

自然特徴点によるランドマークデータベースを用いたカメラ位置・姿勢推定 Camera Position and Posture Estimation Using Feature Landmark Database

大江 統子 佐藤 智和 横矢 直和
Motoko Oe Tomokazu Sato Naokazu Yokoya

1. はじめに

近年、動画像を用いてカメラの位置・姿勢を推定する手法が盛んに研究されており、拡張現実感やロボットナビゲーションに応用されている。これらの分野では、カメラの絶対位置・姿勢が必要とされるため、動画像とGPSや磁気センサを併用する手法 [1]、環境内に配置された人工的なランドマークと入力画像中の自然特徴点を同時に追跡する手法 [2] などが提案されている。しかし、前者は利用可能な環境が制限される問題があり、後者は広範囲に人工的なランドマークを多数配置しなければ、カメラ位置の推定誤差が累積するという問題がある。

そこで、本稿では、環境内に広く存在する自然特徴点の三次元位置と画像テンプレートをあらかじめデータベース化し、入力画像とデータベースの照合によってカメラの位置・姿勢を逐次推定する手法を提案する。本手法では、センサ類や人工的なマーカを用いることなく、累積誤差を含まないカメラの絶対位置・姿勢を推定することが可能となる。

2. ランドマークデータベースによるカメラ位置・姿勢推定

提案手法は図 1 に示すように、事前に撮影された全方位動画像から自然特徴点のランドマークデータベースをオフラインで構築する処理 (A) と、用意されたデータベース中の自然特徴点と単眼動画像上の自然特徴点の対応付けによる逐次的なカメラ位置推定処理 (B) によって構成される。以下、各処理について述べる。

2.1 自然特徴点ランドマークデータベースの構築

まず、佐藤らの手法 [3] を用いて三次元位置が既知の少数の基準マーカと多数の自然特徴点を全方位動画像中から自動追跡し、カメラの動きパラメータを推定する。この手法では、動画像全体での投影誤差を最小化することで、累積的な推定誤差を最小化し、カメラパスと自然特徴点の三次元位置を基準マーカによる絶対座標系で求めることができる。これにより、全方位動画像とそのカメラパス、シーン内の自然特徴点の三次元位置、画像上の座標が得られる。本研究では、これらの情報を用いてそれぞれの自然特徴点ごとに画像テンプレートを自動的に作成し、以下の情報を用いてランドマークデータベースを構築する。

- (a) 自然特徴点の三次元位置
- (b) 自然特徴点を中心とする画像テンプレート
- (c) 画像テンプレート撮影時の全方位カメラの位置・姿勢

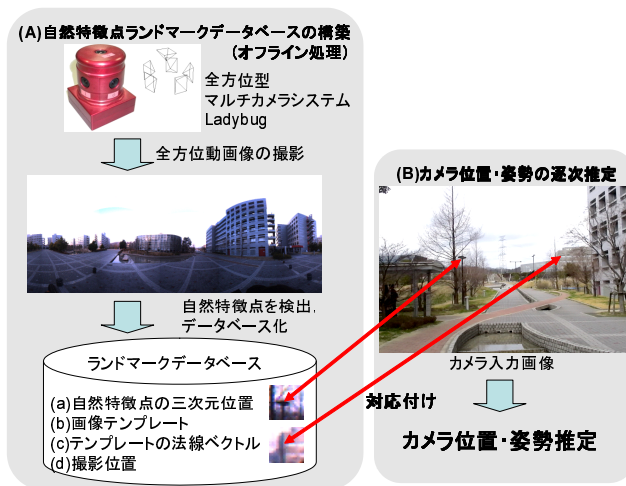


図 1: 提案手法の構成

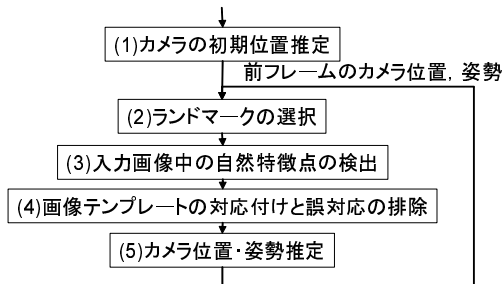


図 2: カメラ位置・姿勢推定処理の流れ

2.2 ランドマークデータベースを用いたカメラ位置・姿勢の逐次推定

前節の手法によって事前にオフラインで作成されたデータベース中の自然特徴点と入力画像から検出される自然特徴点をフレームごとに逐次対応付けることにより、自然特徴点の三次元座標と画像上の二次元座標からカメラ位置・姿勢推定を行う。本手法のカメラ位置・姿勢推定処理の流れは図 2 のようになる。

(1) カメラの初期位置推定

本手法では、現フレームで用いるランドマークを逐次選択するために、初期フレームにおけるカメラの位置・姿勢が必要となる。手法としてはランドマークデータベースを用いたオフラインでの推定やユーザによる大まかな位置の入力などが考えられるが、本稿では初期カメラ位置・姿勢は何らかの手法によって取得済みとする。

(2) ランドマークの選択

ランドマークと入力画像中の自然特徴点を対応付けるために、まずランドマークを選択し、そのランドマークに対応付けられる自然特徴点を入力画像から探索する。本稿では、まず前フレームのカメラ位置に近い地点から撮影されたランドマークをデータベースから選択する。これらのランドマークを前フレームの情報を用いて入力画像面上に投影し、一定の枠内に投影された点を次ステップ以降で用いるランドマークとして選択する。ただし、異なる地点から撮影された、同一の三次元座標を持つランドマークに関しては、ランドマークの三次元位置とランドマークの画像テンプレート撮影時のカメラ位置を結ぶ直線と、ランドマークの三次元位置と前フレームのカメラ位置を結ぶ直線のなす角が最小となるものを選択する。

(3) 入力画像中の自然特徴点の検出

ステップ(2)で選択されたランドマークに対応する入力画像中の自然特徴点の候補を抽出する。まず、前フレームのカメラ位置・姿勢を用いてランドマークを入力画像上に投影する。投影された画像上の座標周辺にランドマークに対応する自然特徴点が存在すると仮定し、投影座標を中心とする一定サイズのウィンドウ内で Harris オペレータにより自然特徴点を検出する。これにより、ひとつのランドマークに対して複数の入力画像中の自然特徴点を対応付けの候補とする。

(4) 画像テンプレートのマッチングによる対応付けと誤対応の排除

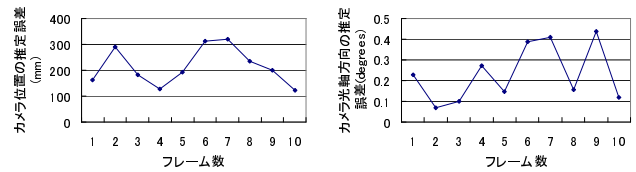
前ステップで検出された自然特徴点を中心とする画像パターンとランドマークの画像テンプレートを正規化相互相関により比較し、最も相関が高い自然特徴点をランドマークと対応付ける。また、相関値が閾値以下の場合には対応付けしない。これにより、データベース内の自然特徴点の三次元座標と入力画像上の二次元座標を対応付ける。次に、RANSAC[4]により誤推定を排除する。まず、対応付けられた三次元座標と二次元座標の組からランダムに6組以上を選択し、暫定的なカメラパラメータを推定する。次に、全ての組を用いて投影誤差を算出する。これらの処理を複数回繰り返し、投影誤差の中間値が最小となる暫定カメラパラメータを選択する。最後に、選択されたカメラパラメータによって各特徴点の投影誤差を評価し、閾値を超える結果を誤対応として削除する。

(5) カメラ位置・姿勢推定

入力画像とランドマークとの対応付けによって得られた二次元座標と三次元座標の対応関係を用いて PnP 問題を解くことによってカメラ位置・姿勢推定を行う。

3. 予備実験

入力画像の自然特徴点とデータベースとの対応付けによるカメラ位置・姿勢推定精度を検証するために、実際に図1(A)に示すような屋外環境を撮影し、実験を行った。まず、図1(A)に示す全方位型マルチカメラシステム(Point Grey Research 社 Ladybug)を用いて撮影した動画像500フレームから佐藤らの手法[3]によってカメラパスと自然特徴点の三次元座標を推定し、ランドマークデータベースを作成した。次に、ビデオカメラ(SONY DCR-PC300)を用いて図1(B)に示すような画像(解像



(a) カメラ位置の推定誤差 (b) 光軸の推定誤差

図 3: カメラパラメータの評価結果

度 720x480 画素, プログレッシブスキャン, 30fps) を撮影し、入力として、2.2 節で述べた手法を用いてカメラ位置・姿勢をフレーム毎に逐次推定した。ただし、本実験では初期フレームのカメラ位置・姿勢はあらかじめ既知とし、入力画像は初期フレームから 100 フレーム目までを 10 フレームおきに用いた。また、対応付けに誤推定が多く存在すると仮定し、RANSAC による繰り返し回数を 1200 回とした。逐次推定の結果、1 フレームで用いられたデータベース中のランドマーク数は平均 40 点、そのうち入力画像と正しく対応付けられたものは平均 25 点、1 フレームの処理時間は約 3 分であった。

次に、推定されたカメラパラメータを真値と比較することで推定精度を評価した。ただし、本実験では、各フレームの画像上でランドマークの位置を手動で指定し、PnP 問題を解くことでカメラパラメータの真値を作成した。図 3 に初期フレームから 10 フレーム目までのカメラパラメータの評価結果を示す。カメラ位置推定誤差は平均約 210mm、カメラの光軸の角度誤差は約 0.23 度であった。対応付けに用いたランドマークの三次元位置と推定カメラ位置の平均距離は約 58m であったことから、カメラ位置の推定精度は撮影対象までの距離の 0.4% 程度であり、累積的な推定誤差なしにおおよそのカメラ位置・姿勢が推定されていることがわかる。

4. まとめ

本稿では自然特徴点ランドマークデータベースを事前に構築し、データベースと入力画像中の自然特徴点を対応付けることによりカメラ位置・姿勢をフレームごとに推定する手法を提案した。実験では、実際にカメラ位置・姿勢推定を行い、カメラ位置、カメラ光軸方向の推定精度を確認した。今後の課題として、データベースと入力画像の自然特徴点の対応付け精度の向上、処理速度の向上による逐次推定のリアルタイム化などが挙げられる。

参考文献

- [1] A. Ronald, B. Hoff, H. Neely III, and R. Sarfaty, "A Motion-Stabilized Outdoor Augmented Reality System", Proc. IEEE Virtual Reality, IEEE CS Press, pp 252-259, 1999.
- [2] U. Neumann and S. You, "Natural Feature Tracking for Augmented-Reality", IEEE Transactions on Multimedia. Vol. 1. No.1, pp 53-64, 1999.
- [3] T. Sato, S. Ikeda, and N. Yokoya, "Extrinsic Camera Parameter Recovery from Multiple Image Sequences Captured by an Omni-Directional Multi-Camera System", Proc. European Conf. on Computer Vision, Vol. 2, pp. 326-340, 2004.
- [4] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Comm. of the ACM, Vol 24, pp 381-395, 1981.