

音源到来方向・時間差を用いた 非同期複数マイクロホンアレイ位置のオンライン推定

関口 航平¹ 中村 圭佑² 坂東 宜昭¹ 糸山 克寿¹ 吉井 和佳¹ 中臺 一博²

¹京都大学 大学院情報学研究科 知能情報学専攻 ²(株) ホンダ・リサーチ・インスティテュート・ジャパン

1. はじめに

未知環境でロボットの自律的な行動を可能とするために、周辺地図と自己位置の同時推定を行う Simultaneous Localization And Mapping (SLAM) の研究が行われている。例えば、複数台のロボット間で情報を共有することによる効率的な SLAM の研究 [1] が取り組まれている。本稿では、周辺の音環境理解を行うために、音響信号を用いた SLAM を、マイクロホンアレイを搭載した複数台のロボットを用いて行う。音響信号を用いれば遮蔽物によるオクルージョン問題を解決できるものの [2]、マイクロホン位置推定の従来法では、全てのマイクロホンの同期が必要 [3]、オフライン [4]、音源の移動方向の指定が必要 [5] といった課題があった。本稿では、観測音のロボット間での到達時間差、振幅比、各ロボットのマイクロホンアレイから得られる音源到来方向を用いて、SLAM と同期時刻ずれの推定をオンラインで同時に行うことによりこれらの問題の解決を図る。実録音を用いた実験により提案法の有効性を示した。

2. 非同期複数マイクロホンアレイ位置・同期時刻ずれのオンライン推定

M チャンネルマイクロホンアレイを搭載した R 台のロボットが静止し、移動する音源が 1 つ存在する状況において、二次元平面上での各ロボットの位置・向き $(r_i^x, r_i^y, r_i^\theta, i = 1, \dots, R)$ 、音源の位置 (s^x, s^y) 、ロボット 1 を基準としたときのロボット j の同期時刻ずれ $(\xi_j, j = 2, \dots, R)$ を推定する。状態の推定には、各ロボットから見た音源方向、観測音のロボット間の振幅比・到達時間差を用いる。ただし、どの観測も部屋の残響や反射の影響によりノイズを含んでいる。状態空間モデルを用いてモデル化し、FastSLAM [6] を用いることで、ノイズに頑健な状態の推定を行う。

入力	M チャンネルマイクロホンアレイを搭載した R 台のロボットでの観測音
出力	(1) ロボット i の状態 $(r_i^x, r_i^y, r_i^\theta)$ (2) 音源の位置 (s^x, s^y) (3) ロボット 1 と j の同期時刻ずれ ξ_j
仮定	各ロボットは静止、音源数は 1 つで移動音源

1 音源を対象としたマイクロホンの位置・同期時刻ずれ推定の関連研究には graph-SLAM [4] による手法と Extended-Kalman-Filter-SLAM による手法 [5] が挙げられる。これらの研究では、非同期な複数のマイクロホンを用いて、各マイクロホンの位置とマイクロホン間の同期時刻ずれを推定しているという点で本研究と異なる。また、前者はオフラインの手法であり、後者は音源の移動方向の指定が必要となる。提案法では、同期マイクロホンアレイを各ロボットに搭載することにより、ロボット間での同期のみを行えばよく、また各ロボットで音源

Online Estimation of Asynchronous Multiple Microphone Array Positions Using Directions and Time Differences of Arrivals: Kouhei Sekiguchi, Keisuke Nakamura, Yoshiaki Bando, Katsutoshi Itoyama, Kazuyoshi Yoshii, Kazuhiro Nakadai

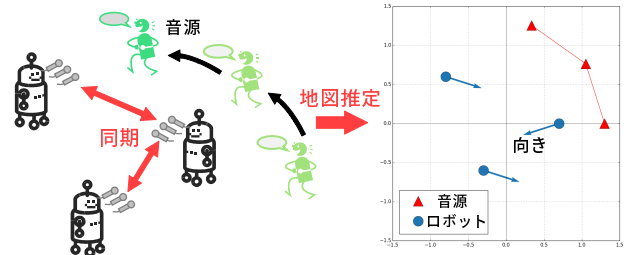


図 1: ロボット・音源位置とロボット間の同期時刻ずれの同時推定システム

到来方向を推定できるため、音源に対する制約が少なくオンラインで推定することができる。

2.1 観測

潜在状態の推定に用いる観測について説明する。
音源到来方向: 各ロボットでそれぞれ同期録音した観測音に対して Multiple Signal Classification (MUSIC) 法 [7] を用いることで、 1° 単位で音源到来方向の推定を行う。ただし、得られる角度はロボットの正面を基準とした角度である。
振幅比: 振幅は音源とマイクロホン間の距離に反比例するため、音源の振幅が分かれば音源までの距離を推定できる。音源の振幅は実際には分からないため、他ロボットの観測音との振幅比を取ることで距離比を得る。振幅比は観測音の 1 フレームの平均値を用いる。
到達時間差: ロボットと音源の位置によってロボットごとに音の到達時間が異なることから、マイクロホンアレイ間の時刻が同期できている場合、到達時間差からロボット間での音源との距離差を得ることができる。実際には完全に同期しておらず、同期時刻ずれがあるため、到達時間差は距離差と同期時刻ずれの情報の混合となっている。到達時間差の推定には雑音に頑健な相関関数である Generalized Cross Correlation with Phase Transform (GCC-PHAT) [8] を用いる。

2.2 状態空間モデル

状態空間モデルの潜在状態 z_t は、出力するロボットの位置・向き、音源位置、同期時刻ずれに加えて、音源の移動先を予測するために音源の速度と進行方向 (s^v, s^θ) を追加した $4R + 3$ 次元のベクトルとして定義する。
状態遷移モデル: ロボットは静止しているため、状態遷移は音源についてのみ扱う。各状態がガウス分布に従うと仮定すると、状態遷移は次の式で表される。

$$s_{k+1} \sim \mathcal{N}(s_k + [s_k^v \cos(s_k^\theta) \Delta t, s_k^v \sin(s_k^\theta) \Delta t, 0, 0]^T, Q)$$

ここで s_k は時刻 k での音源状態、 Δt は 1 つ前の観測からの経過時間、 Q はモデル誤差を表す共分散行列である。
観測モデル: 観測は各ロボットから見た音源到来方向 $\phi_{i,k}$ ($i = 1, \dots, R$)、観測音の振幅比 $\alpha_{j,k}$ ($j = 2, \dots, R$)、到達時間差 $\tau_{j,k}$ であり、各観測がガウス分布に従って分布すると仮定すると、観測モデルは以下のように表される。



図 2: 実験環境

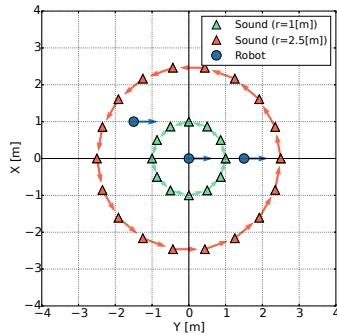


図 3: ロボットと音源の正解位置

$$\phi_{i,k} = \mathcal{N} \left(\arctan \left(\frac{s_k^y - r_{i,k}^y}{s_k^x - r_{i,k}^x} \right) - r_{i,k}^\theta, \sigma_\phi^2 \right)$$

$$\alpha_{j,k} = \mathcal{N} (l_{j,k}/l_{1,k}, \sigma_\alpha^2)$$

$$\tau_{j,k} = \mathcal{N} ((l_{j,k} - l_{1,k})/C, \sigma_\tau^2)$$

ここで, $\sigma_\phi^2, \sigma_\alpha^2, \sigma_\tau^2$ は各観測の分散パラメータ, $l_{i,k}$ は音源とロボット i の距離を表し, $l_{i,k} = \sqrt{(s_k^x - r_{i,k}^x)^2 + (s_k^y - r_{i,k}^y)^2}$ である.

2.3 状態推定アルゴリズム

MUSIC法で音源到来方向が推定される度に振幅比と到達時間差を計算し, FastSLAMを用いて潜在状態をオンライン推定する. 一般的なSLAMでのロボット位置が音源状態, 環境地図がロボット状態に対応している. FastSLAMでは, 時刻 k までの実際の観測 $y_{1:k}$ を用いて, 潜在状態 z_k の事後確率 $p(z_k|y_{1:k})$ をサンプル集合 $D_k = \{d_k^{[1]}, \dots, d_k^{[N]}\}$ で近似する. 時刻 $k-1$ での各サンプル $d_{k-1}^{[i]}$ について, 状態遷移モデルに基づいて音源状態 $s_k^{[i]}$ をサンプリングし, ロボット状態・同期時刻ずれを $s_k^{[i]}$ を用いてEKFにより更新する. 各サンプルについて重みの計算を行い, 重みに基づいてリサンプリングを行うことで時刻 k でのサンプル集合 D_k を決定する.

3. 評価実験

実験条件 8chマイクロホンアレイを搭載したロボットを3台配置し(図2), 半径 r [m] の円周上を人(音源)が手を叩きながら移動した場合での提案法の推定結果を評価した. 音源は半径1.0[m]のとき30°間隔で, 半径2.5[m]のとき20°間隔で25回音を出力した(図3). 音源到来方向の推定に失敗したロボットが存在する場合には状態の更新を行わないものとした. ロボット1を座標の中心とし, ロボット2がx軸の正に来るようにx軸を定め, ロボットは全て0°方向(x軸の正の方向)に向けた. 録音は24bit量子化, 16kHzサンプリングで行った. 非同期で録音すると正解の同期時刻ずれがわからないため, 各ロボットを有線で接続することにより同期録音を行った後, ロボット2, 3の観測音にそれぞれ10, -5[ms]のずれを加えた. 実験室の残響時間は $RT_{60} = 800$ [ms] である. 音源到来方向が推定される度に状態の更新を行い, ロボットと音源の位置, ロボットの向き, 同期時刻ずれについてそれぞれ推定結果と正解との誤差の平均を評価した. 実験結果 図4(a)に音源が半径1.0[m]の円周上を動いたときの結果を, 図4(b)に半径2.5[m]のときの結果を示

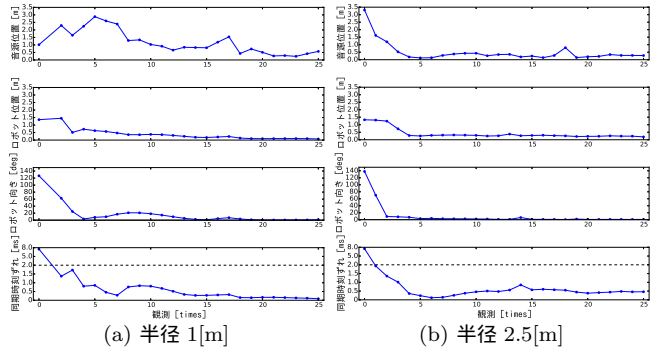


図 4: 各推定結果の観測ごとの誤差

す. 上から音源位置, ロボット位置, ロボットの向き, 同期時刻ずれの誤差を表しており, 0回目の値は初期値の誤差を示している. 25回音を出力した後の推定精度は, 前者では音源位置が0.27[m], ロボット位置が0.19[m], 向きが1.2°, 同期時刻ずれが0.46[ms] (7.3サンプル), 後者では音源位置が0.57[m], ロボット位置が0.07[m], 向きが1.3°, 同期時刻ずれが0.088[ms] (1.4サンプル)となった. どちらも数サンプルのずれで同期できており提案法が動作していることが確認できた. 半径1.0[m]のとき, 音源・ロボット位置の精度は高いが, 同期時刻ずれの精度が低いのは, 図2のような音源とロボットの高さの違いが, 音源がロボットに近い場合に影響するためだと考えられる. 提案法を三次元に拡張することにより解決できると期待できる.

4. まとめ

本稿では, 1音源が移動する状況で, マイクロホンアレイを搭載した複数台のロボットでの観測音から音源到来方向, 振幅比, 到達時間差を求め, ロボットの位置・向き, 音源位置, 同期時刻ずれの推定をFastSLAMを用いて同時に行う手法について述べた. 800[ms]程度の残響時間がある環境で提案法が動作することを確認した. 問題の簡単化のために音源数が1つであるという仮定を置いたが, 今後は複数音源が存在する場合でのSLAMと同期時刻ずれの推定に取り組む.

謝辞 本研究の一部は, JSPS 科研費 24220006, JST ImPACT TRC の支援を受けた.

参考文献

- [1] S. Saeedi *et al.* Neural network-based multiple robot simultaneous localization and mapping. *IEEE Trans. Neural Networks*, 22:2376–2387, 2011.
- [2] Y. Bando *et al.* Microphone-accelerometer based 3d posture estimation for a hose-shaped rescue robot. In *IEEE/RSJ IROS*, 5580–5586, 2015.
- [3] T. K. Le and N. Ono. Reference-distance estimation approach for TDOA-based source and sensor localization. In *IEEE ICASSP*, 2549–2553, 2015.
- [4] D. Su *et al.* Simultaneous asynchronous microphone array calibration and sound source localisation. In *IEEE/RSJ IROS*, 5561–5567, 2015.
- [5] H. Miura *et al.* SLAM-based online calibration of asynchronous microphone array for robot audition. In *IEEE/RSJ IROS*, 524–529, 2011.
- [6] M. Montemerlo *et al.* FastSLAM: A factored solution to simultaneous mapping and localization. In *AAAI*, 2002.
- [7] R. Schmidt *et al.* Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas and Propagation*, 34(3):276–280, 1986.
- [8] C. Zhang *et al.* Why does phat work well in lownoise, reverberative environments? In *IEEE ICASSP*, 2565–2568, 2008.