

ARカーナビゲーションのためのランドマークデータベースを用いた 高精度な自車位置・姿勢の推定

武富貴史*, 佐藤智和 (奈良先端科学技術大学院大学), 山口晃一郎 ((株)豊田中央研究所),
横矢直和 (奈良先端科学技術大学院大学)

Accurate car position and posture estimation using a feature landmark database for AR car navigation
Takafumi Taketomi*, Tomokazu Sato (Nara Institute of Science and Technology)
Koichiro Yamaguchi (Toyota Central R&D Labs., Inc.)
Naokazu Yokoya (Nara Institute of Science and Technology)

Abstract

In this report, we propose a landmark-based car localization method for augmented reality (AR) based car navigation system. In the AR car navigation system, guide information is overlaid on the real scene and drivers can understand guide information more intuitively than 2-D map based car navigation system. To realize the AR car navigation system, car position and posture should be estimated in order to registrate guide information into images of the real scene. To realize accurate geometric registration, in this study, extrinsic camera parameters of a camera placed in front of a vehicle are estimated by using a feature landmark database. The feature landmark database is constructed by using structure-from-motion before using AR car navigation system. To show the effectiveness of the proposed method, the result of geometric registration between guide information and real scene is demonstrated in an outdoor environment.

キーワード：ランドマークデータベース, ARカーナビゲーション, 拡張現実感, カメラ位置・姿勢推定
(landmark database, AR car navigation, augmented reality, camera position and posture estimation)

1. はじめに

現在, 一般的に用いられているカーナビゲーションシステムでは, GPS および内界センサによって取得した位置と地図情報を利用して自車位置を推定し, 二次元の地図上でのナビゲーションを行っている. このような二次元の地図に基づくカーナビゲーションシステムでは, 実環境と案内情報の対応関係を瞬時に把握することが難しく, ナビゲーション画面を注視することによる交通事故の発生が問題となっている. これに対して, 実環境と案内情報の対応関係の把握を容易にするために, 交差点などにおいて三次元のCGモデルを表示する機能を有する製品が販売されているが, CGモデルの作成コストが大きく都市部などの主要な交差点でのみ利用が可能な状態となっている. このような問題に対して, 近年, 車載カメラにより実環境を撮影した画像上に案内情報を合成提示する拡張現実感技術を用いたカーナビゲーションシステム(以下, ARカーナビ)に関する研究が行われている⁽¹⁾⁽²⁾. ARカーナビでは, 実際の走行環境を撮影した画像上に案内情報の提示を行うため, 実環境と案内情報の対応関係を直感的かつ瞬時に把握することができるという特長を持つ. このようなARカーナビを実現するためには, 車載カメラの絶対位置・姿勢を推定する必要がある. 従来から様々なカメラ位置・姿勢推定手法が提案されているが, 本報告ではARカーナビでの

利用を想定し, 実環境と案内情報の画像上での位置合わせの精度を高めやすい, 画像に基づくカメラ位置・姿勢推定手法に着目する.

一般に, 画像からカメラの絶対位置・姿勢を推定するためには, 利用環境の事前知識が必要となる. 利用環境の事前知識として, 環境の三次元モデルを用いる方法⁽³⁾⁽⁴⁾が提案されている. しかし, 広域で複雑な屋外環境では環境の三次元モデル作成の人的コストが膨大となるため, これをARカーナビで利用することは難しい. 一方, これまで我々は, 環境中に存在する自然特徴点の三次元位置と自然特徴点周辺の局所的な画像情報をランドマークとしてデータベースへ登録しておき, ランドマークと入力画像中の自然特徴点を対応付けることでカメラ位置・姿勢の推定を行う手法を提案している⁽⁵⁾. ランドマークデータベースを用いる手法⁽⁵⁾では, structure-from-motion法を利用することで自動でランドマークデータベースを構築することができるため, 広域で複雑な屋外環境においても, 比較的容易にランドマークデータベースを構築することができる. 本研究では, ランドマークデータベースを用いる手法を基礎とし, カメラ位置・姿勢推定に車両の動き予測モデルを導入することによって, 推定されるカメラ位置・姿勢を安定化し, ARカーナビを実現する.

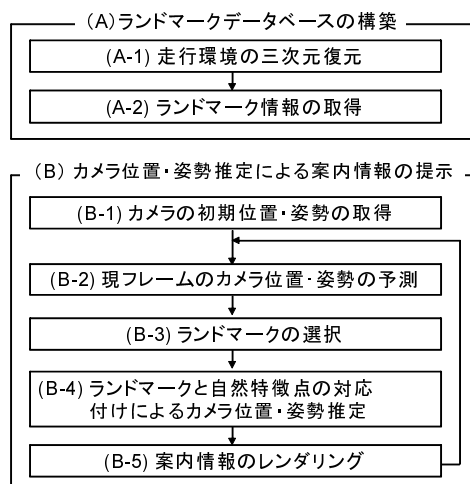


図1 全体の処理の流れ

2. AR カーナビのためのカメラ位置・姿勢推定

本研究では、ランドマークデータベースを用いて車載カメラの位置・姿勢を推定することにより、ARカーナビを実現する。提案手法の処理の流れを図1に示す。ランドマークデータベースを用いたカメラ位置・姿勢推定処理は、オフラインでのデータベースの構築処理(A)とオンラインでのカメラ位置・姿勢推定による案内情報の提示処理(B)からなる。

2.1 ランドマークデータベースの構築

本節では、ランドマークデータベース構築処理(A)について述べる。ここでは、ランドマークデータベースを構築するために、まず、全方位画像を用いた走行環境の三次元復元を行う。次に、三次元復元結果を用いてランドマーク情報を取得する。以下では、各処理について詳しく述べる。

(A-1) 走行環境の三次元復元

まず、走行環境内を移動しながら全方位カメラを用いて撮影する。次に、動画中の自然特徴点を自動追跡することで、structure-from-motionによって自然特徴点の三次元位置、および全方位カメラの外部パラメータを推定する。最後に、カメラパラメータと自然特徴点の三次元位置の累積的な推定誤差を最小化するために、三次元位置が既知の少数の基準点⁽⁶⁾またはGPSから得られる絶対位置⁽⁷⁾を利用して、動画全体での最適化処理を行う。

(A-2) ランドマーク情報の取得

本研究で用いるランドマークデータベースには図2に示すように、(a)自然特徴点の三次元位置、(b)撮影地点ごとの情報、(c)ランドマークの優先度の情報が保持される。以下ではそれぞれの情報の取得方法について述べる。

(a)ランドマークの三次元位置：オンライン処理では、ランドマークの三次元位置と画像上の二次元座標を対応付けることによりカメラの位置・姿勢を推定するため、ランドマークの三次元位置情報が必要となる。ランドマークの三次元位置は(A-1)で述べた処理によって得られるものであり、環境に固定された世界座標系で保持されている。

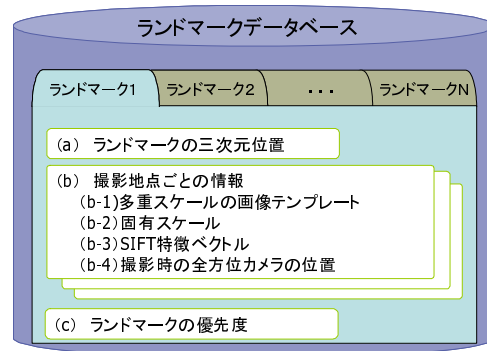


図2 ランドマークデータベースの構成要素

(b)撮影地点ごとの情報の取得：撮影地点によるランドマークの見え方の違いに対応するため、撮影地点ごとにカメラの撮影姿勢に依存しない多重スケールの画像テンプレート(b-1)を作成し、データベースへ登録する。また、カメラの初期位置・姿勢を推定するためにHarris-Laplacian⁽⁸⁾、SIFT-descriptor⁽⁹⁾を用いて、自然特徴点の固有スケール(b-2)、SIFT特徴ベクトル(b-3)を算出し、データベースへ登録する。さらに、オンライン処理でのランドマーク選択処理(B-3)において、撮影可能なランドマークを選択するために、ランドマーク撮影時の全方位カメラの位置をデータベースへ登録する(b-4)。

(c)ランドマークの優先度：オンライン処理でのランドマークの選択処理(B-3)において、正しく対応づく可能性の高いランドマークをデータベース中から効率的に選択するために、ランドマークに優先度を付加する。ここでは、事前にARカーナビを利用する環境でランドマークの優先度を学習するための動画を撮影しておき、撮影された学習用の動画に対して、後述する処理(B)と同様のカメラ位置・姿勢推定を行うことで、各ランドマークの優先度を決定する。本研究では、ランドマーク i の優先度 P_i として、ランドマーク i が推定に利用される確率 $P_i = E_i/D_i$ を用いる。ここで、 D_i 、 E_i は、それぞれ、学習データに対するカメラ位置・姿勢推定の処理結果について、ランドマーク i がデータベースから選択された回数とカメラ位置・姿勢推定の際に誤対応として排除されずに推定に用いられた回数を表す。

2.2 カメラ位置・姿勢推定による案内情報の提示
オンラインでのカメラ位置・姿勢推定による案内情報の提示処理(B)について述べる。オンライン処理では、図1に示したように、まず初期フレームでのカメラ位置・姿勢を取得し(B-1)、続いて処理(B-2)~(B-5)を繰り返す。本研究では、ランドマークの画像上の探索範囲を車両の動きを予測することによって限定することで、カメラ位置・姿勢推定処理の安定化を図る(B-2)。以下では、各処理について詳しく述べる。

(B-1) カメラの初期位置・姿勢の取得

本研究では、初期フレームのカメラ位置・姿勢は、ランドマークデータベースを用いた静止画像からのカメラ位置・姿勢推定手法⁽¹⁰⁾によって自動で取得する(B-1)。初期フ

レームのカメラ位置・姿勢推定処理では、まず、入力画像中の自然特徴点の固有スケール、SIFT 特徴ベクトルを算出した後、これらをデータベース中の全てのランドマークの SIFT 特徴ベクトルと照合することで自然特徴点とランドマークを対応付ける。次に、ランドマークの観測地点の整合性を利用して、対応付けられたランドマークから誤対応を排除する。最後に、得られた対応関係を用いて、PnP 問題⁽¹¹⁾を解くことで 6 自由度のカメラ位置・姿勢を推定する。

(B-2) 現フレームのカメラ位置・姿勢の予測

車両の位置・姿勢は連続フレーム間で急激に変化しないことから、車両の運動を局所的には等速運動で近似できるものとする。すなわち、現フレーム t のカメラ位置・姿勢 $M_{predict}$ を以下の式で予測する。

$$M_{predict} = M_{motion} M_{t-1} \dots \dots \dots (1)$$

ここで、 M_{t-1} は $t-1$ フレーム目で推定されたカメラ位置・姿勢を表し、 M_{motion} はフレーム間でのカメラの相対的な運動を表す。本研究ではフレーム間での細かな振動を抑制するため、 M_{motion} として、 $t-1$ フレームから $t-N$ フレームまでの各フレームに対するカメラの相対運動を用いる。

(B-3) ランドマークの選択

ランドマークの選択処理 (B-3) では、まず処理 (B-2) で予測されたカメラ位置・姿勢 $M_{predict}$ とランドマークが撮影された全方位カメラの位置の幾何学的な関係を用いて、予測されたカメラ位置から撮影可能なランドマークをデータベース中より選択する。次に、選択されたランドマークの候補から優先度の高い順にランドマークを K 個選択する。

(B-4) ランドマークと自然特徴点の対応付けによるカメラ位置・姿勢推定

処理 (B-3) で選択したランドマークと入力画像中の自然特徴点を対応付け、カメラ位置・姿勢の推定を行う (B-4)。ここではまず、予測されたカメラパラメータ $M_{predict}$ を用いて、処理 (B-3) で選択されたランドマークを入力画像上へ投影する。次に、投影したランドマークの画像上の位置を中心とする一定ウィンドウ W 内に存在する全ての自然特徴点周辺の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法によって照合する。最後に、ランドマークと対応付けられた自然特徴点を用いて、現フレーム t のカメラパラメータ M_t を推定する。現フレーム t のカメラパラメータ M_t は処理 (B-1) と同様に、誤対応の排除後に、PnP 問題を解くことによって推定する。

(B-5) 案内情報のレンダリング

推定された現フレームのカメラパラメータ M_t を用いて、カメラで撮影した画像上に事前に三次元位置が与えられた案内情報をレンダリングする (B-5)。本研究では、案内情報の画像上での提示位置を安定させるために、現フレーム t からフレーム $t-I$ までの案内情報の投影位置の重心に案内情報を提示する。



図 3 全方位画像

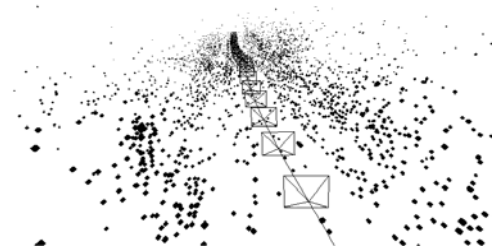


図 4 三次元復元結果

3. 実験

ここでは、従来手法⁽⁵⁾との比較により、提案手法の有効性を示す。本実験では、自動車に全方位型マルチカメラシステム (Point Grey Research 社 Ladybug2) を固定し、屋外環境 (約 120m) を走行しながら撮影した全方位動画像 (500 フレーム) を用いてランドマークデータベースの構築を行った。図 3 に、データベースの構築に用いた全方位動画像の一部を示す。また、データベース構築時のカメラパスと復元されたランドマークの三次元位置を図 4 に示す。本実験では、全方位動画像撮影の際に、助手席に単眼カメラ (SONY DSR-PD-150) を固定することで動画像 (解像度 720×480 画素, プログレッシブスキャン, 15fps, 400 フレーム) を撮影し、評価用として用いた。また、提案手法において、カメラの動きの算出 (B-2) に用いるフレーム数を $N = 10$ とし、案内情報のレンダリング (B-5) では $I = 5$ フレーム分の平均位置を案内情報の提示位置として用いた。なお、従来手法、提案手法ともにデータベースより選択するランドマーク数 K を 70 個とし、ランドマークと自然特徴点の対応付け処理 (B-4) のウィンドウ W を従来手法では 120×60 画素、提案手法では 30×30 画素とした。

従来手法 (予測無し) により推定されたカメラパスを図 5 に、提案手法 (予測有り) により推定されたカメラパスを図 6 に示す。本実験では、データベースの構築用の全方位動画像と案内情報合成用の入力画像を同時に取得したため、

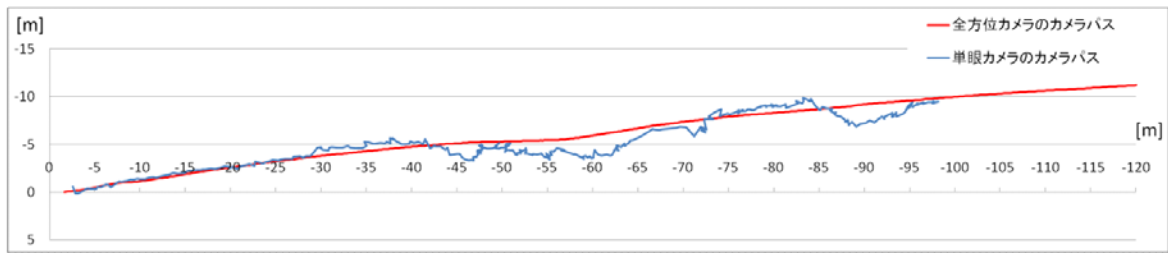


図 5 推定されたカメラパス（予測無しの場合）

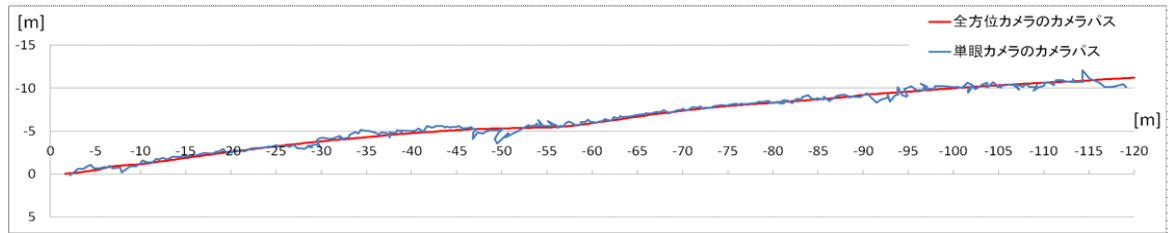


図 6 推定されたカメラパス（予測有りの場合）



130 フレーム目 160 フレーム目

図 7 案内情報の提示結果

本来は全方位カメラのカメラパスと単眼カメラのカメラパスがおおむね一致する．これに対して，従来手法を用いて推定されたカメラパスは全方位カメラのカメラパスから大きく外れており精度が低い．一方，提案手法を用いて推定されたカメラパスは細かな振動はあるが，全方位カメラのカメラパスと同様の軌跡を描いていることが分かる．このことから，提案手法は，従来手法よりも安定したカメラ位置・姿勢推定処理を実現できていることが分かる．

図 7 に，提案手法によって推定されたカメラ位置・姿勢を用いて案内情報を提示した結果を示す．生成された動画から案内情報が位置ずれなく幾何学的に正しい位置に描画されていることを確認した．

4. まとめ

本報告では，ランドマークデータベースを用いたカメラ位置・姿勢推定手法⁽⁵⁾に予測モデルを導入することで，カメラ位置・姿勢推定処理の安定化を行い，AR カーナビへの応用が可能であることを示した．今後は，よりロバストなカメラ位置・姿勢推定処理を実現するために，車載センサ情報との統合を図る．

謝 辞

本研究の一部は，科学技術振興機構 (JST) の戦略的創造研究推進事業 (CREST) 「デジタルメディア作品の制作を支援する基盤技術」プログラム，総務省・戦略的情報通信研究開発推進制度 (SCOPE) の支援による．

参考文献

- (1) 赤穂, 加藤, 西田: “実写ベースカーナビにおける交差点案内の検討と評価”, 信学技報 MVE, pp. 1-6, 2006.
- (2) 澤野, 岡田: “車載カメラによる実時間画像処理とその AR 技術に基づく表示方式によるカーナビへの応用”, 芸術科学会論文誌, Vol. 5, No. 2, pp. 57-68, 2006.
- (3) V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: “Fully automated and stable registration for augmented reality applications,” Proc. ISMAR, pp. 93-102, 2003.
- (4) L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3D camera tracking,” Proc. ISMAR, pp. 48-57, 2004.
- (5) T. Taketomi, T. Sato and N. Yokoya: “Real-time camera position and posture estimation using a feature landmark database with priorities,” CD-ROM Proc. ICPR, 2008.
- (6) 佐藤, 池田, 横矢: “複数動画からの全方位型マルチカメラシステムの位置・姿勢パラメータの推定”, 信学論 (D-II), Vol. J88-D-II, No. 2, pp. 347-357, 2005.
- (7) S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: “Construction of feature landmark database using omnidirectional videos and GPS positions,” Proc. Int. Conf. on 3DIM, pp. 249-256, 2007.
- (8) K. Mikolajczyk and C. Schmid: “Scale & affine invariant interest point detectors,” Int. J. of Computer Vision, Vo. 60, No. 1, pp. 63-86, 2004.
- (9) D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int. J. of Computer Vision, Vol. 60, No. 2, pp. 91-100, 2004.
- (10) 薄, 中川, 佐藤, 横矢: “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”, 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 161-170, 2008.
- (11) R. Klette, K. Schluns and A. Koschian Eds: Computer Vision: Three-dimensional Data from Image, 1998.