

全天球画像の2視点間位置姿勢推定のための適応的閾値を用いた相互最近傍マッチング

Mutual Nearest Neighbor Matching Using Adaptive Threshold for Two-View Pose Estimation with Spherical Images

東京大学 ○安藤 大生, 小松 廉, Lee Junwoon
株式会社クボタ 篠崎 充, 北島 利浩
東京大学 浅間 一, 安琪, 山下 淳

Abstract This study aims to develop a method that achieves high accuracy in estimating pose of spherical camera between two viewpoints, even under the constraint of limited numbers of feature points. We propose a mutual nearest neighbor matching method using adaptive threshold for filtering. Specifically, the distances are calculated among all feature points, and the threshold is determined from the distribution of the distances using percentiles. This study results in robust matching outcomes, consequently enabling high-accuracy pose estimation.

1. 序論

近年, 日本の農業従事者は減少傾向にあり, 農作業の更なる効率化が求められている. 中でも自動運転技術は大きな生産性向上をもたらすものとして期待されている.

農場間の無人移動には Global Navigation Satellite System (GNSS) や Light Detection and Ranging (LiDAR) を使用する Simultaneous Localization and Mapping (SLAM) を用いることができるが, GNSS や LiDAR が効果的でない場面においてカメラ映像を基にする Visual SLAM が有用である. 特に, すべての方向からの情報を得ることができる全天球カメラを用いて Visual SLAM を行うことで, より効果的に自己位置推定および地図生成を行うことができる.

ただし, 全天球カメラを用いた既存の Visual SLAM システム, OpenVSLAM [1]を農場にて適用した結果, 自己位置推定において移動を観測できなかった [2]. これは農場において, 異なる視点で撮影された画像中の特徴点同士が類似しているために対応を正しく認識できていないことが原因であった.

位置姿勢推定において, 2つの異なる視点から相対的な位置と姿勢を推定することは最も基本的な要素技術である. Murrugarra-Llerena らは全天球画像を用いた2視点間位置姿勢推定における特徴点検出手法の比較分析を行った [3]. しかし, この研究では特徴点個数の制約下での考察が行われておらず, リアルタイム性を必要とする場合の議論が不十分である. また, 特徴点個数の制限下での位置姿勢推定を行った研究 [2]に関しては, 実環境での精度評価が行われていないという問題がある.

以上より, 本研究では2視点間での位置姿勢推定に関して高い精度および適度な速度を有する手法の開発を目的とする.

2. 提案手法

2.1 提案手法の方針および概要

本研究においては, SuperPoint [4]の特徴量に着目したマッチング手法を提案する. マッチング手法の中でも, 相互最近傍マッチングは精度において優れている. 相互最近傍マッチングを行う場合, 画像に含まれるノイズなどにより発生する, 誤ったマッチングを除去するためにフィルタリングが有用である.

提案手法では, 画像ペアについて全特徴点間で特徴量に基づく内積を計算し, 得られた値の符号を反転させた行列を作成する. 次にその行列に対して, 行列要素の分布を基にして正規化することで正規化距離行列を作成する. そして正規化距離行列の要素を用いて適応的に閾値を設定し, 行列の要素に対してフィルタリング処理を施す. 最後に, 相互最近傍マッチングを行う.

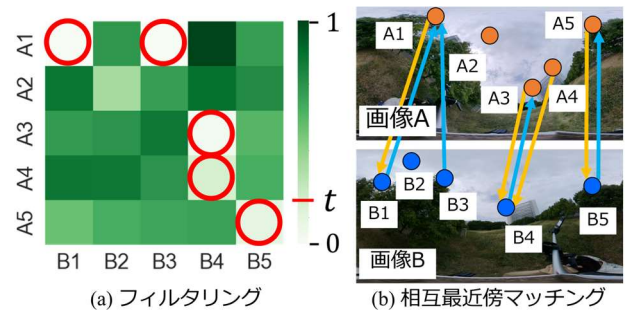


図1 フィルタリングとマッチングの模式図

2.2 適応的閾値によるフィルタリング

画像ペアについて SuperPoint で取得した特徴点同士から正規化した距離の分布を算出し, 分布に対してパーセンタイルで区切ることで閾値を決定する. そして, 閾値以下の距離を持つ特徴点のペアのみをマッチングの候補とする.

パーセンタイルとは分布における位置を示す指標である. 例えば, 50パーセンタイルは分布の中央値を意味する. 先行研究 [2]においては閾値の決定に関して統計的な妥当性がないという問題があった. そこで本研究では, パーセンタイルを用いることで処理に関して統計的な妥当性を高めつつ, フィルタリングの機能を維持することでその問題を解決する.

閾値の決定式およびフィルタリングの式をそれぞれ式(1)および式(2)に示す. ただし \mathbf{D} は正規化距離行列, $P_n(\mathbf{D})$ は \mathbf{D} の n パーセンタイルの値である. また, 図1(a)にフィルタリングの模式図を示す. ただし t は正規化した距離の値から決定された閾値であり, t 以下の距離を持つペアを赤丸で示している.

$$t = P_n(\mathbf{D}), \quad (1)$$

$$d_{ij} = \begin{cases} 1 & \text{if } d_{ij} < t \\ d_{ij} & \text{otherwise} \end{cases} \quad (2)$$

2.3 相互最近傍マッチング

相互最近傍マッチングは双方の特徴点が互いに最も近い場合のみマッチングを認める手法であり, 先行研究 [2]で用いられている.

図1(b)に相互最近傍マッチングの模式図を示す. 距離が閾値 t 以下となるペアに対して, 相互に最近傍となっているか確認をする. 図1においては, フィルタリングの結果としてマッチング候補となった5つの特徴点ペアについて, その内の(A1, B1), (A3, B4)および(A5, B5)のペアが相互に最近傍となっている. この手法によって, 類似した特徴点が多く存在する環境においても高精度なマッチング結果を提供できる.

表 1 特徴点個数の上限を 500 に設定した場合の位置姿勢推定結果

Methods	仮想環境				農場			
	位置推定		姿勢推定		位置推定		姿勢推定	
	MAE [°]	MSE [° ²]	MAE [°]	MSE [° ²]	MAE [°]	MSE [° ²]	MAE [°]	MSE [° ²]
Proposed	0.154	0.199	0.111	0.166	0.601	0.682	0.0438	0.00591
ORB	0.909	1.35	0.772	1.50	1.31	2.38	0.0655	0.0477
TORB	0.619	0.897	0.482	0.874	1.26	2.11	0.0643	0.0449
SPHORB	0.618	0.889	0.535	1.07	1.25	2.26	0.0746	0.0646
SIFT	0.731	1.13	0.434	0.681	0.878	1.46	0.0488	0.00691
TSIFT	0.369	0.513	0.247	0.466	0.614	0.835	0.0445	0.00613
SuperPoint	1.29	3.77	1.20	3.62	0.790	1.08	0.0448	0.00588
ALIKE	3.02	9.45	3.03	9.49	0.755	1.25	0.0629	0.0261
TALIKE	2.88	8.94	2.86	8.92	0.738	1.19	0.0695	0.0231

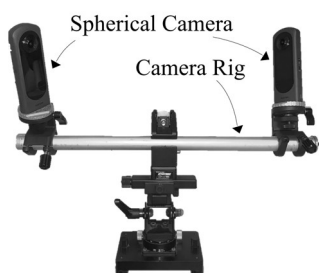


図 2 RICOH THETA X2 台とカメラリグ

3. 実験

提案手法の有効性を検証するために、仮想環境および農場で撮影された全天球画像のペアに対して位置姿勢推定を行った。農場のデータセットを取得する際に用いた全天球カメラ 2 台およびカメラリグを図 2 に示す。このカメラリグにより、2 台のカメラの相対位置姿勢を固定して同時に動画を撮影することができる。また、画像の解像度は 1024×512 とした。

本研究では画像について相対的な位置姿勢推定を行って精度評価を行う。ただし、2 台のカメラでの画像ペアから並進ベクトルおよび回転ベクトルを算出して別々に評価を行う。また、デバイスの構成およびプログラミング言語については、先行研究 [2] と同様に設定する。

4. 実験結果・考察

本研究における位置姿勢推定の結果の一部を表 1 に示す。ただし、各カテゴリにおいて最も精度が高くなった値を赤色、その次に精度が高くなった値を青色で表記した。また、接平面投影手法を用いた場合は TORB のように先頭に T と表記した。

表 1 の平均絶対誤差 (MAE) および平均二乗誤差 (MSE) を見ると、農場での姿勢推定の MSE を除いて、提案手法を用いた場合の精度が最良という結果が得られた。また、特徴点検出アルゴリズムを SuperPoint に固定して、様々なマッチング手法を用いた場合においても、提案手法の精度が最良だった。さらに、提案手法の計算時間は約 0.010 秒であり、リアルタイム処理の妨げにならない程度であった。

また、農場におけるマッチング結果の例を図 3 に示す。ただし既存手法とは、接平面投影、SuperPoint および提案手法を組み合わせた手法からマッチングのみを k 近傍マッチングと Lowe's ratio test の組み合わせに変更した手法である。これを見ると、既存手法、ORB を用いた手法および接平面投影と SIFT を用いた手法と比較して、提案手法が RANSAC によって判定されたインライアの増加に貢献していることがわかる。

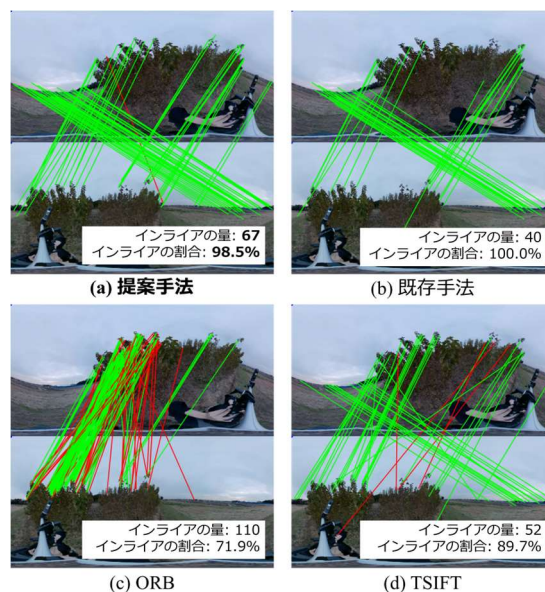


図 3 農場でのマッチング結果の例

5. 結論

本研究では、適応的な閾値を用いてフィルタリングをかけた相互最近傍マッチング手法を提案した。そして、全天球画像での 2 視点間位置姿勢推定において、仮想環境および農場のシーンに対して高精度かつ適度な速度での位置姿勢推定を行うことに成功した。

今後の展望として、全天球画像の歪みが存在することを前提とした学習ベースの特徴点検出手法を作成することを検討している。

参考文献

- [1] Shinya Sumikura, Mikiya Shibuya, Ken Sakurada: "OpenVSLAM: A Versatile Visual SLAM Framework," Proceedings of the 27th ACM International Conference on Multimedia, pp. 2292–2295, 2019.
- [2] 安藤 大生, 小松 廉, Junwoon Lee, 篠崎 充, 北島 利浩, 浅間 一, 安 規, 山下 淳: "適応的閾値を用いた相互最近傍マッチングによる全天球画像の 2 視点間位置姿勢推定精度向上," 第 24 回計測自動制御学会システムインテグレーション部門講演会講演論文集, pp. 2448-2453, 2023.
- [3] Richard J. Jeffri Murrugarra-Llerena, Thiago L. T. Da Silveira and Claudio R. Jung: "Pose Estimation for Two-View Panoramas based on Keypoint Matching: a Comparative Study and Critical Analysis," Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 5198–5207, 2022.
- [4] Daniel DeTone, Tomasz Malisiewicz and Andrew Rabinovich: "SuperPoint: Self-Supervised Interest Point Detection and Description," arXiv:1712.07629, 2018.