

色情報を用いた日照変化に対してロバストな カメラの自己位置姿勢推定手法の提案

Self-Pose Estimation using Color Information Robust to Change in Illumination

潮平寛弥[†]
Hiroya Shiohira

赤嶺有平[‡]
Yuhei Akamine

遠藤聡志[‡]
Satoshi Endo

根路銘もえ子[§]
Moeko Nerome

1. はじめに

ARを利用する際には、画面上に適切な電子情報を重畳表示するためにカメラの位置姿勢を推定する必要がある。また、屋外で利用する場合には、日照変化に対するロバスト性が求められる。推定する方法はいくつかあるが、近年は自然特徴点を用いた推定手法が注目されている。

本研究で想定している位置姿勢推定プロセスは、特徴点マッチング手法を基に対象シーンを様々なアングルで撮影したフレーム群(以下、事前フレーム)からカメラ入力内の第1フレームに合うフレームを選抜し、利用者がシーン内のどの位置で、どの向きにカメラを使用しているかを推定する大域的自己位置推定と、カメラ入力内の微小なカメラ位置姿勢の更新を行う局所的自己位置姿勢推定の2段階処理により実現される。

局所的自己位置姿勢推定では特徴点の位置更新だけを行うので高速に処理できるが、大域的自己位置姿勢推定では何も無い状態からカメラの位置を推定するため、処理に時間が掛かる。しかし、大域的自己位置姿勢推定を高速に行えないとリアルタイム性が損なわれる。

そこで本研究では、特徴点の色情報や2次元位置に基づく特徴量を基に事前フレーム選抜と特徴点マッチング時の外れ値除去を行う、リアルタイムなARアプリケーション利用のための高速な大域的自己位置探索手法を提案する。

2. 提案手法

本研究では、大域的自己位置姿勢推定処理における事前フレーム選抜処理手法及び特徴点マッチング時の外れ値除去手法を提案する。

以下、上述した2つのアルゴリズムを説明する。

2.1. 事前フレーム選抜

事前フレームを選抜するために、予め事前フレームの特徴量を抽出し、入力画像の特徴量と比較することで適切な事前フレームを選抜する。以降は事前フレーム及び入力画像の特徴量を画像特徴量とし、以下に画像特徴量抽出の手順を示す。

画像特徴量抽出

1. 全事前フレームから特徴点を検出し、検出した特徴点の座標における色情報を取得し、k-means法によりクラスタリング
2. クラスタリングの結果を基に、画像毎に抽出

[†]琉球大学大学院理工学研究科

[‡]琉球大学工学部

[§]沖縄国際大学経済学部

した特徴点を分類

3. 対象シーンにおいて、要素が少ないクラスに属している特徴点と、要素が多いクラスに属している特徴点を結ぶ直線を2本算出し、その2本の直線を基に外積を算出
4. 3.で算出した外積の大きさに応じて、出現頻度をクラス別に記録

※ 手順2~4の処理は事前フレームと入力画像の両方に対して行う

画像特徴量抽出時の手順1で取得する色情報は、 $L^*a^*b^*$ 表色系の要素a, bの値が適切だと考えられる。 $L^*a^*b^*$ 表色系は、「輝度」「赤寄りか緑寄りか」「青寄りか黄色寄りか」という3つの要素で構成されている表色系であり、事前実験により、輝度に比べて日照変化の影響を受けにくいという結果が得られている。

入力画像に対する適切な事前フレームを選抜する際は、入力画像と事前フレームで画像特徴量の差分を取り、その差分が小さいものほど適切な事前フレームとなる。

本手法では、画像中の特徴点群から算出できる外積の出現頻度を数えることで画像特徴量としているため、画像1枚あたりの情報量が少なく、非常に高速に選抜処理を行うことができる。

2.2. 特徴点マッチング時の外れ値除去

本手法では、特徴点マッチング時に特徴点群から算出できる外積を比較することにより、適切な特徴点同士がマッチングしているかどうかを判定する。

以下に特徴点マッチング時の外れ値除去の手順を示す。

特徴点マッチング時における外れ値除去処理

1. 既存手法(SIFT[1]など)を用いて入力画像と事前フレーム間で特徴点マッチングした後、特徴量距離が比較的近い特徴点ペア1組(以下、ペアA)と適当な特徴点ペア2組(以下、ペアBとペアC)をランダムに選択
2. 2次元座標の位置関係からペアAを頂点とした外積を算出
3. 算出した2つの外積の差分を算出し、閾値に応じて特徴点ペア3組の正解数もしくは誤答数を加算
4. 1.~3.の処理を十分に行った後、特徴点ペア毎に誤答数が閾値以下の組を数え、その数が多いほど、その事前フレームが入

力画像に対して適切にマッチングしているとする

※ 位置姿勢を推定する際には、特徴点ペア毎に正解数が閾値以上の組を用いる

本手法を用いることにより、特徴点群の幾何構造を基にした外れ値除去を行うことができる。

本手法はリアルタイムなARの利用を目的としているため、上述の手順4で、誤答数が閾値以下の組の数が、予め設定していた閾値よりも大きい場合、特徴点マッチング処理を終えるようにしている。

3. 実験及び結果

本実験では、画像検索などで用いられる visual words 方式の BoF[2](以下、VW) と提案手法で、事前フレーム選抜に要した時間や精度、事前フレーム選抜から再投影誤差を算出するまでに要した時間(以下、合計処理時間)、そして再投影誤差の比較を行った。

本実験において、教師データと入力データには、日照方向の変化を伴う5つの屋外シーンを撮影した動画を用い、特徴点の検出にFAST、特徴量の記述にSIFTを用いた。

VWにおいて、シーン毎に算出する visual words の数を1000個とし、visual words を用いて画像特徴量を記述する際には、各 visual words に類似した特徴量が画像中の何%か記述した。また、特徴点マッチング時には、提案手法を用いて外れ値の除去を行なった。

再投影誤差を算出する際、シーンモデル(3次元特徴点群)と入力画像の2次元特徴点を対応付けることにより、カメラの位置姿勢を推定する必要がある。その後、推定したパラメータでシーンモデルを入力画像上に2次元投影し、その誤差を計測する。

本実験では再投影誤差の算出を2回行った。1回目は外れ値除去を行うためであり、推定したパラメータでシーンモデルを2次元投影し、特徴点ペア間の距離が全特徴点ペア間の中央値以上に離れている特徴点ペアの除去を行なった。そのため、2回目の処理では外れ値をある程度除いた状態で再投影を行っている。

図1で各項目の実験結果を示している。事前フレームの選抜精度は、適切な事前フレームだと判断された時点(特徴点マッチング処理を途中で抜ける、もしくは事前フレーム全てと特徴点マッチングした後)での事前フレーム選抜時の順位の平均を示している。ここでは、6位以降の事前フレームと最終マッチングした際には6位として計算している。

図1左上の事前フレーム選抜精度を見ると、提案手法では2~5位内に適切な事前フレームがあるのに対し、VWでは1~2位内という結果になっている。また、事前フレーム選抜時間(右上)を見ると、VWに比べて提案手法の方が約6倍以上短く、合計処理時間も提案手法の方が短いことが分かる。これは、VWでは画像特徴量を算出する際に、特徴量記述手法の中でも処理が重いSIFTを用いており、特徴点数の増加に伴って抽出時間も増加するためである。それに対し、提案手法では、高速に処理するためにほぼランダムに特徴点を

選択し、外積算出時の処理回数に制限を設けているため、抽出処理に掛かる時間がほぼ同じとなる。そのため、特徴点マッチング処理に時間がかかっても、合計処理時間が短くなる状況が生まれる。

次に、再投影誤差(右下)を見ると、ほとんどのシーンでどの手法もほぼ同じ値だが、シーンEは比較的誤差が大きい。これは、他のシーンに比べて色飛びや似た模様がが多く、また特徴点数も少なかったために、誤差の大きい事前フレームとマッチングしたと考えられる。

現時点において提案手法による事前フレームの選抜精度は良くないが、今以上の精度で事前フレームを絞り込むことができた場合、VWなどを用いて高速かつ高精度に事前フレームを選抜することができると思われる。

青は提案手法、赤はvisual words式BoFの結果

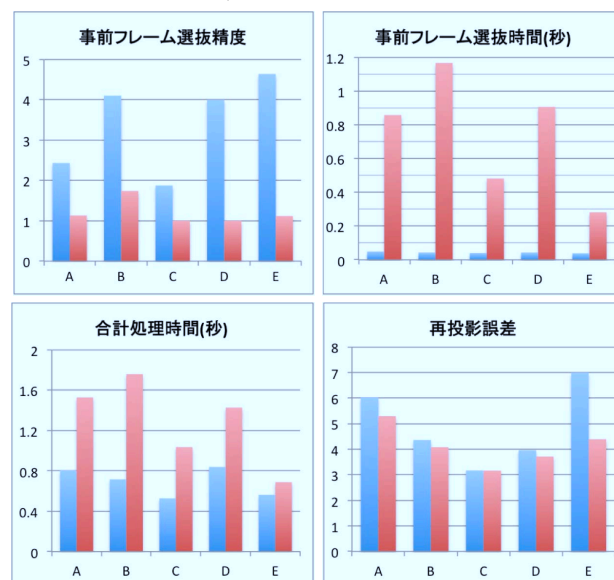


図1: 実験結果

4. おわりに

本稿では、色情報や2次元位置に基づく特徴量を用いることで日照変化に頑健かつVWよりも高速に位置姿勢推定が可能であることを示した。

実験結果より、シーンによって精度にばらつきがあるため、より多くのシーンで実験を行う必要がある。また、より識別性の高い画像特徴量のための幾何構造算出方法や、特徴点マッチング時における外れ値除去の確実性の向上が必要だと考えられる。

参考文献

- [1] D.G.Lowe, "Object Recognition from Local Scale-Invariant Features", International Conference on Computer Vision, pp.1150-1157, 1999
- [2] G.Csurka, C.R.Dance, L.Fan, J.Willamowski and C.Bray: "Visual categorization with bags of key-points", ECCV In Workshop on Statistical Learning in Computer Vision, pp.1-22, 2004