Vol.2011-CVIM-177 No.27
2011/5/19

俊敏な移動体のための画像を用いた自己位置推定

柳澤　悼[†1]　岡谷貴之[†1]　出口光一郎[†1]

本研究では，俊敏に運動する移動体の位置姿勢をその移動体に搭載したカメラのみから得られる画像のみを用いて推定する方法を示す．俊敏な運動に対応するため，SfMを用いた位置姿勢推定のように，撮影された画像列における類似度や連続性を用いることはせず，事前に構築した周囲環境の３次元構造およびその見かけの情報によって構築されたデータベースと，撮影した画像１枚１枚独立に対応をとり，位置姿勢の推定を行う．

３次元モデルを事前に構築して用いるため，対象空間が大きくなればなるほどそのデータベースは大きくなるため，計算量の削減や，効率のよい計算方法が課題となる．

Image-based Self-localization for Agile Mobile Robots

Jun Yanagisawa ,[†1] Takayuki Okatani [†1]
and Koichiro deguchi[†1]

This study considers a problem of estimating the 3D pose of an agile, mobile robot from the images obtained by a camera mounted on it. Unlike SfM (Structure from Motion), the agile motion of the robot prevents us from assuming temporal similarity or continuity of the image sequence captured, Thus, we estimate the pose of the robot at an instant time directly from the single image captured by establishing the correspondence between the image and the database of the surrounding scene. The database contains the 3D structure and appearance of the scene and is generated beforehand by some method. Since the 3D database grows with the size of the target scene, the central issue to be solved is the reduction in computational time and the development of an efficient computation method.

†1 東北大学
Tohoku University

1. はじめに

近年，様々なロボットが開発され実際に利用されるようになってきている．中でも，自然災害で被災した市街地において，倒壊した建物内部など，人が入るのが困難な場所での要救助者の有無や内部の状態を探索するロボット，火星や月といった人を送り込むのが難しい未開拓地を探索するロボット，また身近な環境における建物の警備や掃除，介護などを行うロボットなどがある．今後は自律移動型のものが期待されているこれらのロボットが，様々な環境，状況において自律制御を行うためには，移動時の自身の位置姿勢を周囲環境とともに知ることは必要不可欠であり，これまでにコンピュータビジョンの分野においても様々な研究が行われてきた．SfM(Structure from Motion)，あるいはロボティクスにおける Visual SLAM（Simultaneous localization and mapping）のように，センサとしてカメラを用いて，撮影した画像列からカメラの位置姿勢を推定する手法は，他のセンサが抱える問題のいくつかから解放される上，一般に非常に高精度な推定が実現できる可能性があることから，現在多くの研究が行われている．

本研究では，移動体の自己位置姿勢をカメラのみを用いて推定する問題を，特に「俊敏」に運動する移動体を対象に考える．ここでいう「俊敏」な運動とは，ロボットに搭載したカメラから得られる画像列において，SfM や Visual SLAM で行われる画像系列間における特徴点の追跡を行うのが困難となるような，フレーム間での画像の類似度や連続性が小さくなる運動とする．このようなロボットを対象に，カメラのみを用いて位置姿勢推定を行うために，本研究ではカメラから時々刻々得られる画像１枚ずつを独立に処理し，その各画像取得時点での移動体（カメラ）の空間での位置姿勢を推定する．その際，あらかじめ構築してある周囲環境の３次元構造およびその見かけ（アピアランス）の情報を利用する．

この背景には，最近のこの分野のいくつかの研究の進展がある．まず，最近の SfM 技術の発展により，大規模な３次元空間が容易にモデル化できるようになったことが挙げられる．代表的なものとして Snavely による Bundler[1)] がある．Bundler は Phototourism や Photosynth のベースとなっている SfM ソフトウェアで，インターネット上の旅行者が撮影した未整列である写真コレクションを元に，観光地や景勝地の 3 次元空間を計算する技術を実現した．次に，一般物体認識に代表される，物体のアピアランスに基づく画像認識の技術の進展がある．特に，画像の局所特徴と機械学習の手法に基づく物体認識手法が大きく発展した．そして，これら２つの要素を組み合わせた，３次元空間の認識方法が近年，現実のものとなりつつある．本研究では，これら２つの要素を組み合わせ研究の延長線上に

上述の問題—既知環境において俊敏に運動する移動体の位置姿勢を，カメラで撮影した画像 1 枚のみから推定する問題—を捉え，その解決策を探る．

2. 関 連 研 究

一般的に，SfM や VisualSLAM はカメラが空間を制約なく移動し，画像を時々刻々撮影し，その画像列から毎フレームのカメラ位置姿勢推定（トラッキング）と環境の 3 次元復元（地図生成，マッピング）を同時に実時間で行う．本研究のように位置姿勢推定と地図生成を分離して行う手法は，SfM や VisualSLAM の現在までの発展を考えれば，むしろ古典的である．しかし，トラッキングとマッピングを分離する方法は本研究の目的でも述べたように大規模な 3 次元空間のモデル化が容易になったことや，物体のアピアランスに基づく画像認識の技術の進展を背景に再び注目されるようになってきている．

カメラ位置姿勢推定と 3 次元復元を分離した方法の広大な空間を対象としたときの大きな問題のとしては，マッピングする対象空間が大きくなればなる程，3 次元モデルのデータベースが膨大になることがある．そのデータベースを対象にトラッキングすれば必然的に計算量が非常に大きくなる．この問題を解決するために，計算量やデータベースの削減について様々な研究が行われている．一つの方法として，Nister らの Vocabulary tree[2] を基にしたデータベースの階層構造化がある．Vocabulary tree は，Sivic らによって提案された画像を visual words の集合として表現する方法[3] を階層構造にしたものである．

Schindler らは，Vocabulary tree の階層構造により柔軟性を持たせることで，Nister らが類似画像探索に用いたデータベースより大きな 3×10^4 枚もの道沿いの画像をもとにしたデータベースを対象として位置推定を行った[4]．

また Arnold Irschara らは，理想的な最小限のデータで構成された 3 次元モデルデータベースを再構築することで，データベースの削減を行った[5]．対象となる空間を満遍なく撮影したの大量の写真から，すべてを用いて SfM によって 3 次元モデルデータベースを構築した場合，空間上の同一点が，複数の写真で異なる角度，大きさで撮影されることにより，同一点に対し特徴点の対応をとるときに区別する必要のないレベルに異なる特徴量を重複して持たせることになる．そこで Irschara らは，実際の写真における特徴量を基にした仮想的なイメージ（写真）を作成し，実際の写真とその仮想的なイメージによって構成する 3 次元の情報を満遍なく，かつ最小数のイメージで構成したデータベースを構築した．

また，実際の都市空間を対象とした自己位置推定を行う場合，規則的に整列した窓枠を持つビルが立ち並ぶ空間などは，特徴点が少ないうえに，似たような外観が多い．そのような状況では，特徴点抽出が難しく，かつそれに続く特徴点対応にも誤対応が生じやすいため，正確な位置推定は非常に難しい．

そこで Zach らは，より誤対応のすくない対応関係を生成するために，特徴点の対応をグラフ構造にした場合におけるループを利用することで，それらの誤対応を事前に除去することを可能にした[6]．対応関係のグラフに内在するループグラフにおいて，画像の対応関係が正しいならば，あるノードからスタートして，画像ペアマッチングによって得られたカメラ間の回転や homography 行列をエッジに沿ってかけ合わていくと，スタートノードに戻ることができるが，誤対応を含むループでは戻ることができないことを利用する．

3. カメラの位置姿勢の決定

本手法では，まず，移動体の走行空間をカメラを使って異なる方向から複数枚撮影し，それらの画像から SfM により，オフラインで 3 次元モデルを構築する．この 3 次元モデルの構築時に，各画像の局所特徴量を空間の同一点に対し得て，データベースを構築しておく．

移動体の自己位置姿勢推定時には，搭載カメラから得られるある時刻の画像 1 枚（以下クエリ画像と呼ぶ）の特徴点における局所特徴量を抽出し，これと上述のデータベースを用いて，クエリ画像の特徴点と 3 次元空間モデルの点とを対応付け，カメラの自己位置姿勢を推定するのに必要な情報を得る．手順は以下のような流れとなる．

(1) クエリ画像からの（既知 3 次元モデル構成用画像に用いた同一の）局所特徴量の抽出
(2) 既知 3 次元モデルの特徴点とクエリ画像の特徴点の最近傍探索
(3) カメラの位置姿勢の決定

4. 実　　験

本研究で提案した手法を用いて，実際に自己位置推定を行った結果を示す．自己位置推定を行う空間の 3 次元モデルの復元（既知 3 次元モデルデータベースの構築）は，デジタルカメラで撮影した画像を元に標準的な SfM[1] により行った．

現段階では移動体のカメラから得られる各クエリ画像に対しフレームレートでの位置推定を行うことができていない．本研究では，既知 3 次元空間でカメラを搭載した移動体を走行させ撮影を行った後，得られた画像列を基にオフラインでカメラの位置推定を行った．ただし，本研究で用いたカメラのフレームレート（本研究に用いた USB 無線カメラは最大 30fps）と本研究の位置推定全プロセスの所要時間にそれほど差があるわけではない．フレームレート内での位置姿勢推定，すなわち実時間での推定を行える見込みは十分にある．

Vol.2011-CVIM-177 No.27
2011/5/19

4.1 実験環境

移動体として図 1 に示したラジコン（シー・シー・ピー社製 G-bound）に USB 無線カメラを搭載したものを用いた．このラジコンは，非常に俊敏な動作が特長である．

図 1 本研究で用いたラジコンカーと搭載無線カメラ.

4.2 実験結果

自己位置推定結果は図 2 に示す．カメラの自己位置姿勢を一定の精度で得ることができているが，明らかに間違った位置姿勢推定が行われているのも確認できる．この結果を大きく乱した原因の一つには，本研究で用いた無線カメラがローリングシャッタを使用していたため，撮影された画像が幾何学的に非線形の変形をしてしまっていたことがある．この問題は，グローバルシャッタを使用したカメラを用いることで改善できる．他の原因としては，本実験では既知３次元モデルを周囲 360° 全てを復元せずに行ったため，ラジコンの制御上，復元されていない空間が撮影されてしまい，位置推定が行なえなくなったことがある．位置姿勢推定システムの高速化やデータベースの削減，計算量の削減により，周囲 360° 全てを復元した，より大きなデータベースを対象に位置姿勢推定が行えるにする必要がある．

5. 今後の課題

今後の課題としては，まず本手法における自己位置推定方法をカメラのフレームレートにおさめ，実時間においての自己位置推定を実現することがある．そのためには既知３次元モデルのデータベースの削減，各ステップでの計算量の削減による高速化を行う必要がある．

参考文献

1) N.Snavely, S.M.Seitz, R.Szeliski. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision,* Vol.80, No.2, pp.189-210, 2008.
2) D.Nister H.Stewenius. Scalable recognition with a vocabulary tree. Proceedings of *Computer Vision and Pattern Recognition*, 2006.
3) J.Sivic, A.Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. Proceedings of *Computer Vision and Pattern Recognition,* pp.1470-1477, 2003.
4) G.Schindler, M.Brown, R.Szeliski. City-Scale Location Recognition. Proceedings of *Computer Vision and Pattern Recognition,* pp.17-22, 2007.
5) A. Irschara, C. Zach,J-M. Frahm, H. Bischof. From Structure-from-Motion Point Clouds to Fast Location Recognition. Proceedings of *Computer Vision and Pattern Recognition,* 2009.
6) C. Zach, M. Klopschitz, M. Pollefeys. Disambiguating visual relations using loop constraints. Proceedings of *Computer Vision and Pattern Recognition,* pp. 1426-1433, 2010.
7) D.G.Lowe, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision,* Vol.60, No.2, pp.91-110, 2004.
8) R.M.Haralick, C.N.Lee, K.Ottenberg, M.Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision,* Vol.13, No.3, pp.331-356, 1994.
9) G.Klein, T.Drummond. A single-frame visual gyroscope. Proceedings of *British Machine Vision Conference,* pp.5-8, 2005.

Vol.2011-CVIM-177 No.27
2011/5/19

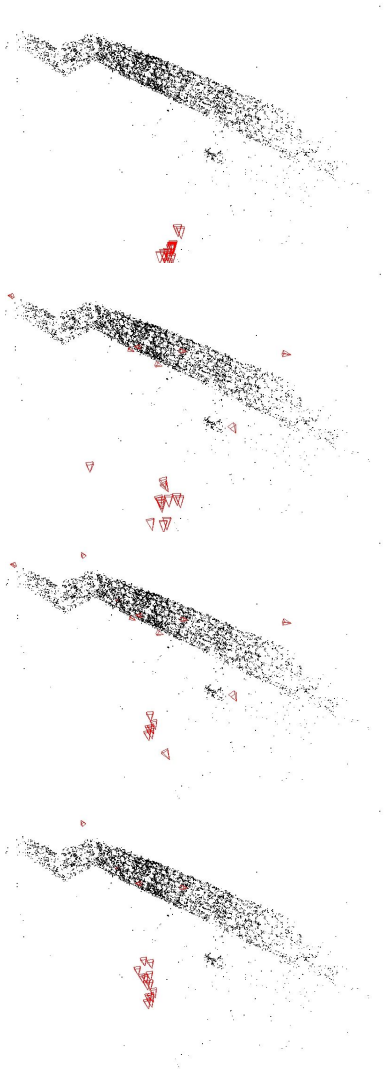

図 2 カメラをラジコンに搭載した自己位置姿勢推定の結果