

ランドマークデータベースを用いた投票による 静止画像からのカメラ位置・姿勢推定

中川 知香 佐藤 智和 横矢 直和

奈良先端科学技術大学院大学 情報科学研究科

あらまし: カメラ付き携帯機器におけるユビキタス AR の実現には、カメラの位置・姿勢情報の取得が必要である。また計算コストなどの点から入力データとして静止画像を用いることが望ましい。本稿では、特徴点追跡に基づく三次元復元によって得られる広域環境の自然特徴点ランドマークデータベースを用いることで、1枚の静止画像からカメラの位置・姿勢を推定する手法を提案する。提案手法では、まず入力画像内の特徴点と類似度が高いランドマークを1対多で対応付け、各ランドマークと同じ見え方で撮影できる領域に投票することにより入力画像が撮影された可能性が高いカメラ位置候補を特定し、誤対応を排除する。次に、カメラ位置候補に投票したランドマークと入力画像の特徴点の組を複数用いて、カメラの位置・姿勢を推定する。実験では、屋外環境のランドマークデータベースを構築し、データベースの撮影経路付近で撮影した画像を用いて提案手法の推定精度を評価した。

Camera Position and Posture Estimation for a Still Image Based on a Voting Approach Using Feature Landmark Database

Tomoka NAKAGAWA Tomokazu SATO Naokazu YOKOYA

Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: Position and posture estimation for a camera embedded in a mobile device is very important for realizing ubiquitous augmented reality. In this report, we propose a novel method for estimating camera position and posture from a still image based on feature landmark database. In the proposed method, firstly, candidates of corresponding points of feature landmarks are searched in an input image. Secondly, camera position from which candidate points are visible is roughly computed by a voting approach. Finally, camera position and posture is determined by using pairs of landmark and feature point voted for the roughly computed camera position. The validity of the proposed method has been shown through experiments for an outdoor environment.

1 はじめに

近年、携帯電話や PDA などのモバイル端末によるヒューマンナビゲーションシステムが実用化されている。一般的に広く普及しているナビゲーションシステムは、携帯電話に内蔵されている GPS から取得した位置情報を用いて 2 次元的地図上での道案内を実現している。しかし、これらのシステムでは、複雑な交差点等で地図と現実の交差点の関係を正しく把握することが難しく、システムの案内に正しく従って移動す

ることは必ずしも容易ではない。このため、カーナビゲーションの分野においては、一部の交差点などで 3 次元的な CG を表示することで直感的な進路情報の理解を助けるシステムが開発され、市販されている。しかし、3 次元地図情報の作成・更新には膨大な人的コストがかかるため、現在 3 次元地図を利用可能な範囲は極めて狭く、利用できる場所が限定されてしまう。このような問題を解決するために、現実環境を撮影した画像にナビゲーション情報などを表す仮想物体を CG

で重畳表示することで、利用者に直感的な案内情報を提供する拡張現実感技術 (Augmented Reality; AR) と呼ばれる分野の研究が近年盛んに行われている。AR においては仮想物体を現実環境の正しい位置に重畳表示するために、正確なカメラの絶対位置・姿勢情報が必要である。このような AR(ユビキタス AR) 実現のためのカメラ位置・姿勢推定を、PC よりも計算能力が劣る反面どこにでも持ち運べる携帯端末上で実行するためには、次の 3 点を満たすことが要求される。

- (1) 簡素な機器構成で実行できること
- (2) 広域環境で利用可能なこと
- (3) 携帯端末における計算コストが低いこと

以下では、要求事項 (1)~(3) に基づき、既に提案されているカメラ位置・姿勢推定の従来手法におけるユビキタス AR への利用可能性を検証する。

従来のカメラ位置・姿勢推定手法は、GPS やジャイロなどのセンサを用いる手法、画像を用いる手法、それらのハイブリッドに分類できる。センサを用いる手法 [1, 2] は、一般に複数のセンサを用いることで広域環境での利用に対応可能であり、広く研究されているが、カメラとセンサの同期を取ることが難しいという問題や、システムが複雑になるという問題がある。

画像を用いる手法には、(a) 三次元位置関係が既知の人工的なマーカを用いる手法 [3, 4]、(b) 環境中を事前に撮影した画像とその撮影位置・姿勢情報から成る画像データベースを用いる手法 [5, 6]、(c) 環境中の建造物の角などの自然特徴点の 3 次元位置と撮影地点情報をランドマークとして格納した自然特徴点ランドマークデータベースを用いる手法 [7, 8] などがある。これらの手法は、一般に入力画像と環境の事前知識を格納したデータベースを照合することでカメラの絶対位置・姿勢を推定するため、利用者が持つ端末にはカメラ以外のセンサを必要とせず、システム構成が簡素になる利点がある。しかし、人工的なマーカを用いる手法 (a) は広域環境へのマーカの配置に多大な人的コストがかかるという問題や、マーカによって景観が損なわれるという問題がある。画像データベースを用いる手法 (b) は入力画像に類似した画像をデータベース内から探索することで、カメラのおおよその位置・姿勢を特定できるが、カメラの位置・姿勢を 6 自由度で正確に推定することは難しく、精度の高い位置・姿勢情報が要求される AR には向かない。一方、自然特徴点ランドマークデータベースを用いる手法 (c) は、事前に動画像として環境を撮影し、structure-from-motion による三次元復元を行うことでランドマークデータベースを作成しておき、入力画像からデータベースに格納

されている自然特徴点ランドマークと対応する特徴点を複数対応付けることで、カメラの位置・姿勢を 6 自由度で推定できる。しかし、この手法はデータベースの探索範囲を限定するために時系列的な情報を利用しており、動画像を入力とすることを前提としているため、端末に対して大容量のメモリや高い計算コストを要求する。

他方、センサとランドマークを用いるハイブリッド手法 [9, 10] は、センサから得られるカメラ位置・姿勢を、画像を用いる手法を利用することで補正するというアプローチを採っているため、センサの組み合わせによって利用可能な範囲が限定されるという問題や、センサとカメラの同期を取ることが難しいという問題が残されている。

このように、現状提案されている手法において、前述の要求事項 (1)~(3) をすべて満たす手法は存在しない。そこで、本研究では要求事項 (1)~(3) を同時に満たすユビキタス AR のためのカメラ位置・姿勢推定の開発を目的とし、静止画像 1 枚と安価な GPS から取得できる誤差数十 m の位置情報を入力する自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定手法を提案する。本手法では、簡素な機器構成を実現するために、利用者の端末として市販のカメラ付き GPS 携帯を想定し、1 枚の静止画像と GPS による位置情報を入力とする (要求事項 (1) に対応)。また、広域環境に対応するために、大江らの手法 [7] と同様に全方位画像を用いた三次元復元により広域環境に対応したランドマークデータベースを作成し使用する (要求事項 (2) に対応)。さらに、携帯端末の計算コストを抑えるために、サーバ・クライアント型システムを想定し、サーバ側でカメラ位置・姿勢推定処理を実行する (要求事項 (3) に対応)。

提案手法は、オフライン処理によるランドマークデータベースの構築とオンライン処理によるランドマークデータベースを用いたカメラ位置・姿勢推定処理の二段階から構成される。全体の処理の流れを図 1 に示す。ランドマークデータベースの構築 (オフライン処理) では、事前に環境内を全方位動画像として撮影し、動画像から検出された自然特徴点の三次元位置と撮影地点情報をランドマーク情報としてランドマークデータベースに登録する。カメラ位置・姿勢推定処理 (オンライン処理) では、まず利用者がクライアントである携帯端末で撮影した画像を、ネットワークを通してサーバに送信する。サーバは受け取った画像の特徴点と対応づくランドマークをデータベースから探索し、これらのランドマークと入力画像上の自然特徴点との対応付けを行うことで、カメラ位置・姿勢を推定する。本研究では、データベース中の膨大な数のランドマークから

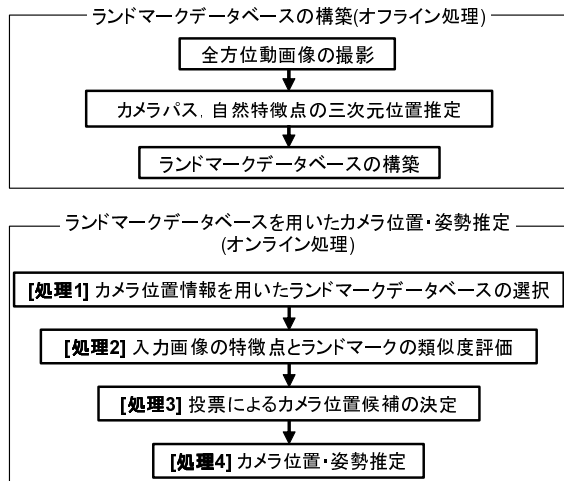


図 1: 全体の処理の流れ

入力画像の特徴点と対応付くランドマークを高速に検索するため、(1)GPS から取得したカメラの位置情報、(2) 入力画像の特徴点とランドマークの類似度、(3) 同一視点から観測できるランドマーク数、を順に用いて 3 段階の処理で入力画像上に存在すると考えられるランドマークを絞り込む。最後に、サーバは推定されたカメラ位置・姿勢に基づき、入力画像に CG でナビゲーション情報を重畳表示し、クライアントに送信する。

以下、本稿では、2 章でオフラインで事前に作成しておくランドマークデータベースの構成要素とその構築手法について述べ、3 章でランドマークデータベースを用いた投票による静止画像からのオンライン・カメラ位置・姿勢推定手法について提案する。また、4 章では、屋外環境での実験結果を示し、提案手法の推定精度について評価する。最後に、5 章でまとめと今後の課題について述べる。

2 ランドマークデータベースの構築

2.1 ランドマークデータベースの構成要素

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 (A) 自然特徴点の三次元位置と (B) 撮影地点ごとの情報から成る。撮影地点ごとの情報はランドマークと入力画像の特徴点に対応付けるために保持する。ランドマークの見え方は撮影地点によって異なるので、各ランドマークに複数の撮影地点ごとの情報を登録する。なお、本研究で用いるランドマークデータベースは、文献 [7] で用いられているデータベースを基礎としているが、高速かつロバストなランドマークの探索を実現するため、ランドマー

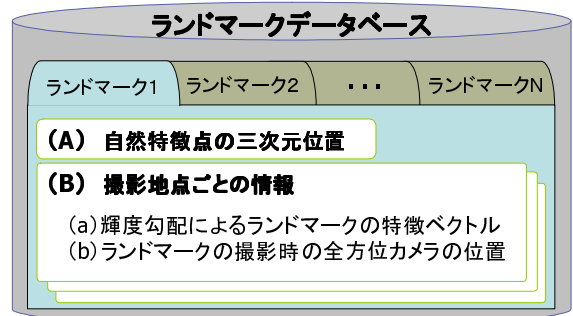


図 2: ランドマークデータベースの構成要素

ク情報としてテンプレート画像の代わりに輝度勾配情報から成る特徴ベクトルを用いる。撮影地点ごとの情報は次の 2 つの要素で構成される。

- (a) 輝度勾配によるランドマークの特徴ベクトル: 各ランドマークの画像上の位置を中心とする SIFT 記述子 [11] による多次元ベクトル (特徴ベクトル) を保持する。SIFT 記述子を用いることでランドマークの画像上での見え方の特徴を輝度勾配から成る特徴ベクトルとして表し、回転などの変化に対してロバストな対応点探索を実現する。ただし、特徴点位置の抽出には文献 [7] と同様、三次元復元に適した Harris オペレータ [12] を用いるため、SIFT で用いられているスケール情報はそのまま利用できず、本稿ではスケールを固定とする。
- (b) ランドマーク撮影時の全方位カメラの位置: 各ランドマーク撮影時の全方位カメラの位置を世界座標系で保持する。3 章で述べる投票処理において、ランドマークを同じ見え方で撮影できる領域を算出するために用いる。

2.2 ランドマークデータベースの構築

ランドマークデータベースは、対象となる環境を全方位カメラで撮影し、structure-from-motion に基づく三次元復元処理を行うことで作成する。本研究では、佐藤らの手法 [13] を用いて、移動を伴って撮影された全方位動画画像上の自然特徴点を追跡し環境の粗な三次元復元を行い、自然特徴点の三次元位置と画像上の座標、および全方位動画画像のカメラパスを取得する。次に、以下の手順で特徴ベクトル [11] をランドマークの撮影地点情報ごとに算出する。

1. 画像の傾き補正: 特徴点を中心とする画像情報から、撮影時のカメラ姿勢の変化によるランドマー

クの画像上での見え方の変化を補正することで、カメラ姿勢やレンズ歪みの影響を取り除いた画像テンプレートを生成する。ここでは、大江らの手法 [7] と同様に、世界座標系において、カメラの投影中心とランドマークの 3 次元位置を結ぶ直線に対して垂直な面に画像上でのランドマーク位置周辺のパターンを投影することで、幾何学的に補正されたランドマークのテンプレートを作成する。

2. 勾配方向ヒストグラムの生成: 回転不変なマッチングを可能にするために、1. で作成した画像テンプレートから検出する輝度勾配情報を用いて勾配方向ヒストグラムを生成し、勾配方向の頻度が最も多い方向に画像テンプレートを回転する。次に、SIFT 記述子を用い、輝度勾配に基づき D 次元の特徴ベクトル $f(v_1, v_2, \dots, v_D)$ を生成する [11]。

また、3.1 節で述べる GPS の位置情報を用いたランドマークデータベースの選択のために、ランドマークの撮影地点情報を基にデータベースを数 m 間隔のブロック単位に分割しておく。

3 投票による静止画像からのカメラ位置・姿勢推定

本章では、2 章で述べたランドマークデータベースを用いて、静止画像からのオンライン・カメラ位置・姿勢推定を実現する手法について述べる。本研究では、サーバ・クライアント型システムを想定し、画像処理によるカメラ位置・姿勢推定はサーバ側で行うものとする。また、ランドマークデータベースはすべてサーバに格納されている。以下では、サーバにおけるカメラ位置・姿勢の推定手順について詳述する。先に図 1 に示したように、本手法では、まず GPS から得られる誤差数十 m の位置情報を用いて大規模なデータベースから推定に用いるデータベースの範囲を限定する (処理 1)。次に、入力画像から検出した特徴点周辺の画像情報から特徴ベクトルを抽出し、ランドマークとの類似度を評価することで、データベース中の大量のランドマークから入力画像と類似性の高いランドマークを絞り込む (処理 2)。さらに、各ランドマークを同じ見え方で撮影することが可能な領域を算出し、その領域に投票することで入力画像が撮影された可能性が高いカメラ位置候補を複数決定する (処理 3)。最後に、カメラ位置候補に投票したランドマークと入力画像の特徴点との対応関係からカメラ位置・姿勢推定を行う (処理 4)。以下、それぞれの処理について述べる。

3.1 カメラ位置情報を用いたランドマークデータベースの選択

携帯機器に内蔵された GPS から取得した誤差数十 m 程度の位置情報を基に広域環境のデータベースから推定に用いるデータベースを決定する (処理 1)。ここでは、計測誤差が 100m を超えることはないと仮定し、GPS による計測位置との二乗距離が 100m 以内のブロック内の撮影地点情報を持つランドマークデータベースを以降の処理で用いるデータベースとして選択する。

3.2 入力画像の特徴点とランドマークの類似度評価

入力として与えられた 1 枚の静止画像から特徴点を検出し、前節で選択されたデータベースから類似度の高いランドマークを選択する (処理 2)。ただし、1 つのランドマークには複数の撮影地点情報が格納されているため、ここでは入力画像上の特徴点と見え方の類似度が最も高い 1 つの撮影地点情報を選択する。

まず、入力画像からデータベース作成時と同じ Harris オペレータ [12] を用いて多数の特徴点を検出し、ランドマーク作成時と同様に特徴点周辺の画像情報から 2.2 節で述べた手順で特徴ベクトルを抽出する。次に、入力画像の各特徴点から得られた特徴ベクトル f_{in} と、処理 1 で選択されたデータベース内のすべてのランドマークの撮影地点ごとの特徴ベクトル f_{LM} について、総当たりで特徴空間上における二乗距離 (非類似度) $S = \sum_{d=1}^D \{f_{in}(v_d) - f_{LM}(v_d)\}^2$ を算出する。ここで算出される二乗距離 S は値が小さいほど類似度が高いため、最後に二乗距離 S を昇順に並び替え、二乗距離が一定閾値以下の上位 α 個のランドマークを 1 対多で対応付ける。ただし、本研究ではデータベース構築時と入力画像撮影時のカメラ光軸回りの回転角が大きく異ならないという仮定の下で、特徴点に対してランドマークが光軸方向に一定以上回転している対応を誤対応とみなして排除する。これを入力画像のすべての特徴点に対して行うことで、入力画像上の特徴点と類似度の高いランドマークを絞り込む。

3.3 投票によるカメラ位置候補の決定

処理 2 で選択されたランドマークには、入力画像上の特徴点と真に対応するランドマーク以外に撮影対象とは全く異なる位置のランドマークが多数含まれている。このような誤対応は、見え方がほぼ同一となるラ

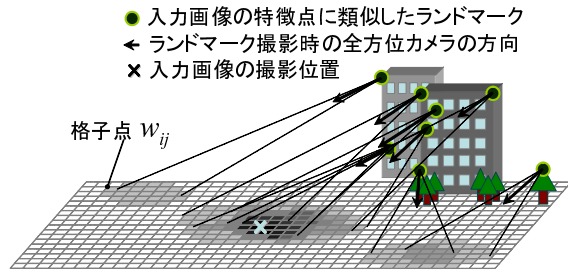


図 3: 入力画像の撮影位置と入力画像の特徴点に類似したランドマークの関係

ランドマークがデータベース内に複数存在するために発生するが、見え方の類似は画像上の局所的なものであることが大半で、大局的な情報を用いれば誤対応を排除できる。そこで本研究では、処理 2 で選択されたすべてのランドマークについて、各ランドマークを同じ見え方で撮影できる領域を算出し、その領域に投票することで、入力画像が撮影された可能性が高いカメラ位置候補を複数決定する (処理 3)。具体的には、図 3 に示すように、GPS から得られる計測地点を中心に、一定距離内の空間を地面に対して水平方向に格子状に分割し、処理 2 で選択されたランドマーク p の撮影地点ごとの情報に基づき格子点への投票を行う。まず、GPS の計測位置 (g_x, g_y, g_z) を中心とする世界座標系における $(2n + 1) \times (2n + 1)$ 個の格子点の xy 座標 $w_{ij} (-n \leq i \leq n, -n \leq j \leq n)$ を以下のように定義する。

$$w_{ij} = \begin{pmatrix} w_i \\ w_j \end{pmatrix} = \begin{pmatrix} g_x + L \times i \\ g_y + L \times j \end{pmatrix}$$

ただし、 L は格子間隔を表す。次に、処理 2 で選択された各ランドマーク p に対して、以下の条件を同時に満たすすべての格子点 w_{ij} に投票を行う。

[条件 1] 格子点 w_{ij} と、ランドマーク p に対応する全方位カメラの撮影位置の xy 座標 c_p の距離 $|w_{ij} - c_p|$ が閾値以下

[条件 2] ランドマーク p の三次元位置 S_p と全方位カメラの位置 C_p を結ぶ直線、および格子点 w_{ij} に C_p の高さ c_z を与えた点 (w_i, w_j, c_z) とランドマーク p の三次元位置 S_p を結ぶ直線の成す角度が閾値以下

次に、投票数が空間的に極大になる格子点位置を、入力画像を撮影した可能性の高いカメラ位置候補とする。また、カメラ位置候補は複数存在するため、投票数が多い順にカメラ位置候補の順位を決定しておく。

3.4 カメラ位置・姿勢推定

最後に、前節で得られたカメラ位置候補に投票したランドマークと入力画像の特徴点との対応関係からカメラ位置・姿勢推定を行う (処理 4)。カメラ位置・姿勢情報は、ランドマークの 3 次元位置と入力画像上の特徴点の 2 次元位置の組を複数用いて PnP 問題 [14] を解くことで推定される。しかし、カメラ位置候補に投票したランドマークと入力画像上の特徴点の組の中には多数誤対応が含まれるために、それを取り除く必要がある。そこで、本手法では、投票数が多いカメラ位置候補から順に、誤対応を排除してカメラ位置・姿勢を推定する処理を繰り返し、尤もしいカメラ位置・推定情報が得られた時点で処理を完了する。具体的には、誤対応の排除のために RANSAC [15] を用いる。ただし、RANSAC では、入力とする対応の組に正しい対応関係が 50% 未満しか存在しない場合、うまく機能しない。そこで、本手法では、カメラの水平方向の画角を利用し、カメラの方位を変えながら画角内に存在するランドマークのみを用いて、RANSAC による誤対応除去を行い、逐次カメラ位置・姿勢推定の尤もらしさを検証する。そのため、本研究では、カメラ位置・姿勢推定に使用された特徴点数 l と再投影誤差 R に着目し、以下の 2 つの条件を満たすカメラ位置・姿勢推定が行われた場合に、尤もしい推定結果であると判定する。

[条件 1] 再投影誤差 R が閾値以下

[条件 2] 推定に用いられた特徴点数 l が閾値以上

ただし、再投影誤差 R は、入力画像の特徴点の座標 (u, v) と推定するカメラ位置・姿勢を用いてランドマークの三次元座標を画像上に投影した座標 (\hat{u}, \hat{v}) の二乗距離の和 $R = \sum_{i=1}^l \{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2\}$ で定義される。また、これらの条件は予備実験により、経験的に得られたものである。なお、最後のカメラ位置候補までカメラ位置・姿勢を推定しても尤もしい結果が得られなかった場合は、カメラ位置・姿勢推定に失敗したとみなし、システムは利用者に画像を取り直しを要求する。

4 実験

提案手法の有効性を示すため、屋外環境のランドマークデータベースを構築し、ランドマークデータベースの撮影経路付近で撮影した画像を用いてカメラ位置・姿勢推定を行った。また、入力画像の特徴点と 3 次元位置を手動で対応付けて撮影位置・姿勢を推定し、比



図 4: 全方位型マルチカメラシステム Ladybug と撮影された全方位画像

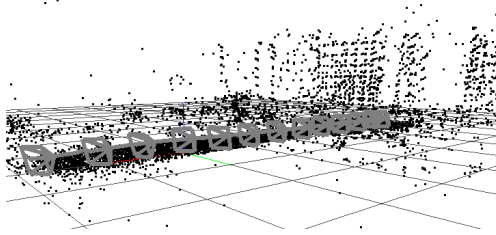


図 5: ランドマークデータベース構築時のカメラパスとランドマークの三次元位置

較することによる推定精度の評価を行った。ただし、本実験では、実際にサーバ・クライアント型システムの構築は行わず、携帯端末からの画像を想定した入力画像を使用した。また、カメラ内部パラメータはあらかじめ Tsai の手法 [16] によって校正し、撮影時は内部パラメータを固定した。さらに、データベースの規模が比較的小さいため、GPS による位置情報を用いたランドマークデータベースの利用範囲の限定 (処理 1) は行わない。

4.1 屋外環境における実験

本研究で用いたランドマークデータベースは図 4 左に示す全方位マルチカメラシステム (Point Grey Research 社 Ladybug) で屋外環境 (約 75m) を歩きながら撮影した全方位動画像 (図 4 右) を用いて構築した。ランドマークデータベース構築時のカメラパスとランドマークの三次元位置を図 5 に示す。ここでは 2.2 節で述べた手法により約 12400 個のランドマークがデータベースに登録された。また、各ランドマークに対して平均 8 地点で特徴ベクトルが生成された。本実験では、特徴ベクトル作成のためのテンプレートサイズを 31×31 、この次元数を 128 次元とした。特徴ベクトル作成時のスケールは、入力画像とランドマーク作成時のカメラ画像上での大きさが同一地点からの撮影においてほぼ等しくなるように設定した。

表 1: カメラ位置・姿勢推定の各処理における閾値

(処理 2) 一つの特徴点に対応付けるランドマークの数 (個)	5
(処理 2) 光軸回りの回転制約 (度)	30
(処理 3) 条件 1 の閾値 (m)	5
(処理 3) 条件 2 の閾値 (度)	5
(処理 4) 条件 1 の閾値 (画素)	15
(処理 4) 条件 2 の閾値 (個)	10

次に、66 枚の静止画像 (解像度 720×480 画素、水平方向の画角 57 度) を撮影し、各画像に対して 3 章で述べた手法によるカメラ位置・姿勢推定を行った。本実験において使用したカメラ位置・姿勢推定処理の各手順における閾値を表 1 に示す。また、入力として用いた画像の例を図 6 に、入力画像上に存在するランドマークを図 7 に、3.4 節で述べたカメラ位置・姿勢推定処理において尤もらしいカメラ位置・姿勢結果が得られたときに推定に用いられた RANSAC による誤対応除去後のランドマークの画像上の位置を図 8 に示す。図 7 と図 8 を比較すると、提案手法でカメラ位置・姿勢推定に用いられた特徴点数が比較的少ないことが分かる。今後、3.2 節で述べた入力画像の特徴点とランドマークの類似度評価の処理をスケール変化に対応させることでより多くのランドマークが検出されると考えられる。

世界座標系において地表面と平行となる XY 平面を 50cm 単位の格子に分割し、3.3 節で述べた投票処理を行った結果を図 9 に示す。図中の濃度値は投票数を表しており、明るいほど投票数が多い。なお、印は入力画像のカメラ位置の真値を、印は最終的な推定結果を出力したカメラ位置候補を表す。実験結果から、入力画像の撮影位置付近に投票数の極大値が出現していることが分かる。なお、カメラ位置の真値は、各入力画像上の特徴点と事前にトータルステーションを用いて測定した環境内の特徴点の位置を手動で対応付け、PnP 問題 [14] を解くことで作成した。また 3.4 節で述べた手法によってシステムが自動的に推定結果の妥当性を判断した結果、本実験においては、66 枚中 64 枚がカメラ位置・姿勢推定に成功したと判断された。

本実験で推定されたカメラ位置、真値、およびランドマークデータベース構築時のカメラパスを図 10 に示す。図 10 において、太い実線はランドマークデータベース構築時のカメラパス、印が真値、印が推定結果を表す。また 印と 印を結ぶ細い実線が成功した推定結果と真値の関係を表し、×印は失敗した推定結果の真値を表す。システムが成功として出力したカ

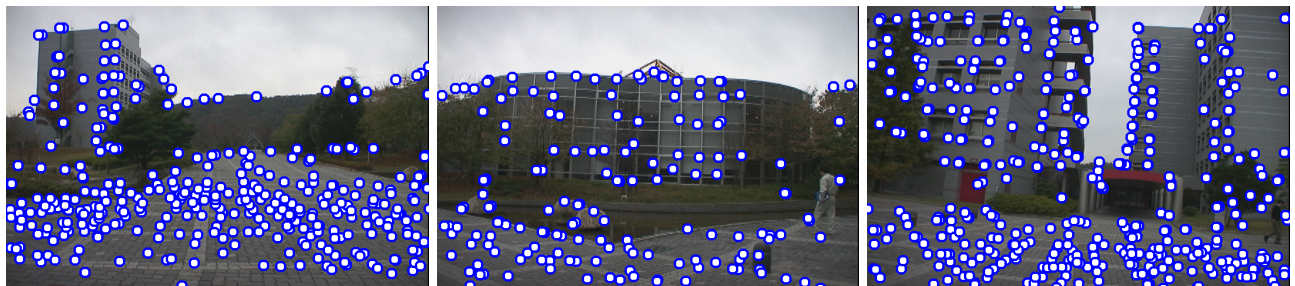


場面 1

場面 2

場面 3

図 6: 入力画像



場面 1

場面 2

場面 3

図 7: 環境中に存在するランドマークの画像上の位置



場面 1

場面 2

場面 3

図 8: カメラ位置・姿勢推定に使用した RANSAC による誤対応除去後のランドマークの画像上の位置

カメラ位置・姿勢推定結果の真値との位置誤差の平均は約 50cm, 姿勢誤差は約 1 度であり, カメラ位置・姿勢推定に用いられた特徴点数は平均 22 個, 再投影誤差は平均 3.8 画素であった. また, 提案手法の処理の 1 枚の入力に対する処理時間は PC(CPU:Pentium4 3GHz, Memory:1.5GB) を用いて平均 64 秒であった.

実験結果から, カメラ位置から 20m 離れた位置に CG を重畳表示した場合, 投影誤差は最大約 15 画素となる. このことから注釈などを重畳表示するにはそのまま利用可能であるが, 景観シミュレーションや match move などにそのまま適用することは難しい. ただし, 本手法で推定したカメラ位置・姿勢情報を基に空間中のランドマークの探索範囲を限定し, 再度対応付けを行うことで, 今後さらに精度の高い推定を実現できると考えられる.

また, 平均処理時間の内訳は, 入力画像の特徴点とランドマークの類似度評価と投票処理に 30 秒, カメラの位置・姿勢推定処理に 34 秒であった. ここで, 3.2 節で述べた入力画像の特徴点とランドマークの類似度評価では, 入力画像の各特徴点とデータベース内のすべてのランドマークの撮影地点情報について, 総当りで比較している. 従って, 特徴ベクトルの計算において, データベース内で KD-Tree[8, 11] を構築しておくことで, 処理の高速化が見込める.

5 まとめ

本稿では, ランドマークデータベースを用いた投票による静止画像からのカメラ位置・姿勢推定手法を提案した. 提案手法は, GPS から取得した誤差数十 m

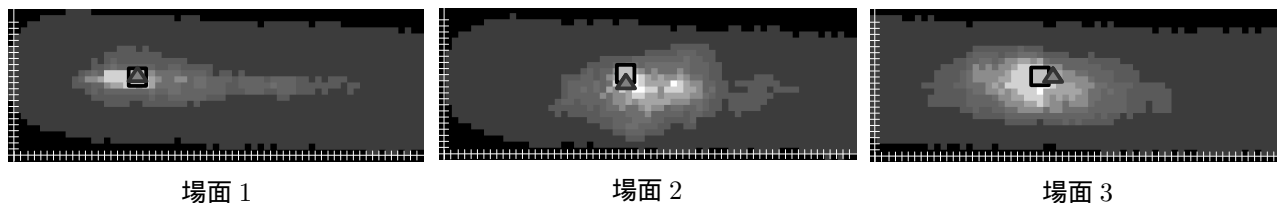


図 9: 投票結果

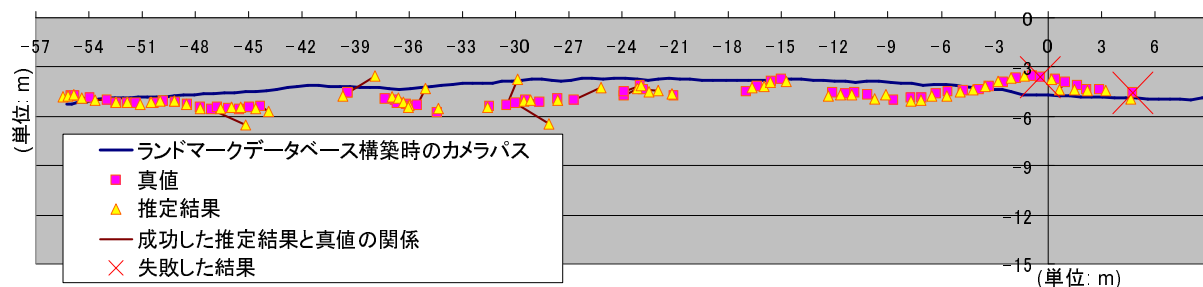


図 10: ランドマークデータベース構築時のカメラパスと入力画像の撮影位置の真値および推定されたカメラ位置

の位置情報と、1枚の静止画像からデータベース内のランドマークを段階的に絞り込むことでカメラ位置・姿勢推定を行う。

実験により、注釈などを重畳表示するアプリケーションへの利用が可能な程度の精度であることを確認した。今後は、入力画像の特徴点とランドマークの類似度評価処理において Harris-Laplacian 法 [17] を応用してスケールを自動で決定することで高精度化を図り、特徴ベクトルの計算において KD-Tree[8, 11] を利用することで処理の高速化を行う。

参考文献

- [1] T. Hollerer, S. Feiner and J. Pavlik: "Situated documentaries: Embedding multimedia presentations in the real world," Proc. Int. Symp. on Wearable Computers, pp. 79–86, 1999.
- [2] A. Davison, Y. G. Cid and N. Kita: "Real-time 3d slam with wide-angle vision," Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [3] H. Kato and H. Billinghurst: "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," Proc. 2nd IEEE/ACM Int. Workshop on Augmented Reality, pp. 85–94, 1999.
- [4] 羽原, 町田, 小川, 竹村: "画像マーカを用いた屋内位置検出機構とその評価", 電子情報通信学会 技術研究報告, Vol. 102, No. 216, pp. 65–70, 2002.
- [5] R. Cipolla, D. Robertson and B. Tordoff: "Image-based localization," Proc. 10th Int. Conf. Virtual Systems and Multimedia, pp. 22–29, 2004.
- [6] 岩佐, 粟飯原, 横矢, 竹村: "全方位画像を用いた記憶に基づく位置推定", 電子情報通信学会論文誌 (D-II), Vol. J84-D-II, No. 2, pp. 310–320, 2001.
- [7] 大江, 佐藤, 横矢: "幾何学的位置合わせのための自然特徴点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", 日本バーチャルリアリティ学会論文誌, Vol. 10, No. 3, pp. 285–294, 2005.
- [8] I. Gordon and D. G. Lowe: "Scene modelling, recognition and tracking with invariant image features," Proc. Int. Symp. on Mixed and Augmented Reality, pp. 110–119, 2004.
- [9] M. Kourogi and T. Kurata: "Personal positioning based on walking locomotion analysis with self-contained sensors and wearable camera," Proc. 2nd IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp. 103–112, 2003.
- [10] Y. Kameda, T. Takemasa and Y. Ohta: "Outdoor see-through vision utilizing surveillance cameras," Proc. Int. Symp. on Mixed and Augmented Reality, pp. 151–160, 2004.
- [11] D. G. Lowe: "Distinctive image features from scale-invariant keypoints," Int. Journal of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.
- [12] C. Harris and M. Stephens: "A combined corner and edge detector," Proc. Alvey Vision Conf., pp. 147–151, 1988.
- [13] 佐藤, 池田, 横矢: "複数動画画像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", 電子情報通信学会論文誌 (D-II), Vol. J88-D-II, No. 2, pp. 347–357, 2005.
- [14] R. Klette, K. Schluns and A. Koschan Eds: "Computervision: Three-dimensional data from image," Springer, 1998.
- [15] M. A. Fischler and R. C. Bolles: "A paradigm for model fitting with applications to image analysis and automated cartography," Comm. of the ACM, Vo. 24, pp. 381–395, 1981.
- [16] R. Y. Tsai: "An efficient and accurate camera calibration technique for 3d machine vision," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 364–374, 1986.
- [17] K. Mikolajczyk and C. Schmid: "Scale & affine invariant interest point detectors," Int. Journal of Computer Vision, Vo. 60, No. 1, pp. 63–86, 2004.