

特微量辞書を用いた屋外環境下における 自己位置推定手法の提案

神山 朗^{1,a)} 赤嶺 有平²

概要：単眼カメラのみを用いた自己位置推定を屋外で行う場合、照度変化に対しロバストである局所特微量を使用する手法があげられるが、局所特微量は大きなアングル変化を含むカメラの位置姿勢の変化に対し、空間上の同一特徴点から同じ様な局所特微量を得ることが困難である。本研究では、予め3次元の点群で構成されるマップとマップを構成する各点に対し、複数のカメラ位置姿勢から得られた局所特微量を付加する（以下、特微量辞書と呼ぶ）。さらに自己位置推定時のカメラ位置姿勢に対し、特微量辞書から特微量を選抜することでカメラ入力内の特徴点との同定処理を高速に行い、処理をリアルタイムに行う手法を提案する。

1. はじめに

近年、実世界の映像上にCG等の情報を重畳表示するAR(Augmented Reality: 拡張現実感)技術の研究が盛んである。AR技術にはコンテンツを提示する情報端末の位置姿勢の推定処理が不可欠であり、GPSや電子コンパス等のセンサを用いるセンサベースの自己位置推定や、カメラ映像より得られる画像を処理して推定を行うビジョンベースの自己位置推定など様々な手法が提案されている。

単眼カメラのみを用いて既知の屋外シーンでARコンテンツを提示する場合、あらかじめシーンより計測した特徴点の3次元位置とシーンを撮影するカメラ映像内の特徴点とを対応付け、情報端末の位置姿勢を推定する方法がある。

画像の拡大縮小や照度変化にロバストであるSIFT[1]やSURF[2]等の特微量を上記の対応付け処理に用いた時、1つのカメラ位置姿勢から得られた特微量を付加しただけでは、シーンに存在する曲面や凹凸面を持った物体上の特徴点から同一の特微量を抽出することが難しく、対応付け処理の誤対応により自己位置推定が破綻するといった問題が生じた。

我々はこの問題に対して、事前に作成したシーン内の特徴点の3次元位置(以下、マップと呼ぶ)と、マップ内の点に対して一定の角度ごとにSURF特微量を付加した特微量辞書を用いて、マップの点とカメラ映像内の特徴点との同

定処理を行い、情報端末の自己位置推定を行う手法[3]を提案した。

本研究では、特微量辞書内の特微量をバイナリデータで特微量ベクトルを記述するBRIEF特微量[4]に変更し、さらに特微量辞書を最適化することで提案手法の高速化及び高精度化を図った。

2. 関連研究

2008年にCastleらが、高精度で高速なマーカーレス型ARを実現する"Video-rate Localization in Multiple Maps for Wearable Augmented Reality"(以下、PTAMMと呼ぶ)[5]を発表した。

PTAMMでは特定のシーンを映したカメラの映像より特徴点を抽出し、シーンに対するマップを作成する。その後、作成したマップを用いての自己位置推定とマップの拡張処理を並列に行うことで、リアルタイムで高精度な自己位置推定を実現している。また、自己位置推定に用いたマップを保存することが可能であり、マップを作成したシーンに対して、再度自己位置推定を行うことが可能である。

しかし、我々が行った屋外での予備実験において、ARを重畳表示したい現実空間のオブジェクトが曲面や凹凸面を持っていた場合、ユーザーがそのオブジェクトを注視しながらオブジェクトを回りこむ動きをすると、位置姿勢推定時の誤差が蓄積して自己位置推定が破綻するといった問題が生じた。

¹ 琉球大学 大学院 理工学研究科
Grad. School of Engr. and Sci. University of the Ryukyus

² 琉球大学 工学部 情報工学科
Dept. of Info. Engr. University of the Ryukyus

a) k118581@ie.u-ryukyu.ac.jp

3. 提案手法

本研究では、自己位置推定を行う前にあらかじめシーンを適切に撮影した動画像を用いて、自己位置推定に必要な情報群(以下、シーンモデルと呼ぶ)を作成する。作成するシーンモデルの詳細を以下で述べる。

3.1 シーンモデル

A マップ

自己位置推定の対象となるシーンを適切に撮影した動画像より特徴点の検出を行い、それを3次元復元したものをマップとする。今回、特徴点はHarrisオペレータ[6]により検出し、マップはVoodooCameraTracker[7]を用いて作成した。VoodooCameraTrackerは画像シーケンスよりカメラパラメータの推定を行い、特徴点の3次元復元を行うソフトウェアである。

B 特徴量辞書

マップ作成時の動画像内に映るマップの点から得られる特徴量が、一定のしきい値以上変化した時のカメラ位置姿勢の範囲を保存し、マップの点に複数個付加したものをマップの点を持つ特徴量辞書とする。図1にマップの点と特徴量辞書の関係を示す。

C キーフレーム

マップ作成時に動画像のフレーム間で特徴点の特徴量が大きく変わるフレームを抜き出し、そのフレームを撮影したカメラ位置姿勢とキーフレームに映るマップの点を持つ特徴量を共に保存したものをキーフレームとする。後述する自己位置推定プロセスの大域的自己位置探索で用いる。

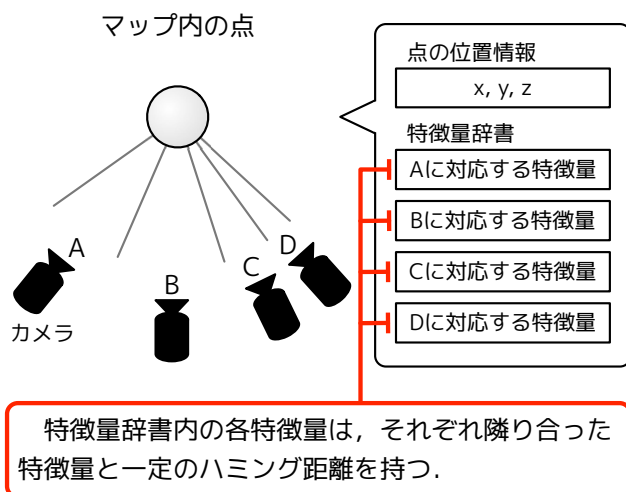


図1 マップ内の点と特徴量辞書の関係

Fig.1 Relation between Map Point and Feature Dictionary

3.2 自己位置推定プロセス

本研究の自己位置推定プロセスは、カメラ映像の第1フレームからシーン内のどこに情報端末が位置するかを探索する大域的自己位置探索(3.2.1)と、カメラ映像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探索(3.2.2)で実現される。プロセス中の特徴点はマップ作成時と同じHarrisオペレータで検出し、特徴量はBRIEFアルゴリズムで64bitの特徴量ベクトルを抽出する。

3.2.1 大域的自己位置探索

この処理は自己位置推定の開始時と自己位置推定が破綻した際に呼ばれる。後述する局所的自己位置探索に比べて、処理が呼ばれる頻度は少ないが、シーン内のどこに情報端末が位置するかという大きな範囲の探索を行うため、処理に時間がかかる。

- (1) カメラ映像のフレームから特徴点を検出し、特徴点位置より特徴量を抽出する。
- (2) (1)の特徴量と各キーフレームの持つ特徴量をハミング距離で比較(以下、特徴量マッチングと呼ぶ)し、距離がしきい値以下のマッチング結果が多いキーフレームをN個選抜する。
- (3) (2)で選抜されたキーフレームが持つカメラ位置姿勢でマップの点をカメラ映像のフレームに投影し、その2次元位置を開始点としてLK法[8]でフレーム内のマップの点の2次元位置を探索する。
- (4) (3)で探索されたマップの点の2次元位置と対応するマップの点の3次元位置(以下、画像内の特徴点とマップの点の対応を特徴点对応と呼ぶ)を用いて自己位置推定を行う。推定された位置姿勢は再投影誤差の2乗の平均を用いて評価される。再投影誤差は(3)で探索されたマップの点の2次元位置と、推定された位置姿勢により再投影されたマップの点の2次元位置の距離の差であり、再投影誤差が少ない位置姿勢ほど評価が高い。
- (5) 選抜されたN個のキーフレームに対し(3)(4)の処理を行った後、最も評価の高かったカメラ位置姿勢を自己位置とする。
- (6) (5)で決定したカメラ位置姿勢でマップの点をカメラ映像のフレームに投影し、その2次元位置に近い特徴点を対応付けて、そのフレームにおける最終的な特徴点对応を作成する。

3.2.2 局所的自己位置探索

この処理はカメラ映像のフレーム間の位置姿勢の微小な動きを推定し、更新する処理である。前述した大域的自己位置探索に比べて処理の頻度が高いが、カメラ映像の前フレームにおけるカメラ位置姿勢を基に現在のフレームの位置姿勢を探索するため処理時間は短い。

- (1) 前フレームで作成した特徴点对应の2次元位置を開始点として、LK法で現在のフレーム内の特徴点を探索する。
- (2) (1)の処理で探索された特徴点以外の新しい特徴点を現フレームから検出し、その特徴量を抽出する。
- (3) 前フレームのカメラ位置姿勢を用いて、現在のフレーム内に映るマップの点の特徴量辞書より特徴量を取り出す。図2に特徴量辞書から特徴量を取り出す処理の例を示す。
- (4) (2)(3)を特徴量マッチングし、ハミング距離がしきい値以下のマッチング結果を持つマップの点と特徴点の対応を現在のフレームの特徴点对应に追加する。
- (5) 現在のフレームの特徴点对应には(1)の探索や(4)の特徴量マッチングにより誤対応が出るのでRANSAC法[9]を用いて外れ値の影響を除去し、現在のフレームの位置姿勢を推定する。
- (6) 推定したカメラ位置姿勢でマップの点をカメラ映像のフレームに投影し、その2次元位置に近い特徴点に対応付けて、現在のフレームにおける最終的な特徴点对应を作成する

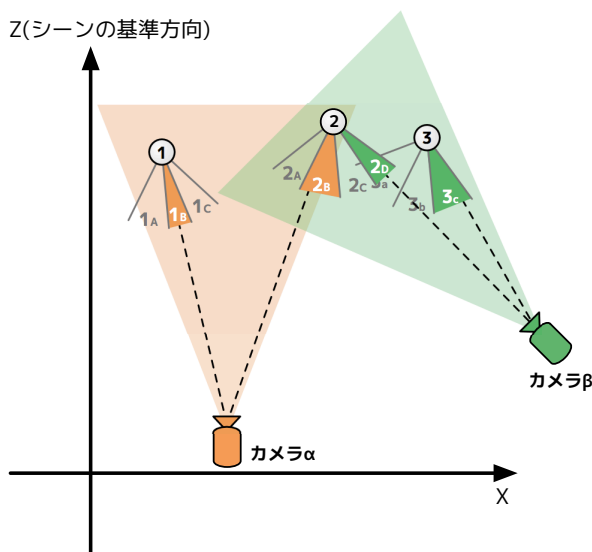


図2 特徴量辞書を用いる処理の例
 Fig.2 Usage Example of Using Feature Dictionary

マップの各点を持つ特徴量辞書から抽出される特徴量は、マップの点とカメラ位置を結んだ線分とマップ作成時に任意に決定したシーンの基準方向が作る角度によって決定される。図2の例ではカメラαにはマップ内の点1, 点2が映り、点1からは特徴量 1_B を、点2からは特徴量 2_B を抽出する。カメラβにはカメラα同様に点2が映るが、カメラ位置が大幅に変わったことで点2の持つ特徴量 2_D が抽出され、アフィン変換では対応できない特徴量の変化に対応し、より精度の高い特徴量マッチングが可能になる。

movieA



movieB



図3 実験に使用した movieA, movieB の一部
 Fig.3 movieA, movieB

表1 計算機の仕様

Table1 Specification of Computer

OS	Mac OS X 10.7.5
CPU	2.66GHz Quad-Core Intel Xeon
RAM	3GB 1067MHz DDR3

4. 実験

4.1 実験方法

本実験では提案手法の評価として、同一の屋外シーンを日照が異なる時間帯に、シーン内の凹凸面を持ったオブジェクトをユーザーが注視しながら回りこむような動きで撮影した動画 movieA, movieB(図3)を用意し、movieAを用いてシーンモデルを作成した後、movieBに対して自己位置推定を行った。そして、movieBを撮影した際のカメラ位置姿勢をオフラインで高精度に推定した結果との比較を行い、さらに関連研究で挙げた PTAMM で movieAを用いてマップを作成した後、movieBの自己位置推定を行った結果との比較を行った。

今回、大域的自己位置探索の処理(2)で選ぶキーフレームの数は、対象のシーンではない複数のキーフレームを混在したキーフレーム群より対象のシーンのキーフレームが選ばれる確率を調査した予備実験から、その確率と処理速度のトレードオフの関係を考慮し、5つとした。

表1に実験に使用した計算機のスペックを示す。

4.2 結果と考察

表2に提案手法と PTAMM の処理時間を示す。また図4の左に movieB を撮影した際のカメラ位置姿勢の推移をオフラインで高精度に推定した結果を示し、右に提案手法で movieB に対し、自己位置の推定を行った結果を示す。そして、図5にも同様にオフラインで推定したカメラ位置姿勢の推移を左に示し、右に PTAMM で movieB に対し、自己位置の推定を行った結果を示す。各図の赤い点はマップの点を表し、青い点と線分はカメラの軌跡を表す。今回の

表 2 提案手法と PTAMM の処理時間

Table2 Computation Time of Estimating Camera Pose

提案手法の処理	処理時間 (ms)
最初の自己位置の推定	1403.58
その後の自己位置の推定 (1 フレームの平均)	68.45
PTAMM の処理	
自己位置の推定 (1 フレームの平均)	34.58

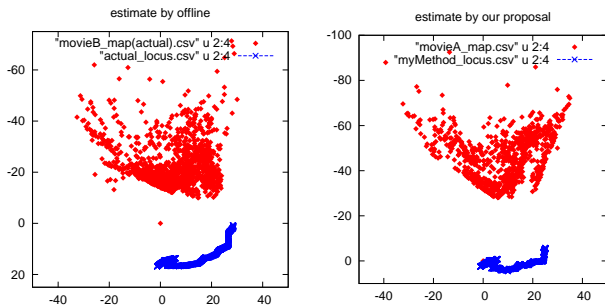


図 4 提案手法による自己位置の推定結果 (X-Z 平面)

Fig.4 Estimated Trajectory by Our Proposal

実験では、マップのスケールが手法により異なるため、比較が行いやすいように手でスケールを調整している。

オフラインで *movieB* を撮影した際のカメラ位置姿勢の軌跡は、X 軸方向に進行しながらマップの点 (オブジェクト) を回りこむような動きをしている。その動きを PTAMM では直線で推定し、さらに進行方向に進むに連れて位置姿勢推定時の誤差が蓄積している。提案手法のカメラ位置姿勢の軌跡はオフラインの推定結果と同じようにマップの点を回り込んだ軌跡を推定できており、PTAMM に比べ推定誤差が少ない。シーン内の凹凸面を持ったオブジェクトに対し、提案手法の特徴量辞書の効果が出ていることが確認できた。

また、提案手法の処理時間は最初の自己位置の推定に約 1.5 秒の時間を要するものの、その後の自己位置の推定は約 14.7fps で動作することを確認できた。実際の使用では、ユーザーが 1 秒程度カメラを自己位置推定の対象シーンに向けて静止し、初期の自己位置の推定が終わった後に AR 提示を行うといった形になる。

5. まとめ

本研究では、自己位置推定に使用するマップ内の点に対し、それが映されたカメラ位置姿勢の範囲と対応付けた BRIEF 特徴量を複数付加した特徴量辞書を用いて、マップの点と自己位置推定時のカメラ映像から得られる特徴点との同定処理を行い、情報端末の位置姿勢の推定を行う手法を提案した。

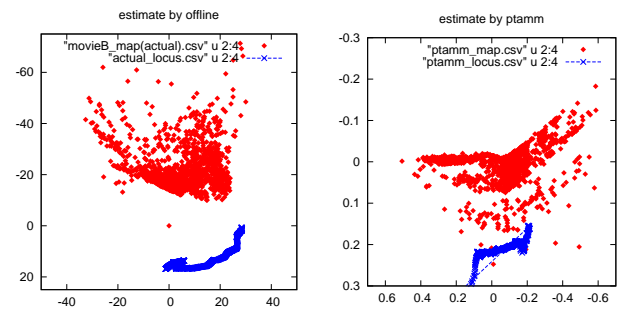


図 5 PTAMM による自己位置の推定結果 (X-Z 平面)

Fig.5 Estimated Trajectory by PTAMM

今後の課題として、自己位置推定時の LK 法により探索するマップの点の 2 次元位置を Harris オペレータにより特徴点候補とした特徴点から探索する機構の実装や、自己位置推定の状況に応じてカメラ映像内の特徴点との同定処理に使用するマップの点数を適宜変更していく機構の実装、さらに同一の屋外シーンに対して、マップ生成時と自己位置推定時の日照が大きく異なる環境下での自己位置推定の実験を行うことが挙げられる。

参考文献

- [1] D. Lowe : "Distinctive image features from scale-invariant keypoints", Proc. International Journal of Computer Vision, pp.91-110, 2004.
- [2] H. Bay, T. Tuytelaars, and L. V. Gool : "SURF: Speeded Up Robust Features", Proc. 9th European Conference on Computer Vision, 2006.
- [3] 神山 朗, 赤嶺 有平, 他 : "特徴量辞書を持つ 3 次元マップを用いた屋外における自己位置推定手法の提案", FIT2011 第 10 回情報科学技術フォーラム, 2011.
- [4] M. Calonder, V. Lepetit, C. Strecha, and P. Fua : "Brief: Binary robust independent elementary features", Proc. European Conference on Computer Vision, pp.778-792, 2010.
- [5] R. Castle, Georg Klein, and D. Murray : "Video-rate Localization in Multiple Maps for Wearable Augmented Reality", Proc. 12th IEEE International Symposium on Wearable Computers, 2008.
- [6] C. Harris and M. Stephens : "A combined corner and edge detector", Proc. 4th Alvey Vision Conference, pp.147-151, 1998.
- [7] digi-lab : "Voodoo Camera Tracker", 入手先 (<http://www.ipsj.or.jp/journal/submit/manual/j-manual.html>), 2008.
- [8] B. D. Lucas and T. Kanade : "An Iterative image registration technique with an application to stereo vision", Proc. Imaging Understanding Workshop, pp.121-130, 1981.
- [9] M. A. Fischer and R. C. Bolles : "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", Commun. ACM, no.24, vol.6, pp.381-395, 1981.