NAIST-IS-MT0451083

修士論文

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

中川 知香

2006年3月10日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である.

中川 知香

審査委員： 横矢 直和 教授 (主指導教員)
千原 國宏 教授 (副指導教員)
山澤 一誠 助教授 (副指導教員)

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

中川 知香

内容梗概

近年サービスが開始された携帯電話におけるヒューマン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 次元的な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，システムの案内に従って正しく移動することは必ずしも容易ではない．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対してナビゲーション情報などを表す仮想物体を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) の研究が近年盛んに行われている．ビデオシースルー型 AR では，仮想物体を幾何学的に正しい位置に重畳表示するために，カメラの正確な絶対位置・姿勢情報が必要である．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は，多数のセンサを用いることやリアルタイムで動画像を処理することが難しく，また広域環境に対応させる必要があるために，従来のカメラ位置・姿勢推定手法をそのまま利用することは難しい．そこで本研究では，特徴点追跡に基づく三次元復元によって事前に得られる自然特徴点の三次元位置と撮影地点ごとの画像情報をランドマークとして登録した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，市販されているカメラ付き携帯端末でも容易に取得可能な 1 枚の静止画像からのカメラ位置・姿勢推定手法を提案する．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

* 奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT0451083, 2006 年 3 月 10 日.

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，ランドマークと入力画像の特徴点の誤対応を排除する．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・屋内環境のランドマークデータベースを構築し，屋外・屋内で撮影された画像からのカメラ位置・姿勢推定実験を通して提案手法の有効性を検証する．

キーワード

カメラ位置・姿勢推定, 静止画像, 自然特徴点, ランドマークデータベース, 拡張現実感

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

Tomoka Nakagawa

Abstract

Recently, a human navigation service for cellular phones has been started by several cellular phone providers. In this service, the position of a user is measured by GPS, and the user's position and guiding information are shown on a 2-D map displayed in the cellular phone. However, it is not easy for every user to match a 2-D map with a real complicated environment. To realize intuitive guiding, a novel technique called Augmented Reality (AR) has been proposed and investigated. In video see-through AR, guiding information for the user is drawn as virtual objects on an image which captures a real environment. To overlay virtual objects at geometrically correct positions on real images, absolute position and posture information of the camera are required. However, it is difficult to determine absolute positions and postures accurately on commercial mobile devices, because existing complex sensors used in conventional research cannot be employed as standard equipments due to weight, size and power consumption problems. Additionally, video image processing is not available on cellular phones due to their limited computational resources. In this research, I propose a novel method for estimating camera position and posture from a still image which is acquired by a camera embedded in a commercial mobile device based on a feature

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0451083, March 10, 2006.

landmark database for a large environment. The proposed method is composed of offline and online stages. In the offline stage, the feature landmark database is constructed by omni-directional 3-D reconstruction of a target scene. In the online stage, camera position and posture of a mobile device are estimated using the feature landmark database. In this stage, first, candidates of corresponding points of feature landmarks are searched in an input image. Second, camera positions from which the candidate points are visible are roughly computed by a voting approach. Finally, camera position and posture are determined by using pairs of landmarks and feature points voted for in roughly computed camera positions. The validity of the proposed method has been shown through experiments in both outdoor and indoor environments.

Keywords:

camera position and posture estimation, still image, natural features, landmark database, augmented reality

目 次

図 目 次

表 目 次

1. はじめに

近年，携帯電話によるヒューマンナビゲーションシステムなどの位置依存情報を利用したサービスが実用化されている．一般に広く普及しているナビゲーションシステムでは，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 次元的な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では，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) の研究が近年盛んに行われている．ビデオシースルー型 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・姿勢情報が必要とされる．

カメラの位置・姿勢推定を目的とする研究は既に数多く行われており，GPS やジャイロなどのセンサを用いる手法 [1, 2, 4, 5, 6, 7, 8, 9]，画像を用いる手法 [3, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]，それらのハイブリッド [20, 21, 22, 23, 24] に分類できる．センサを用いる手法は，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上に，システムが複雑になるという問題がある．さらに，携帯端末上で多数のセンサを利用することはサイズやコストの面から難しい．画像を用いる手法は，(a) 人工的なマーカを用いる手法，(b) 画像データベースを用いる手法，(c) 自然特徴点ランドマークデータベースを用いる手法などに分類できるが，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，従来手法では利用可能な範囲の制約や，大きな計算コストのために，現

状の携帯端末を用いてカメラ位置・姿勢を高精度・広範囲に推定することは難しい．他方，センサと画像を用いるハイブリッド手法は，誤差の累積や計測レート，計算コストといった各手法の欠点を互いに補うことで誤差の蓄積を防ぎ，推定のロバスト性を向上させる．しかし，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，カメラ付き携帯端末を用いて広域環境に対応したカメラ位置・姿勢推定を実現する手法は従来存在しない．そこで，本論文では特徴点追跡に基づく三次元復元によって事前に得られる自然特徴点の三次元位置と撮影地点ごとの画像情報をランドマークとして登録した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，市販されているカメラ付き携帯端末でも容易に取得可能な 1 枚の静止画像からカメラ位置・姿勢推定を行う手法を提案する．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，ランドマークと入力画像の特徴点の誤対応を排除する．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．本研究では，屋内外環境のランドマークデータベースを構築し，屋内・屋外で撮影された画像からのカメラ位置・姿勢推定実験を通して提案手法の有効性を検証する．

以降，2 章ではカメラ位置・姿勢推定に関する従来研究と本研究の位置付け・方針について述べる．3 章では，ランドマークデータベースの構築要素とランドマーク情報の獲得方法について述べる．4 章では，3 章で構築したランドマークデータベースを用い，入力画像上の特徴点と類似したランドマークを同じ見え方で撮影できる領域に投票することによるカメラ位置・姿勢推定について述べる．5 章では，屋外・屋内環境のランドマークデータベースを構築し，実環境を撮影した画像を用いたカメラ位置・姿勢推定実験について報告する．最後に，6 章でまとめと今後の課題について述べる．

2. カメラ位置・姿勢推定に関する従来研究と本研究の位置付け

本章では，本研究に関連する従来研究と本研究の位置づけについて述べる．まず，従来のカメラ位置・姿勢推定手法を，(1)GPS やジャイロなどのセンサを用いる手法，(2) 画像を用いる手法，(3) それらのハイブリッド手法に分類し，それぞれの手法の特徴と問題点について述べ，ユビキタス AR への利用可能性を検証する．次に，本研究の位置づけと方針，および提案手法の概要について述べる．

2.1 センサを用いたカメラ位置・姿勢推定

センサを用いたカメラ位置・姿勢推定は，絶対位置を取得可能な GPS などのインフラと相対位置・姿勢などを取得可能なセンサを組み合わせる手法 [4, 5, 6, 7, 8] と環境内に人工的に配置した超音波センサなどのインフラを用いる手法 [1, 2, 9] に分類することができる．インフラとしては，GPS，超音波センサ，赤外線ビーコン，無線 LAN などが用いられ，絶対位置や姿勢情報を取得するために利用されている．また，相対位置・姿勢などを取得可能なセンサとしては，加速度センサやジャイロセンサ，歩数計などが利用されている．また，特別なインフラを用いないが，絶対方位情報を取得可能なセンサとして電子コンパスが挙げられる．

一般的に，センサを用いたカメラ位置・姿勢推定では，絶対位置を取得可能なインフラと相対位置・姿勢センサを組み合わせる手法 [4, 5, 6, 7, 8] が主流となっており，特に屋外では位置情報を GPS，姿勢情報をジャイロセンサで獲得する手法が一般的である．また，RTK-GPS などのように高精度な位置情報を取得可能なセンサと，推定精度の高い姿勢センサを用いることで高精度なカメラ位置・姿勢推定手法が実現できる．しかし，GPS には計測周期が長いために間欠的な位置情報しか取得できないという問題があり，一般的なジャイロセンサには長時間使用すると蓄積誤差が発生してしまうという問題がある．これに対して，神原ら [6] は，RTK-GPS と複数の加速度センサから構成された慣性航法装置を用いたハイブリッドセンサによる屋外型拡張現実感システムを提案している．この手法は，長時間使用しても誤差が蓄積せず広範囲な屋外環境において利用できる．しかし，

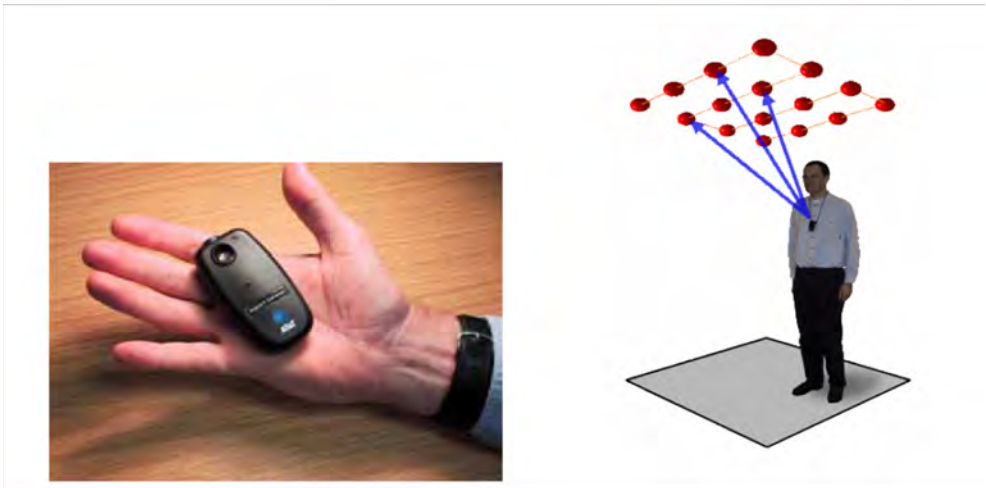


図 1　ActiveBat と天井に設置した超音波センサによる測位 [1, 2]

高価なセンサを使用しているためコストが高い．また，屋内で利用するためには別の手法と組み合わせる必要がある．Tenmoku ら [8] は，屋外では GPS，屋内では赤外線ビーコン (IrDA) を用いて利用者の絶対位置を計測し，ジャイロセンサや電子コンパスなどから構成された姿勢センサを用いて絶対姿勢を計測，さらに歩数計や姿勢センサを組み合わせて相対的な移動位置を計算することによって位置・姿勢を推定している．しかし，これらの手法は利用者が持つシステムに複数のセンサを組み込まなければならないため，システムが複雑になるという問題がある．

一方，環境中にインフラとしてセンサを埋め込む手法は，図 1 に示すような超音波センサや赤外線ビーコンなどのマーカを環境に設置することによって，絶対位置・姿勢を推定する [1, 2, 9]．これらの手法は，利用者の装備が簡易なものとなり，かつ比較的高精度に位置・姿勢を計測できるという特徴がある．しかし，環境に大量のセンサを設置しなければならず，またその幾何学的な位置関係を計測する必要があるために，広範囲な環境を想定した場合，このようなインフラを用いる手法は人的コストが膨大になるという問題がある．

2.2 画像を用いたカメラ位置・姿勢推定

カメラからの入力画像を用いる手法には，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, 10, 11, 12, 13, 14, 15]，環境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 [16, 17]，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 [18, 19] 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．以下では，それぞれの手法について詳述する．

2.2.1 人工的なマーカを用いる手法

人工的なマーカを用いたカメラ位置・姿勢推定手法としては，ARToolkit[11] に代表されるパターンや形状，色などが既知の画像マーカを利用する手法 [3, 10, 12] や，人工的なマーカと自然特徴点追跡を併用する手法 [13, 14, 15] などが挙げられる．これらの手法は，三次元位置が既知のマーカを撮影した画像から，マーカに対するカメラの相対的な位置・姿勢を決定する．

画像マーカを用いる手法は環境中に多数の画像マーカを配置する必要があるため，一般的な画像マーカを用いた場合には環境の景観を損ねてしまうという問題がある．そこで，中里ら [10] は，天井に半透明の再帰性反射材を用いた不可視マーカを配置し赤外線カメラで撮影することによって，景観を損ねないウェアラブル拡張現実感のためのカメラ位置・姿勢推定手法を提案している．他方，Wagner ら [3] は，図 2 に示すように，画像マーカを利用することで PDA に取り付けられたカメラの位置・姿勢をスタンドアロンで推定する AR システムを開発しており，携帯端末でもカメラ位置・姿勢推定ができることを実証した．しかし，人工的なマーカを用いる手法は，前節で述べたインフラとして多数のセンサを設置する手法と同様に，広域環境に多数のマーカを配置するために，多大な人的コストがかかるという問題がある．

図 2 Wagner らの AR システム [3]

Neumann ら [13] や Davison ら [14] は，動画像を入力とし，人工的なマーカと環境内の自然特徴点追跡を併用することによってカメラ位置・姿勢推定を行う手法を提案している．しかし，このような手法では，人工的なマーカが写らない状態が長時間続くと推定誤差が累積するという問題や，初期フレームにおいて人工的なマーカが入力画像上に写っていなければならないという問題があり，広域環境への適用にはやはりマーカ配置・計測のための人的コストが膨大となるという問題がある．

佐藤ら [15] は三次元位置が既知の少数の基準点と多数の自然特徴点を全方位動画像中で自動追跡し，動画像全体での最適化処理を行うことでカメラ位置・姿勢を推定する手法を提案している．この手法では，三次元位置が既知の基準点を用い，動画像全体での投影誤差の最小化を行うことで，複雑で広範囲な環境を撮影した場合でも累積的な推定誤差を最小化したカメラ位置・姿勢推定が可能である．しかし，動画像全体を用いた最適化によって累積誤差を最小化するために，オフライン処理を前提としており，携帯端末上での使用を想定した場合，大容量のメモリや高い計算コストが必要となるため，そのまま手法を適用することは困難である．

2.2.2 画像データベースを用いる手法

画像データベースを用いたカメラ位置・姿勢推定手法は，環境を事前に撮影した画像とその撮影位置・姿勢情報を画像データベースに登録しておき，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・姿勢を特定する手法である．この手法には，入力画像と最も類似度の高いデータベース画像の撮影位置を出力する手法 [16, 25] や，入力画像と最も類似度の高いデータベース画像の位置・姿勢からの相対的なカメラの位置・姿勢を推定する手法 [17, 20] などがある．

岩佐らの手法 [16] は，事前に複数の地点で全方位カメラを用いて撮影された画像からカメラ位置に固有な自己相関画像を生成し，自己相関画像の集合から識別に有用な特徴軸の抽出を行い，データベースとして固有空間を構築しておく．次に入力画像を固有空間に射影することによって類似性を評価し，入力画像と類似度の高い画像の撮影位置を入力画像の撮影位置として推定する．この手法は，位置推定に有効な大局的な情報を含む全方位画像から自己相関画像を生成することによって，センサの向きに依らない位置に固有な情報を抽出している．また，興梠ら [25] は映像上への注釈情報の重ね合わせを実現するため，パノラマ画像とその画像上に添付された注釈を情報源として用いるパノラマベースト・アノテーション手法を提案している．この手法は，事前に環境中の複数地点で撮影されたパノラマ画像群を用意しておき，入力画像と最も近い視点位置で撮影されたパノラマ画像を選び出すことにより，入力画像の大まかな位置を推定できる．これらの手法は，入力画像の撮影位置を入力画像と最も類似度の高いデータベース画像の撮影位置とするため，精度の高い位置・姿勢情報が要求される AR を行うためには，センサなどを用いてデータベース画像の撮影位置からの相対的な位置・姿勢を推定する必要がある．興梠らは，この手法 [25] の発展として，画像とセンサの組み合わせによる手法についても提案しているが，これはハイブリッドな手法に分類されるため，2.3 節で述べる．

Cipolla ら [17] は，画像処理を用いてデータベース画像の撮影位置からの相対的な位置・姿勢を推定する手法を提案している．この手法は，画像中の縦・横方向の直線とその消失点を用いて入力画像に写っている建造物が画面に対して垂直

に見えるように画像の垂直化を行い，入力画像と対応するデータベース画像を探索し，対応関係を求めることにより，データベース画像を撮影したカメラの絶対位置・姿勢からの相対的なカメラの位置・姿勢を推定する．この手法は，静止画像 1 枚を入力としており，サーバ・クライアント型システムを想定しているため，携帯端末でも利用可能である．しかし，建造物を対象とし，平行直線が画像内に複数存在していることを前提としているため，利用可能な環境が限定されるという問題がある．また，カメラの位置・姿勢を 6 自由度で推定できず，精度の高い位置・姿勢情報が要求される AR には向かない．

2.2.3 自然特徴点ランドマークデータベースを用いる手法

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・姿勢推定手法としては，Skrypnyk ら [18] や大江ら [19] の手法が挙げられる．これらの手法は，環境中の建造物の角などの自然特徴点の三次元位置，および自然特徴点の画像テンプレートなどの撮影地点情報をランドマークとして事前にデータベースに格納しておく．次に，入力画像上の二次元特徴点と対応するランドマークをデータベースから探索し，入力画像の特徴点とランドマークの複数の組からカメラの位置・姿勢を 6 自由度で推定する．

Skrypnyk らの手法では，まず物体を撮影した複数枚の画像からスケールや回転に不変な特徴点と特徴点間のマッチングに利用できる SIFT 記述子を算出し，特徴点の三次元復元を行いうことで SIFT 記述子と特徴点の三次元位置をデータベースに登録する．次に入力画像から抽出した SIFT 記述子との対応付けを行うことで，カメラ位置・姿勢推定を行う．この手法は，撮影画像のみでデータベースを構築できるが，広域環境を対象とした場合，三次元復元の推定誤差が累積するという問題があり，小物体，小領域への適用にとどまっている．従って，広域環境で利用される携帯端末に適用することは難しい．

これに対して，大江らは広域環境に対応したデータベースを構築し，利用する手法を提案している．大江らの手法は，事前に全方位動画像として広域環境を撮影し，structure-from-motion による三次元復元によって推定し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としてデータベースに登録する．次に，

入力画像から抽出した特徴点との対応付けをオンラインで行うことで，カメラの位置・姿勢推定を行う．この手法は，動画像を入力としており，前フレームのカメラ位置・姿勢情報を用いてデータベースの探索範囲を限定するため，高速なカメラ位置・姿勢推定が可能である．しかし，動画像のリアルタイム処理は，携帯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．また，初期フレームのカメラ位置・姿勢が既知であることを前提としており，実際には何らかの推定手法を用いて初期位置・姿勢を推定する必要がある．

2.3 センサと画像を用いるハイブリッドなカメラ位置・姿勢推定

センサと画像を用いるハイブリッドなカメラ位置・姿勢推定手法 [20, 21, 22, 23, 24] は，計測レート，計算コストといった各手法の欠点を互いに補うことで誤差の蓄積を防ぎ，推定のロバスト性を向上させるアプローチを採っている．

Kourogi ら [20] は，ウェアラブル型拡張現実感システム Weavy を提案している．このシステムでは，2.2.2 項で述べた画像データベースによる手法を用いて，利用者が装着したカメラの絶対的な位置・方位を画像から推定する．これに加えて，加速度計，およびジャイロセンサから取得される相対的な計測情報を組み合わせることで，利用者の位置・姿勢を推定している．この手法は，画像データベースから絶対位置・方位が推定されるため，センサの誤差の蓄積は防止される．しかし，入力画像と最も類似度の高いデータベース画像の位置・方位情報を用いるため，高精度なカメラ位置・姿勢推定のためには利用者が移動する範囲を密に撮影しておく必要がある．そのため，広域環境をカバーする膨大な量の画像データベースが必要となり，同じ範囲の環境の自然特徴点ランドマークデータベースなどと比較し，多くの記憶容量が必要となる．さらに，端末において多数のセンサを用いるため，システムが複雑になるという問題がある．

内山ら [21] は，AR のための位置合わせ方法として，6 自由度センサの情報をマーカの識別および繰り返し演算の初期値として用いる，ICP アルゴリズムによる位置・姿勢推定手法を提案している．この方法は，カメラの位置姿勢計測を途切れなく安定に求めることができる．しかし，マーカを用いているため，広域環境で利用するためには，多数のマーカを配置しなければならない．

横地ら [22] は，データベースなどの事前知識を用いず，動画像中の特徴点の追跡 (structure-from-motion) と GPS からの絶対位置情報を用いた最適化に基づくカメラ位置・姿勢推定手法を提案している．この手法は，屋外環境下で取得した動画像と RTK-GPS による位置情報を用いてカメラの位置・姿勢推定を行っており，画像上で定義される誤差と GPS の位置情報から定義される誤差を同時に最小化することで，カメラ位置・姿勢の推定結果に誤差が蓄積することを防ぐ．しかし，動画像全体での最適化処理を行うため，佐藤ら [15] の手法と同様に，オフライン処理を前提としており，携帯端末上での使用を想定した場合，携帯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．

以上で述べたように，従来提案されているハイブリッド手法には，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．

2.4 本研究の位置付けと方針

前節までで概観したように，従来のカメラ位置・姿勢推定手法において，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にユビキタス AR を直接適用できる手法は存在しない．ユビキタス AR 実現のためのカメラ位置・姿勢推定を携帯端末上で実行するためには，次の 3 点を同時に満たすことが重要である．

(1) 簡素な機器構成で実行できる

(2) 広域環境で利用できる

(3) 携帯端末における計算コストが低い

そこで，本研究ではユビキタス AR 実現のための要求事項 (1)〜(3) を同時に満たすユビキタス AR のためのカメラ位置・姿勢推定手法の開発を目的とし，静止画像 1 枚と GPS 情報を入力と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．提案手法では，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 枚の静止画像と誤差数十 m 程度の GPS 情報を入力とする (要求事項 (1) に対応)．また，広

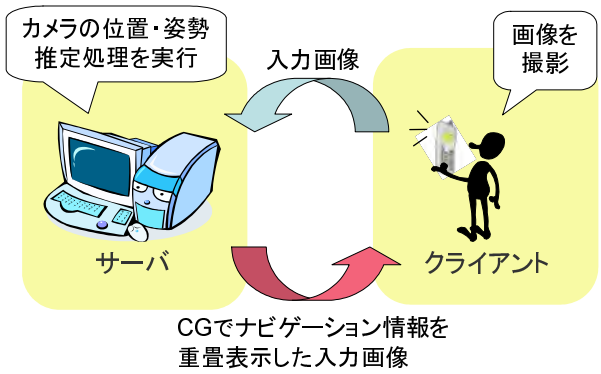


図 3 本研究で想定するサーバ・クライアント型システム

域環境に対応するために，大江らの手法 [19] 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)．また，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 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アプローチをとる．

なお，本研究では図 3 に示すサーバ・クライアント型システムを想定する．このシステムでは，まず利用者は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を用いてカメラ位置・姿勢推定を行い，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．

2.5 提案手法の概要

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

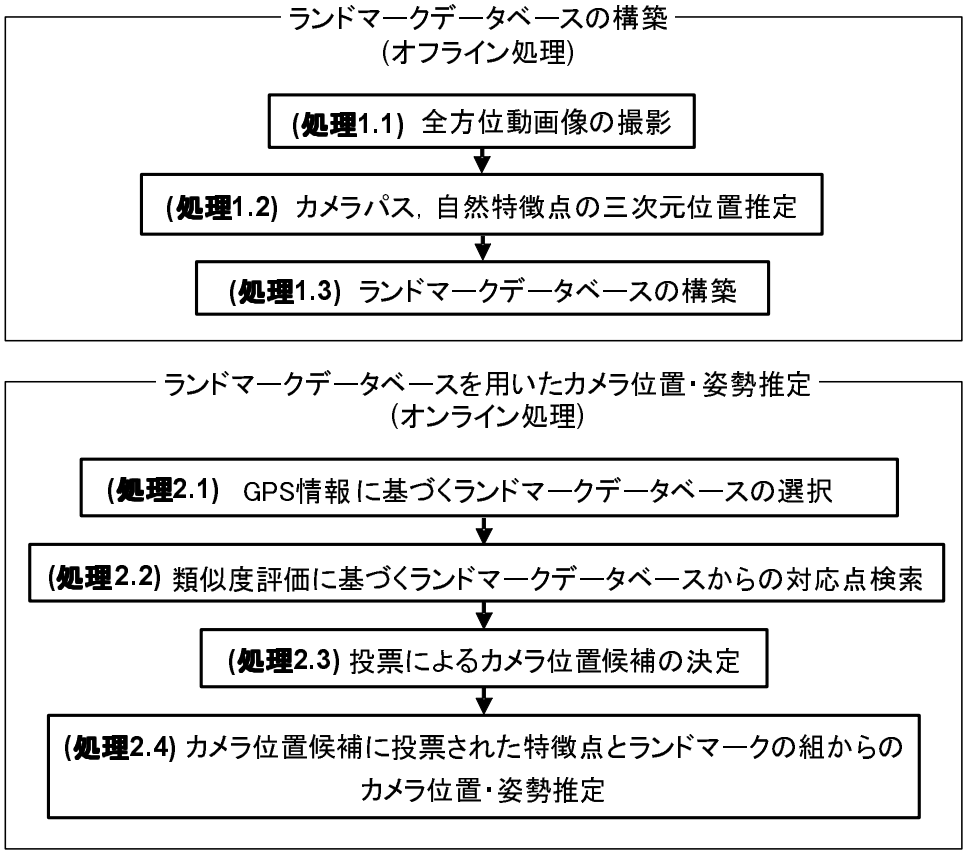


図 4 全体の処理の流れ

二段階から構成される．全体の処理の流れを図 4 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，まず事前に環境内を全方位動画像として撮影し，環境の粗な三次元復元を行う．次に，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，入力画像の特徴点と対応付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．なお，本研究では全方位カメラおよび利用者の携帯端末に取り付けられたカメラの内部パラメータはあらかじめ校正済みとする．

3. ランドマークデータベースの構築

本章では，次章で述べるカメラ位置・姿勢推定に必要なランドマークデータベースの構築方法 (オフライン処理) について詳述する．本手法では，環境内を撮影した全方位動画像から自然特徴点の 3 次元位置と撮影地点情報を抽出し，ランドマークとして用いる．ランドマークデータベースの構築では，図 4 に示すように，まず広範囲を一度に撮影可能な全方位カメラを用いて環境内を移動しながら撮影する (処理 1.1)．次に，全方位動画像上の自然特徴点追跡による三次元復元を行い，自然特徴点の三次元位置と動画像撮影時のフレームごとのカメラ位置・姿勢情報を推定する (処理 1.2)．最後に，全方位動画像，および推定した自然特徴点の三次元位置とフレーム毎のカメラ位置・姿勢情報から，各自然特徴点のランドマーク情報を取得し，ランドマークデータベースを構築する (処理 1.3)．また，4.1 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十 m 間隔のブロック単位に分割しておく．

以下では，まずランドマークデータベースの構成要素について述べ，次に各ランドマークの情報を取得するための手順について述べる．

3.1 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 5 に示す．ランドマークは，次章で述べ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 (オンライン処理) において，入力画像中の特徴点との対応付けに用いる．データベース内には多数のランドマークが登録されるため，入力画像上の特徴点と対応するランドマークをデータベース中から効率よく探索する必要がある．本研究で用いるランドマークデータベースは，文献 [19] 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．ランドマークは，1 つの (A) 自然特徴点の三次元位置と複数の (B) 撮影地点情報から成る．撮影地点情報は，(a) ランドマーク撮影時の全方位カメラの

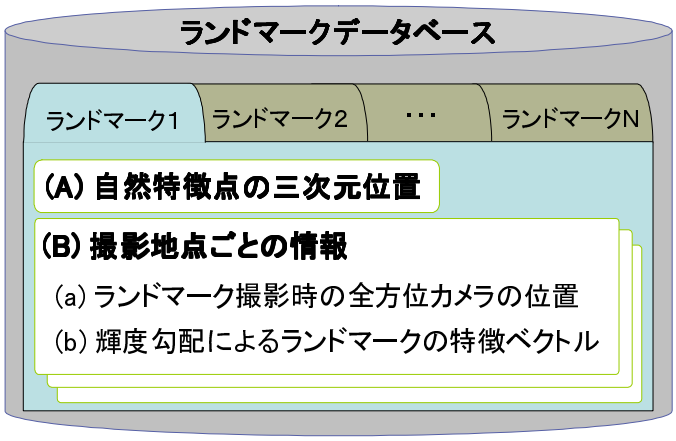


図 5 ランドマークデータベースの構成要素

位置と (b) 輝度勾配によるランドマークの特徴ベクトルによって構成される．以下にそれぞれの要素について詳述する．

(A) 自然特徴点の三次元位置

自然特徴点の三次元位置は，ランドマークの三次元位置と入力画像中の自然特徴点の二次元座標の組から端末のカメラ位置・姿勢を推定するために用いる．この三次元位置は，次節で述べる環境の三次元復元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る．世界座標系はＸ軸，Ｙ軸が実環境における地面に対して水平，Ｚ軸が地面に対して垂直な座標系であるとする．

(B) 撮影地点情報

撮影地点情報は，ランドマークと入力画像の特徴点を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情報を登録する．撮影地点情報は次の 2 つの要素で構成される．

(a) 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，自然特徴点の三次元

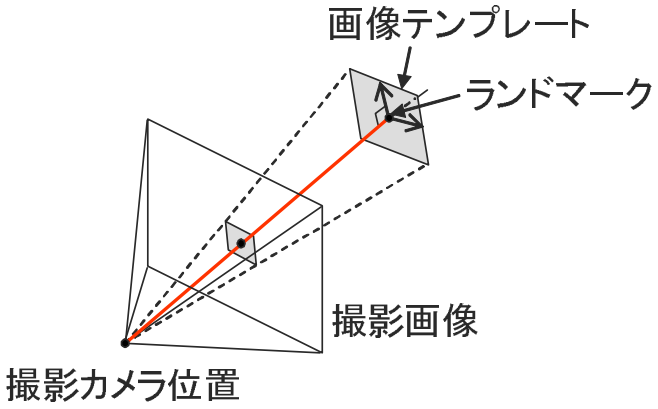


図 6 特徴ベクトルを抽出するためのランドマークの画像テンプレート

位置 (A) と同様の世界座標系で保持する．4 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．

(b) 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 [26] 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．また，各ランドマークの画像は，図 6 のように，世界座標系においてカメラの投影中心とランドマークを結ぶ直線に対して垂直な面に撮影画像を投影することで作成する．これにより，撮影時のカメラ姿勢やレンズ歪みの影響を排除した画像テンプレートを生成し，その画像情報から特徴ベクトルを抽出する．

3.2 全方位動画像からの環境の三次元復元によるランドマーク情報の獲得

本節では，全方位動画像からの三次元復元を利用して，ランドマーク情報を獲得し，ランドマークデータベースを構築する方法について述べる．ランドマークデータベースの構成要素は先に図 5 に示した通りである．はじめに，全方位動画像から環境の三次元復元を行うことによって，自然特徴点の三次元位置と画像上の座標，および全方位動画像のカメラパスを取得する．次に，三次元復元で得られた情報を用いて，輝度勾配によるランドマークの特徴ベクトルをランドマークの撮影地点情報ごとに算出する．

以下では，自然特徴点の三次元位置 (A) と撮影地点情報であるランドマーク撮影時の全方位カメラの位置 (B-a) を取得するための三次元復元手法，輝度勾配によるランドマークの特徴ベクトル (B-b) の作成方法について述べる．

3.2.1 全方位動画像による環境の三次元復元

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 [15] を用いて，三次元位置が既知の少数の基準特徴点と，移動を伴って取得した全方位動画像上から Harris オペレータ [27] によって検出された自然特徴点を追跡することによって，自然特徴点の三次元位置と画像上の座標，および全方位動画像のカメラパスを取得する．Harris オペレータでは，入力画像上の座標 $\mathbf{x} = (x, y)$ の特徴量 $F(\mathbf{x})$ 算出のために，まずガウシアンオペレータによる入力画像の平滑化処理を行う．次に一定の大きさの正方形窓 $\mathbf{W}$ において，画像上の輝度 I の勾配 I_x，I_y を用いて以下に示す行列 $\mathbf{A}$ を算出する．

$$\mathbf{A} = \sum_{\mathbf{x} \in \mathbf{W}} \begin{pmatrix} I_x(\mathbf{x})^2 & I_x(\mathbf{x})I_y(\mathbf{x}) \\ I_x(\mathbf{x})I_y(\mathbf{x}) & I_y(\mathbf{x})^2 \end{pmatrix} \tag{1}$$

この行列 $\mathbf{A}$ を用いて，特徴量 $F(\mathbf{x})$ を以下の式により算出する．

$$F(\mathbf{x}) = \mathbf{det}(\mathbf{A}) - \alpha \mathbf{trace}(\mathbf{A})^2 \tag{2}$$

ただし，α はオペレータの感度を表し，本研究では経験的に $\alpha = 0.06$ を用いる．画像内のすべての画素 $\mathbf{x}$ において特徴量 $F(\mathbf{x})$ を算出した後に，一定サイズのウィンドウ内で特徴量 $F(\mathbf{x})$ が極大値となる点を画像特徴点として検出する．このようにして窓 $\mathbf{W}$ 内で検出された 2 次元特徴点を用いる．

佐藤らの手法では，まず基準特徴点の三次元位置をトータルステーションと呼ばれる三次元計測機材を用いて計測し，少数のキーフレームの画像上で基準特徴点を指定する．次に，全方位画像中の三次元位置が未知の自然特徴点と基準特徴点を同時に自動追跡し，動画像全体での最適化処理を行うことで，カメラパラメータの累積的な推定誤差を最小化する．これにより，複雑で広範囲な環境を撮影した場合でも，全方位動画像のカメラパスと自然特徴点の三次元位置を基準特徴点による絶対座標系で求めることができる．また，この三次元復元から，自然特徴点の三次元位置 (A) とランドマークの撮影地点情報であるランドマーク撮影時の全方位カメラの位置 (B-a) を取得する．

3.2.2 特徴点の輝度勾配による特徴ベクトルの抽出

輝度勾配によるランドマークの特徴ベクトル (B-b) は，ランドマークと入力画像の特徴点の類似性を評価するために用いられる．特徴ベクトルは以下の手順で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
   図 7(i) に示すような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．図 6 に示したように，ここでは大江らの手法 [19] 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. 特徴ベクトルの抽出
   特徴ベクトルの抽出処理を図 7(ii)，(iii) に示す．本研究では，SIFT 記述子

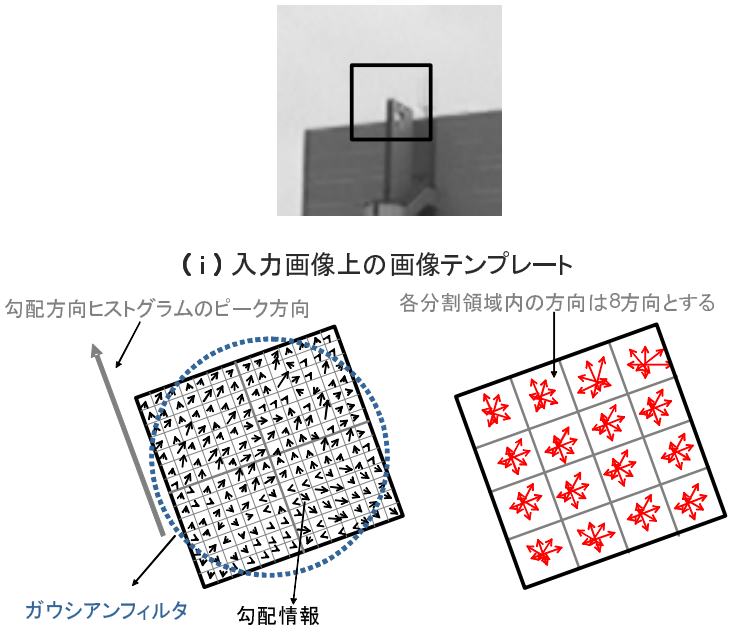


図 7　入力画像上の画像テンプレートと特徴ベクトルの抽出処理

[26] を用いて特徴ベクトルを生成する．SIFT 記述子では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する．次に，画像テンプレート上の各輝度勾配に対してテンプレートの中心を原点とするガウシアンによって重み付けを行い，勾配方向ヒストグラムにおいて頻度が最も多い方向に画像テンプレートを回転させる (図 7(ii))．さらに，画像テンプレートを $M \times M$ 領域に分割，各分割領域内における勾配方向を 8 方向に分類し，各方向に分類された勾配方向の勾配の大きさを加算することで，$D = M \times M \times 8$ 次元の特徴ベクトル $\mathbf{f} = (v_1, v_2, \cdots, v_D)$ を生成する (図 7(iii))．このとき，分割領域の幅 s_w は画像テンプレートの幅を M で割った値であり，小数点以下は切り捨てる．さらに，画像テンプレートの中心から s_w を用いて $M \times M$ 分割することで特徴ベクトルの生成を行う．

4. 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，3 章で述べた手法で構築したランドマークデータベースを用いて，静止画像からのカメラ位置・姿勢推定を実現する手法について述べる．先にも述べたように，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．図 4 に示したように，本手法では，まず GPS から得られる数十 m 程度の誤差を含む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 2.1)．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 2.2)．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 2.3)．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 2.4)．以下，それぞれの処理について述べる．

4.1 GPS 情報に基づく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 から取得した数十 m 程度の誤差を含む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 2.1)．この処理では，まず，GPS 情報とランドマークの撮影地点情報を基に数十 m 間隔で分割されたランドマークデータベースから推定処理に用いるブロックを複数選択する．次に，GPS から得られる計測誤差が最大 γ[m]，携帯機器の位置に最も近いランドマークデータベース構築時のカメラ位置と携帯機器の間の距離が最大 Γ[m] であると仮定し，GPS による計測位置とのユークリッド距離が $\gamma + \Gamma$[m] 以内の撮影地点情報を持つランドマーク群を以降の処理で用いるデータベースとして選択する．

4.2 類似度評価に基づくランドマークデータベースからの対応点探索

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 2.2)．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 [27] を用いて多数の特徴点を検出する．次に，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 3.2.3 節で述べた手順で特徴ベクトルを抽出する．ただし，画像の傾き補正処理では入力画像上の特徴点の 3 次元位置が不明なため，図 8 に示すように，カメラ座標系において，カメラの投影中心と特徴点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，傾き補正された画像テンプレートを生成する．次に，入力画像の各特徴点から得られた特徴ベクトル $\mathbf{f}_{IN} = (v_{IN1}, \cdots, v_{IND})$ と，処理 2.1 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 $\mathbf{f}_{LM} = (v_{LM1}, \cdots, v_{LMD})$ について，総当りで次式に示す特徴空間上における二乗距離 (非類似度)S を算出する．

$$S = |\mathbf{f}_{IN} - \mathbf{f}_{LM}|^2 = \sum_{d=1}^{D} (v_{INd} - v_{LMd})^2 \tag{3}$$

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 個までのランドマークを 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．

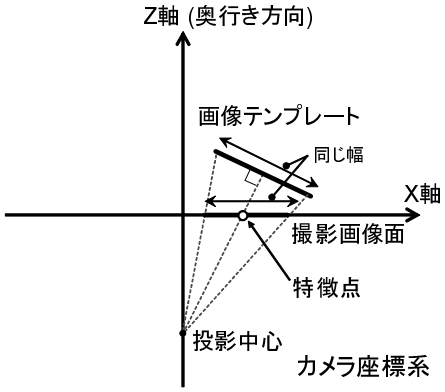


図 8 入力画像における画像傾き補正処理

4.3 投票によるカメラ位置候補の決定

前節で述べた処理 2.2 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 2.2 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 2.3)．

具体的には，図 9 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 2.2 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 $\mathbf{w}_{ij}(-n \leq i \leq n, -n \leq j \leq n)$ を以下のように定義する．

$$\mathbf{w}_{ij} = \begin{pmatrix} w_i \\ w_j \end{pmatrix} = \begin{pmatrix} g_x + L \times i \\ g_y + L \times j \end{pmatrix} \tag{4}$$

ただし，L は格子間隔を表す．次に，入力画像のすべての特徴点について処理 2.2

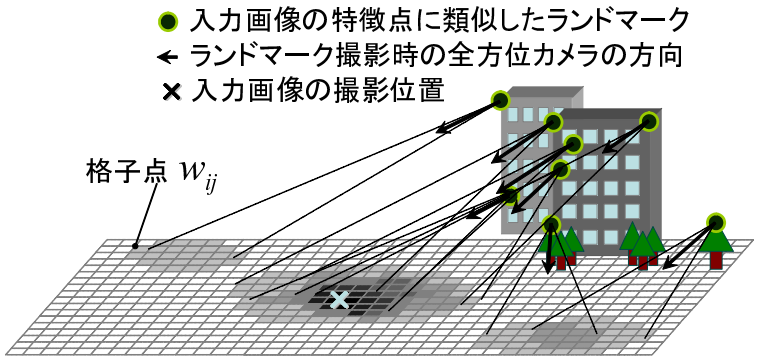


図 9　入力画像の撮影位置と入力画像の特徴点に類似したランドマークの関係

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 $\mathbf{w}_{ij}$ に投票する．

[**条件 1**] 格子点 $\mathbf{w}_{ij}$ 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 $\mathbf{c}_p$ の距離 $|\mathbf{w}_{ij} - \mathbf{c}_p|$ が閾値 T_1 以下

[**条件 2**] ランドマーク p の三次元位置 $\mathbf{S}_p$ と全方位カメラの位置 $\mathbf{C}_p$ を結ぶ直線，および格子点 $\mathbf{w}_{ij}$ に $\mathbf{C}_p$ 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 $\mathbf{S}_p$ を結ぶ直線の成す角度が閾値 T_2 以下

ただし，入力画像の同一特徴点に対して選択された複数のランドマークからの投票について，同一の格子点 $\mathbf{w}_{ij}$ への重複した投票は行わない．

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．ただし，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．

4.4 カメラ位置候補に投票された特徴点とランドマークの組からのカメラ位置・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 2.4)．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 6 点以上用いて PnP 問題 [28] 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，尤もらしいカメラ位置・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[29] を用いる．RANSAC による誤対応の排除では，以下の処理を行う．

1. 次の処理を k 回繰り返す.

   (a) i 回目の繰り返し処理において，対応付けられたランドマークの三次元座標と入力特徴点の二次元座標の組からランダムに P 点 (6 点以上) を選択し，暫定的なカメラ位置・姿勢を推定する．

   (b) 入力画像の特徴点の座標 (u_{ip}, v_{ip}) と，暫定的に推定したカメラ位置・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 $(\hat{u}_{ip}, \hat{v}_{ip})$ との距離の二乗誤差である再投影誤差 R_{ip} と，再投影誤差の中間値 RM を以下の式から求める．

$$R_{ip} = (u_{ip} - \hat{u}_{ip})^2 + (v_{ip} - \hat{v}_{ip})^2 \tag{5}$$

$$RM_i = med(R_{i1}, R_{i2}, \cdots, R_{iQ}) \tag{6}$$

   ここで，Q は対応付けられたランドマークの数とする．

2. ステップ 1 で得られた複数個の再投影誤差の中間値 RM が最小となる暫定的なカメラ位置・姿勢情報を選択する．

3. 選択されたカメラ位置・姿勢情報を用いて各ランドマークの投影誤差を評価し，閾値を超える結果を誤対応として削除する．

最後に，得られた誤対応排除後の結果から，以下の式で定義される再投影誤差の和 E(以下，再投影誤差) が最小となるカメラ位置・姿勢を求める．

$$E = \sum_{j=1}^{l} \{(u_j - \hat{u_j})^2 + (v_j - \hat{v_j})^2\} \tag{7}$$

ここで，l は誤対応排除後の特徴点数である．ここでは線形最小二乗法によってカメラ位置・姿勢の初期値を算出し補正を行った後に，Levenberg-Marquardt 法によって再投影誤差の非線形最小化 [30] を行う．このようにして得られたカメラ位置・姿勢情報を最終的な推定結果とする．

処理の繰り返し回数 k は，誤対応を含まない特徴点の組のみで暫定的なカメラ位置・姿勢が推定される確率 ζ を 1 に近づけるよう設定する．ζ は，以下の式で算出される．

$$\zeta = 1 - (1 - (1 - \epsilon)^P)^k \tag{8}$$

なお，ε はステップ 1 で暫定的なカメラ位置・姿勢推定に使用されるランドマークの中に，誤対応が含まれる確率である．これにより，例えば，$\varepsilon = 0.2, P = 10$ と設定すれば，$\zeta = 0.99998$ となるための繰り返し回数は $k = 100$ となる．

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 E 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，尤もらしい推定結果であると判断する．

[**条件 1**] 再投影誤差 E 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

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ら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処理の失敗を通知する．

5. 実験

提案手法の有効性を検証するため，屋外・屋内環境における実験と定量的な精度評価を行った．まず，本実験ではランドマークデータベースを構築するために，図 10 左に示すような水平方向に 5 個，上方向に 1 個の CCD カメラを配置した全方位マルチカメラシステム (Point Grey Research 社 Ladybug) を用いて屋外・屋内環境を移動しながら撮影した．さらに，佐藤らの手法 [15] を適用し，取得した全方位動画像 (図 10 右) のカメラパスと自然特徴点の三次元座標を推定し，ランドマークデータベースを構築した．次に，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，定量的な評価実験では，入力画像の特徴点とその 3 次元位置を手動で対応付けて撮影位置・姿勢を推定した結果を正解データとし，本手法によるカメラ位置・姿勢推定結果との比較を行った．ただし，本実験では，サーバ・クライアント型システムの構築は行わず，携帯端末による撮影画像を想定した複数の静止画像を入力として用いた．また，入力画像を撮影した単眼カメラの内部パラメータはあらかじめ Tsai の手法 [31] によって校正し，撮影時は内部パラメータを固定した．なお，各実験における入力画像については，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自然特徴点の位置を手動で対応付け，PnP 問題 [28] を解くことでカメラ位置・姿勢の正解データを作成した．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 画素，特徴ベクトルの次元数を 128 とした．特徴ベクトル作成時の画像スケールは，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．

5.1 屋外環境における実験

5.1.1 ランドマークデータベースの構築 (屋外実験)

本実験では，屋外環境 (本学キャンパス:約 75m) を図 10 に示す全方位マルチカメラシステムで 1,251 フレームの動画像 (解像度 768×1024 画素 $\times 6$，15fps) として撮影した．次に，三次元復元手法を適用し，10 フレームおきの画像上に存在す

図 10　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

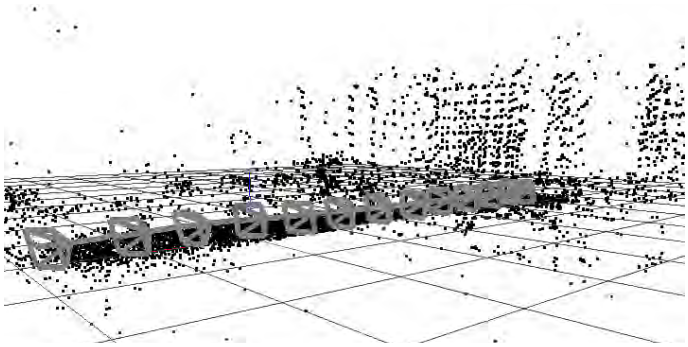


図 11　推定されたカメラパスとランドマークの三次元位置

る自然特徴点を用いて，3.2 節で述べた手法によりランドマークデータベースを構築した．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 11 に示す．図中の曲線は推定されたカメラパスを，四角錘は 100 フレーム毎のカメラ位置・姿勢を表す．ここでは 3.2 節で述べた手法により約 12,400 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 つの撮影地点ごとのランドマーク情報が生成された．

5.1.2　提案手法によるカメラ位置・姿勢推定 (屋外実験)

164 枚の静止画像 (解像度 720 × 480 画素) を撮影し，各画像に対して先に構築したデータベース内のランドマークすべてを用いたカメラ位置・姿勢推定を行っ

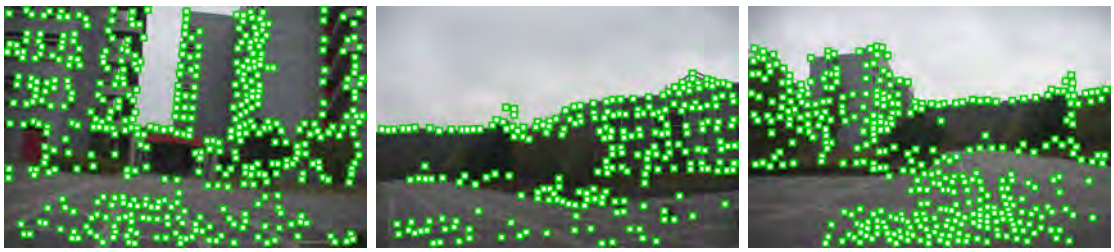
表 1 カメラ位置・姿勢推定の各処理における閾値 (屋外実験)

処理 2.2	1 つの特徴点に対応付けるランドマークの数 α	3
処理 2.2	光軸回りの回転制約 (度)	15
処理 2.3	格子間隔 L(cm)	50
処理 2.3	条件 1 の閾値 T_1(m)	5
処理 2.3	条件 2 の閾値 T_2(度)	5
処理 2.4	RANSAC の繰り返し回数 k	300
処理 2.4	条件 1 の閾値 (画素)	5
処理 2.4	条件 2 の閾値 (個)	10

た．ただし，ランドマークデータベース構築時のカメラパスが短いため，データベース内のすべてのランドマークを用いる．本実験で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．まず，処理 2.4 においてシステムが自動的に推定結果の尤もらしさを判断した結果，本実験では，164 枚中 102 枚が尤もらしい推定結果である (成功した) と判断された．図 12 に，システムが成功したと判断したときの，(A) 入力画像，(B) 入力画像から Harris オペレータを用いて検出した入力特徴点，(C) 入力画像上に存在するランドマークの位置，および (D)4.4 節で述べた処理 2.4 のカメラ位置・姿勢推定処理において最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係，を示す．ただし，図 12 におけるランドマークの位置は入力画像の正解データを基に出力した．ここで，図 12(D) 上の○印が前者，□印が後者を表しており，この二つが重なっていれば，正しい対応関係でカメラ位置・姿勢推定を行ったことを表す．図 12 の (C) と (D) 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．これは，環境中に類似パターンの多い地面などのランドマークが多数存在しているため，4.2 節で述べた処理 2.2 の入力画像の特徴点とランドマークの類似度評価処理で，真に対応付くべきランドマークが上位に順位付けされなかったことなどが原因として考えられる．ただし，図 12(D) からほぼ正しい対応関係を用いてカメラ位置・

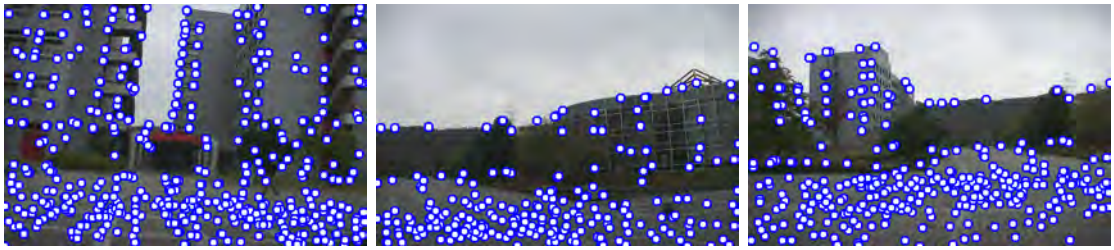
場面 a-1　　場面 a-2　　場面 a-3

(A) 入力画像



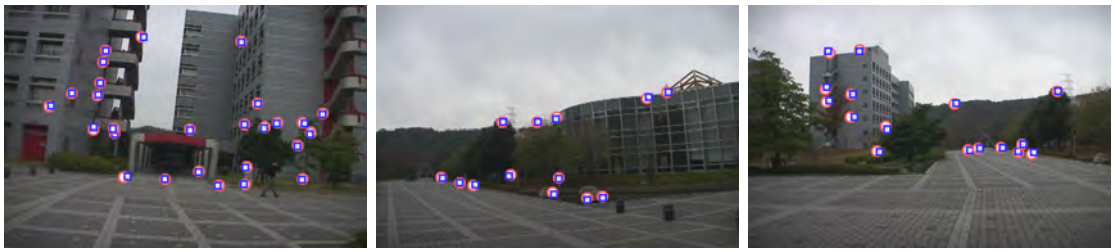
場面 a-1　　場面 a-2　　場面 a-3

(B) 入力特徴点の位置



場面 a-1　　場面 a-2　　場面 a-3

(C) 環境中に存在するランドマークの画像上の位置



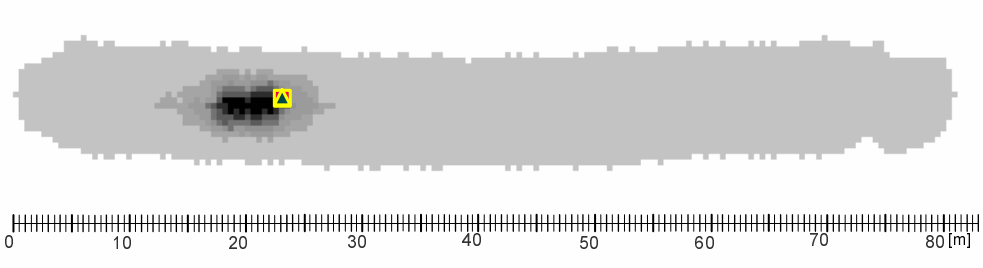
場面 a-1　　場面 a-2　　場面 a-3

(D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの画像上の位置と
そのランドマークと対応する入力特徴点の関係

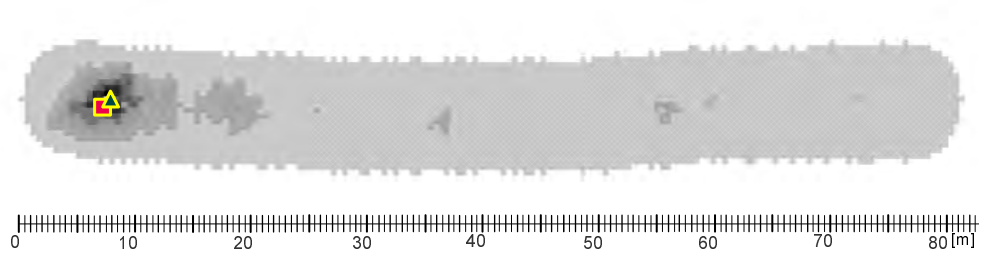
図 12　屋外実験：成功判断時の結果

姿勢が推定できていることが分かる．また，図 13 に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，4.3 節で述べた処理 2.3 の投票処理を行った結果を示す．図中の濃度値は投票数を表しており，暗いほど投票数が多い．また，図中の 1 グリッドは実環境の 50cm を表している．なお，図中の□印は入力画像のカメラ位置の正解データを，△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．投票結果から，場面 a-1，場面 a-2 では，入力画像の撮影位置付近に投票数の極大値が出現しており，最終的な推定結果を出力したカメラ位置候補も入力画像の撮影位置付近にあることが分かる．場面 a-3 では，投票数の極大値が分散しており，最大投票数の領域が入力画像の撮影位置から 60m 程度離れた位置に出現しているが，最終的な推定結果を出力したカメラ位置候補は入力画像の撮影位置付近に出現している．最大投票数の領域が入力画像の撮影位置から離れた理由は，入力画像の特徴点と対応付くべきではないランドマークの投票によって生じる誤投票が多数発生したためと考えられる．ただし，本手法では誤投票を考慮し，複数のカメラ位置候補を設定して順次探索するため，最終的な推定結果を出力したカメラ位置候補が正解データ付近に現れ，結果的にカメラ位置・姿勢推定に成功したと判断された．

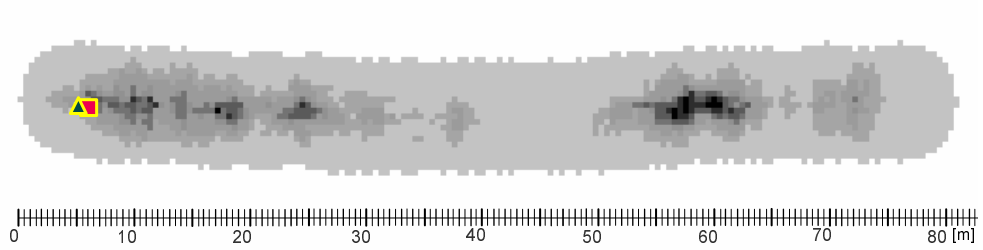
また，図 14 に，システムが失敗したと判断したときの，(A) 入力画像，(B) 入力特徴点，(C) ランドマークの位置，および (D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係を，図 15 に投票結果を示す．図 14(D) から，推定に用いられたランドマークと入力特徴点の関係に誤対応がある場合や，ほぼ正しい対応関係でもその推定に用いられた特徴点数自体が少ない，もしくは少なすぎて推定されない場合などがあることが分かった．図 15 の場面 b-1，場面 b-2 に示すように，正解データ付近に極大値が現れる場合も見られるが，最終的なカメラ位置・姿勢推定に用いることができる特徴点数が少ないために，システムは失敗という判断を行っている．



場面 a-1



場面 a-2

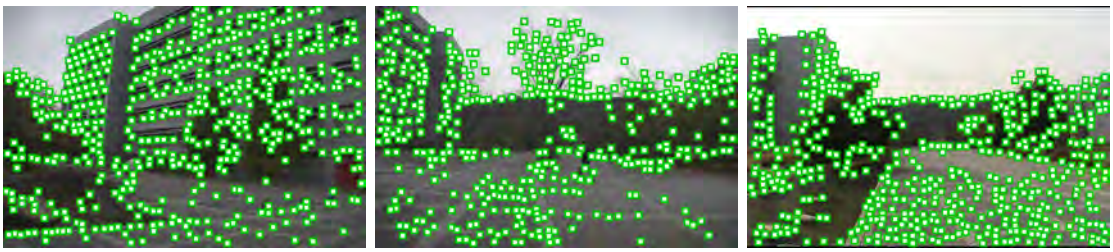


場面 a-3

図 13　処理 2.3 の投票結果 (屋外実験：成功判断時)

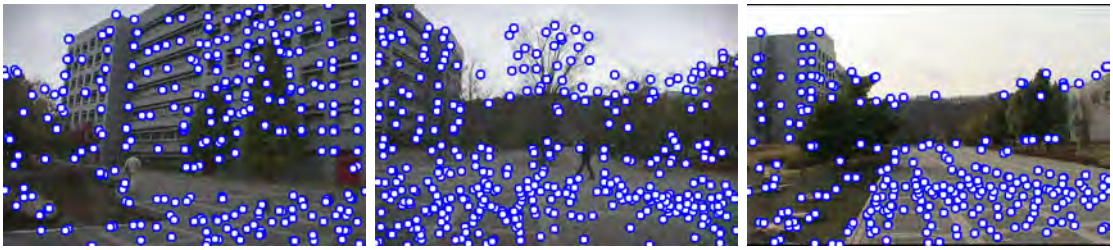
場面 b-1　場面 b-2　場面 b-3

(A) 入力画像



場面 b-1　場面 b-2　場面 b-3

(B) 入力特徴点の位置



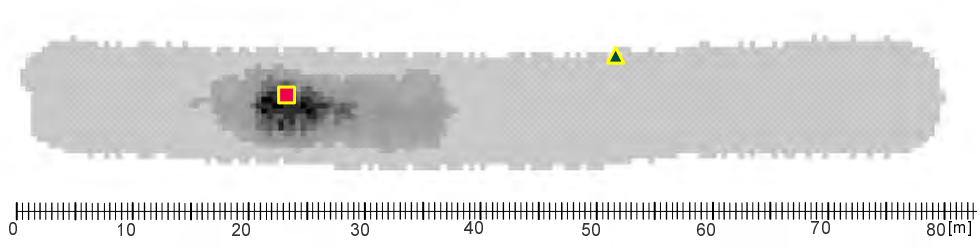
場面 b-1　場面 b-2　場面 b-3

(C) ランドマークの位置

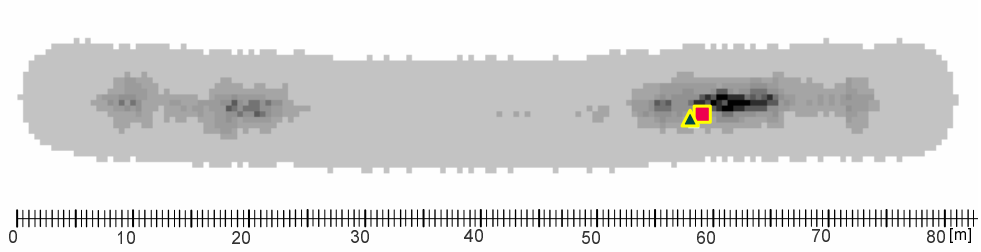
場面 b-1　場面 b-2　場面 b-3

(D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの画像上の位置と
そのランドマークと対応する入力特徴点の関係

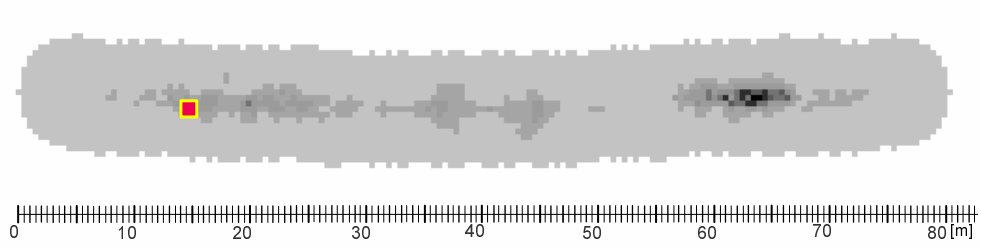
図 14　屋外実験：失敗判断時の結果



場面 b-1



場面 b-2



場面 b-3

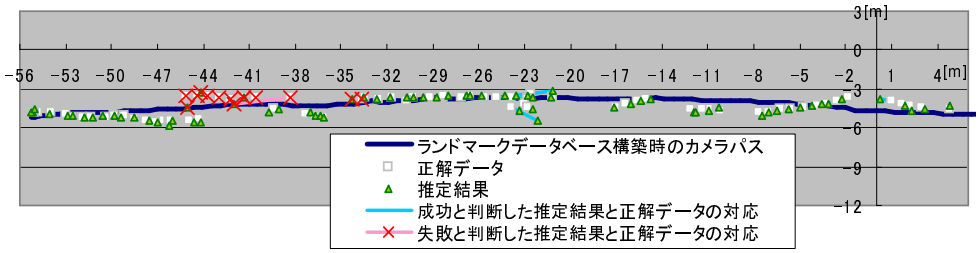
図 15　処理 2.3 の投票結果 (屋外実験：失敗判断時)

5.1.3 定量的な評価 (屋外実験)

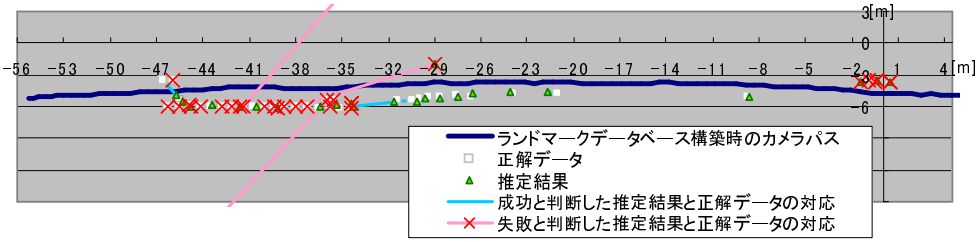
本節では，屋外において推定されたカメラ位置・姿勢に関する定量的な評価について述べる．まず，本実験で使用した正解データと推定結果，およびランドマークデータベース構築時のカメラパスについて，入力画像撮影時のカメラとランドマークデータベース構築時のカメラ間の距離 (カメラ間距離) が，(i)0〜1m 以内，(ii)1〜2m 以内，(iii)2〜3m 以内，(iv)3〜4m 以内の結果を図 16 に示す．図 16 において，太い実線はランドマークデータベース構築時のカメラパス，□印が正解データ，△印が推定結果を表す．また□印と△印を結ぶ細い実線が成功した推定結果と正解データの対応を表し，×印は失敗した推定結果の正解データを表す．システムが成功したと判断したカメラ位置・姿勢推定結果に関する位置推定の誤差は平均約 1172mm，姿勢の推定誤差は平均約 0.89 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 17.2 個，再投影誤差は平均 3.1 画素であった．ただし，位置誤差が大きいにも関わらず，成功したと判断されている結果も見られた．

また，より詳細な結果の分析を行うため，推定結果をカメラ間距離に応じて複数のグループに分割し，それぞれのグループについて評価した．図 17，図 18 にグループごとの成功率，位置誤差の平均を示す．図 17 において，カメラ間距離が離れるほど成功率が下がるが，2.5m 付近を超えると増加する傾向が見られた．また図 18 において，カメラ間距離 2m 付近から位置誤差が大きくなる傾向が見られた．これらの図からカメラ間距離が離れ過ぎると推定結果の尤もらしさを判断する処理がうまく働かなくなることが分かった．なお，カメラ間の距離が 2m 以内のカメラ位置・姿勢推定結果に関する位置推定の誤差は平均 420mm であり，姿勢の推定誤差は平均 0.65 度であった．この結果から，本実験の屋外環境では，ランドマークデータベース作成時のカメラ位置から 2m 以内の範囲であれば，注釈提示などには十分な精度の結果が得られることが分かった．

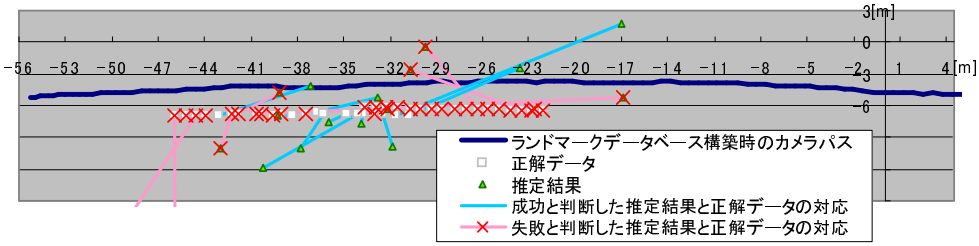
次に，図 19 にカメラ間距離と推定されたすべての結果に対する位置誤差の関係を示す．図中の○印は成功と判断された推定結果と正解データの位置誤差，×印は失敗と判断された推定結果と正解データの位置誤差を表す．ここで，カメラ位置・姿勢推定に使用された特徴点数が少なすぎるなどして推定結果が出力され



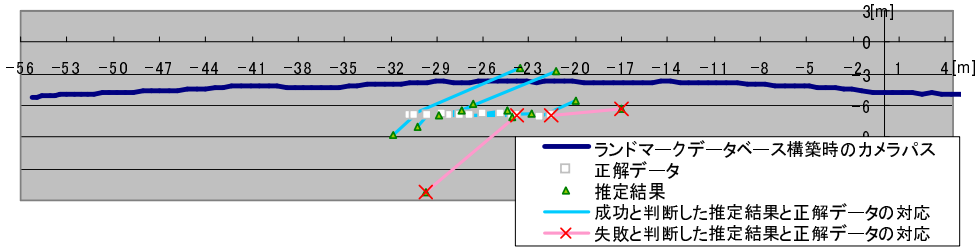
(i) カメラ間距離が 0〜1m 以内の結果



(ii) カメラ間距離が 1〜2m 以内の結果



(iii) カメラ間距離が 2〜3m 以内の結果



(iv) カメラ間距離が 3〜4m 以内の結果

図 16　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正解データおよび推定されたカメラ位置 (屋外実験)

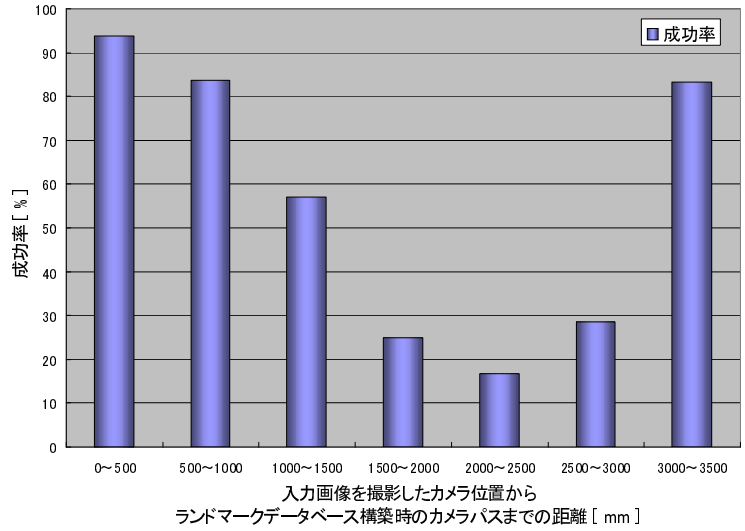


図 17　カメラ間距離と成功率の関係 (屋外実験)

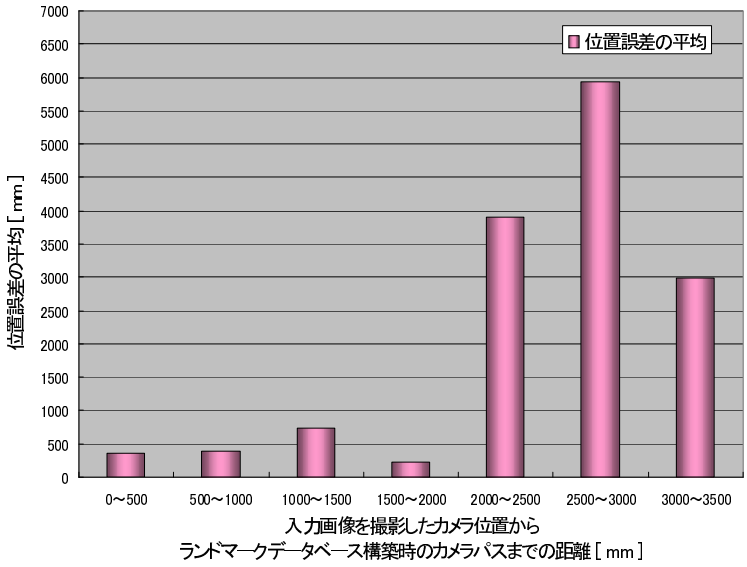


図 18　カメラ間距離と位置誤差の関係 (屋外実験)

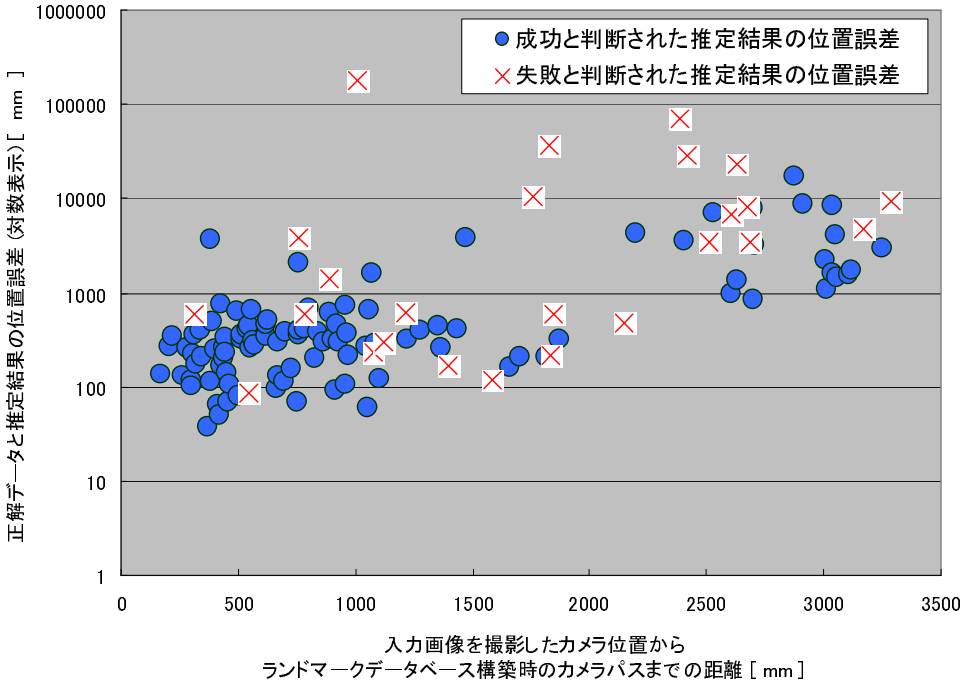


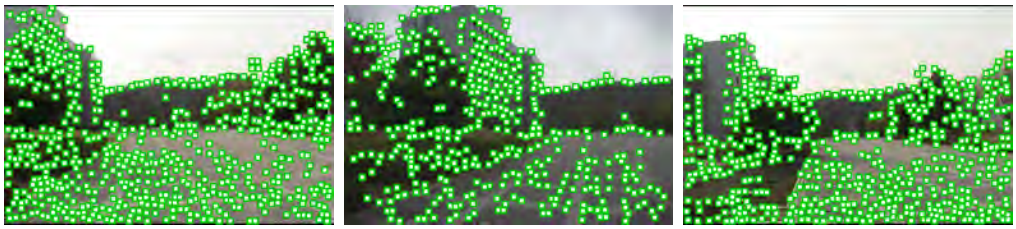
図 19　カメラ間距離とすべての推定結果の関係 (屋外実験)

なかった結果は表示していない．図 19 において，カメラ間距離が小さいにも関わらず 1m 以上の大きな位置誤差で成功と判断された結果が存在することが分かる．図 20 に，位置推定誤差が大きいにも関わらず推定が成功と判定された結果に対応する，(A) 入力画像，(B) 入力特徴点の位置，(C) ランドマークの位置，および (D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係を示す．図 20 から，入力画像内の大半が自然物や類似パターンで占められているため，複数の誤対応が発生していることが分かった．

また，提案手法における 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 45 秒であり，その内訳は，処理 2.2 の入力画像の特徴点とランドマークの類似度評価と処理 2.3 の投票処理に 32 秒，処理 2.4 のカメラ位置・姿勢推定処理に 13 秒であった．

場面 c-1　場面 c-2　場面 c-3

(A) 入力画像



場面 c-1　場面 c-2　場面 c-3

(B) 入力特徴点の位置

場面 c-1　場面 c-2　場面 c-3

(C) ランドマークの位置

場面 c-1　場面 c-2　場面 c-3

(D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係

図 20　成功と判断されたが大きな推定誤差が生じていた結果

5.2 屋内環境における実験

5.2.1 ランドマークデータベースの構築 (屋内実験)

屋外環境と同様に，全方位マルチカメラシステムを用いて屋内環境 (本学キャンパス情報研究科 B 棟 3 階) を移動しながら 701 フレームの全方位動画像として撮影し，ランドマークデータベースを構築した．ここでは 5 フレームおきの画像上の自然特徴点を用いることで，3.2 節で述べた手法により約 2,100 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 9.4 地点で撮影地点ごとのランドマーク情報が生成された．

5.2.2 提案手法によるカメラ位置・姿勢推定 (屋内実験)

屋外環境に対する実験と同様に表 2 に示す閾値を用いて提案手法によるカメラ位置・姿勢推定実験を行った．ただし，ランドマークデータベース構築時のカメラパスが短いため，本実験においても屋外実験と同様に GPS は用いず，データベース内のすべてのランドマークを用いる．まず，処理 2.4 によりシステムが自動的に推定結果の尤もらしさを判断した結果，本実験においては，14 枚中 10 枚がカメラ位置・姿勢推定に成功したと判断された．図 21 に，システムが成功したと判断したときの，(A) 入力画像，(B) 入力特徴点の位置，(C) ランドマークの位置，および (D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係を示す．屋外実験と同様に，図 21(D) 上の○印が前者，□印が後者を表しており，この二つが重なっていれば，正しい対応関係でカメラ位置・姿勢推定を行ったことを表す．図 21 の (C)，(D) から，屋内実験においてもカメラ位置・姿勢推定に用いられる特徴点数が少ないことが分かった．また，図 22 に投票処理を行った結果を示す．図 22 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．

また，図 23 に，システムが失敗したと判断したときの，(A) 入力画像，(B) 入力特徴点の位置，(C) ランドマークの位置，および (D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの位置とそのランドマークと対応する入力特徴点の関係を，図 24 に処理 2.3 の投票処理を行った結果を示す．システムが失敗と

表 2 カメラ位置・姿勢推定の各処理における閾値 (屋内実験)

処理 2.2	1 つの特徴点に対応付けるランドマークの数 α	3
処理 2.2	光軸回りの回転制約 (度)	15
処理 2.3	格子間隔 L(cm)	50
処理 2.3	条件 1 の閾値 T_1(m)	3
処理 2.3	条件 2 の閾値 T_2(度)	15
処理 2.4	RANSAC の繰り返し回数 k	300
処理 2.4	条件 1 の閾値 (画素)	5
処理 2.4	条件 2 の閾値 (個)	7

判断したすべての場面において，環境中に存在する特徴的なランドマークが少ないため，投票に用いることができるランドマークの数が少なく，正しくカメラ位置・姿勢を推定することができなかった.

場面 d-1　　場面 d-2　　場面 d-3

(A) 入力画像

場面 d-1　　場面 d-2　　場面 d-3

(B) 入力特徴点の位置

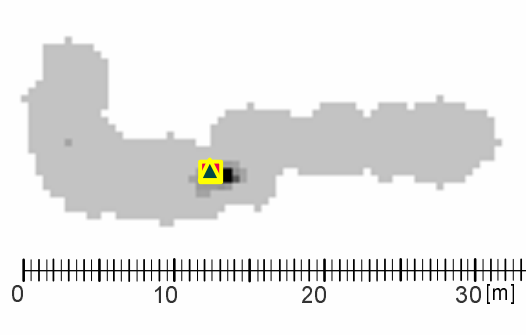
場面 d-1　　場面 d-2　　場面 d-3

(C) ランドマークの位置

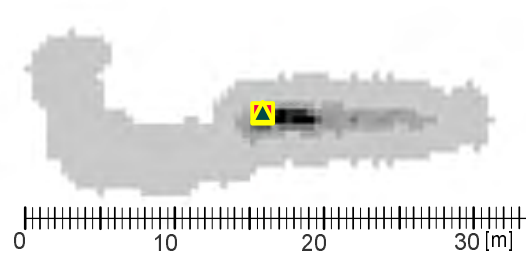
場面 d-1　　場面 d-2　　場面 d-3

(D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの画像上の位置とそのランドマークと対応する入力特徴点の関係

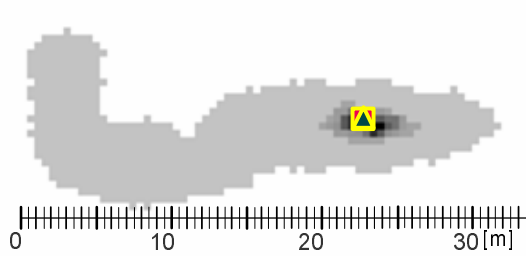
図 21　屋内実験：成功判断時の結果



場面 d-1

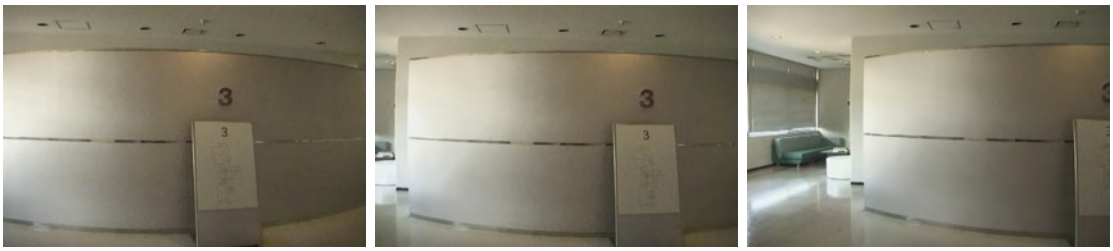


場面 d-2



場面 d-3

図 22　投票結果 (屋内実験：成功判断時)



場面 e-1　　場面 e-2　　場面 e-3

(A) 入力画像

場面 e-1　　場面 e-2　　場面 e-3

(B) 入力特徴点の位置

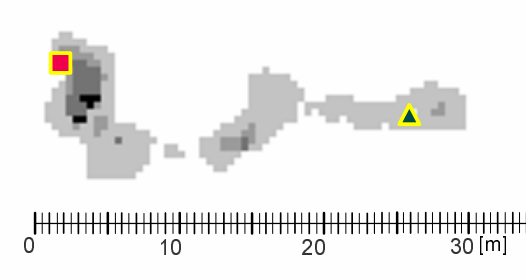
場面 e-1　　場面 e-2　　場面 e-3

(C) ランドマークの位置

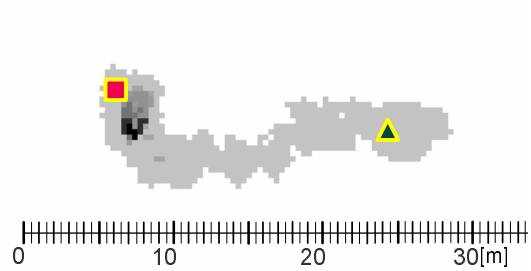
場面 e-1　　場面 e-2　　場面 e-3

(D) 最終的なカメラ位置・姿勢推定に用いられたランドマークの画像上の位置とそのランドマークと対応する入力特徴点の関係

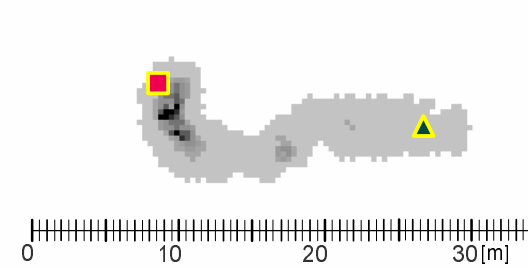
図 23　屋内実験：失敗判断時の結果



場面 e-1



場面 e-2



場面 e-3

図 24　投票結果 (屋内実験：失敗判断時)

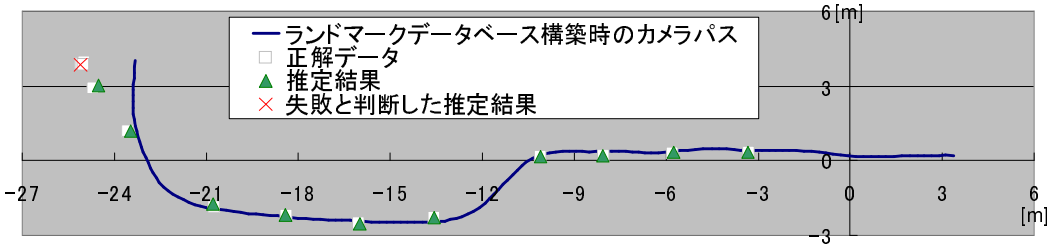


図 25　ランドマークデータベース構築時のカメラパスと入力画像の撮影位置の正解データおよび推定されたカメラ位置 (屋内実験)

5.2.3　定量的な評価 (屋内実験)

本実験で推定されたカメラ位置，正解データ，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 25 に示す．図 25 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，□印が正解データ，△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正解データの対応を表し，×印は失敗した推定結果の正解データを表す．ただし，屋内実験では，ランドマークデータベース構築時のカメラパスに近い位置で撮影した入力画像を使用した．

図 23(C) に示すように，屋内環境で失敗した入力画像中には環境中に存在するランドマークの数が少なかった．システムが成功として出力したカメラ位置・姿勢推定結果の正解データとの位置誤差の平均は約 72mm，姿勢誤差は約 0.7 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 12.5 個，再投影誤差は平均 1.8 画素であった．また，提案手法における 1 枚の入力に対する処理時間は，屋外実験と同じ PC を用いて平均 29 秒であり，その内訳は，処理 2.2 の入力画像の特徴点とランドマークの類似度評価と処理 2.3 の投票処理に 4.3 秒，処理 2.4 のカメラ位置・姿勢推定処理に 24.7 秒であった．

5.3 考察

屋外・屋内環境における実験により，広域環境のランドマークデータベースを構築し，入力画像の特徴点周辺の画像情報との類似度の評価，および空間への投票によるカメラ位置候補の決定によって，データベース中の膨大なランドマークから入力画像の特徴点と対応付くランドマークを効率よく絞り込むことができることを確認した．また，入力画像撮影時のカメラとランドマーク構築時のカメラ間の距離が近ければ，多くの場合において静止画像 1 枚からのカメラ位置・姿勢推定が可能であり，提案手法により利用者にカメラ位置・姿勢推定結果を提供することを想定したシステム側で，尤もらしいカメラ位置・姿勢推定結果を自動的に判断できることを確認した．ただし，カメラ間距離が離れるほど，システムによる推定結果の尤もらしさの判断処理が失敗することが分かった．これは，カメラ間距離によって入力画像とランドマーク内の画像情報のスケールが変化したため処理 2.2 の類似度評価処理で真に対応付くランドマークを選択できず，さらに処理 2.4 のカメラ位置・姿勢推定処理で誤対応をすべて除去できなかったために起こったものであると考えられる．また，図 20 に示したように，入力画像内の大半を占める自然物や類似パターンによる誤対応が原因でカメラ間距離に関係なく大きな誤差が生じる例も見られた．投票処理においては，図 13，図 22 に示したように，入力画像の撮影位置付近に投票数の極大値が集中する投票結果が多数を占めたが，図 13 の場面 a-3 のように，極大値が広域にわたって分散する投票結果もあった．ただし，本手法では投票数が多いカメラ位置候補から逐次カメラ位置・姿勢推定結果の尤もらしさを検証するため，結果的にはカメラ位置・姿勢推定に成功するものが多かった．しかし，図 15，図 24 に示すように，入力画像の撮影位置付近に投票数の極大値が存在しているにも関わらず，カメラ位置・姿勢推定に失敗するものや，入力画像の撮影位置から非常に離れた位置に極大値が発生したものもあった．これらが発生した原因は，入力画像の特徴点に類似するパターンがデータベース内に多数存在することによる誤対応・誤投票が発生したため (原因 1) と考えられる．また図 20 のようにカメラ位置・姿勢推定において成功と判断されても誤対応により大きな推定誤差が発生するものもあった．この原因は，カメラ間距離の類似度評価処理への影響 (原因 2)，および入力画像内の自然

物や類似パターンが画像内の大半を占めることによる誤対応の発生 (原因 3) が考えられる．(原因 1) と (原因 3) による誤対応に対しては，環境中に多数の類似パターンが存在するランドマークに関する投票の制限や，1 つの特徴点に対応付けるランドマークの数 α の環境に対する動的な調整，類似度評価の精度向上などの対策が考えられる．また (原因 2) に対しては，今後，Harris-Laplacian 法 [32] 等を利用することでスケール変化にも対応することによりある程度解決できると考えられる．また，現在のところ，図 12(D)，図 21(D) に示すように，提案手法によるカメラ位置・姿勢推定に用いられる特徴点数は，図 12(C)，図 21(C) に示す環境中に存在するランドマークと比較して大幅に少ない．そこで，今後，推定したカメラ位置・姿勢推定を基に環境中のランドマークを限定し再度探索を行い正しい対応付けを行うことで，推定精度が向上することが見込まれる．

また，本手法は環境に特徴的なランドマークが少ない場合には推定処理ができないという原理的な問題が残されている．これついては，環境中の特徴的なランドマークの存在を自動的に判断し，ランドマークが少ない環境であれば，利用者に取り直しを要求することなどが必要となる．今回の実験において，これらの問題が顕著に現れた例が，屋内環境における実験である．図 25 に×印で表された推定に失敗した結果は，環境中のランドマークの数が少なかったことが失敗の主な原因であると考えられる．さらに，屋内環境において失敗した結果は，ランドマークデータベース構築時のパスから 2m 程度離れており，かつ入力画像の撮影位置から撮影対象までの距離が短いためにスケール変化の影響を受けやすかったことも要因として挙げられる．

最後に，現在得られているカメラ位置・姿勢推定の応用範囲について考察する．屋外環境の実験結果から，例えばカメラ間距離が 2m 以内であった場合，その位置誤差は 420mm，姿勢誤差は 0.65 度であり，カメラ位置から 20m 離れた位置に仮想物体を重畳表示した場合，投影誤差は最大約 14 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，本手法によって得られるカメラ位置・姿勢情報を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ことが分かる．ただし，先にも述べたように，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，さらに精度の高い推定を実現できるため，上述した用途への利用も期待できる．ここで，ランドマークデータベース構築時の撮影経路については，道幅や撮影対象までの距離などを考慮する必要がある．例えば，今回の屋外実験のような環境で道幅が 4m であった場合，道の中央を全方位カメラで撮影すればよいことが分かる．また，処理時間については，1 枚の入力に対して PC(CPU:Pentium4 3GHz, Memory:1.5GB) を用いて，屋外実験で平均 45 秒，屋内実験で平均 29 秒であった．ここで，4.2 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における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[18, 26] を構築しておくことで，処理の高速化が可能となるため，今後，サーバ・クライアント型システムの実装を行うことで，実用的な携帯型 AR システムを構築することができる見込みである．

6. まとめ

本論文では，自然特徴点ランドマークデータベースを事前に構築し，GPS から取得した誤差数十 m 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・姿勢を推定する手法を提案した．ランドマークデータベース構築時には，全方位動画像を用い，半自動で多数のランドマークデータベースを構築する．また，入力画像とランドマークの類似度を比較するために，回転にロバストな対応点探索が可能で，正規化相互相関などに比べて比較的高速に類似度を判断できる特徴ベクトル (SIFT 記述子) を用いた．提案手法では，入力画像の特徴点と対応付くランドマークを効率よく探索するため，入力画像の特徴点と類似した多数のランドマークの位置関係を利用した投票処理を行った．投票処理では，ランドマークを同じ見え方で撮影できる領域を算出し投票することで，入力画像が撮影された可能性の高いカメラ位置候補を決定する．また，提案手法は，サーバ・クライアント型システムを想定しているため，市販のカメラ付き GPS 携帯やカメラ付き PHS などの携帯端末においてもカメラの絶対位置・姿勢推定が可能である．

実験では，実際の屋外・屋内環境における実験と精度評価を行った．まず，環境内の全方位動画像を撮影し，自然特徴点ランドマークデータベースを構築した．次に，構築したランドマークを用いて複数の静止画像のカメラ位置・姿勢推定を行い，手動で作成した正解データと比較することによって，入力画像撮影時のカメラ位置がランドマークデータベース構築時の撮影経路に近ければ，正解データに近いカメラ位置候補が得られていることや，システム側が自動的に推定結果の尤もらしさを判断した結果がおおよそ正しいことを確認した．また，提案手法による推定結果が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．

今後の課題としては，GPS を用いた広域環境での実験，ランドマークの絞込みにおける各処理の精度向上，処理の高速化，サーバ・クライアント型システムの構築などが挙げられる．まず，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は，Harris-Laplacian 法 [32] を応用してスケールを自動で決定することが考えられる．これによりランドマークデータベース撮影経路の間隔を現状よ

り広くすることができると考えられる．また，処理の高速化では，特徴ベクトルの計算においてKD-Tree[18, 26]を利用することが考えられる．このような課題を解決することによって，広域環境におけるカメラ付き携帯機器を用いたARによるヒューマンナビゲーションシステムなどに利用できる．さらに，提案手法はランドマークデータベースを用いた動画像からのカメラ位置・姿勢推定手法において，初期位置・姿勢を与えることなどにも応用可能である．

謝辞

本研究を進めるにあたり, その全過程において細やかな御指導, 御鞭撻を頂いた視覚情報メディア講座横矢直和教授に心より感謝申し上げます. また，本研究の遂行にあたり，有益な御助言，御鞭撻を頂いた像情報処理学講座千原國宏教授に厚く御礼申し上げます．そして，本研究の全過程を通して温かい御指導をして頂いた視覚情報メディア講座山澤一誠助教授に深く感謝申し上げます．さらに，本研究の遂行に的確な御助言を頂いた視覚情報メディア講座神原誠之助手に深く御礼申し上げます．研究活動の全過程を通して多くの御助言，御指導賜りました視覚情報メディア講座佐藤智和助手に心より感謝致します．特に，佐藤智和助手には本研究のテーマの設定から本論文の執筆，その他の発表論文の添削，発表練習に至るまで細やかな御指導を頂きました．また，研究室での生活を支えて頂いた視覚情報メディア講座事務補佐員守屋智代女史に心より感謝申し上げます．最後に，物心両面において常に温かい御支援を頂いた視覚情報メディア講座の諸氏に深く感謝致します.

参考文献

[1] A. Harter, A. Hopper, P. Steggles, A. Ward and P. Webster: “The anatomy of a context-aware application,” Proc.ACM/IEEE Int. Conference on Mobile Computing and Networking, pp. 59–68, 1999.

[2] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A.Ward and A. Hopper: “Implementing a sentient computing system,” IEEE Computer Magazine, Vo. 34, No. 8, pp. 50–56, 2001.

[3] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. IEEE Int. Symp. on Wearable Computers, pp. 21–23, 2003.

[4] T. H*ö*llerer, S. Feiner and J. Pavlik: “Situated documentaries: Embedding multimedia presentations in the real world,” Proc. Int. Symp. on Wearable Computers, pp. 79–86, 1999.

[5] 小田島太郎, 神原誠之, 横矢直和: “拡張現実感技術を用いた屋外型ウェアラブル注釈提示システム”, 画像電子学会誌, Vo. 32, No. 6, pp. 832–840, 2003.

[6] 神原誠之, 横矢直和: “RTK-GPS と慣性航法装置を併用したハイブリッドセンサによる屋外型拡張現実感システム”, 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 933–938, 2005.

[7] 李欣洙, 間瀬憲一, 阿達透, 大沢達哉, 中野敬介, 仙石正和, 日高裕敏, 品川準輝, 小林岳彦: “GPS，歩数計及び方位計を用いた歩行者移動経路追跡法”, 電子情報通信学会論文誌 (B), Vol. J84-B, No. 12, pp. 2254–2263, 2001.

[8] R. Tenmoku, M. Kanbara and N. Yokoya.: “A positioning method combining specification of users absplute position and dead reckoning for wearable augmented reality system,” Proc. CREST/ISWC Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing, pp. 19–22, 2004.

[9] N. B. Priyantha, A. Chakraborty and H. Balakrishnan: “The cricket location-support system,” Proc. ACM/IEEE Int. Conference on Mobile Computing and Networking, pp. 32–43, 2000.

[10] 中里祐介, 神原誠之, 横矢直和: “ウェアラブル拡張現実感のための不可視マーカと赤外線カメラを用いた位置・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. 295–304, 2005.

[11] H. Kato and H. Billinghurst: “Marker tracking and hmd calibration for a video-based augmented reality conferencing system,” Proc. IEEE/ACM Int. Workshop on Augmented Reality, pp. 85–94, 1999.

[12] 羽原寿和, 町田貴史, 小川剛史, 竹村治雄: “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　MVE, pp. 65–70, 2002.

[13] U. Neumann and S. You: “Natural feature tracking for augmented-reality,” IEEE Transactions on Multimedia, Vo. 1, No. 1, pp. 53–64, 1999.

[14] A. Davison, Y. G. Cid and N. Kita: “Real-time 3D slam with wide-angle vision,” Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.

[15] 佐藤智和, 池田聖, 横矢直和: “複数動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.

[16] 岩佐英彦, 粟飯原述宏, 横矢直和, 竹村治雄: “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.

[17] R. Cipolla, D. Robertson and B. Tordoff: “Image-based localization,” Proc. Int. Conf. Vertual Systems and Multimedia, pp. 22–29, 2004.

[18] I. Skrypnyk 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.

[19] 大江統子, 佐藤智和, 横矢直和: “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.

[20] M. Kourogi and T. Kurata: “Personal positioning based on walking locomotion analysis with self-contained sensors and wearable camera,” Proc. IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp. 103–112, 2003.

[21] 内山晋二, 山本裕之, 田村秀行: “複合現実感のためのハイブリッド位置合わせ手法－６自由度センサとビジョン手法の併用－”, 日本バーチャルリアリティ学会論文誌, Vol.8, No. 1, pp. 119–125, 2003.

[22] 横地裕次, 池田聖, 佐藤智和, 横矢直和: “動画像とGPSによる位置情報を用いたカメラ外部パラメータの推定”, 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 650–657, 2005.

[23] 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.

[24] L. Naimark and E. Foxlin: “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker,” Proc. IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp. 27–36, 2002.

[25] 興梠正克, 蔵田武志, 坂上勝彦, 村岡洋一: “パノラマ画像群を位置合わせに用いたライブ映像上への注釈提示とその実時間システム”, 電子情報通信学会論文誌 (D-II), Vol. J84-D-II, No.10, pp. 2293–2301, 2001.

[26] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int. Journal of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.

[27] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.

[28] R. Klette, K. Schluns and A. koschan Eds: "Computer vision: Three-dimensional data from image," Springer, 1998.

[29] 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.

[30] 出口光一郎: "射影幾何学による PnP カメラ補正問題の統一的解法", 情処シンポジウム, Vol. 90, pp. 41–50, 1990.

[31] 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.

[32] K. Mikolajczyk and C. Schmid: "Scale & affine invariant interest point detectors," Int. Journal of Computer Vision, Vo. 60, No. 1, pp. 63–86, 2004.