf{y}_{1:k-1})$  は Unscented 変換を行わずに、1時刻前の事後分布  $p(\xi_{k-1} | \mathbf{y}_{1:k-1})$  の平均  $\hat{\xi}_{k-1}$  と分散  $\hat{\mathbf{P}}_{k-1}$  を用いて、次のように計算を簡略化できる。

$$p(\xi_k | \mathbf{y}_{1:k-1}) = \mathcal{N}(\xi_k | \xi_k^-, \mathbf{P}_k^-) \quad (10)$$

$$\xi_k^- = \mathbf{P}_k^- ((\mathbf{P}_k^*)^{-1} \hat{\xi}_{k-1} + \mathbf{P}^{-1} \xi) \quad (11)$$

$$\mathbf{P}_k^- = ((\mathbf{P}_k^*)^{-1} + \mathbf{P}^{-1})^{-1} \quad (12)$$

$$\mathbf{P}_k^* = \mathbf{F}^T \hat{\mathbf{P}}_{k-1} \mathbf{F} + \mathbf{P}_k \quad (13)$$

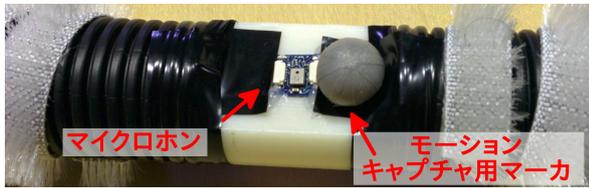
$\mathbf{F} \in \mathbb{R}^{L \times L}$  は遷移関数を行列に置き換えたものである。

## 4. 実験

駆動機構を有するプロトタイプホース型ロボットを用いて、駆動しているロボットの姿勢を推定し提案法を評価した。

### 4.1 プロトタイプ・ホース型ロボット

図1に、駆動機構を搭載したプロトタイプ・ホース型ロボットの写真を示す。本体は、直径38mmのコルゲートチューブからなり、全長3mである。本ロボットは、Namari らの Tube-type Active Scope Camera [3] と同じ織毛と振動モータを用いた駆動で前進する。振動モータはロボット内に40cm間隔で7つ直列に装着



(a) マイクロホン



(b) 小型スピーカ

図4 プロトタイプホース型ロボットに装着されたマイクロホンと小型スピーカ

されている。また、 $M = 8$ つのマイクロホン(図4(a))と $N = 7$ つの小型スピーカ(図4(b))がロボット表面に $l = 20$ cm 間隔で交互に装着されている。両端のマイク間の距離は2.8mである。

#### 4.2 実験条件

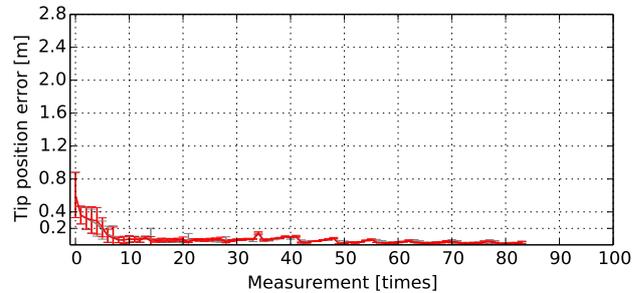
姿勢変化速度推定を行う提案法と、行わない手法を比較した。残響時間( $RT_{60}$ )が800msの実験室で、C字、S字、直線の三種の姿勢を初期姿勢としてロボットを駆動させながら姿勢推定し、その精度を評価した。試験音には信号対雑音比が良好な試験音であるTSP信号[10](512ms)を用い、各小型スピーカから順に再生し録音した。録音はロボット聴覚ソフトウェアHARK[11]を用いて8ch同期、24bit量子化、16kHzサンプリングで行った。到達時間差推定には残響に頑健なGCC-PHAT[12]を用いた。姿勢推定結果の計算には、CPUがi7-3517U(コア数:2, 1.9GHz)でメモリが4.0GBのノートパソコンを用い、リアルタイムに動作することを確認した。

状態変数の初期値 $\xi_0 = [\zeta_0, \dot{\zeta}_0]$ は、姿勢 $\zeta_0$ には正解データを平均として分散15degの正規分布からサンプリングし、姿勢の変化速度 $\dot{\zeta}_0$ は正解を与えた。制約パラメータの平均 $\xi$ は、 $l_{n,k}$ が20cmで直線となる姿勢が静止している状態とした。その他パラメータは実験的に与えた。

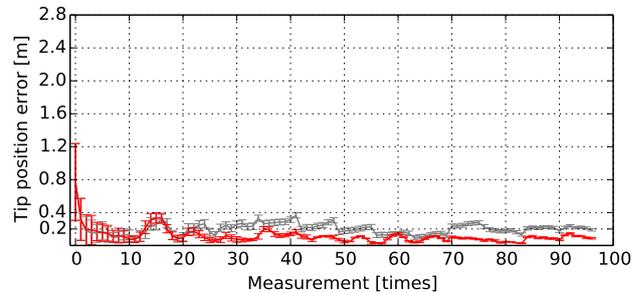
32種の異なる初期値を用いて姿勢推定し、その先端位置誤差を評価した。先端位置誤差は、正解データと推定姿勢の手元側のマイクロホンと小型スピーカの座標 $u_{1,k}, v_{1,k}$ を一致させたときの、先端のマイクロホン $mic_M$ の位置誤差である。マイクロホンと小型スピーカの正解位置はモーションキャプチャシステム(OptiTrack, NaturalPoint Inc.)を用いて計測した。

#### 4.3 実験結果

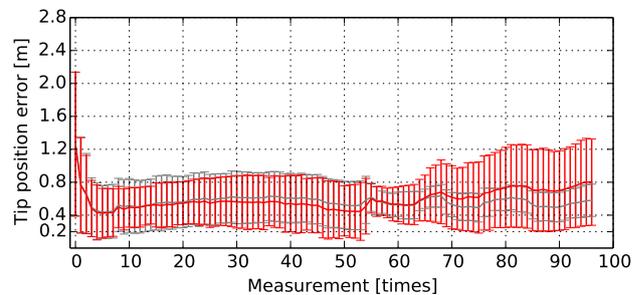
図5に示すように、C字型とS字型では、先端位置誤差が観測20回目以降で抑制され一定値以下になっている。また、姿勢の推定結果の一例(図6, 7)からも、動的に変動する姿勢を推定結果が追従している。特にS字型では提案法が姿勢の変化速度予測を行わない手



(a) C字



(b) S字



(c) 直線

図5 プロトタイプロボットを用いた姿勢推定結果の先端位置誤差。赤が動作予測あり、灰が動作予測なし。先端位置誤差の平均と標準偏差をそれぞれ折れ線とエラーバーを示す。

法と比較し、23回目の観測以降で先端位置誤差をより強く抑制しており、提案法の有効性を確認できる。

直線型では、7回目程度の観測までは先端位置誤差が低下しているが、その後80cmまで上昇している。これは、直線状の姿勢ではC字やS字状の姿勢と比較し姿勢の変動に対する到達時間差の変化が小さいため、正解の湾曲した姿勢から離れ、制約パラメータの平均 $\xi$ である直線状の姿勢に収束したためと考えられる(図8)。この問題は、内界センサや曲げセンサといった他のセンサと統合したマルチモーダル姿勢推定により解決できると考えている。

#### 5. まとめ

駆動機構を有するホース型ロボットの姿勢推定を行うため、姿勢の変化速度を推定する音を用いた姿勢推定法を開発し、その有効性を評価した。駆動機構を有しマイクロホンと小型スピーカを搭載した3mのプロトタイプ・ホース型ロボットを用いて実験を行い、動的な姿勢の変動が生じても初期値の誤差を一定値以下に抑圧できることを確認した。特に、S字型の姿勢を

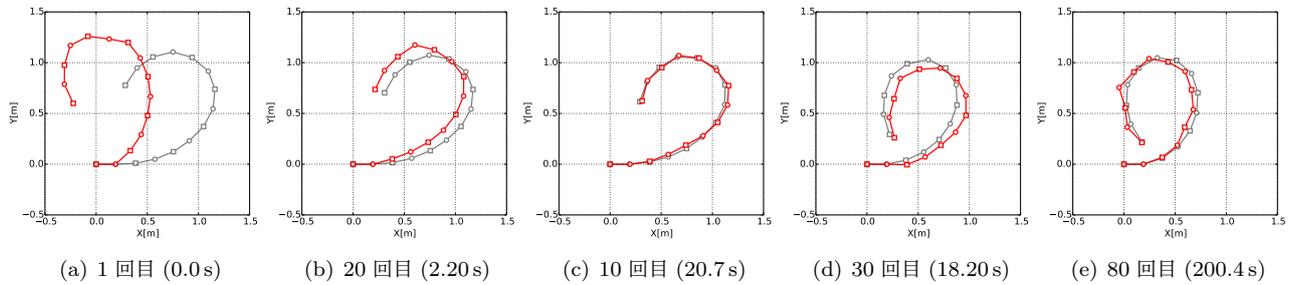


図6 C字型の初期姿勢での各観測時の姿勢推定結果. 赤が推定姿勢を灰色が正解を示す. ()内は推定開始からの時刻を表す.

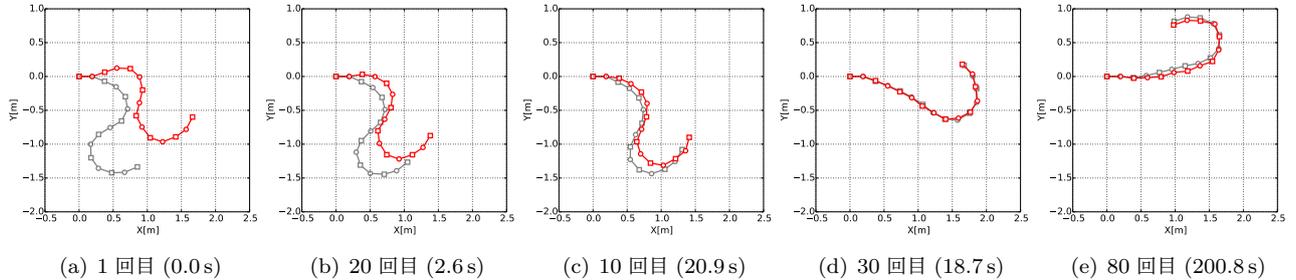


図7 S字型の初期姿勢での各観測時の姿勢推定結果. 赤が推定姿勢を灰色が正解を示す. ()内は推定開始からの時刻を表す.

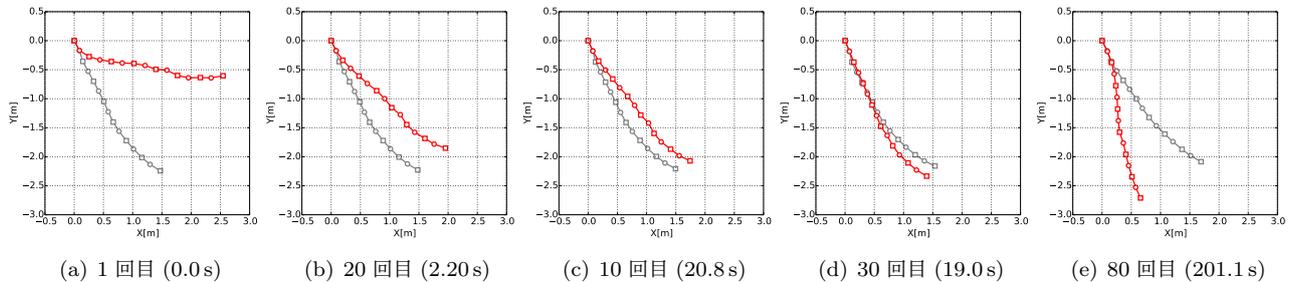


図8 直線型の初期姿勢での各観測時の姿勢推定結果. 赤が推定姿勢を灰色が正解を示す. ()内は推定開始からの時刻を表す.

初期姿勢とした場合、姿勢の変化速度を推定しない手法より先端位置誤差を強く抑制し、最大50.3%誤差を抑制できることを確認した。一方で、直線状の姿勢など姿勢変動に対する到達時間差の変化が小さい姿勢では他の姿勢より大きな誤差が生じることを確認した。

今後はプロトタイプ・ロボットに内界センサを搭載し、マルチモーダル化による精度向上を行う。また、模擬環境による評価を行い、搜索効率化の観点から音を用いた姿勢推定法の有効性を確認する。

謝辞 本研究は科研費基盤(S) No.24220006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. *Journal of Robotics and Mech.*, Vol. 15, No. 5, pp. 474–481, 2003.
- [2] K. Hatazaki, et al. Active scope camera for urban search and rescue. In *IEEE/RSJ IROS 2007*, pp. 2596–2602.
- [3] H. Namari, et al. Tube-type active scope camera with high mobility and practical functionality. In *IEEE/RSJ IROS 2012*, pp. 3679–3686.
- [4] M. Ishikura, et al. Shape estimation of flexible cable. In *IEEE/RSJ IROS 2012*, pp. 2539–2546.
- [5] 坂東宜昭ほか. マイクロホンアレイの位置推定によるホース型ロボットの姿勢推定. 情報処理学会第76回全国大会, No. 3, pp. 189–190, 2013.
- [6] N. Ono, et al. Blind alignment of asynchronously recorded signals for distributed microphone array. In *WASPAA 2009*, pp. 161–164.
- [7] H. Miura, et al. Slam-based online calibration of asynchronous microphone array for robot audition. In *IEEE/RSJ IROS 2011*, pp. 524–529, 2011.
- [8] S. Julier, et al. New extension of the kalman filter to nonlinear systems. In *AeroSense 1997*, pp. 182–193.
- [9] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, Vol. 14, No. 8, pp. 1771–1800, 2002.
- [10] 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, p. 1119, 1995.
- [11] K. Nakadai, et al. Design and implementation of robot audition system HARK—open source software for listening to three simultaneous speakers. *Advanced Robotics*, Vol. 24, No. 5-6, pp. 739–761, 2010.
- [12] C. Zhang, et al. Why does that work well in lownoise, reverberative environments? In *IEEE ICASSP 2008*, pp. 2565–2568.