

自然特徴点ランドマークデータベースに基づく

静止画像からのカメラ位置・姿勢推定

奈良先端科学技術大学院大学 情報科学研究科

佐藤 智和, 薄 充孝, 中川 知香, 横矢 直和

デジタルカメラや携帯電話に内蔵されたカメラで撮影された 1 枚の静止画像を入力として、カメラの位置・姿勢を推定する手法について解説する。本研究では、あらかじめ全方位型マルチカメラシステムを用いて撮影した全方位動画画像から自然特徴点を抽出し、これをランドマークとしてデータベースに登録しておくことで、静止画像 1 枚を対象としたカメラ位置・姿勢推定を実現する。また、推定されたカメラパラメータを用いて入力画像上に道案内情報などの注釈を重畳表示しユーザに提示することで、拡張現実感ナビゲーションへの応用が可能であることを示す。

〇はじめに

携帯電話上で動作するヒューマンナビゲーションシステムや、ユーザの現在位置に応じた情報配信を行うサービスが既に実用化されている。現在、これらのサービスでは、携帯機器に内蔵された GPS や複数の基地局からの電波強度を利用してユーザの位置を特定することで、2 次元の地図上での道案内や位置に依存した情報配信を実現している。しかし、GPS や電波強度によるユーザ位置の測定結果には 10m から 100m 程度の誤差が含まれており、また屋内環境などのように GPS による電波が遮蔽される環境においては測位精度が大幅に低下することが問題となる。これに対して、本研究では 1 枚の画像中から、事前に登録された自然特徴点ランドマークを抽出することで、屋内外を問わずユーザが持つカメラの位置・姿勢を高精度に推定する手法を提案する。本手法を用いれば、カメラの位置だけでなく姿勢の情報も計測可能となるため、2 次元地図と比較してより直感的な道案内を実現する拡張現実感ナビゲーション¹⁾²⁾への応用が可能となる。

拡張現実感ナビゲーションを実現するためのカメラ位置・姿勢推定手法としては、GPS に加えてジャイロやコンパスなどの複数のセンサを組み合わせる方式が提案されている¹⁾²⁾が、センサの実装コスト・サイズの問題から一般的な携帯電話で利用することは難しく、また屋内での利用が難しいという問題も残されている。一方、本研究と同様に、入力画像中の自然特徴点やエッジなどの自然特徴をあらかじめ構築されたデータベース内の情報と照合することで、特殊なセンサを用いることなくカメラ位置・姿勢を推定する手法³⁾⁴⁾も既に提案されているが、従来手法の多くは動画画像による入力を前提としており、計算機リソースの限定された携帯電話上において計算機負荷の大きいこれらの手法を動作させることは難しい。

これに対して、本研究では以下のような枠組みで静止画像 1 枚を入力としたカメラ位置・姿勢推定を行うことで、ユーザが持つ端末上での計算コストを最小限に抑える。

- (1) ユーザは携帯電話に内蔵されたカメラで静止画を撮影し、GPS や電波強度による位置情報と共に写真をサーバに送信する。
- (2) サーバは事前に構築されている環境のデータベースと写真を照合し、カメラ位置・姿勢を推定する。
- (3) サーバは受信した写真にナビゲーション情報および各種ガイド情報を合成し、ユーザに返送する。

上記の枠組みにより、現在既に市販されているカメラ付き携帯電話や、将来においても大勢を占めるであろう比較的安価で低機能な携帯端末上においても拡張現実感によるナビゲーションを実現することができる。

本手法の処理の流れを図 1 に示す。本研究では、まず対象となる環境を全方位カメラで移動撮影し、オフライン処理で structure from motion 法(以下、SFM 法)による三次元復元を行うことで、自然特徴点の三次元位置とその見え方の情報をランドマークとしてデータベースに登録する。オンライン処理では、データベースに登録された膨大な数のランドマークから、入力画像中の自然特徴点に対応する正しいランドマークを検索するために、GPS また

は電波強度による位置情報、ランドマークの類似度、ランドマーク観測時のカメラ位置の整合性、を順に用いてカメラ位置・姿勢推定に利用するランドマークを段階的に絞り込み、最終的に誤対応を排除した上で6自由度でのカメラ位置・姿勢推定を行う。

○ランドマークデータベースの構築

図1に示したように、本研究では、まず撮影対象となる環境を全方位カメラを用いて撮影し、structure from motion法によって自然特徴点の三次元位置と全方位カメラのカメラパラメータを算出する(A-1)。次に、三次元復元結果に基づきランドマークデータベースを構築する(A-2)。

全方位カメラによる環境の三次元復元

本研究では、対象となる環境を多眼型の全方位カメラを用いて移動撮影し、Harris オペレータを用いて動画像中から物体の角などの自然特徴点を抽出する。次に、抽出された自然特徴点を画像間で対応づけ、動画像全体に対して再投影誤差の総和を最小化することで、自然特徴点の三次元位置と全方位カメラの位置・姿勢を推定する。ここで、実環境と自然特徴点の三次元位置の間の幾何学的な位置関係は、基準マーカと自然特徴点を同時に画像中で追跡する手法⁵⁾またはGPSによる位置計測情報と画像からの三次元復元情報を併用する手法⁶⁾を用いて求める。

ランドマーク情報の獲得

ランドマークデータベースの構成要素を図2に示す。ランドマークデータベースは(I)自然特徴点の三次元位置と(II)撮影地点ごとの情報から成る。ランドマークの見え方は撮影地点によって異なるので、本研究では各ランドマークに複数の撮影地点情報を登録する。撮影地点情報は、(II-a)ランドマーク観測時のカメラ位置、(II-b)自然特徴点の固有スケール、(II-c)SIFT特徴ベクトルから成る。本研究では、全方位カメラによる環境の三次元復元結果を、自然特徴点の三次元位置(I)およびランドマーク観測時のカメラ位置(II-a)としてそのまま利用する。また、自然特徴点の固有スケール(II-b)、SIFT特徴ベクトル(II-c)の算出には、それぞれHarris-Laplacian⁷⁾およびSIFT-descriptor⁸⁾を用いる。

○静止画像からのカメラ位置・姿勢推定

全方位カメラを用いて構築した自然特徴点ランドマークデータベースを利用して、1枚の静止画像から撮影時のカメラ位置・姿勢を推定する。ここでは、まずGPSや携帯電話の電波強度を用いておおよそのカメラ位置を特定することでデータベースの選択を行う(B-1)。次に、SIFTを用いた類似度評価により入力画像上の自然特徴点と対応付くランドマークの候補を複数選択する(B-2)。更に、選択された各ランドマークが、それぞれ対応付いた入力画像上の自然特徴点と同じ見え方で観測できるカメラ位置・姿勢を算出し、投票によって1地点から最も多くのランドマーク候補を観測できるカメラ位置・姿勢の候補を決定する(B-3)。最後に、決定されたカメラ位置・姿勢の候補に投票したランドマーク群を用いて最終的に6自由度のカメラ位置・姿勢推定を行う(B-4)。

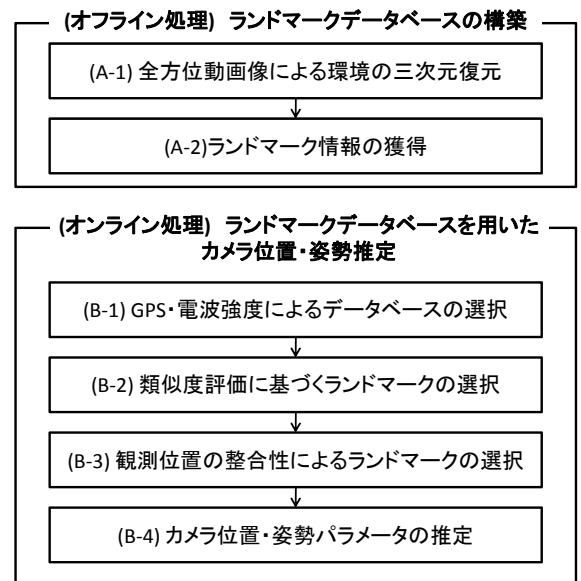


図1 カメラ位置・姿勢推定処理の流れ

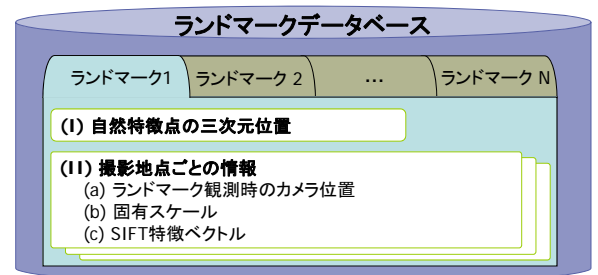


図2 ランドマークデータベースの構成要素

GPS・電波強度によるデータベースの選択

本研究では、あらかじめ多数の地点・地域においてランドマークデータベースが構築されていることを想定し、GPS または電波強度を用いて検索に用いるデータベースを選択する。ここでは、ランドマーク観測時のカメラ位置を基準に、あらかじめランドマークデータベースが 100m×100m 程度の単位に分割されているものとし、まず GPS・電波強度によって得られる観測位置周辺に存在するデータベースを全て選択する。

類似度評価に基づくランドマークの選択

入力画像上で検出された自然特徴点と見え方が類似したランドマークをデータベースから選択する。ここではまず、入力画像上の自然特徴点を抽出し、抽出された各自然特徴点に対する SIFT 特徴ベクトルをデータベース構築時と同様の手法によって算出する。次に、入力画像上の自然特徴点とランドマークの類似度を算出する。最後に、各自然特徴点に対して算出された類似度を昇順に並び替え、類似度が上位の一定個数のランドマークを自然特徴点と対応づける。これにより、画像上の各自然特徴点との類似度が高い複数のランドマークをデータベース中から選択する。

観測位置の整合性によるランドマークの選択

類似度に基づいて対応付けられた自然特徴点とランドマークの組み合わせには、自然特徴点と真に対応するランドマーク以外の誤対応が多数存在する。提案手法では、このような誤対応を排除するために、入力画像が環境中の単一の位置・姿勢で撮影されているという事実に着目し、前節の処理で選択されたランドマークを最も多く観測可能なカメラ位置・姿勢を投票によって算出する。またこれにより、投票値が最大となったカメラ位置・姿勢以外に投票したランドマークを排除する。

カメラ位置・姿勢パラメータの推定

投票により抽出されたランドマークと各ランドマークに対応付けられた入力画像上の自然特徴点を用いて 6 自由度のカメラの位置・姿勢を推定する。ここでは、各ランドマークを画像上に投影した座標と、各ランドマークに対応する自然特徴点の画像座標の間の二乗距離(再投影誤差)の総和を最小化することで 6 自由度のカメラ位置・姿勢を推定する。ただし、ここでは投票結果からランダムに自然特徴点・ランドマークの組を繰り返し抽出し、LMedS 基準⁹⁾を最小とする自然特徴点・ランドマークの組を抽出することで、投票結果から誤対応を含まない自然特徴点・ランドマークの組を抽出しカメラ位置・姿勢を推定する。

○実データを用いた実験

提案手法の有効性を示すために屋外環境においてランドマークデータベースを構築し、提案手法によるカメラ位置・姿勢推定の精度と推定成功率を検証した。

データベースの構築

本実験ではまず、図 3 左に示す全方位型マルチカメラシステム (PointGreyResearch 社製 Ladybug) を用いて、実験対象となる環境内の 2 本の経路上を移動しながら撮影し、ランドマークデータベースを構築した。Ladybug は水平方向に 5 つ、上方向に 1 つの

合計 6 つのカメラユニットから構成されており、各カメラユニットはそれぞれ 768×1024 画素の解像度で動画を同期撮影できる。ここでは、図 3 右に示す 6 枚の画像を含む、合計 7,200 枚(1,200 フレーム)の撮影画像を入力として用い、動画像上の自然特徴点の動きから自然特徴点の三次元位置と全方位カメラのカメラパラメータを復元した。次に、得られた三次元復元結果を用いてランドマークデータベースを構築した。図 4 に、データベースに登録



図 3 データベース構築に用いた全方位型マルチカメラシステムと入力画像の一部

されたランドマークの三次元位置と全方位カメラの位置の関係を、地表に対する上面図として示す。同図中の太い実線は全方位カメラのカメラパスを、点群はランドマークの位置を表している。本実験では、データベースに約 12,500 個のランドマークが登録され、各ランドマークに対して平均 7.5 地点の撮影地点ごとの情報が登録された。

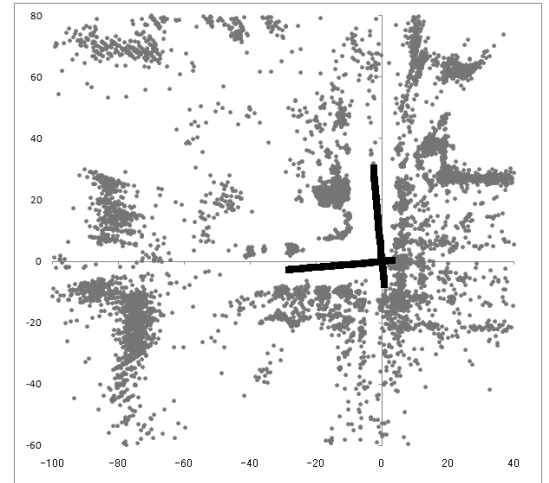


図 4 データベースに登録された自然特徴点と全方位カメラの位置

カメラ位置・姿勢の推定

市販の GPS・カメラ付き携帯電話(Casio 社製 GzOne W42CA)を用いて撮影した静止画像を用い、提案手法によるカメラ位置・姿勢の推定結果を真値と比較することで推定精度を評価する。カメラ位置・姿勢の真値は、あらかじめ環境内の自然特徴点をトータルステーションで計測し、入力画像上で位置を手動で指定した上で再投影誤差の最小化によりカメラ位置・姿勢を算出することで作成した。本実験ではサーバ・クライアント型システムは構築せず、携帯電話による画像撮影後に PC 上に画像を転送した上でカメラ位置・姿勢推定処理を行った。また、用いたデータベースの規模が比較的小さいため、今回は登録されたすべてのランドマークを用いて実験を行った。まず、図 5 に示すように、データベース構築時の全方位カメラの撮影経路周辺において、5m 間隔の格子点上(6×6=36 地点)で、異なる 2 方向(方向 1、方向 2)に対して 640×480 画素の 72 枚の静止画像を撮影した。本実験では、72 枚の画像のうち、カメラ位置・姿勢の真値を作成することが可能な 65 枚の画像を入力画像として用いた。

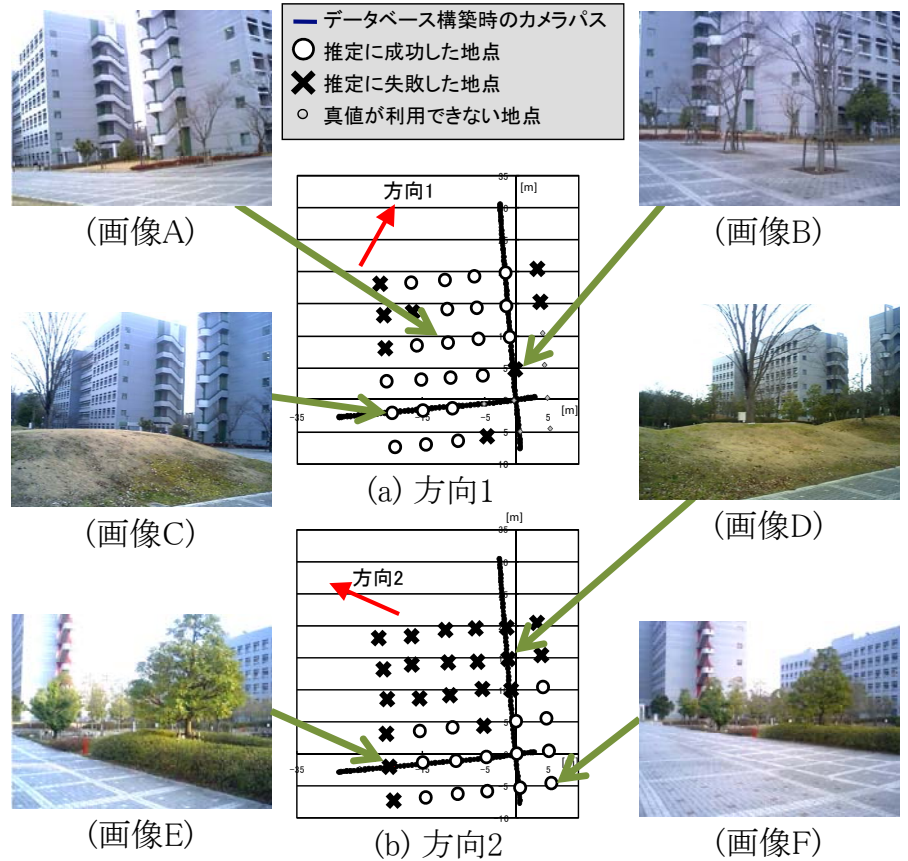


図 5 入力画像の撮影地点、撮影画像例と各地点における推定成功・失敗の判定結果

図 5 に、再投影誤差が 5 画素を超えた場合を失敗とみなした場合の、撮影地点ごとの推定成功・失敗の結果を示す。画像中に人工物を主に捉えることができる方向 1 に関しては、真値が利用可能な撮影地点について 72.4%の地点で推定に成功した。方

向 1 に対して成功と判定された全ての画像に対する推定誤差の平均は、位置誤差が 1.4m、姿勢誤差が 1.4度であり、これは静止画像による拡張現実感ヒューマンナビゲーションを行うために問題ない精度であると考えられる。これに対して方向 2 では、撮影された半数以上の画像上で画像中に自然物が大きく写り込んでおり、推定成功率は 41.7%となった。方向 2 に対して推定が成功と判定された全ての画像に対する推定誤差の平均は、位置誤差が 6.8m、姿

勢誤差が 3.9 度であり、方向 1 に比べて推定精度が低い。この原因として、方向 1 の場合と比較して、カメラ位置から建物上のランドマークまでの距離が、方向 2 では 3 から 5 倍程度離れていること、また自然物による遮蔽によりカメラ位置・姿勢の推定に利用できるランドマーク数が少ないことが挙げられる。図 6 に、推定が成功した地点において、拡張現実感によるナビゲーション情報の提示を行った例を示す。同図中には、目的地や周辺の建物への案内情報を入力画像上に重畳表示しており、ユーザへの直感的な道案内報の提供を実現している。



図 6 推定結果を用いた拡張現実感ナビゲーションの実現例

○まとめ

サーバ・クライアント方式による携帯電話上での拡張現実感への応用を想定し、事前に構築したランドマークデータベースを用いる静止画像 1 枚からの新たなカメラ位置・姿勢推定手法を紹介した。本手法では、静止画像 1 枚からのカメラ位置・姿勢推定が可能であるため、現在普及しているカメラ付き携帯電話をそのまま用いることができるという特長を持つ。実験により、人工物を入力画像中に十分捉えている場合には、データベース構築時のカメラ位置から離れた地点においても、静止画像へのナビゲーション情報を重畳するために問題ないと考えられる精度でカメラ位置・姿勢推定を行えることを確認した。今後は、環境内に類似したランドマークが存在しない特徴的なランドマークを優先的に用いることで推定のロバスト性を高める手法を開発する。また、同一地点から異なる環境条件で撮影された複数の画像をデータベースの構築時に用いることで推定のロバスト性を高める手法についても検討する。

謝辞

本研究は、総務省戦略的情報通信研究開発推進制度(SCOPE)により実施したものである。

参考文献

- 1) 天目隆平, 神原誠之, 横矢直和: ``「平城宮跡ナビ」マルチメディアコンテンツを利用したモバイル型観光案内システム'', 第 1 回デジタルコンテンツシンポジウム講演予稿集, No. S3-6, 2005.
- 2) M. Kourogi, N. Sakata, T. Okuma and T. Kurata: ``Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/Self-contained sensor system'', Proc. Int. Conf. on Artificial Reality and Telexistence, pp.1310-1321, 2005.
- 3) L. Vacchetti, V. Lepetit and P. Fua: ``Stable real-time 3D tracking using online and offline information'', Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, No. 10, pp.1385-1391, 2004.
- 4) 武富貴史, 佐藤智和, 横矢直和: ``優先度情報を付加した自然特徴点ランドマークデータベースに基づく実時間でのカメラ位置・姿勢推定'', 画像の認識・理解シンポジウム(MIRU2008)講演論文集, pp. 860-865, 2008.
- 5) 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.
- 6) S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: ``Construction of feature landmark database using omnidirectional videos and GPS positions'', Proc. Int. Conf. on 3-D Digital Imaging and Modeling, pp.249-256, 2007.
- 7) K.Mikolajczyk and C.Schmid: ``Scale & affine invariant interest point detectors'', Int. Journal of Computer Vision, Vo. 60, No. 1, pp.63-86, 2004.
- 8) D.G. Lowe: ``Distinctive image features from scale-invariant keypoints'', Int. Journal of Computer Vision, Vo. 60, No. 2, pp.91-100, 2004.
- 9) P. J. Rousseeuw: ``Least median of squares regression'', J. of American Statistical Association, Vol. 79, pp.871-880, 1984.

キーワード

カメラパラメータ 画像撮影時のカメラの状態を表すパラメータ。内部パラメータと外部パラメータから成り、内部パラメータは焦点距離や歪み係数などのカメラの特性を、外部パラメータは撮影時のカメラの位置・姿勢を表す。
自然特徴点 人為的に設置・設定される点をマーカと呼ぶことに対して、画像上のエッジの交点や先端などのような、自然に存在する特徴的な点を表す。

著者紹介

佐藤 智和 奈良先端科学技術大学院大学 情報科学研究科 助教	薄 充孝 2007 年奈良先端科学技術大学院大学 情報科学研究科 博士前期課程修了。現在、トヨタ自動車株式会社勤務	中川 知香 2006 年 奈良先端科学技術大学院大学 情報科学研究科 博士前期課程修了。現在 富士ゼロックス株式会社勤務	横矢 直和 奈良先端科学技術大学院大学 情報科学研究科 教授
--------------------------------------	---	--	--------------------------------------

