5R-09

動的物体の三次元復元のための Audio-Visual Structure from Motion の検討

紺野 隆志[1], 西田 健次[1], 糸山 克寿[1], 中臺 一博[1,2]
1 東京工業大学工学院システム制御系
2 (株) ホンダ・リサーチ・インスティチュート・ジャパン

1 はじめに

SfM (Structure from Motion) は, 物体やシーンに対して様々な視点で撮影した画像群から, カメラの位置と姿勢および, 物体の三次元構造を復元する手法 [1] である. 対象物体やシーンが静止していることを前提としており, 人や電車, 風に揺られる木など静止していない物体は, 特徴点マッチングにおいて除外される. そのため復元物の点群が存在しない領域には, もともとそこには何も存在しないのか, それとも動的物体が存在しているが除外されてしまったのかを判別することはできない.

この問題に対し, 動いている物体は音がするという仮定をおくことにより, 音響信号を手掛かりに動いている物体の三次元位置を推定, 結果を SfM の復元結果へと統合する方法を検討した [2]. この手法では, 1 つのマイクロホンアレイを動かすことで仮想的に複数のマイクロホンアレイがあるとみなし, 各位置で得られる音源方向を三角測量に基づき統合することにより, 音源の三次元位置推定を行う. 計測対象は動的物体であるため得られる推定結果は, 動的物体が存在する可能性のある領域, つまり物体の三次元上の動作領域となる. そのため, 首を振って周期的に動作をする扇風機や風に揺られる木など, 動いてはいるが同じ場所に留まる物体に対しては, この手法は有効であるが, 人や電車など, 広範囲を動作する物体への適用は難しい. 本稿では, マイクロホンアレイ処理と動的物体に関する事前知識から, 各時刻ごとにその三次元位置を推定する. 各時刻の位置推定結果をトラッキングし動的物体の運動過程を推定することにより, 時間的に変動する三次元構造を復元する手法を提案する.

2 提案手法

図 1 に, 提案手法のフローチャートを示す. 画像情報を用いて静的物体の三次元復元を行い, 音響情報を用いて時間的に変動する動的物体の復元を行う. 最後にこれらの結果を統合することにより, 三次元構造復元の性能改善を図る.

2.1 画像情報による静的物体の三次元復元

画像処理として, まず SfM によりカメラ姿勢推定と疎な三次元形状復元を行う. ここで推定した復元物をもとに, ワールド座標系に対するマイクロホンアレイ座標系の座標変換の推定および, 平面や線路上などの動的物体が存在する領域の推定を行う. また, 推定したカメラ姿勢を用いて MVS (Multi View Stereo) [3] により密な三次元形状復元を行う. 本稿では, オープンソースソフトウェア (OSS) の COLMAP [1] をベースにして行った. 特徴点マッチングや三角測量の際に, RANSAC [4] を用いて Outlier の除去をしているため, 動いている物体の復元はされず, 静止している物体のみ復元される.

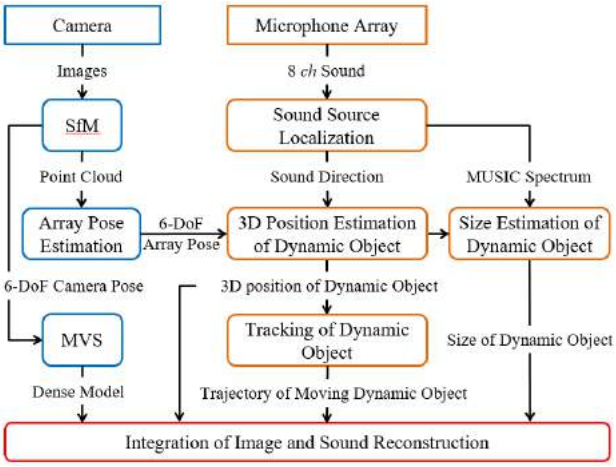

Fig. 1 Flowchart of the proposed system

2.2 音響情報による動的物体の三次元位置追跡

マイクロホンアレイは, 音源の方位角のみが推定可能とする. まず, MUSIC 法 [5] により音源定位を行い, マイクロホンアレイ座標系における音源の方向 θ を得る. MUSIC 法の実装にはロボット聴覚 OSS である HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [6] を用いる. 動的物体は点音源ではなく大きさを持つと考えられるため, MUSIC スペクトルのパワーの大きさにしきい値を設け, しきい値を超える方向は音源とすることにより, 音源の方向に幅 $[\theta_{min}, \theta_{max}]$ をもたせる. この幅は, 動的物体の大きさに対応すると考えることができる. θ はパワーが最も大きい方向とし, 以下では θ についてのみ議論する. 音源定位では仰角が得られないため, マイクロホンアレイに対する法線ベクトルを $\boldsymbol{n}$, マイクロホンアレイの中心 $X_M \in \mathbb{R}^3$ を通る定位方向 θ のベクトルを $\boldsymbol{\theta}$ とすると, $\boldsymbol{n}$ と $\boldsymbol{\theta}$ の外積である $\boldsymbol{N}$ を法線とする平面上に音源は存在する. この音源の存在平面と, SfM により推定した動的物体が存在する領域を用いて, 三角測量的に音源の三次元位置を推定する. 最後に, パーティクルフィルタにより推定した音源の三次元位置をトラッキングし, 動的物体の運動過程を推定する.

3 評価実験

円形のレールの上を車両が時計回りに走行するプラレールを用いて, 提案手法の評価を行った. レールおよび背景は静的物体であるため, 画像により三次元復元を行い, 車両は動的物体であるため, 音により三次元位置追跡を行った.

3.1 実験設定

SfM では, プラレールを一周するように動画を撮影し, キーフレームのみを抽出した画像を用いた. 画像の画素数は, 5472 × 3648 とした. 音響信号の収録には, 8

1 Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology
2 Honda Research Institute Japan Co., Ltd.

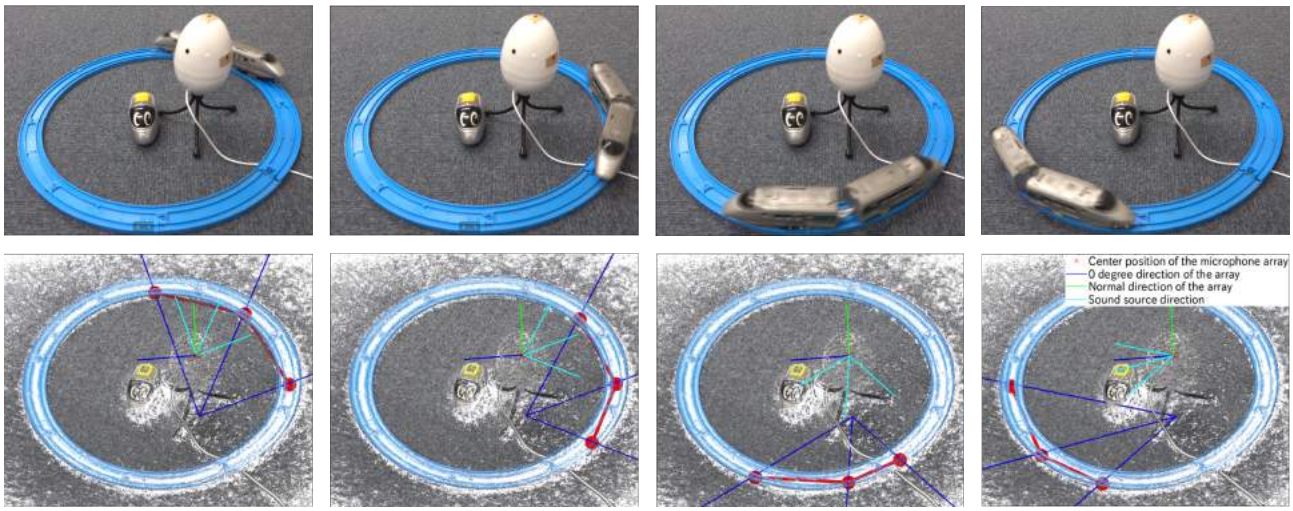

Fig. 2　時間的に変化する動的物体の存在領域推定. (上段) 撮影した画像. (下段) 上段に対応する復元結果

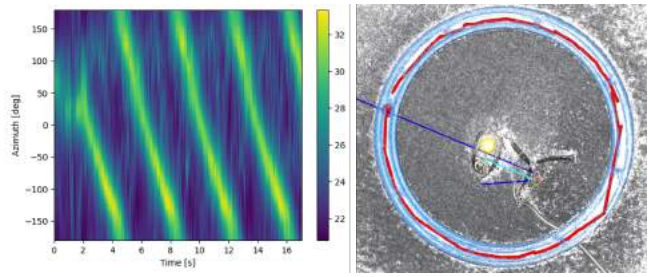

Fig. 3　(左図) 計測時間すべての MUSIC スペクトル. (右図) パーティクルフィルタによる動的物体の運動軌跡推定

個のマイクロホンが同一平面上に円状に配置されているマイクロホンアレイを床に 1 個固定し行った. 計測時間は, 列車がレールをおよそ 5 周する 17 秒とした. マイクロホン平面の法線ベクトルが床面の法線ベクトルと平行になるようにし, 0 度方向は任意の方向を向けて配置した. マイクロホンアレイの表面に複数のマーカを取り付け, SfM でこのマーカの三次元座標を推定することにより, マイクロホンアレイ座標系を推定した. また, 音源はレール上にあると仮定をし, 音源の三次元位置は, 音源定位により求めた音源の存在平面とレールの交点により推定をした. 動的物体の運動過程は, この交点をパーティクルフィルタにより追跡し推定をした. パーティクルフィルタは, モデルには以下の式 (1), (2) で表される 1 次階差モデルを, プロセスノイズ v_k と観測ノイズ w_k にはガウスノイズを用いた.

$$x(k+1) = x(k) + v_k, \quad v_k \sim N(0, V) \tag{1}$$

$$y(k+1) = x(k+1) + w_k, w_k \sim N(0, W) \tag{2}$$

$x(k) \in \mathbb{R}^3$ は, 動的物体の位置ベクトルであり, $y(k) \in \mathbb{R}^3$ は, 音源定位を用いた三角測量により推定した動的物体の位置ベクトルである.

3.2　実験結果

図 2 に, SfM により復元した 3 次元構造上に, 本手法により推定をした時間的に位置が変化する動的物体を統合した結果を示す. 赤色の線が動的物体の存在領域を示し, 三個ある赤丸のうち真ん中の赤丸は MUSIC スペクトルのパワーが最も大きい位置である. 実際の画像と比べて, 動的物体の位置と大きさがよく推定できていることがわかる. 図 3 左図は, 計測時間すべての MUSIC スペクトルを示しており, 図からパワーのしきい値を 30 に設定した. 図 3 右図は, MUSIC スペクトルのパワーが最も大きい位置をパーティクルフィルタにより追跡した結果を示しており, 赤線がその軌跡となる. 動的物体の運動軌跡もよく推定できていることがわかる.

4　まとめ

本稿では, SfM では復元することができない動的物体に対して, 音響信号を手がかりに物体の三次元位置および大きさ, 運動軌跡を推定する方法について述べた. また, プラレールを用いて提案手法の評価を行い, 音響信号を用いることによって, SfM の性能を向上できる可能性があることを示した. 今回は, 計測にマイクロホンアレイを 1 つだけ用いたため, 音源の存在領域を仮定したが, 複数個用いることにより, 存在領域を仮定せずに音源の 3 次元位置を推定できると考えられる. また, 動的物体の位置と大きさを線で表現したが, 画像と音の意味的な対応関係を用いた物体検出を併用することにより, より高度な三次元再構成が可能になると考えられる.

謝辞　本研究は, JSPS 科研費 16H02884, 16K00294,17K00365 および, JST ImPACT タフロボティクスチャレンジの助成をうけた.

参考文献

[1] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016.
[2] T. Konno et al. Improvement of sfm using sound information. In *SICE SI2018*.
[3] J. L. Schönberger et al. Pixelwise view selection for unstructured multi-view stereo. In *IEEE European Conference on Computer Vision (ECCV)*, pages 501–518, 2016.
[4] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
[5] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.
[6] K. Nakadai et al. Design and implementation of robot audition system 'hark'-open source software for listening to three simultaneous speakers. *Advanced Robotics*, 324:739–761, 2010.