5R-06

室内環境理解のための 分散型カメラ付きマイクアレイの位置推定・同期

須村 允亮[1]　関口 航平[2]　坂東 宜昭[3]
Aditya Arie Nugraha[2]　Yicheng Du[4]　吉井 和佳[4,2]
[1]京都大学 工学部情報学科　[2]理化学研究所 AIP　[3]産業技術総合研究所　[4]京都大学 大学院情報学研究科

1. はじめに

室内に配置した複数のマイクアレイおよびカメラを用いて視聴覚統合環境理解を行ううえで，それらの位置を推定し，同期ずれを是正するキャリブレーションが重要である．従来は，カメラ情報を用いない非同期複数マイクロホンのキャリブレーション [1] や非同期複数マイクアレイの位置・同期ずれ推定の手法 [2] について提案されてきた．

カメラ情報を併用した従来研究では，カメラの位置が既知であり、室内に音源がひとつ存在するという条件下で，マイクアレイおよび音源の位置を推定する手法が提案されている [3]．しかし，カメラ位置の計測が必要であることから，システムの可搬性に問題があった．

本研究では，視聴覚センサとしてカメラ付きマイクアレイを複数使用したキャリブレーション法を提案する．ここで，各デバイスに搭載されたマイクアレイとカメラの相対位置が既知であることを活用する．カメラ情報を用いて，各デバイスから他のデバイス・音源への距離・方向を推定する．同時に，マイク情報を用いて，各デバイスから音源方向を推定し，デバイス間の到達時間差を求めておく．視聴覚情報を統一的に表す確率的生成モデルを定式化し，GraphSLAM [4] を用いて推定パラメータの事後確率を最大化することにより，デバイスおよび音源の位置を推定すると同時に，デバイス間の同期を行う手法を提案する．

2. 提案法

本章では，複数台のカメラ付きマイクアレイを用いた位置推定・同期の手法について具体的に述べる．

2.1 問題設定

N (≥ 2) 台のカメラ付きマイクアレイを室内に配置する．各デバイス内のマイクは同期しており，デバイス間は非同期である．深度センサ付きカメラをデバイス 1 とし，他の全てのデバイスが視野に入るカメラを持つデバイス 2 は空間の中央付近に配置する．このような環境において，参照音源を 1 つ用いた推定を以下の設定で行う．

入力	N (≥ 2) 台のカメラ付きマイクアレイでの (1) 観測音　(2) 映像
出力	(1) 各時刻 t における音源座標 $\mathbf{s}_t = (s_t^x, s_t^y)$ (2) デバイス n の位置 $\mathbf{m}_n = (m_n^x, m_n^y, m_n^\theta)$ (3) デバイス 1 とデバイス n の同期ずれ ξ_n
仮定	デバイスは移動しない，移動音源が 1 つ存在

Localization and Synchronization of Distributed Camera-Attached Microphone Arrays for Indoor Environment Understanding: *Y. Sumura, K. Sekiguchi, Y. Bando, A. A. Nugraha, Y. Du, K. Yoshii*

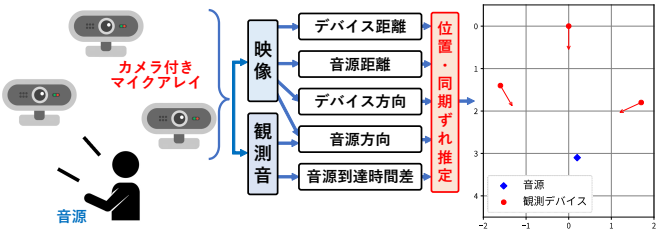

図 1: 音源位置とデバイス位置・同期ずれの推定システム

ここで $\mathbf{s}_t$ について $t = 1, ..., T$，$\mathbf{m}_n$ について $n = 1, ..., N$，ξ_n について $n = 2, ..., N$ である．

2.2 状態空間モデル

時刻 t における状態空間モデルの潜在変数を表す $4N+1$ 次元ベクトルを $\mathbf{y}_t = (\mathbf{s}_t, \mathbf{m}_1, ..., \mathbf{m}_N, \xi_2, ..., \xi_N)$，観測ベクトルを $\mathbf{z}_t$ とする．

状態遷移モデル: デバイスはすべて静止しているため，各デバイスの位置 $\mathbf{m}_n$ および同期ずれ ξ_n は時間変化しない．音源は現在位置からガウス分布に従ってランダムに移動すると仮定する．音源の状態遷移は次式で表される．

$$\mathbf{s}_{t+1} \sim \mathcal{N}\left(\mathbf{s}_t,\ \mathrm{Diag}\left(\sigma_x^2, \sigma_y^2\right)\right) \tag{1}$$

ここで，$\mathrm{Diag}\left(\sigma_x^2, \sigma_y^2\right)$ は音源の移動に関する対角共分散行列である．

観測モデル: 観測変数として，デバイス 1 から時刻 t における音源の距離 $k_{1,t}$，他のデバイス n ($n = 2, ..., N$) までの距離 q_{1n}，各デバイス n ($n = 1, ..., N$) からの音源方向 $\theta_{n,t}^*$ ($* \in \{\mathrm{c}, \mathrm{m}\}$)，デバイス 2 から他のデバイス n ($n = 1, 3, ..., N$) の方向 ϕ_{2n}，デバイス 1 に対するデバイス n ($n = 2, ..., N$) の音源到達時間差 $\tau_{1n,t}$ を用いる．観測 $\mathbf{z}_t$ がガウス分布に従うと仮定すると，観測の要素は次の式で表される．

$$k_{1,t} \sim \mathcal{N}\left(l\left(\mathbf{m}_1, \mathbf{s}_t\right),\ \sigma_{k_1}^2\right),\ q_{1n} \sim \mathcal{N}\left(l(\mathbf{m}_1, \mathbf{m}_n),\ \sigma_{q_1}^2\right)$$

$$\theta_{n,t}^* \sim \mathcal{N}\left(\arctan\frac{s_t^y - m_n^y}{s_t^x - m_n^x} - m_n^\theta,\ \sigma_{\theta_n^*}^2\right)$$

$$\phi_{2n} \sim \mathcal{N}\left(\arctan\frac{m_n^y - m_2^y}{m_n^x - m_2^x} - m_2^\theta,\ \sigma_{\phi_2}^2\right)$$

$$\tau_{1n,t} \sim \mathcal{N}\left(\frac{l\left(\mathbf{m}_n, \mathbf{s}_t\right)}{v} + \xi_n - \frac{l\left(\mathbf{m}_1, \mathbf{s}_t\right)}{v},\ \sigma_{\tau_1}^2\right)$$

ここで，$\sigma_{k_1}^2, \sigma_{q_1}^2, \sigma_{\theta_n^*}^2$ ($* \in \{\mathrm{c}, \mathrm{m}\}$), $\sigma_{\phi_2}^2, \sigma_{\tau_1}^2$ は各観測の分散パラメータ，$l\left(\mathbf{a}, \mathbf{b}\right)$ は座標 $\mathbf{a}, \mathbf{b}$ 間の距離である．$\theta_{n,t}^*$ および ϕ_{2n} に対して本来は von-Mises 分布を仮定すべきだが，分散が十分小さい時はガウス分布に近似可能である．

2.3 観測

位置・同期ずれ推定に用いる観測を得る方法について述べる．デバイス距離 q_{1n} および音源距離 $k_{1,t}$ は，深度センサ付きカメラを備えたデバイス 1 を用いて測定する．各デバイスの方向 ϕ_{2n} は，他のすべてのデバイスが視野に入るようなカメラを含むデバイス 2 の映像を用いて，その正面方向を基準として測定する．各デバイスに対する音源の方向 $\theta^{*}_{n,t}$ は，カメラ映像と観測音の両方を用いて観測する．カメラでは，映像からその正面方向を基準とし，観測値 $\theta^{\mathrm{c}}_{n,t}$ を得る．マイクアレイでは，観測音に対して multiple signal classification (MUSIC) 法 [5] を用いることにより，$\theta^{\mathrm{m}}_{n,t}$ を推定する．デバイス 1 に対するデバイス n $(n = 2, ..., N)$ の音源到達時間差 $\tau_{1n,t}$ は各デバイスの観測音によって測定する．デバイス間で同期している場合，音源到達時間差はデバイスと音源の距離差から幾何的に算出できるが，非同期の場合は同期ずれの時間が加算される．同期ずれの推定には音源到達時間差の観測が必要であり，観測音に対して generalized cross correlation with phase transform (GCC-PHAT) 法 [6] を用いて推定する．

2.4 位置・同期ずれ推定アルゴリズム

時刻 T までの観測ベクトルの列 $\mathbf{z}_{1:T}$ を入力として，時刻 T までの音源座標，センサ位置・同期ずれを表す $2T + 4N - 1$ 次元のベクトル $\mathbf{y}_{1:T} = (\mathbf{s}_1, ..., \mathbf{s}_T, \mathbf{m}_1, ..., \mathbf{m}_N, \xi_2, ..., \xi_N)$ を GraphSLAM を用いて推定する．GraphSLAM では，推定パラメータの事後確率 $P(\mathbf{y}_{1:T}|\mathbf{z}_{1:T})$ が最大となる $\mathbf{y}_{1:T}$ を求める．推定パラメータの事後確率は以下のように表される．

$$P(\mathbf{y}_{1:T}|\mathbf{z}_{1:T}) \propto \prod_t P(\mathbf{z}_t|\mathbf{y}_t)P(\mathbf{s}_t|\mathbf{s}_{t-1})$$

ここで，$P(\mathbf{m}_n)$ および $P(\xi_n)$ は一様分布を仮定している．時刻 t における観測ベクトルの事後確率 $P(\mathbf{z}_t|\mathbf{y}_t)$ は，潜在状態から幾何的な計算により得られる観測の理論値を求める関数 $\mathbf{z}_t = h(\mathbf{y}_t)$ を，初期値 $\boldsymbol{\mu}_t$ について一次近似することによりガウス分布で表す．これと式 (1) を用いて推定パラメータの事後確率をガウス分布で表す．推定結果を再度初期値として用いて推定を行い，推定結果が収束するまでこれを繰り返し行う．

3. 評価実験

本章では，シミュレーションを用いた提案手法の実験について述べる．

3.1 実験条件

シミュレーションで 6 台の観測デバイスの正解位置を設定し，参照音源の位置を移動させながら観測値を得た場合での推定手法の有効性について評価を行った．デバイス 1 の座標を原点，正面方向を x 軸方向に設定し，デバイス 2 は空間の中央付近に，残り 4 つのデバイスは空間の四隅に配置した (図 2)．また，同期ずれはデバイス 2 から順に 5.0, 5.0, -5.0, -5.0, 5.0[ms] に設定した．観測値には，デバイスと音源の正解位置から幾何的に計算した観測の理論値に対し，ガウス分布に従ったノイズを加えることにより作成したデータを用いた．音源は，図 2

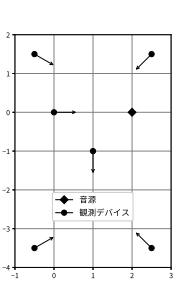

図 2: 配置

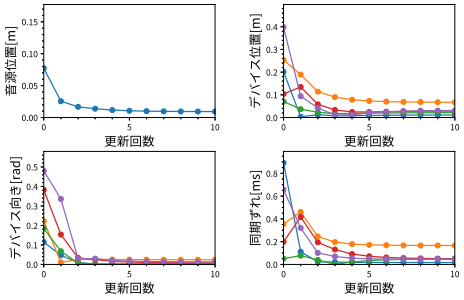

図 3: 推定結果

を初期位置として式 (1) に基づき $T = 100$ 回ランダムに移動させた．分散パラメータ σ_x, σ_y は共に 0.10[m] とした．これらの観測から得られた推定結果 $\mathbf{y}_{1:T}$ を再び初期値として推定し，これを 10 回繰り返した．更新ごとに得られたデバイス位置，同期ずれ，音源座標の推定結果について，正解との誤差を評価した．

3.2 実験評価

図 3 に 100 個の音源観測を用いて 10 回更新を行ったときの結果を示す．図中のグラフは，音源位置，デバイス位置，デバイス向き，デバイスの同期ずれの誤差を表しており，横軸は初期値の更新回数を表す．デバイスの推定結果は，座標および時刻の基準となっているデバイス 1 を除く 5 台すべてのものを示している．推定結果から，初期値の更新回数を重ねるごとに誤差が小さくなっており，提案手法の有効性を確認できた．

4. まとめ

本稿では，カメラ付きマイクアレイを複数配置し，移動可能な参照音源 1 つを用いて音源位置，デバイスの座標・向き・同期ずれを GraphSLAM により推定する手法について述べた．シミュレーションを用いた実験により，提案手法が動作することを確認した．今後は実際に観測デバイスを配置した環境で観測を行い，実環境での提案手法の有効性について検討する．

謝辞 本研究の一部は，科研費 No. 19H04137, 20K21813, 20H00602, 20K19833 の支援を受けた．

参考文献

[1] H. Miura et al. SLAM-based online calibration of asynchronous microphone array for robot audition. In *IEEE/RSJ IROS*, pages 524–529, 2011.

[2] K. Sekiguchi et al. 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. In *IPSJ*, 2016.

[3] A. Plinge and G. A. Flink. Geometry calibration of distributed microphone arrays exploiting audio-visual correspondences. In *EUSIPCO*, pages 116–120, 2014.

[4] S. Thrun and M. Montemerlo. The GraphSLAM algorithm with applications to large-scale mapping of urban structures. *IJRR*, 25:403–429, 2006.

[5] R. Schmidt et al. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas and Propagation*, 34(3):276–280, 1986.

[6] C. Zhang et al. Why does PHAT work well in low noise, reverberative environments? In *IEEE ICASSP*, pages 2565–2568, 2008.