

NAIST-IS-MT0551072

修士論文

スケール不変特徴量を用いた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

薄 充孝

2007年2月1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。

薄 充孝

審査委員：

横矢 直和 教授 (主指導教員)

木戸出 正継 教授 (副指導教員)

スケール不変特徴量を用いた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*

薄 充孝

内容梗概

携帯電話におけるヒューマンナビゲーションは、二次元的な地図上に GPS によって得られるユーザ位置と経路情報を表示することで道案内を実現している。しかし、二次元地図と現実空間の対応関係を直感的に把握することが難しく、システムの案内に従って移動することは必ずしも容易ではない。この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) に関する研究が近年盛んに行われている。案内情報を幾何学的に正しい位置に重畳表示させるためには、カメラの正確な絶対位置・姿勢を推定することが必要となる。従来、携帯機器による AR を用いたヒューマンナビゲーションを想定したシステムはいくつか提案されているが、カメラ位置・姿勢の推定精度が十分でなかったり、動画像による入力を前提としているために携帯機器での処理には計算コストが大きすぎるという問題があった。これらの問題を解決するために、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・姿勢推定手法が提案されている。この手法では、世界座標における三次元位置と見え方が既知の複数のランドマークを入力画像中から探索す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、画像のスケール変化に対応していないため、ランドマークデータベース構築時の撮影経路から大きく離れた地点では推定が失敗するという問題があった。そこで本研究では、この手法を基礎として、スケール不変

* 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-MT0551072, 2007年2月1日.

特徴量を用いた拡張を行うことで推定のロバスト性を向上させる．提案手法ではまず，画像の局所構造に依存する特徴点ごとの固有スケールを算出する．次に，この固有スケールを用いることで撮影距離に依存しないスケール不変特徴量を生成し，これを用いて入力画像中の特徴点とランドマークを対応付ける．さらに，カメラ位置・姿勢候補を特徴点の固有スケールを用いた投票により決定することで，誤対応の排除を行う．これにより，データベース撮影経路から離れた地点で撮影された画像を入力として用いた場合にもカメラ位置・姿勢を推定することが可能となる．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，データベースのカメラパスからの入力画像撮影地点の距離とカメラ位置・姿勢推定精度の関係を従来手法と比較することで，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，ランドマークデータベース，スケール不変特徴量

Camera Position and Posture Estimation for a Still Image Based on a Landmark Database Using Scale-Invariant Features*

Mitsutaka Susuki

Abstract

Human navigation systems on cellular phones can navigate users showing user's position and navigation information on a 2-D map. However, it is sometimes difficult to follow guiding information because 2-D navigation is not intuitive for many users. To realize intuitive guiding, a novel technique called Augmented Reality (AR) has been proposed and investigated. In AR applications, guiding information for a user is drawn on an image of a real environment. To overlay guiding information at geometrically correct positions on real images, absolute position and posture information of the camera are required. Although several methods for mobile navigation system using AR have already been proposed, most of them do not have sufficient accuracy, or the computational complexity is too large for cellular phones. A method for estimating camera position and posture from a still image based on a feature landmark database has also been proposed to solve these problems. However, it still can not estimate camera position and posture parameters accurately if the input image is taken at a position far from the camera path where the landmark database is created. This problem is due to the lack of handling scale changes of input images. In this study, scale-invariant features are used for creating a feature landmark database. With the proposed

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0551072, February 1, 2007.

method, the camera position and posture can be successfully estimated even if the input image is taken at a position far from the camera path where the landmark database is created. In experiments, the effectiveness of the proposed method has been shown by comparing it with the previous work.

Keywords:

camera position and posture estimation , still image , landmark database , scale-invariant features

目次

1. はじめに	1
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	3
2.1 センサを用いたカメラ位置・姿勢推定	3
2.2 画像を用いたカメラ位置・姿勢推定	4
2.2.1 人工的なマーカを用いる手法	4
2.2.2 画像データベースを用いる手法	6
2.2.3 環境の三次元モデルを用いる手法	8
2.2.4 自然特徴点ランドマークデータベースを用いる手法	9
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	11
2.4 本研究の位置づけと方針	12
2.5 提案手法の概要	16
3. ランドマークデータベースの構築	18
3.1 ランドマークデータベースの構成要素	18
3.2 全方位動画像による環境の三次元復元	20
3.3 特徴点スケールの算出	20
3.4 多重スケール特徴ベクトルの生成	22
4. スケール不変特徴量を用いた静止画像からのカメラ位置・姿勢推定	24
4.1 スケール不変特徴量を用いた類似度評価に基づくランドマークと特徴点の対応付け	24
4.2 特徴点スケールを用いた投票によるカメラ位置・姿勢候補の決定	26
4.3 カメラ位置・姿勢推定	28
4.3.1 カメラ位置・姿勢候補に投票されたランドマークと特徴点の組からのカメラ位置・姿勢推定	28
4.3.2 ランドマーク絞込みによるカメラ位置・姿勢の再推定	30
5. 実験	31

5.1	従来手法との比較によるスケール変化への対応の評価 (実験 1) . . .	31
5.2	実験 1 に関する考察	41
5.3	提案手法のロバスト性検証実験 (実験 2)	42
5.3.1	方向 1 に対する実験結果と考察	44
5.3.2	方向 2 に対する結果と考察	51
6.	まとめ	59
	謝辞	61
	参考文献	62

目 次

1	Möhring ら [15] の AR システム	5
2	興相ら [21] の注釈表示システム	7
3	Vacchetti ら [23] の手法で用いる三次元モデルとキーフレーム	8
4	各カメラ位置・姿勢推定手法の位置づけ	13
5	画像を用いたカメラ位置・姿勢推定の各手法の位置づけ	13
6	本研究で想定するサーバ・クライアント型システム	15
7	全体の処理の流れ	17
8	ランドマークデータベースの構成要素	19
9	Harris-Laplacian 適用例	21
10	多重スケール特徴ベクトル生成領域の例	22
11	特徴ベクトル生成手順	23
12	ランドマークと画像上特徴点の比較例	25
13	1つのランドマークからの投票例	26
14	再投影誤差の和の最小化によるカメラ位置・姿勢推定	29
15	全方位型マルチカメラシステム Ladybug と全方位画像	32
16	推定されたカメラとランドマークの三次元位置	32
17	実験 1 における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・方向	33
18	入力画像	35
19	提案手法による位置投票結果	36
20	提案手法による姿勢投票結果	37
21	従来手法による位置投票結果	38
22	データベース構築時のカメラパスからの距離と位置推定誤差	40
23	データベース構築時のカメラパスからの距離と姿勢推定誤差	40
24	実験 2 における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・方向	43
25	実験 2 の方向 1 におけるカメラ位置・姿勢推定の成功位置	45
26	方向 1 における推定に成功した入力画像例	47

27	方向 1 における推定に成功した地点での位置投票例	48
28	方向 1 における推定に成功した地点での姿勢投票例	49
29	方向 1 における推定に失敗した入力画像例	50
30	実験 2 の方向 2 におけるカメラ位置・姿勢推定の成功位置	52
31	方向 2 における推定に成功した入力画像例	54
32	方向 2 における推定に成功した地点での位置投票例	55
33	方向 2 における推定に成功した地点での姿勢投票例	56
34	方向 2 における推定に失敗した入力画像例	57
35	類似構造を持つ建物 (入力画像)	58
36	類似構造を持つ建物 (データベース構築時撮影画像)	58

表 目 次

1	カメラ位置・姿勢推定の各処理における閾値	33
2	従来手法および提案手法によるカメラ位置・姿勢推定の成功率, 推定精度および再投影誤差	39
3	カメラ位置・姿勢推定にかかる処理時間 [単位:秒]	39
4	方向 1 におけるカメラ位置・姿勢推定の成功率, 推定精度および再投影誤差	46
5	方向 2 におけるカメラ位置・姿勢推定の成功率, 推定精度および再投影誤差	53

1. はじめに

近年，携帯電話や PDA などの小型携帯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．このようなヒューマンナビゲーションシステムでは，機器に内蔵されている GPS により取得した位置情報を用い，二次元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，複雑な交差点など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．このため，最近では一部の交差点などで三次元的な CG を表示することでより直感的な道案内をするサービスが提供されはじめている．しかし，CG による三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，三次元地図情報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いる．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，利用者に直感的なナビゲーション情報を提供する拡張現実感 (Augmented Reality; AR) 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，カメラの正確な絶対位置・姿勢情報が必要となる．

カメラの位置・姿勢の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組み合わせて用いる手法 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]，画像を用いる手法 [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]，それらのハイブリッド [30, 31, 32, 33, 34, 35, 36] 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，システムが複雑になるという問題があり，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，一般的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，従来提案されている手法では利用可能な範囲，計算コスト，データベース構築コスト

に問題がある。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，システムの機器構成が複雑になる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スケール情報・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。提案手法は，本研究と同じく携帯端末でのユビキタスARの実現を目的とした静止画像1枚からカメラ位置・姿勢推定が可能な中川らの手法[29]を基礎とし，スケール情報を考慮することで，より広範囲でのカメラ位置・姿勢推定を可能とする。提案手法では，まず，入力画像上の特徴点と類似度が高いランドマークをスケール不変特徴量を用いて1対多で対応付ける。次に，対応付いた各ランドマークが撮影された可能性の高い領域をスケール情報により限定し投票を行うことで，3自由度のカメラ位置・姿勢候補を決定し，誤対応を排除する。また，カメラ位置・姿勢候補に投票されたランドマークと入力画像上の特徴点の組を用いて，6自由度の暫定的なカメラ位置・姿勢を求める。最後に，暫定的なカメラ位置・姿勢を基に推定に利用する入力画像上の特徴点とランドマークを限定した上でカメラの位置・姿勢を推定する。

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，スケール情報を考慮したランドマークデータベースの構築手法について述べる。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でカメラ位置・姿勢を推定する手法について述べる。5章では，屋外において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付けについて述べる。まず、従来のカメラ位置・姿勢推定手法に関して、ARでの利用を前提としたオンラインでの処理が可能なものを、(1)GPSやジャイロなどのセンサを用いる手法 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]、(2)画像を用いる手法 [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]、(3)それらのハイブリッド手法 [30, 31, 32, 33, 34, 35, 36] 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、本研究の位置付けと方針、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、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, 3] と、絶対位置を取得可能なGPSなどのインフラと相対位置・姿勢など取得可能なセンサを組み合わせる手法 [4, 5, 6, 7, 8, 9, 10, 11] 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用いられる。

環境中にインフラとしてセンサを埋め込む手法 [1, 2, 3] は、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。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、広範囲な環境を想定した場合、インフラ設置・維持のためのコストが膨大になる

という問題がある。

これに対して、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, 9, 10, 11] では、インフラでカメラ位置・姿勢が取得できないような状況に至っても、センサを用いて前回取得した絶対位置・姿勢からの相対位置・姿勢を計測し、カメラ位置・姿勢推定を行う。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 [7]。これらの手法ではインフラの設置密度を下げることでコストの削減を図るために自律計測の研究も多く行われている [10, 11]。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。

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 [12, 13, 14, 15, 16, 17, 18, 19]、環境を事前に撮影した画像とその撮影位置・姿勢情報からなる画像データベースを用いる手法 [20, 21, 22]、事前に作成した環境の三次元モデルを用いる手法 [23, 24, 25]、環境中の建造物の角などの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 [26, 27, 28, 29] 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[12] 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 [13, 14, 15, 16, 17]、人工的なマーカと自然特徴点追跡を併用する手法 [18, 19] などが挙げ



図 1 Möhring ら [15] の AR システム

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．Wagner ら [14] や Möhring ら [15] は，図 1 に示すように画像マーカを利用することで PDA や携帯電話に付いているカメラの位置・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．これらの手法では，ユーザは特殊な機器を用いず，市販の PDA や携帯電話で AR を利用することが可能である．しかし，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 [16] 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．また，齊藤ら [17] は部屋の内装となる模様位置情報を埋め込むことで周囲に調和したマーカを用いる手法を提案している．しかし，これらの手法は，画像内にマーカが一定以上の大きさに写らないとカメラ位置・姿勢を推定できないため，広域での利用に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，マーカの設置・維持に多大なコストがかかるという問題がある．

このようなマーカ設置のコストを低減するために，Neumann ら [18] や Davison

ら [19] は、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。これらの手法は、マーカを画像上に捉えなくても位置・姿勢の推定を継続することができる。しかし、人工的なマーカが観測できない場合には、自然特徴点を追跡することで前フレームからカメラの相対的な動きを推定し利用するため、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があり、誤差の累積を防ぐためにはやはり多くのマーカを設置する必要がある。

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 [20, 21] 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 [22] などがある。

岩佐らの手法 [20] では、事前に複数の地点で全方位カメラを用いて撮影し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、興相ら [21] は図 2 に示すように映像上への注釈情報の重ね合わせを実現するために、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

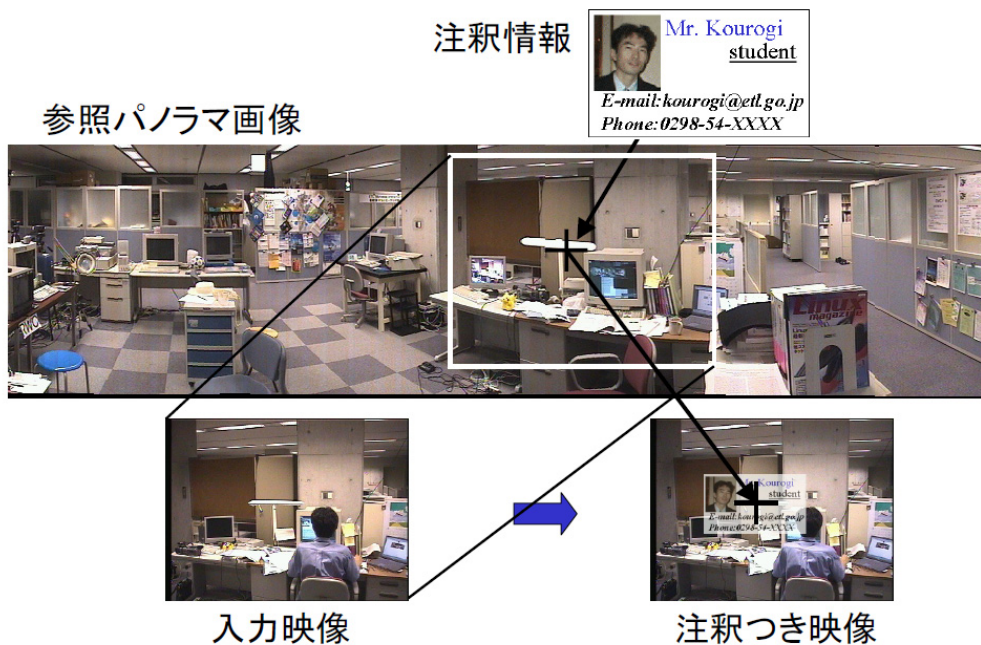


図 2 興相ら [21] の注釈表示システム

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。

Cipolla ら [22] 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

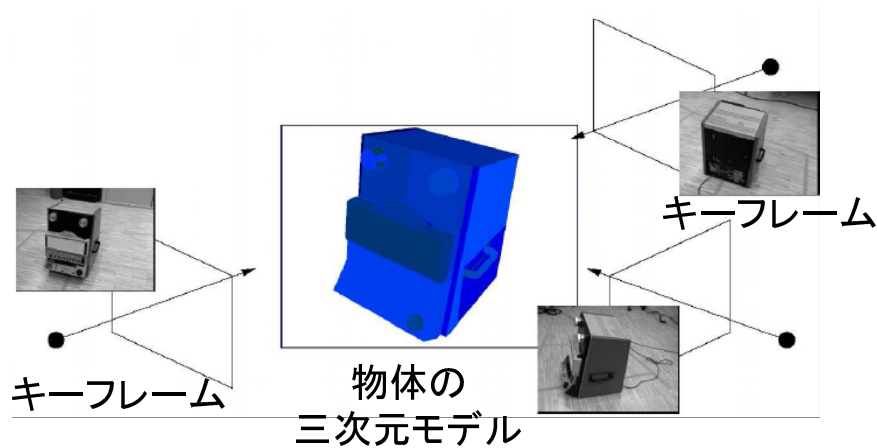


図 3 Vacchetti ら [23] の手法で用いる三次元モデルとキーフレーム

位置・姿勢情報が要求される AR には向かない。

2.2.3 環境の三次元モデルを用いる手法

環境の三次元モデルを用いる手法には、面を持った三次元モデルを用いた手法 [23] や自然特徴点とワイヤフレームモデルを用いた手法 [24, 25] などがある。Vacchetti ら [23] は、図 3 に示すような物体や環境の三次元モデルと複数のキーフレーム画像から得られる特徴点情報を用いて三次元モデルと入力画像の対応関係を求めることで、リアルタイムでカメラ位置・姿勢を推定する手法を提案している。

Vacchetti ら [24] は、自然特徴点とワイヤフレームモデルを用いたカメラ位置・姿勢推定手法も提案している。一般的に、ワイヤフレームモデルを用いたカメラ位置・姿勢推定は、入力画像中のエッジとワイヤフレームモデルの対応をとることで位置・姿勢の推定を行うが、環境中にエッジが多いと、ワイヤフレームモデルと入力画像中のエッジの誤対応を起こしやすくカメラ位置・姿勢推定が不安定となる。一方、自然特徴点のみを用いたモデルでは、特徴点と成り得る模様が少ない場所ではカメラ位置・姿勢推定が不安定になるという問題がある。Vacchetti

らの手法 [24] では，このような問題に対し，環境中の自然特徴点の三次元位置とワイヤフレームモデルを統合した三次元モデルを作成し，そのモデルと入力画像の対応関係を求めることで，環境中の模様の有無にかかわらずロバストなカメラ位置・姿勢の推定を実現している．また Rosten ら [25] は，手法 [24] と同様環境中の自然特徴点の三次元位置とワイヤフレームモデルを統合した三次元モデルを用い，さらに注目画素とその周辺の 16 点の輝度差のみを用いて高速に特徴点検出を行うことで，高速な移動や大きな姿勢変化を伴う入力映像に対してもロバストなカメラ位置・姿勢を実現している．

このように，環境の三次元モデルを用いる手法は，カメラの位置・姿勢を 6 自由度で正確に求めることができ，AR への利用が可能である．しかし，環境の三次元モデルを生成する必要があるため，広域での利用を想定した場合はモデル作成のためのコストが膨大になってしまうという問題がある．

2.2.4 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，Skrypnyk ら [26]，大江ら [27] ら，Snavely ら [28]，中川ら [29] の手法が挙げられる．これらの手法は，まず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，入力画像上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 [26] 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量である SIFT 特徴量を算出し，この SIFT 特徴量を用いて画像間で特徴点マッチングを行い特徴点の三次元復元を行う．このようにして算出された特徴点の三次元位置と SIFT 特徴量をデータベースに登録しておき，入力画像から抽出した SIFT 特徴量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，特別な機材を用いず，一般的なカメラによって撮影される画像のみから自動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には，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，広域環境でのナビゲーションに適用することは難しい．

これに対して大江らは，全方位カメラを用いて動きながら環境を撮影することで広域環境に対応したランドマークデータベースを構築し，これを利用することでカメラ位置・姿勢推定を行う手法を提案している [27]．大江らの手法では，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ことで，比較的高速なカメラ位置・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，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，実際には何らかの手法を用いて初期位置・姿勢を推定する必要がある．

Snively ら [28] の手法では，環境を撮影した複数枚の静止画像から特徴点を対応付け，structure-from-motion により自然特徴点の三次元復元を行い，画像と自然特徴点の三次元位置・SIFT 特徴量をデータベースに保存する．この手法では，静止画像 1 枚からのカメラの位置・姿勢推定を前提としているが，大江らの手法と同様に，データベースの探索範囲を限定するために，比較的高精度な初期位置を与えることが必要となる．

中川ら [29] はカメラの初期位置・姿勢が高精度に得られない場合にも，カメラ位置・姿勢を推定する手法を提案している．この手法ではまず，大江らの手法と同様に全方位動画像から structure-from-motion によって特徴点の三次元復元を行い，推定した自然特徴点の三次元位置と SIFT-descriptor によって算出される特徴点の見え方の特徴量をデータベースに登録しておく．次に，入力画像上から抽出した SIFT-descriptor による輝度勾配特徴ベクトルを用いて特徴点のマッチングを行う．さらに，対応付けられたランドマークが撮影された可能性の高い領域に投票を行い，カメラ位置候補を決定することでランドマークを絞り込む．このようにして絞り込まれたランドマークと入力画像上の特徴点の組を用いてカメラ

位置・姿勢推定を行う。この手法では、画像1枚のみから初期位置設定なしでカメラ位置・姿勢推定を行うことが可能であるという特徴がある。しかし、手法が画像のスケール変化に対応していないため、システムを利用可能な範囲がデータベース構築時のカメラパス付近に限定されてしまう。そのため、広域での利用においては、利用環境の大半の部分で、あらかじめ全方位動画像を取得しておく必要があり、広域環境で利用するにはデータベース構築に多大なコストがかかるという問題がある。

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 [30, 31, 32, 33, 34, 35, 36] は、お互いの欠点を補い合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。

人工的な画像マーカと加速度センサやジャイロセンサを組み合わせる手法 [30, 31, 32] では、画像マーカから絶対位置を取得できるため、マーカを定期的に画角内に捉えることでセンサの蓄積誤差を打ち消すことができる。しかし、定期的に画像中にマーカを捉えるために、広域環境で使用する場合はマーカの設置に人的コストがかかるという問題が残る。

画像データベースと加速度計、およびジャイロセンサを組み合わせる手法 [33] は、画像マーカの代わりに画像データベースを用いて絶対的な位置・方位を推定し、これに加えてセンサから取得される相対的な情報を組み合わせることで、利用者の位置・姿勢を推定する。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ためには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。

環境の三次元モデルと加速度センサを組み合わせる手法 [34] は、モデルと入力

画像の対応付けの際にセンサの情報も用いることで、高速な対応付けを可能にしている。これにより、高速で移動した場合や姿勢が大きく変化した場合にもロバストにカメラ位置・姿勢を求めることができる。また、オクルージョンなどで対応付けに失敗しても、過去のカメラ位置・姿勢からの相対位置をセンサ情報から求めることでカメラの位置・姿勢を推定する。しかし、広域環境に適用した場合には、やはり三次元モデルの作成に膨大なコストがかかってしまうという問題がある。

自然特徴点ランドマークとジャイロセンサを組み合わせた手法 [35, 36] は、ジャイロセンサの誤差をランドマークと画像上の特徴点との対応を取ることで補正する手法である。景観を損なうことなく、高精度にカメラ位置・姿勢を推定することが可能であり、データベース構築のコストも低い。しかし、文献 [35] の手法では、カメラ位置は固定されており、カメラの移動を考慮していない。また、文献 [36] の手法では、ランドマークの探索範囲を GPS とジャイロセンサで限定する必要があり、GPS の計測精度が悪い環境では利用が難しい。

このように、現時点で提案されているハイブリッドなカメラ位置・姿勢推定手法の大半は、センサによって得られる情報を画像を用いて補正するというアプローチを採っており、画像は補助的に利用されている。このため、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が残されている。

2.4 本研究の位置づけと方針

図 4 に、広域環境におけるユビキタス AR での利用を前提とした、従来提案されているカメラ位置・姿勢推定手法の特徴を示す。前節までで概観したように、AR での利用を前提としたオンライン処理での実行が可能なカメラ位置・姿勢推定手法は既に多数提案されており、それぞれ異なった特徴を持つが、従来のカメラ位置・姿勢推定手法を用いて、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上でのユビキタス AR を広域環境で実現することは容易ではない。すなわち、センサやインフラを主として利用するカメラ位置・姿勢推定手法には、システムが複雑になるという問題、使うセンサの種類によって利用可能な範囲が限定されるという問題があり、現在既に広く普及している携帯電話や PDA など

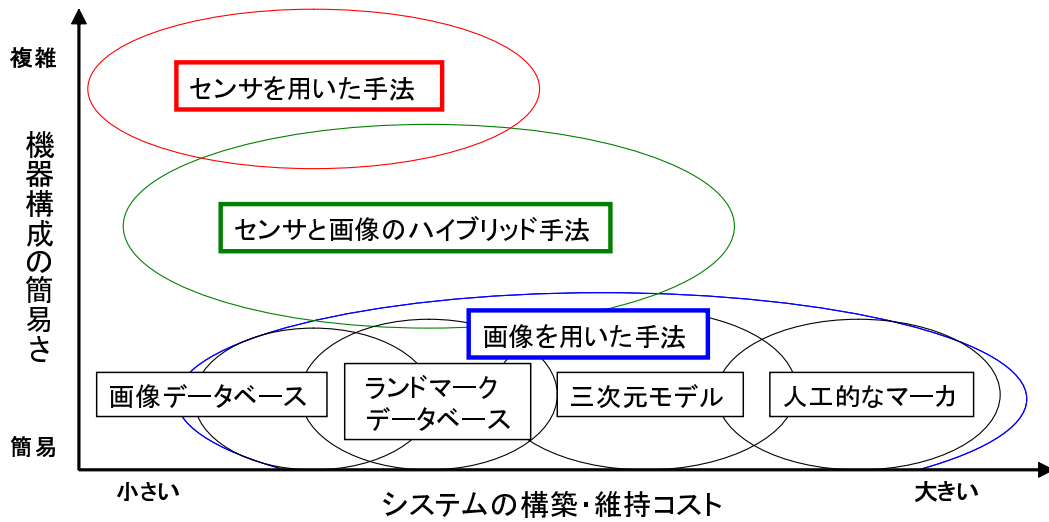


図 4 各カメラ位置・姿勢推定手法の位置づけ

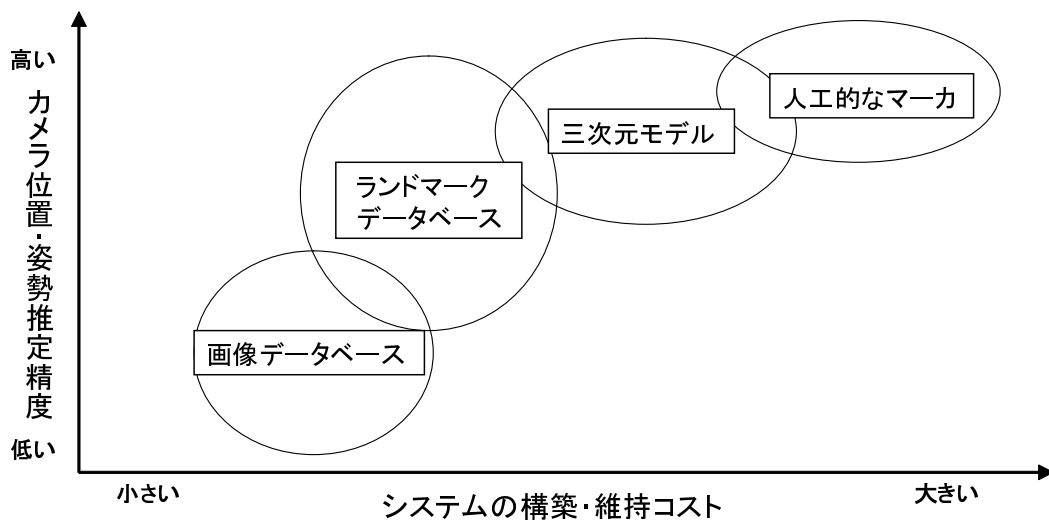


図 5 画像を用いたカメラ位置・姿勢推定の各手法の位置づけ

で利用することが難しい。一方、画像を主として用いる手法は、携帯電話やPDAなどに内蔵されているカメラを利用することで、カメラ位置・姿勢推定を行えるため、現在普及している多くの携帯端末において、利用が可能である。しかし、実際には、図5に示すようにデータベース構築コストと位置・姿勢推定の精度にはトレードオフの関係があり、ヒューマンナビゲーションを行うためのARを実現可能な位置・姿勢推定精度を保ちつつ、なるべくデータベース構築コストを抑える必要がある。従来手法において、中川らの手法[29]は、ヒューマンナビゲーションを行うために十分な精度でカメラの位置・姿勢推定が可能で、かつ市販のカメラ付き携帯電話やPDAによって撮影される静止画像1枚から初期位置設定なしで推定を行うことができる。このため、携帯端末上でARヒューマンナビゲーションを行うために適した手法であると言える。しかし、中川らの手法では、データベース構築時のカメラパスから離れた地点で正しく推定を行えないため、広範囲でのカメラ位置・姿勢推定を可能とするためには、やはりデータベース構築に大きなコストがかかるという問題が残されている。そこで、本研究では中川らの手法を基礎として、データベース構築時のカメラパスから大きく離れた地点でも推定が可能となるよう改良を行うことで、カメラ位置・姿勢の推定可能範囲の拡大とロバスト性の向上を目指す。具体的には、中川らの手法に対して以下の2つの改良を加える。

- ・ 画像スケールに依存しない特徴量を用いた特徴点の対応付け
- ・ スケール情報を利用したカメラ位置・姿勢候補の絞り込み

画像スケールに依存しない特徴量(スケール不変特徴量)を特徴点の対応付けに用いることで、撮影距離が大きく異なる画像間においても正しい対応付けが可能となる。また、中川らの投票方法では誤対応が多いと投票がデータベース構築時のカメラパス付近に集中してしまい、入力画像の撮影位置がデータベース構築時のカメラパスから大きく離れた場合には、正しくカメラ位置候補を推定できないという問題がある。そこで本研究では、スケール情報を考慮した投票を用いてランドマークの絞り込みを行うことで、このような問題を解決し、データベース構築時のカメラパスから大きく離れた地点でもカメラ位置・姿勢推定を可能とする。

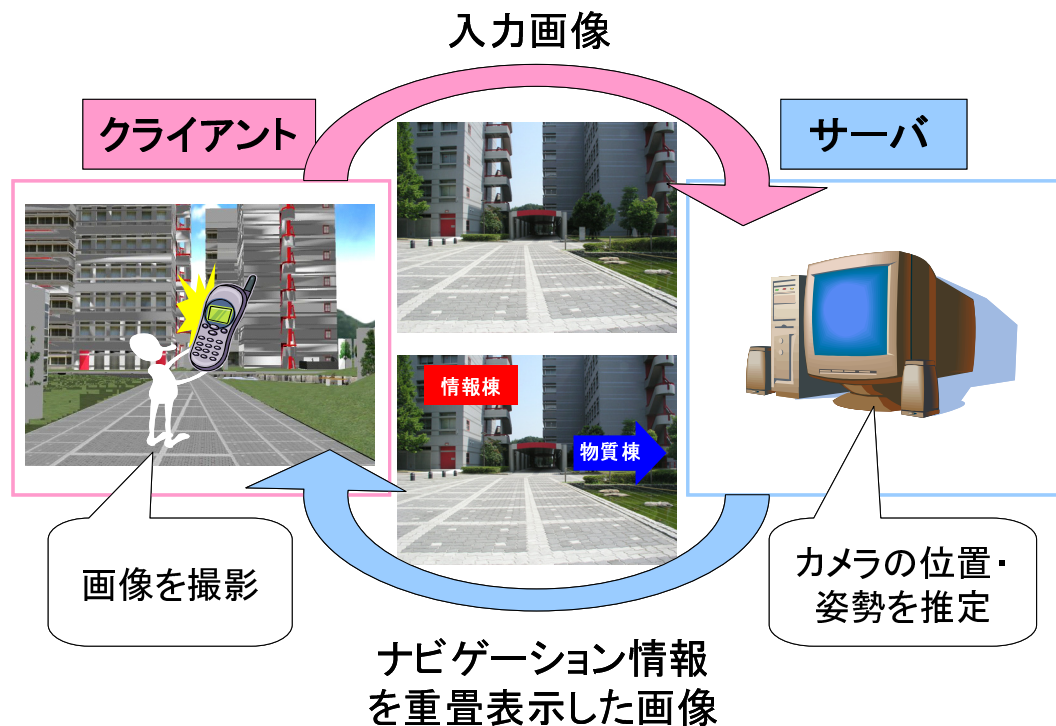


図 6 本研究で想定するサーバ・クライアント型システム

なお、本研究では中川らの手法と同様に、図 6 に示すようなサーバ・クライアント型システムを想定し、カメラ位置・姿勢推定処理はサーバ側で行うものとする。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と GPS 情報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・GPS 情報とサーバ上にあるデータベース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、入力画像上に CG でナビゲーション情報を重畳表示し、クライアントに結果を送信する。

2.5 提案手法の概要

提案手法は、システムの管理者が事前に行うランドマークデータベースの構築処理とユーザがシステムを利用する際に行うカメラの位置・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7に示す。本手法では、文献[29]の手法を基礎とし、スケール不変特徴量生成に必要な特徴点スケールの算出処理(a-3,b-1)を追加する。また、その特徴点スケールを用いた投票による撮影位置・姿勢候補の算出(b-4)を新たに行う。ただし、基本的な処理の流れは手法[29]と同様である。

ランドマークデータベース構築(オフライン処理)では、まず事前に環境内を移動を伴う全方位カメラによって撮影し(a-1)、撮影した全方位動画像から structure-from-motion により環境の粗な三次元復元を行う(a-2)。次に、三次元復元された自然特徴点に対して、全方位画像を用いて特徴点スケールを算出し(a-3)、その特徴点スケールを用いて多重スケール特徴ベクトルの生成を行う(a-4)。最後に、以上のようにして求めた情報をデータベースに登録する(a-5)。

カメラ位置・姿勢推定処理(オンライン処理)では、まずデータベース構築時と同様に入力画像から特徴点スケール算出(b-1)と多重スケール特徴ベクトルの生成(b-2)を行う。次に、多重スケール特徴ベクトルを用いて入力画像の特徴点と類似度の高い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特徴点との対応付けを行う(b-3)。さらに、対応付けられたランドマークの投票によりカメラ位置・姿勢候補を決定する(b-4)。最後に、カメラ位置・姿勢候補に投票されたランドマークと入力画像上特徴点の組を用いてカメラ位置・姿勢を推定する(b-5)。

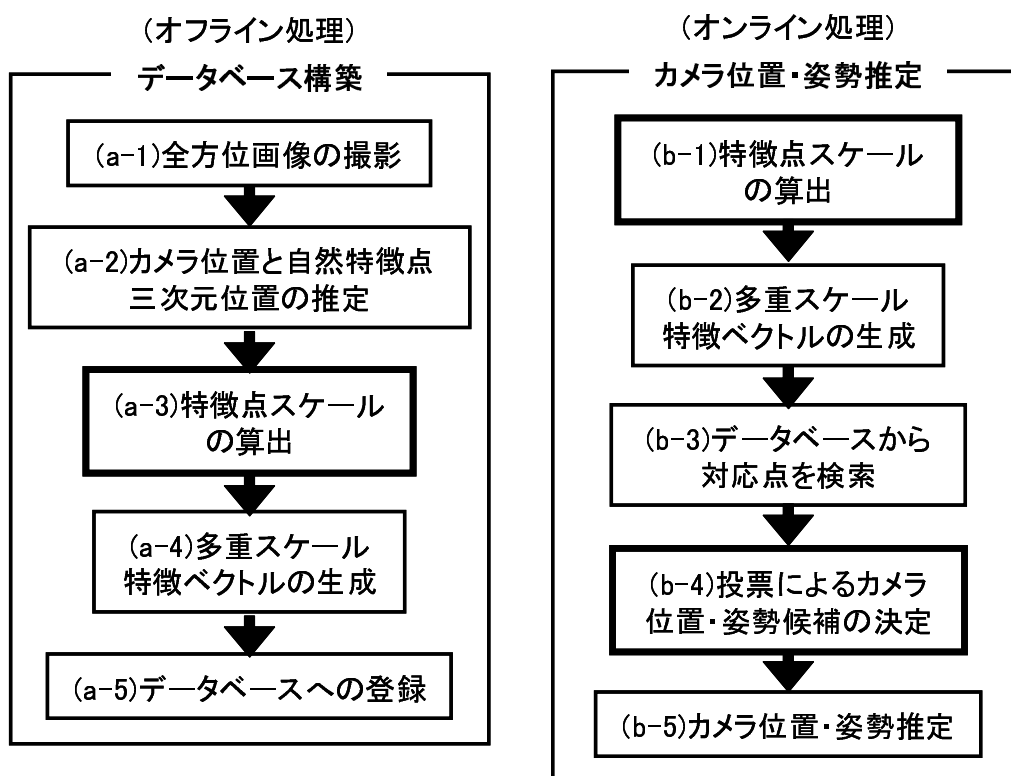


図 7 全体の処理の流れ

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・姿勢推定に用いられるランドマークデータベースの構築方法(オフライン処理)について述べる．ランドマークデータベースの構築では，図7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，システムの利用環境内を移動しながら撮影する(a-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全方位カメラ位置・姿勢情報を推定する(a-2)．また，処理(a-2)で追跡された自然特徴点の特徴点スケールを算出(a-3)し，算出された特徴点スケールを基に多重スケール特徴ベクトルを生成する(a-4)．最後に，自然特徴点の三次元位置，ランドマーク撮影時の全方位カメラの位置，特徴点スケール，多重スケール特徴ベクトルをランドマークとしてデータベースに登録する(a-5)．

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

本研究では，文献[29]で用いられているランドマークデータベースを基礎に，特徴点スケール情報を付加したランドマークデータベースを作成する．また，ランドマークの識別性を向上させるために特徴ベクトルを多重スケールで保持する．本研究で用いるランドマークデータベースには，図8に示すようにランドマークごとに(A)自然特徴点の三次元位置，(B)撮影地点情報，が保持されており，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位置，(b)特徴点スケール，(c)多重スケール特徴ベクトルから成る．以下にそれぞれの要素について述べる．

(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

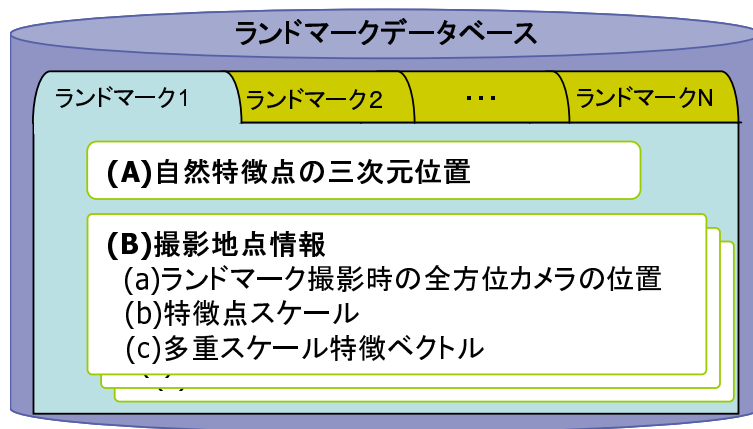


図 8 ランドマークデータベースの構成要素

用いる．この三次元位置は，3.2 節で述べる環境の三次元復元によって得られる．

(B) 撮影地点情報

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 3 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

4.2 節で述べる投票によるカメラ位置・姿勢候補の決定に用いる．自然特徴点の三次元位置と同様，3.2 節で述べる環境の三次元復元 (a-2) によって得られる．

(b) 特徴点スケール

特徴点周辺の局所構造に依存する特徴点ごとの固有のスケールを表し，多重スケール特徴ベクトルの生成 (a-4,b-2) と投票によるカメラ位置・姿勢候補の算出 (b-4) に用いる．この特徴点スケールは，3.3 節で述べる Harris-Laplacian による特徴点スケールの算出処理 (a-3,b-1) によって得られる．

(c) 多重スケール特徴ベクトル

特徴点スケールに対応した特徴点周辺画像領域から得られる SIFT-descriptor[37] による多次元ベクトルを多重スケールで保持し，ランドマークと入力画像上の特徴点の対応付け (b-3) に用いる．3.4 節で述べる多重スケール特徴ベクトルの生成処理 (a-4,b-2) によって得られる．

3.2 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 [38] 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 [39] 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撮影時のカメラパスを取得する．

具体的には，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．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．本研究では，これにより得られた三次元復元結果から，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．

3.3 特徴点スケールの算出

上述の手法で三次元位置が推定されたすべての自然特徴点に対して，各画像上における特徴点の固有スケールを算出する．ここでは，カメラ姿勢の違いによる同一特徴点の見え方の変化の影響を排除するために，まずレンズ歪みを補正し

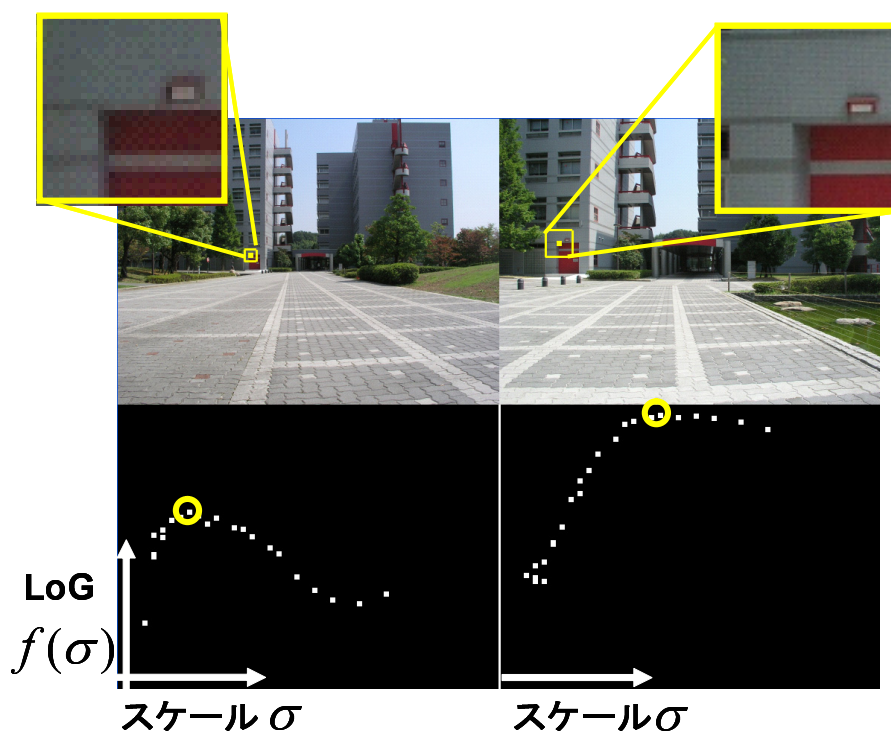


図 9 Harris-Laplacian 適用例

た上で、カメラの投影中心を中心とする無限遠の球面上に、撮影した全方位画像を投影する。次に、この球面上に投影された画像上の各特徴点に対して Harris-Laplacian[40] を適用することによって、画像の局所構造に依存する特徴点スケール ω を算出する。

Harris-Laplacian による固有スケールの算出 [40] は、Harris オペレータによって検出された画像上の特徴点に対し、スケール σ を変化させながら式 (1) で表される Laplacian-of-Gaussian (LoG) オペレータを適用し、その極大位置を求めることを行う。

$$LoG = f(\sigma) = -\frac{r^2 - 2\sigma^2}{2\pi\sigma^6} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1)$$

ここで、スケール σ は Gaussian フィルタの半径を表しており、 r はオペレータ中心から注目画素までの距離を示す。これにより得られる LoG 値は、画像の局所

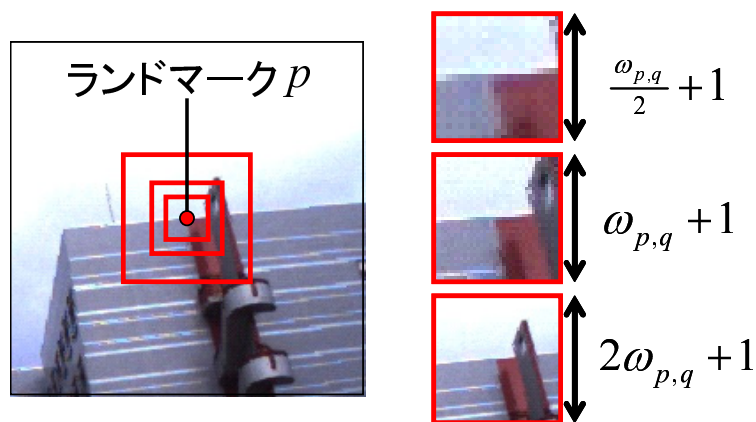


図 10 多重スケール特徴ベクトル生成領域の例

構造に依存し，図 9 に示すように同様の局所構造に対しては同様の値の変化を示す．従って，LoG の極大値に対応するスケールを特徴点スケール ω とすれば，特徴点までの撮影距離の違いにより画像スケールが異なっても，図 9 の上部の 2 つの特徴点領域のように，空間的に同範囲の領域をスケール ω により一意に決定することができる．

3.4 多重スケール特徴ベクトルの生成

本節では，前節で求めた特徴点スケールを用いて，多重スケール特徴ベクトルを生成する手法について述べる．多重スケール特徴ベクトルは以下の手順で撮影地点情報ごとに生成する．

1. 多重スケール特徴ベクトル生成領域の決定

ランドマーク p の q 番目の撮影地点情報における特徴点スケール $\omega_{p,q}$ を用いて，図 10 に示すように，特徴点を中心とした $(\omega_{p,q}/2 + 1) \times (\omega_{p,q}/2 + 1)$ ， $(\omega_{p,q} + 1) \times (\omega_{p,q} + 1)$ ， $(2\omega_{p,q} + 1) \times (2\omega_{p,q} + 1)$ のウィンドウを輝度勾配ベクトル生成領域とする．

2. 輝度勾配ベクトルからの特徴ベクトルの生成

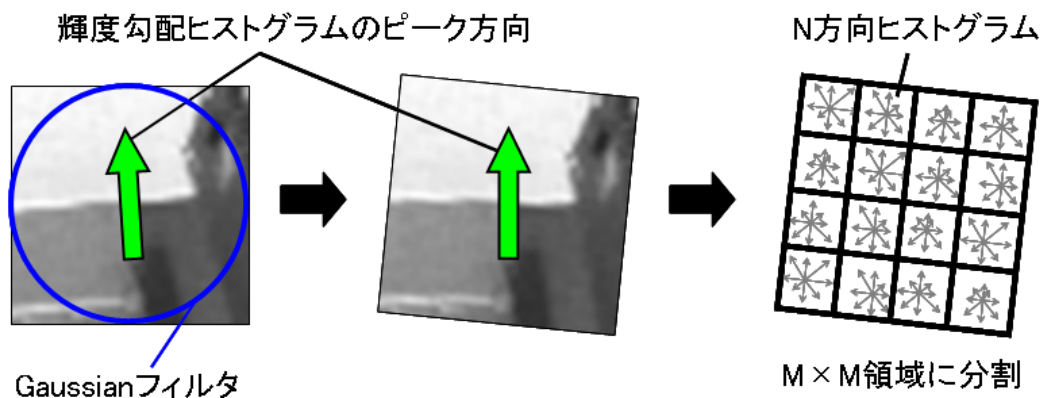


図 11 特徴ベクトル生成手順

特徴ベクトルの生成手順を図 11 に示す．本研究では，SIFT-descriptor[37] を用いて輝度勾配ヒストグラムから特徴ベクトルを生成する．SIFT-descriptor では回転不変なマッチングを可能にするために，まずステップ 1 で求めた特徴ベクトル生成領域に対して Gaussian フィルタをかけ，輝度勾配ヒストグラムを作成する．次に，その輝度勾配ヒストグラムからピークとなる輝度勾配を求め，特徴ベクトル生成領域画像を輝度勾配ピーク方向に回転させる．さらに，画像を $M \times M$ 領域に分割し，各領域ごとに輝度勾配を N 方向に分類したヒストグラムを作成する．このヒストグラムの各値を特徴量の要素とすることで， $D = M \times M \times N$ 次元の特徴ベクトル $f = (v_1, \dots, v_D)$ を生成する [37]．

4. スケール不変特徴量を用いた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、静止画像からカメラ位置・姿勢推定を行う方法について述べる。なお、カメラ位置・姿勢推定処理の基本的な処理の流れは文献 [29] と同じであるが、入力画像特徴点からスケール不変特徴量を生成するために、特徴点スケール算出処理 (b-1) を追加し、また算出された特徴点スケールを考慮した投票処理 (b-4) を新たに行う。

すなわち、本手法では、まず入力画像から検出された画像上の特徴点の特徴点スケールを算出し (b-1)、算出された特徴点スケールを基に多重スケール特徴ベクトルを生成する (b-2)。次に、データベースに登録されているランドマークを GPS 情報により限定した上で、ランドマークと入力画像上特徴点を多重スケール特徴ベクトルを比較することで対応付ける (b-3)。また、対応付けられたランドマークと同じ見え方で撮影可能な領域に投票を行い、カメラ位置・姿勢候補を決定する (b-4)。最後に、処理 (b-4) で決定したカメラ位置・姿勢候補に投票されたランドマークとそれに対応付いている画像上特徴点の組を複数用いてカメラ位置・姿勢推定を行う。以下では、これら各処理について詳しく述べるが、処理 (b-1) におけるスケール算出処理は 3.3 節で述べた手法と同様であるため、説明を省略し、以下では処理 (b-2) ~ (b-5) について順に述べる。

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。

4.1 スケール不変特徴量を用いた類似度評価に基づくランドマークと特徴点の対応付け

まず、対応付けに用いるランドマークを GPS 情報を用いてランドマークデータベースの中から選出する。その上で、入力画像上で検出された特徴点とデータベースに登録されているランドマークを、それぞれの多重スケール特徴ベクトルを比較することで対応付ける。ただし、1つのランドマークには複数の撮影

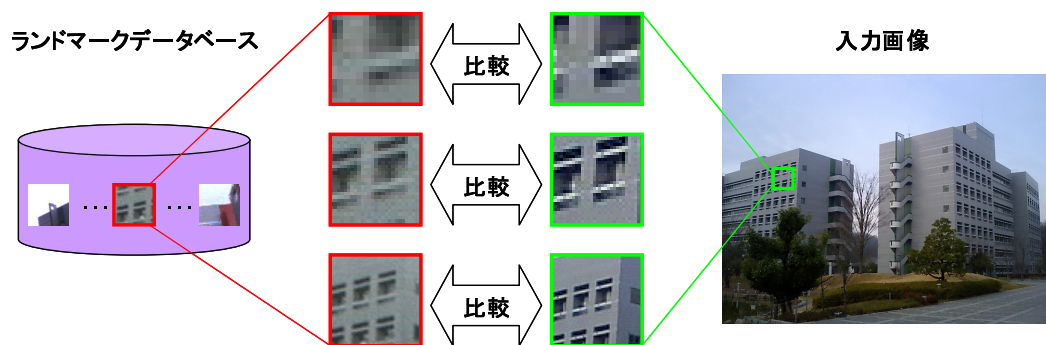


図 12 ランドマークと画像上特徴点の比較例

地点情報が格納されているため，ここでは入力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．各入力特徴点から得られた特徴ベクトル $\mathbf{f}_{IN} = (v_{IN1}, \dots, v_{IND})$ と，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 $\mathbf{f}_{LM} = (v_{LM1}, \dots, v_{LMD})$ の比較は，式 (2) を用いて，総当りで特徴空間上における二乗距離 (非類似度) を算出することで行う．

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

本研究では，図 12 に示すように，各スケールに対応する特徴ベクトルを用いて S を算出し，それぞれの値が全て一定閾値以下の場合にそのランドマークと入力画像特徴点を対応付ける．このような比較を行うことで，大局的な比較と局所的な比較を同時に行い，対応点決定の精度を高める．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るという仮定の下で，特徴点に対してランドマークが光軸回り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．これをすべての入力画像特徴点に対して行うことで，入力画像特徴点と類似度の高いランドマークを絞り込む．

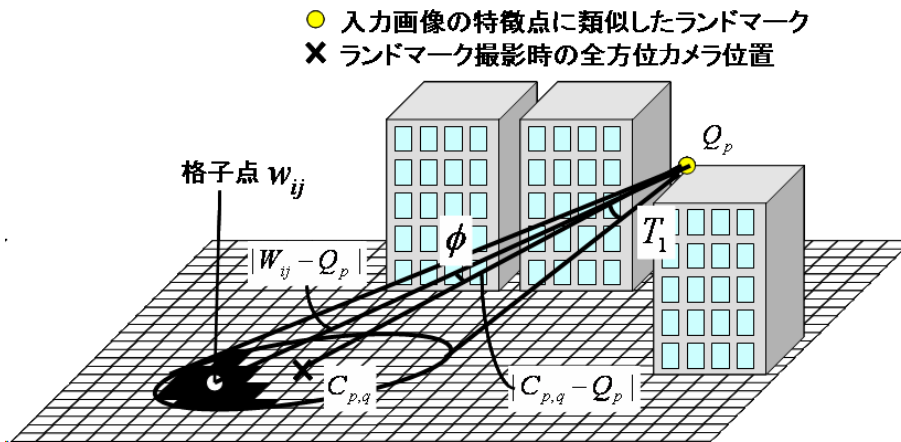


図 13 1つのランドマークからの投票例

4.2 特徴点スケールを用いた投票によるカメラ位置・姿勢候補の決定

4.1 節で述べた処理で対応付けられた入力特徴点とランドマークの組み合わせには、入力特徴点と真に対応するランドマーク以外の全く異なるランドマークと対応付けているものが多数存在する。提案手法では、このような誤対応を排除するために、特徴点スケールを用いた投票によるカメラ位置・姿勢候補の決定を行う。この処理では、選択されたランドマークと同様の見え方で撮影可能な領域に投票を行い、その投票値により入力画像が撮影された可能性が高いカメラ位置・姿勢を決定する。なお、文献 [29] でも投票によるカメラ位置候補の決定処理を行っているが、文献 [29] の投票方法では誤投票が多く存在した場合、投票の極大値がデータベース構築時のカメラ位置周辺に集中してしまうため、正しくカメラ位置候補を決定できない。提案手法では、特徴点スケールを利用して投票位置をさらに限定することでこの問題を回避する。

具体的にはまず、図 13 に示すように、GPS から得られる計測地点を中心とする一定範囲領域を地面に対して水平方向に格子状に分割し、それぞれの格子に、水平方向の姿勢の回転に対応する l 個 (360 度を $360/l$ 度ずつ分割) の投票箱を設

置する．次に，分割された空間に対して 4.1 節で述べた手法により選択されたランドマーク p の撮影地点情報に基づいた位置・姿勢の投票を行う．以下では，投票手法の詳細を述べる．

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．

$$\mathbf{w}_{ij} = \begin{pmatrix} w_i \\ w_j \end{pmatrix} = \begin{pmatrix} g_x + L \times i \\ g_y + L \times j \end{pmatrix} \quad (3)$$

ただし， L は格子間隔を表す．次に，すべての入力特徴点について 4.1 節で述べた手法により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} の m 番目の投票箱に投票する．

[条件 A] ランドマーク p の三次元位置 $\mathbf{Q}_p(Q_x, Q_y, Q_z)$ とランドマーク p を撮影した q 番目の全方位カメラの位置 $\mathbf{C}_{p,q}$ を結ぶ直線，および格子点 \mathbf{w}_{ij} に $\mathbf{C}_{p,q}$ の高さ c_z を与えた点 $\mathbf{W}_{ij}(w_i, w_j, c_z)$ と三次元位置 \mathbf{Q}_p を結ぶ直線の成す角 ϕ が閾値 T_1 以下．

[条件 B] ランドマーク p の三次元位置 \mathbf{Q}_p と \mathbf{W}_{ij} の距離，および \mathbf{Q}_p と全方位カメラの位置 $\mathbf{C}_{p,q}$ の距離の比が一定範囲内．すなわち，
 $(1 - \alpha) \frac{\omega_{LM}}{\omega_{IN}} \leq \frac{|\mathbf{W}_{ij} - \mathbf{Q}_p|}{|\mathbf{C}_p - \mathbf{Q}_p|} \leq (1 + \alpha) \frac{\omega_{LM}}{\omega_{IN}}$.

ただし， ω_{LM}, ω_{IN} はランドマークおよび入力特徴点の特徴点スケールを， α は範囲決定のための定数を表す．なお， m は以下の式により決定する．

$$m = \left\lceil \frac{\theta l}{2\pi} \right\rceil \quad (4)$$

ただし， $[a]$ は a を超えない最大の整数を表し，世界座標系における x 軸と $(Q_x - w_i, Q_y - w_j)$ の成す角を θ [ラジアン] とする．以上の処理により得られた投票結果に対して，投票数が極大になる位置・姿勢を，入力画像を撮影した可能性の高いカメラ位置・姿勢候補とする．ただし，カメラ位置・姿勢候補は複数存在するため，ここでは投票数が多い順にカメラ位置・姿勢候補の順位を決定しておく．なお，ここで求まるカメラ位置・姿勢候補は，位置 2 自由度，姿勢 1 自由度の 3 自由度となるため，次節で述べる手法で 6 自由度のカメラ位置・姿勢を算出する．

4.3 カメラ位置・姿勢推定

前節で述べた手法により算出されたカメラ位置・姿勢候補および投票結果からカメラ位置・姿勢推定を行う (b-5) . ここでは, ランドマークの三次元位置と入力特徴点の二次元位置の組を複数用いて PnP 問題 [41] を解くことでカメラ位置・姿勢を推定する. 提案手法ではまず, 前節で述べた手法で決定した 3 自由度のカメラ位置・姿勢候補に対して, 投票に用いられた入力画像上の特徴点とランドマークの組から 6 自由度の暫定的なカメラ位置・姿勢を求める. 次に, 求めた暫定的なカメラ位置・姿勢に基づいて限定された探索範囲からランドマークを再探索することで, 最終的な出力となるカメラ位置・姿勢を求める. 以下では, 各ステップの処理について詳述する.

4.3.1 カメラ位置・姿勢候補に投票されたランドマークと特徴点の組からのカメラ位置・姿勢推定

カメラ位置・姿勢候補に投票された特徴点とランドマークの組を用いて 6 自由度の暫定的なカメラ位置・姿勢を求める手法について述べる. まず, 4.2 節で述べた手法により得られたカメラ位置・姿勢候補点に投票されたランドマークとそれに対応する入力特徴点の組み合わせを用いて, PnP 問題を解くことでカメラ位置・姿勢を推定する. ただし, 投票されたランドマーク中には多数の誤対応が含まれるため, それを取り除く必要がある. そこで, 投票数が多いカメラ位置・姿勢候補から順に, LMedS 推定 [42] を用いて誤対応を排除する. その後, 図 14 に示すように, 誤対応排除後の結果から以下のような式で定義される再投影誤差の和 R が最小となるカメラ位置・姿勢を求める.

$$R = \sum_{i=1}^k \{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2\} \quad (5)$$

ただし, k は推定に用いるランドマークと入力特徴点の組み合わせの数, (u, v) は入力画像上の特徴点の座標, (\hat{u}, \hat{v}) はランドマークの三次元座標を画像上に投影した座標である. この再投影誤差の和の最小化により, 尤もらしいカメラ位置・姿勢が得られた時点で推定を完了する. 本研究では, 推定結果の再投影誤差 R が

- カメラ位置・姿勢候補に投票されたランドマーク
- カメラ位置・姿勢候補に投票された入力画像上の特徴点
- ランドマークを画像上に投影した点

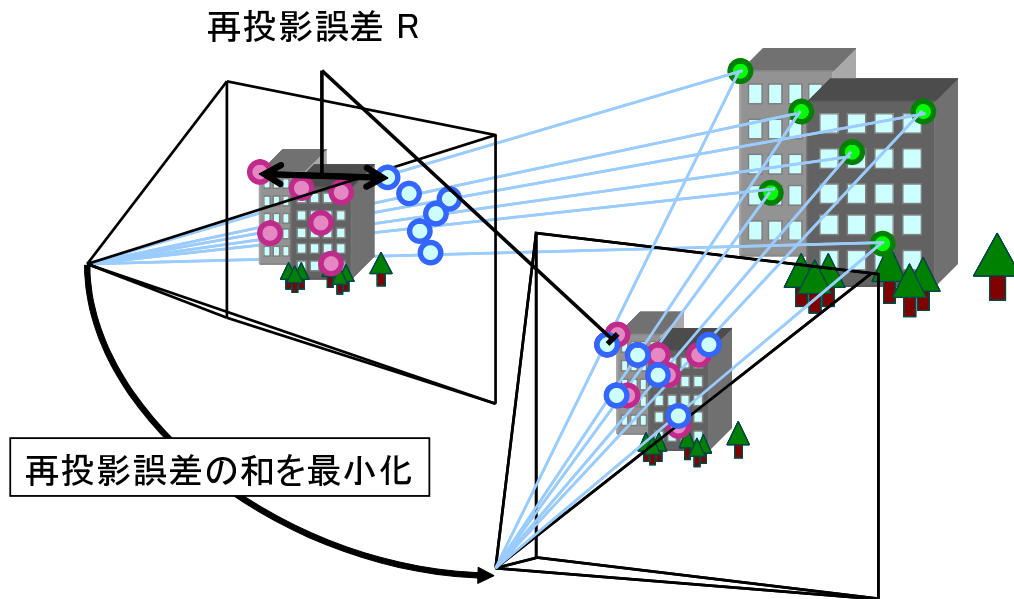


図 14 再投影誤差の和の最小化によるカメラ位置・姿勢推定

閾値以下の場合には，システムが尤もらしい推定結果であると判定し，次項で述べる推定結果の高精度化を行う．本手法では，線形解法によってカメラ位置・姿勢の初期値を求めるため，少なくとも入力画像上にランドマークが6点以上検出される必要がある．そのため，投票されたランドマークが6点未満の場合，もしくは再投影誤差が閾値を上回った場合には推定に失敗したとみなし，システムは利用者に画像の撮り直しを要求する．

4.3.2 ランドマーク絞込みによるカメラ位置・姿勢の再推定

前項で求めたカメラ位置・姿勢を用いてランドマークの探索範囲を限定し，再探索を行うことでカメラ位置・姿勢推定結果の高精度化を行う．本研究における再推定処理は以下のような手順で行う．

(手順1) 暫定カメラ位置・姿勢の画角内に存在するすべてのランドマークを選択する．次に，選択されたランドマークから，ランドマーク p の三次元位置 Q_p と暫定カメラ位置を結ぶ直線，および Q_p とランドマーク p を撮影したデータベース構築時のカメラ位置 C_p を結ぶ直線の成す角が閾値 T_2 以下となるものを選択し，推定に用いるランドマークを限定する．

(手順2) 手順1で限定されたランドマーク p と対応付ける入力画像上の特徴点を，ランドマーク p を暫定カメラ位置・姿勢を用いて画像上に投影した座標位置から d 画素以内にあるものに限定する．

(手順3) 手順1,2で限定されたランドマーク p と入力特徴点の特徴ベクトルの非類似度 S が最小となるものを探索する．さらに，そのときの S が閾値以下となるものを選択し，選択されたランドマークと入力特徴点の組み合わせを用いて再度カメラ位置・姿勢推定を行い，その結果を最終的な出力とする．

5. 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，カメラ位置・姿勢推定を行った．本実験では，まずスケール変化への対応を評価するために従来手法との比較を行い(実験1)，次に提案手法のロバスト性を確認するために多地点において推定精度を検証した(実験2)．実験1では，撮影方向が一定でデータベース構築時のカメラパスからの距離が異なる複数枚の画像を入力としてカメラ位置・姿勢推定を行い，本手法の基礎となっている手法[29]と推定結果を比較することでスケール変化への対応を評価する．実験2では，撮影位置と方向の異なる複数枚の画像を入力としてカメラ位置・姿勢推定を行い，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・方向の関係性を考慮しながら提案手法のロバスト性を検証する．なお両実験ともに，入力特徴点とトータルステーションで測定したランドマークの三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真値として推定精度の評価を行った．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，カメラ内部パラメータはあらかじめ Tsai の手法[43]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，今回は用いたデータベースの規模が比較的小さいため，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．

5.1 従来手法との比較によるスケール変化への対応の評価(実験1)

本実験では，文献[29]で用いられた全方位動画画像を用いて，スケール情報を含むランドマークデータベースを新たに構築した．この全方位動画画像は，図15左に示す全方位マルチカメラシステム(Point Grey Research社Ladybug)を用いて屋外環境(約75m)を移動しながら撮影することで取得された．図15右にデータベース構築に用いられた画像の一部を示す．また，オフライン処理で復元され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16に示す．図中の四角錐は100フレームごとのカメラ位置・姿勢を，四角錐を結ぶ曲線は推定されたカメラパスを表す．ここでは約75mの経路で撮影した全方位動画画像から3章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



図 15 全方位型マルチカメラシステム Ladybug と全方位画像

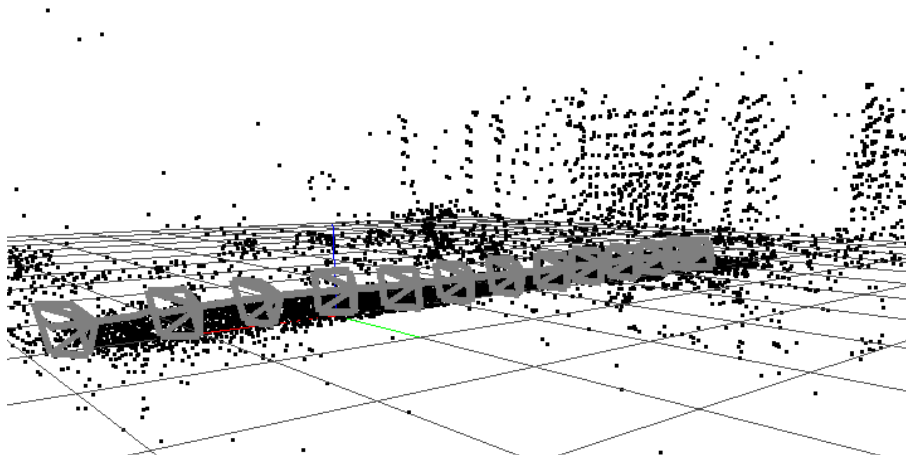


図 16 推定されたカメラとランドマークの三次元位置

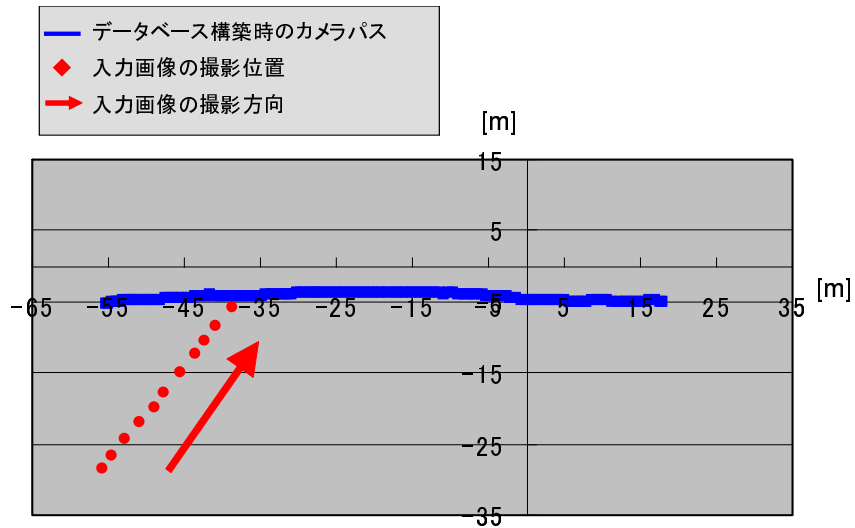


図 17 実験 1 におけるデータベース構築時のカメラパスと入力画像の撮影位置・方向

表 1 カメラ位置・姿勢推定の各処理における閾値

輝度勾配ベクトルの次元数	128
格子の分割数 l	72
条件 A の閾値 T_1 (度)	10
条件 B の係数 α	0.2
尤もらしさを判断する再投影誤差の閾値 (画素)	5
手順 1 の閾値 T_2 (度)	5
手順 2 の閾値 d (画素)	5

登録された。

次に、構築したランドマークデータベースを用いてカメラ位置・姿勢推定を行った。本実験で使用したカメラ位置・姿勢推定処理の閾値を表 1 に示す。本実験では、ランドマーク構築時のカメラパスから 2m ~ 23m 離れた異なる地点から同一の方位・対象を 11 枚の静止画像（解像度 720 × 480 画素）として撮影し、入力として用いた。図 17 に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・方向の関係を示す。図は、環境を水平面に対して上方からの見下ろした図となっており、図中の青の曲線はデータベース構築時のカメラパスを、赤い点・矢印は入力画像の撮影位置・方向を表す。入力画像の一部を図 18 に示す。同図中 (a),(b),(c) はそれぞれランドマーク構築時のカメラパスから 22m,10m,2m, 離れた地点から撮影した画像である。これらを入力画像として、4.2 節で述べた手法により投票を行った結果を図 19,20 に示す。また、文献 [29] の手法による投票結果を図 21 に示す。図 19 および図 21 は環境を水平面に対して上方からの見下ろした図となっており、図中の青の曲線はデータベース構築時のカメラパスを、赤い×印はカメラ位置の真値を、黄色い十字はカメラ位置・姿勢候補点を、緑の十字は推定されたカメラ位置・姿勢を表す。また、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。ただし、図 19 での投票数は、真値の姿勢に対応する投票箱に投票された数を用いている。図 20 は、真値の位置を含む図 19 の横軸に平行な直線位置における投票箱への投票の様子を示している。横軸は図 19 の横軸と同じ軸を表しており、縦軸は水平方位を表す。また、カメラ位置・姿勢推定の結果として表 2 に、従来手法および提案手法によるカメラ位置・姿勢推定の成功率、推定精度および再投影誤差の平均と標準偏差を示す。加えて、図 22 に、データベース構築時のカメラパスからの距離と位置推定誤差の関係を、図 23 に、データベース構築時のカメラパスからの距離と姿勢推定誤差の関係を示す。さらに、カメラ位置・姿勢推定を処理 (b-1,b-2), 処理 (b-3,b-4), 処理 (b-5) の 3 つのステップに分け、提案手法と従来手法のそれぞれの処理時間と処理 (b-1) ~ (b-5) までの合計の処理時間を表 3 に示す。



(a) 22m 地点

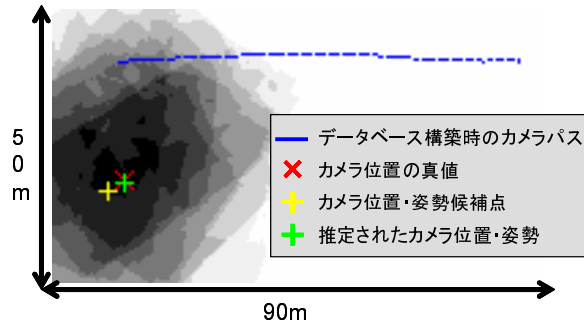


(b) 10m 地点

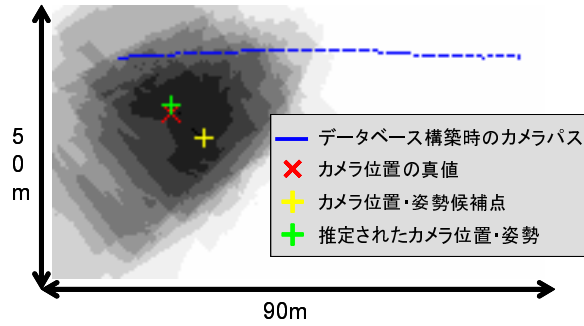


(c) 2m 地点

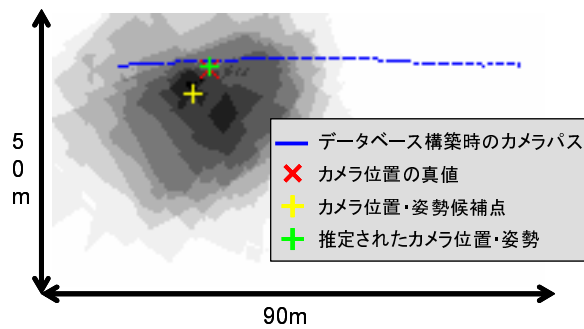
図 18 入力画像



(a) 22m 地点

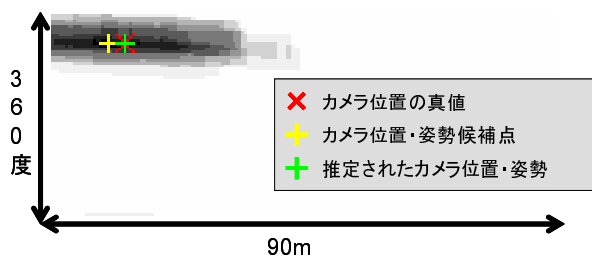


(b) 10m 地点

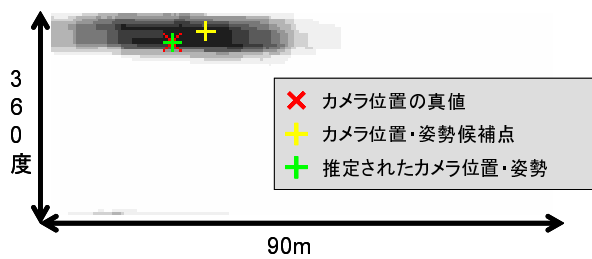


(c) 2m 地点

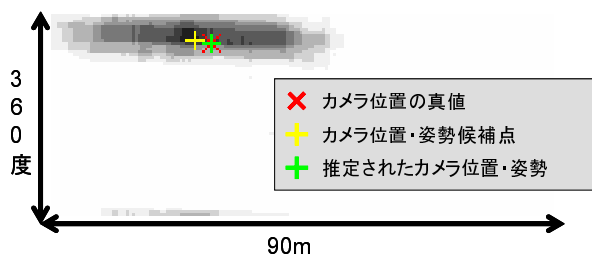
図 19 提案手法による位置投票結果



(a) 22m 地点

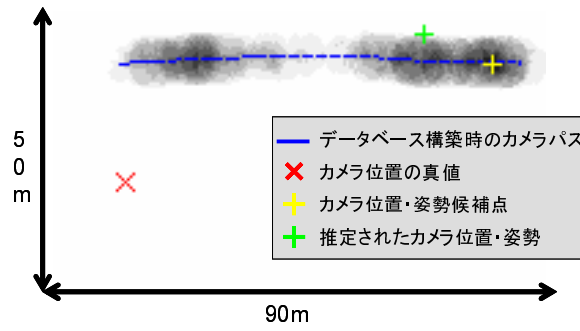


(b) 10m 地点

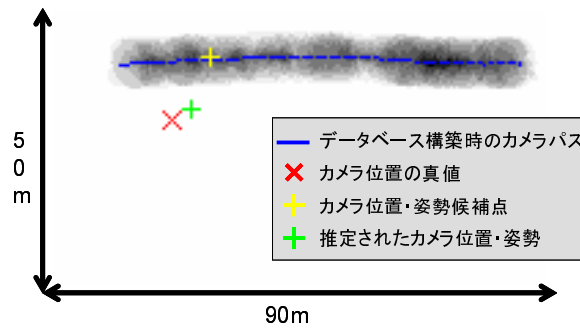


(c) 2m 地点

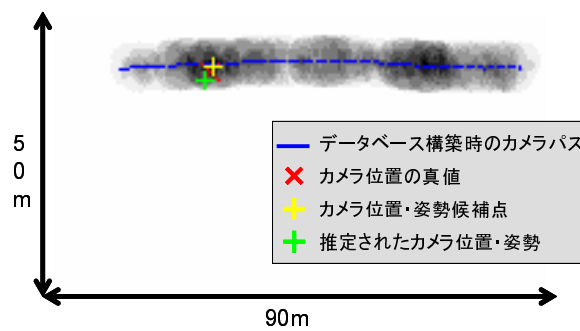
図 20 提案手法による姿勢投票結果



(a) 22m 地点



(b) 10m 地点



(c) 2m 地点

図 21 従来手法による位置投票結果

表 2 従来手法および提案手法によるカメラ位置・姿勢推定の成功率，推定精度
および再投影誤差

	提案手法	従来手法
システムが成功と判定した割合 (%)	100.0	63.6
平均位置誤差 (m)	1.1	4.2
位置誤差標準偏差 (m)	0.7	2.3
平均姿勢誤差 (度)	0.7	2.3
姿勢誤差標準偏差 (度)	0.3	1.3
平均再投影誤差 (画素)	1.0	1.8
再投影誤差標準偏差 (画素)	0.5	0.5

表 3 カメラ位置・姿勢推定にかかる処理時間 [単位:秒]

	提案手法	従来手法
b-1,b-2:特徴ベクトル生成	108.26	0.54
b-3,b-4:対応付け・投票	65.67	25.76
b-5:カメラ位置・姿勢推定	1.33	44.42
合計	175.27	70.71

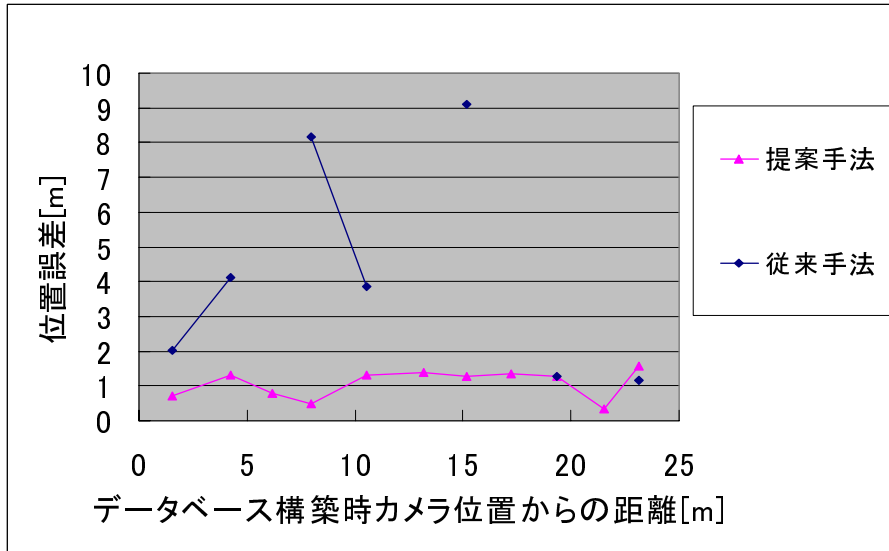


図 22 データベース構築時のカメラパスからの距離と位置推定誤差

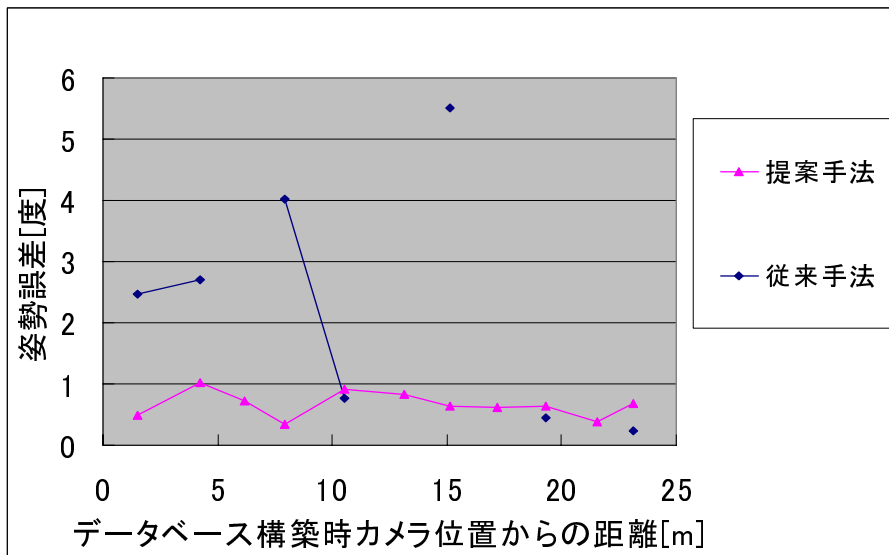


図 23 データベース構築時のカメラパスからの距離と姿勢推定誤差

5.2 実験1に関する考察

図19および図21から、従来手法に比べ提案手法の方が真値付近に投票が集中し、真値付近にカメラ位置・姿勢が推定されていることが確認できる。また、表2から、システムがカメラ位置・姿勢の推定に成功したと判定した割合が従来手法から大幅に向上しており、提案手法では11枚全て成功と判定されたことから、カメラパスから離れた位置で入力画像を撮影した場合における推定のロバスト性が従来手法に比べ向上していることが分かる。また、同表から、提案手法によって得られるカメラ位置・姿勢の推定精度および再投影誤差に関して、従来手法より高精度に求まっていることが確認できる。さらに、それぞれの位置・姿勢誤差の標準偏差を比べると、従来手法に比べ提案手法の方が誤差にバラつきが少ないことが分かる。また、図22および図23から、提案手法の位置誤差・姿勢誤差が従来手法に比べ、データベース構築時のカメラパスからの距離にかかわらず一定であることが分かる。これらの結果から、提案手法を用いることで従来手法に比べ、データベース構築時のカメラ位置から離れた地点におけるカメラ位置・姿勢推定のロバスト性が向上し、かつ高精度にカメラ位置・姿勢を推定できていることが確認できる。

本実験における提案手法の平均位置誤差は1.1m、平均姿勢誤差は0.7度であり、ARナビゲーションへの応用において、建物などへの注釈表示には問題がない精度であると言える。ただし、カメラ位置の近くの物体などにCGを重畳表示する場合には、表示位置に大きなズレが生じる可能性がある。

なお、1枚の入力画像に対するカメラ位置・姿勢推定の平均処理時間は表3に示すようにPC(CPU:Xeon 3GHz, Memory:16GB)を用いて約175秒であり、その内訳は、入力画像からの特徴ベクトルの作成(処理(b-1,b-2))に約108秒、入力画像上の特徴点とランドマークとの対応付けと投票によるカメラ位置・姿勢候補点の決定(処理(b-3,b-4))に約65秒、カメラ位置・姿勢推定(処理(b-5))に約1秒であった。従来手法と比較すると、対応付け・投票処理とカメラ位置・姿勢推定処理の合計処理時間はほぼ同等であった。これは、提案手法では投票で姿勢候補を決定しているのに対し、従来手法ではカメラの姿勢をカメラ位置・姿勢推定時に探索しているため、投票による姿勢決定により、多重スケールによる特徴ベ

クトルの比較回数の増加を補うだけの高速化を実現していることが確認できる。しかし、全体としては、従来手法よりも約 100 秒遅く、その原因は特徴ベクトル生成処理にある。これは、提案手法ではスケール対応を実現するために、特徴点スケールの算出処理 (b-1) が必要となり、加えて特徴ベクトルの多重スケール化により特徴ベクトル生成時間が増加しているためである。これらの結果から、提案手法を実用化するためには、入力画像上の特徴点を全て使うのではなく、まずいくつかの特異な点を対応付け、その上でランドマークの三次元位置と特徴点の画像上の座標を考慮し、利用可能な特徴点のみを選出するなどの改良を加え処理を高速化する必要があると言える。

5.3 提案手法のロバスト性検証実験 (実験 2)

本実験では、実験 1 と異なる位置において複数の撮影経路上で新たなランドマークデータベースを構築し、異なる多数の地点・方位で撮影した入力画像からカメラ位置・姿勢推定を行った。このデータベースは約 80m の経路で撮影した全方位動画から 3 章で述べた手法により構築され、約 12,500 個のランドマークが登録された。図 24 に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・方向の関係を示す。本実験では、ランドマークデータベース構築時にカメラパスが交差するように撮影を行い、そのカメラパス周辺において 5m の格子点上で画像 (解像度 640 × 480 画素) を撮影した (6 × 6=36 地点)。

以下では、撮影方向とカメラ位置・姿勢推定精度の関係について検証するために、まず全ての位置において、画像中に人工物を主に捉えることができる方向 (図 24 の方向 1) に関する実験結果を示し、次に画像中に自然物を多く含む方向 (図 24 の方向 2) に関する実験結果を示す。なお、本実験では撮影された画像 72 枚 (36 地点 × 2 方向) の内、手動で画像特徴点とランドマークを対応付けて真値を取得可能な 61 枚を入力画像として用いた。

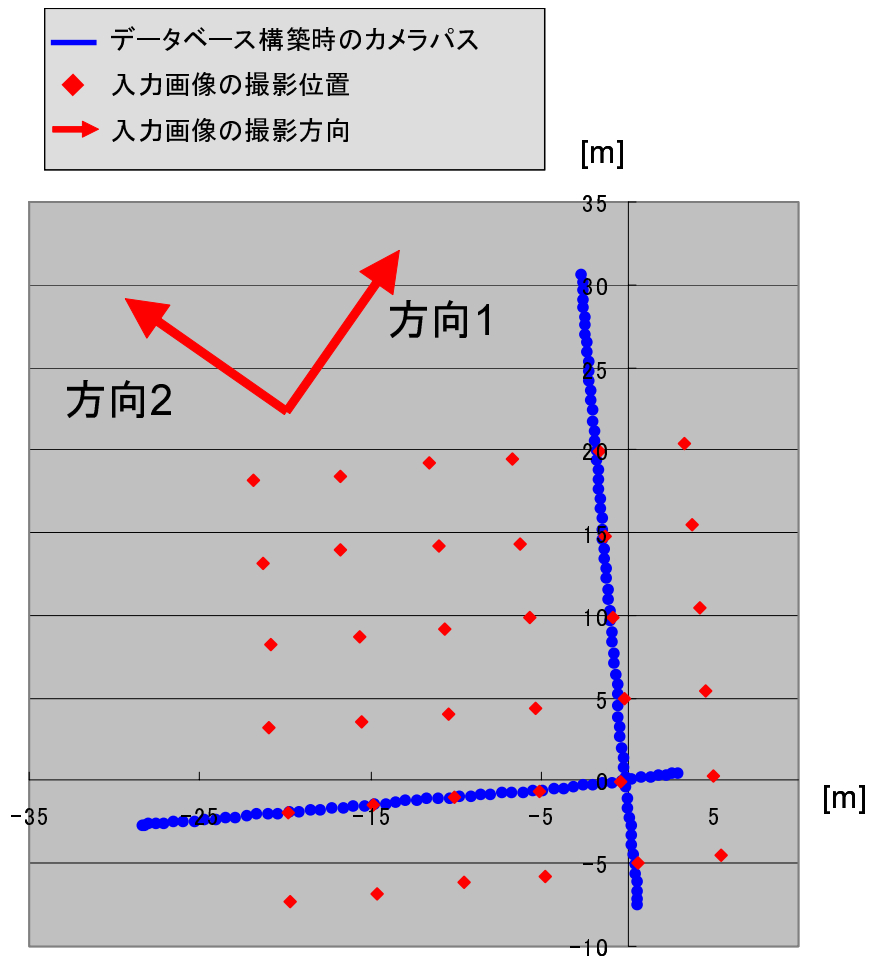


図 24 実験 2 におけるデータベース構築時のカメラパスと入力画像の撮影位置・方向

5.3.1 方向 1 に対する実験結果と考察

ここでは画像中に人工物を主に捉えることができる方向(方向 1)に対するカメラ位置・姿勢推定結果を示し、実験結果に対する考察を述べる。図 25 にシステムがカメラ位置・姿勢推定に成功したと判断した入力画像に対応する真のカメラ位置を示す。図中の青の曲線はデータベース構築時のカメラパスを、赤い点・矢印は入力画像の撮影位置・方向を、緑の点はシステムがカメラ位置・姿勢推定に成功したと判断した位置を示す。推定に成功した入力画像の例として、d,e,f 地点で撮影された画像を図 26 に、位置投票結果を図 27 に、姿勢投票結果を図 28 に示す。また、図 27 および 28 の濃度値は投票数を表しており、暗いほど投票数が多い。図 26 から、d,e,f 地点で撮影された画像は主として人工物を捉えていることが分かる。また位置・姿勢に対する投票結果から、真値付近に投票が集中しており、真値の近くにカメラ位置・姿勢が推定されていることが確認できる。

次に、システムが推定失敗と判定した入力画像の例として、c,g,h 地点撮影された画像を図 29 に示す。図から、これらの画像には大きく木が写り込み、人工物を遮っているのが分かる。そのため、画像中から十分な数のランドマークが検出できず推定に失敗している。また、ランドマークが画像中に写っていたとしても、特徴点スケールを算出する際の LoG 適用領域に木などの自然物が入り込むことで、特徴点スケールを正しく求められず、対応付けに失敗したり、投票場所が真値から外れ、安定して推定が行えない。また、a,b 地点では大きな自然物は写り込んでいないが、撮影方向とデータベース構築時のカメラパスが交わらないために、a,b 地点から見たランドマークと同じ見え方のランドマークがデータベース中に存在せず、推定に失敗している。

方向 1 に関する全体のカメラ位置・姿勢推定の成功率、推定精度および再投影誤差の平均と標準偏差を表 4 に示す。推定成功率は 79.3% であり、先に挙げたように、木などの自然物が画像に大きく写り込んでいる場合や、撮影方向とデータベース構築時のカメラパスが交わらない場合を除いてほとんどの画像で推定に成功している。また、推定精度も平均位置誤差が 1.4m、平均姿勢誤差が 1.4 度で、AR を用いてヒューマンナビゲーションを行うのに問題ない精度である。これらの結果から、入力画像中に主に人工物を捉えており、撮影方向とデータベース構

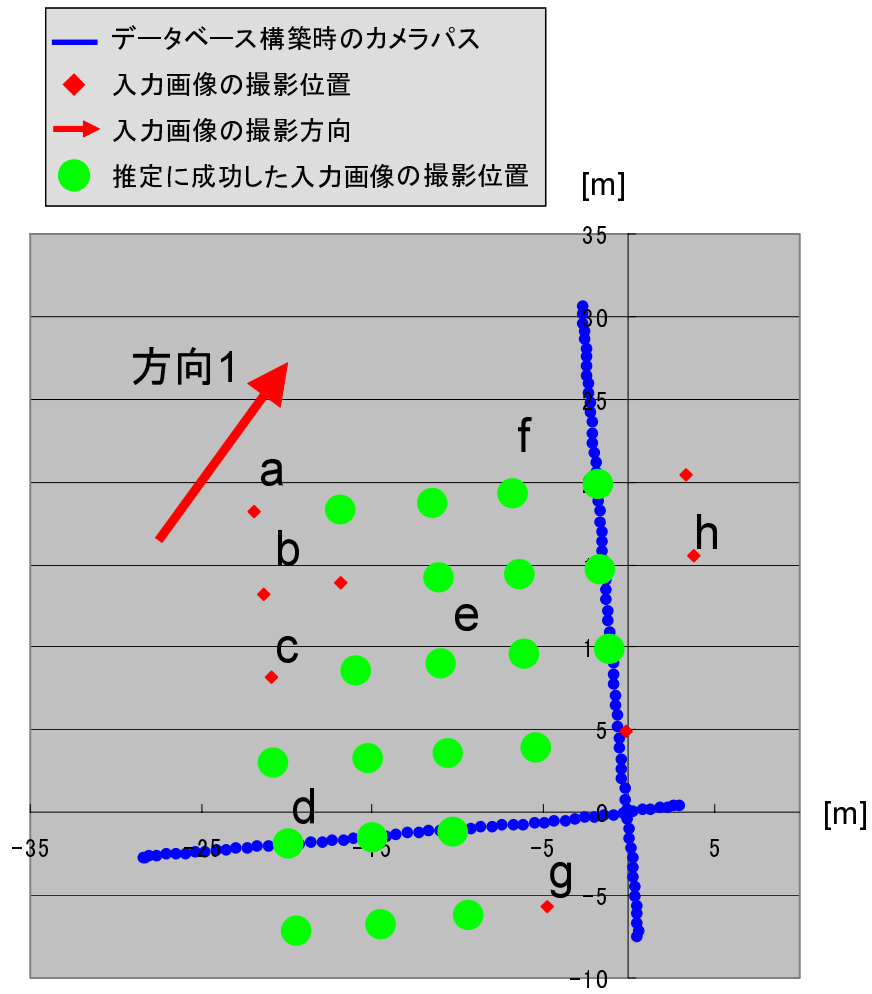


図 25 実験 2 の方向 1 におけるカメラ位置・姿勢推定の成功位置

表 4 方向 1 におけるカメラ位置・姿勢推定の成功率，推定精度および再投影誤差

システムが成功と判定した割合 (%)	79.3
平均位置誤差 (m)	1.4
位置誤差標準偏差 (m)	2.5
平均姿勢誤差 (度)	1.4
姿勢誤差標準偏差 (度)	2.0
平均再投影誤差 (画素)	2.0
再投影誤差標準偏差 (画素)	0.9

築時のカメラパスが交わるような場合においては，データベース構築時のカメラパスから離れた地点でもロバストにカメラ位置・姿勢推定が行えることを確認できた。



(a) d 地点

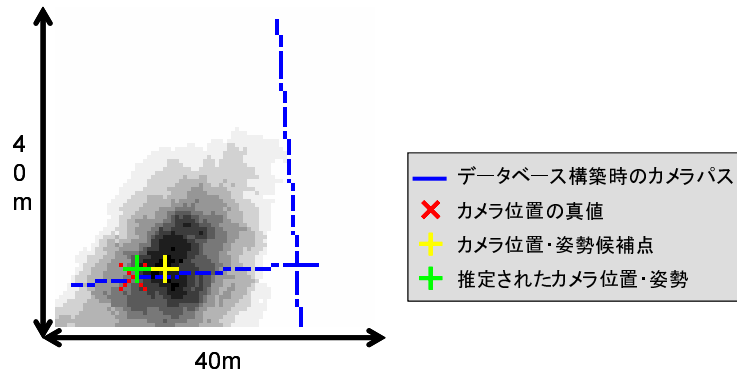


(b) e 地点

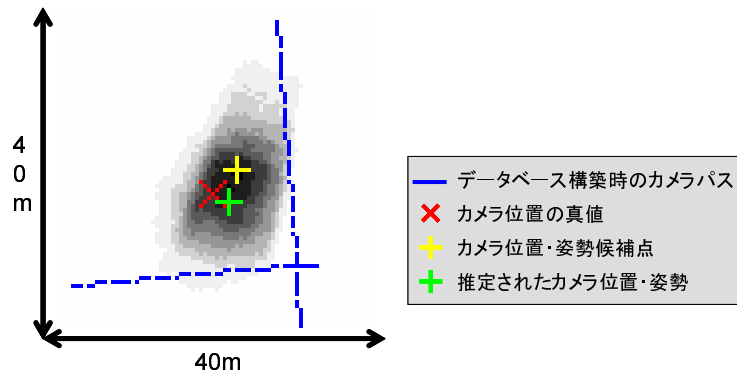


(c) f 地点

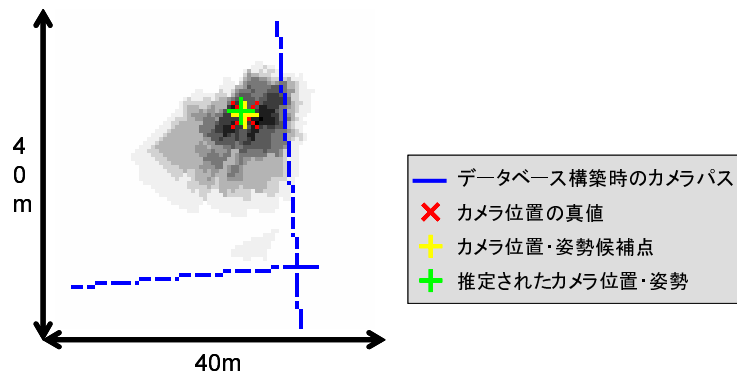
図 26 方向 1 における推定に成功した入力画像例



(a) d 地点

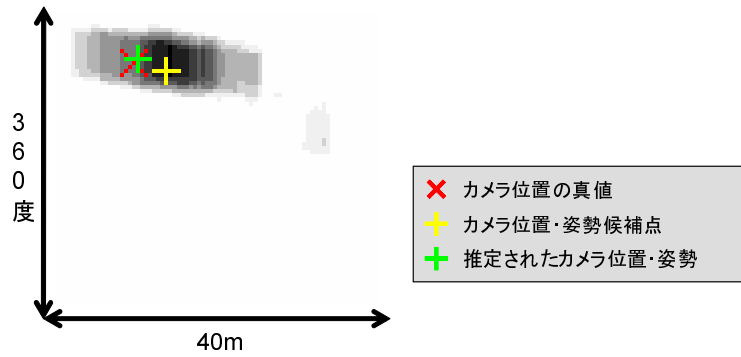


(b) e 地点

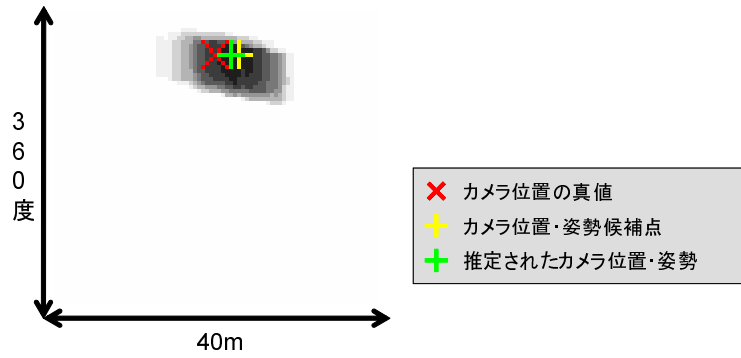


(c) f 地点

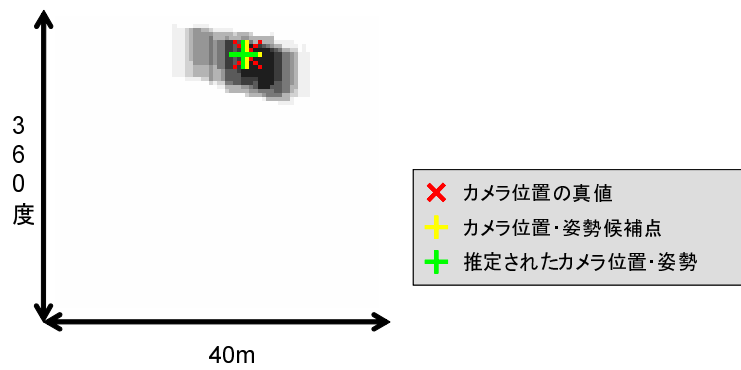
図 27 方向 1 における推定に成功した地点での位置投票例



(a) d 地点



(b) e 地点



(c) f 地点

図 28 方向 1 における推定に成功した地点での姿勢投票例



(a) c 地点



(b) g 地点



(c) h 地点

図 29 方向 1 における推定に失敗した入力画像例

5.3.2 方向 2 に対する結果と考察

ここでは、画像中に自然物を多く含む方向(方向 2)のカメラ位置・姿勢推定結果を示し、実験結果に対する考察を述べる。図 30 にシステムがカメラ位置・姿勢推定に成功したと判断した位置を示す。推定に成功した入力画像の例として、i,k,l 地点で撮影された画像を図 31 に、位置投票結果を図 32 に、姿勢投票結果を図 33 に示すこれらの画像は、比較的画像中に人工物を多く捉えており、推定に成功している。

次に、推定に失敗した入力画像の例として、a,d,j 地点撮影された画像を図 34 に示す。これらの地点で撮影された画像も、方向 1 での失敗例と同様画像中に自然物が大きく写り込んでいるため推定に失敗している。この撮影方向における推定失敗のほとんどは、画像への自然物の写り込みによるものである。その他の失敗の例としては、c 地点などが挙げられる。この地点で撮影された画像を図 35 に示す。画像中にある程度大きく人工物を捉えているが、本実験を行った環境にはこの建造物と類似した構造を持つ建造物が他に 2 つ存在し、画像上の特徴点が違う建造物のランドマークと対応付けられ推定に失敗している。図 36 に、図 35 で捉えている建造物と類似した構造を持つ建造物を捉えたランドマークデータベース構築時の撮影画像を示す。図 36 中の中央左側の建造物と図 35 で捉えている建造物が類似した構造をしており、本実験では誤った対応付けが行われた。このように、環境中に類似した構造が複数存在する場合は、誤った対応により推定が失敗してしまう場合がある。この問題を解決するためには、電子コンパスなどの携帯端末内蔵のセンサ情報を利用してランドマークを限定することが考えられる。また、このようなあいまい性が存在する場合には、複数箇所投票のピークが出現することからシステムがこれを自動検出し、ユーザに方向や位置を変えて入力画像の撮り直しを要求することも解決策として考えられる。

方向 2 に関する全体のカメラ位置・姿勢推定の成功率、推定精度および再投影誤差の平均と標準偏差を表 5 に示す。推定成功率は 41.7% であり、方向 1 に比べ半分程度の成功率である。これは、方向の違いによる画像中の人工物の割合の差だと考えられる。また、推定精度も平均位置誤差が 6.8m、平均姿勢誤差が 3.9 度と悪く、方向 1 に比べ大幅に劣る。この原因としては、方向 2 では画像中に自然

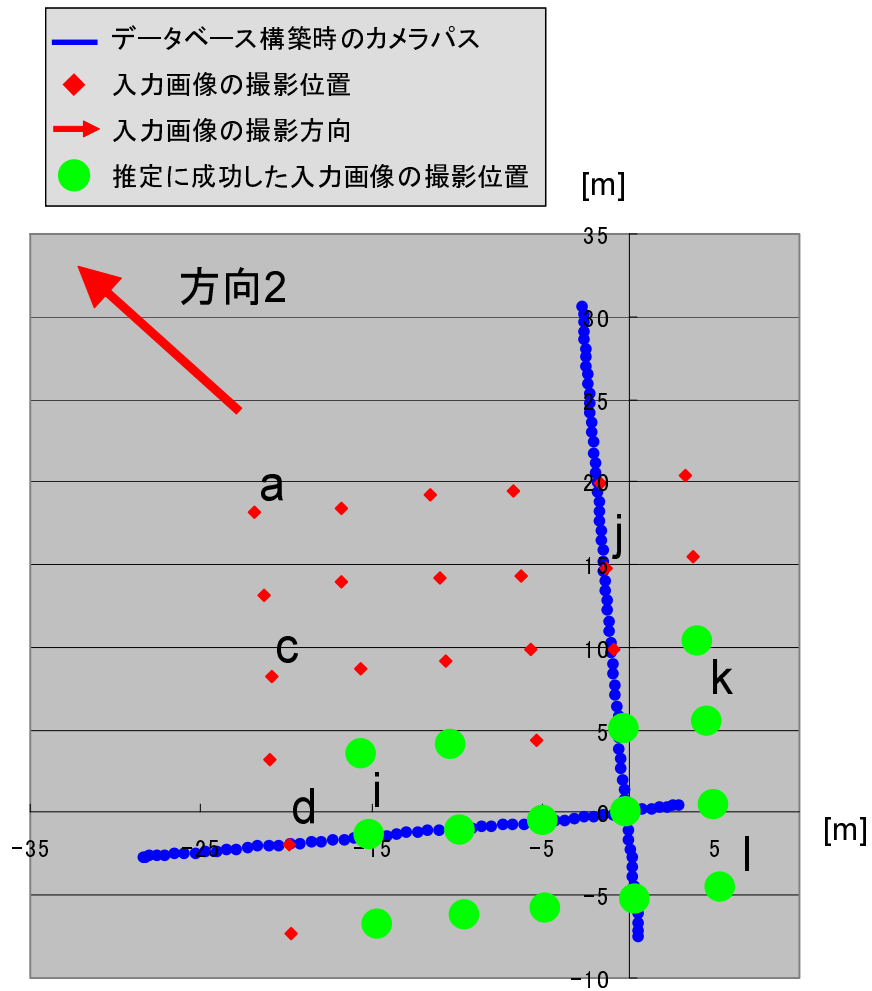


図 30 実験 2 の方向 2 におけるカメラ位置・姿勢推定の成功位置

表 5 方向 2 におけるカメラ位置・姿勢推定の成功率，推定精度および再投影誤差

システムが成功と判定した割合 (%)	41.7
平均位置誤差 (m)	6.8
位置誤差標準偏差 (m)	9.1
平均姿勢誤差 (度)	3.9
姿勢誤差標準偏差 (度)	4.5
平均再投影誤差 (画素)	2.0
再投影誤差標準偏差 (画素)	1.1

物を多く含むため，方向 1 に比べてカメラ位置・姿勢推定に使われる画像特徴点とランドマークの対応組の数が少ないことが挙げられる．また，推定に成功したカメラ位置とランドマークの距離が離れていることも原因の 1 つと考えられる．これらの結果から，安定したカメラ位置・姿勢推定を行うためには，画像中に人工物を多く含むように入力画像を撮影する必要があることが分かった．



(a) i 地点

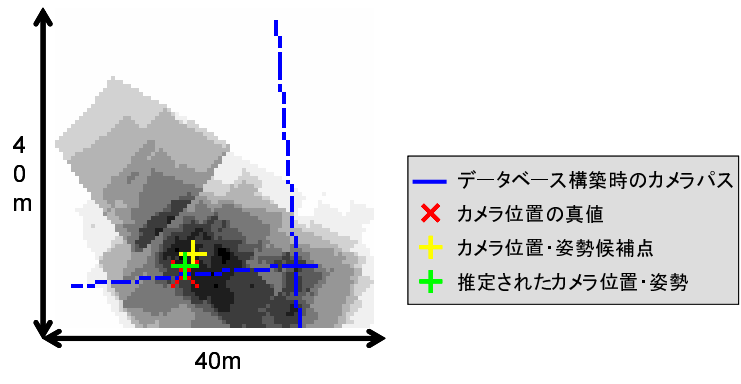


(b) k 地点

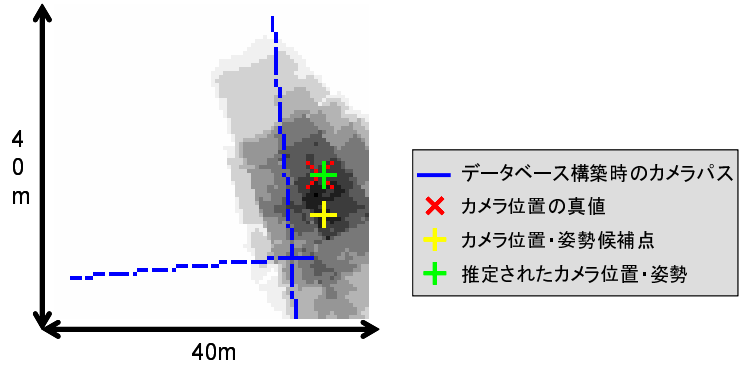


(c) l 地点

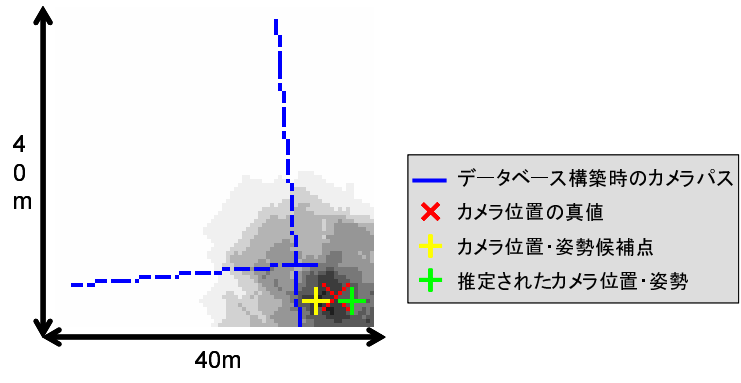
図 31 方向 2 における推定に成功した入力画像例



(a) i 地点

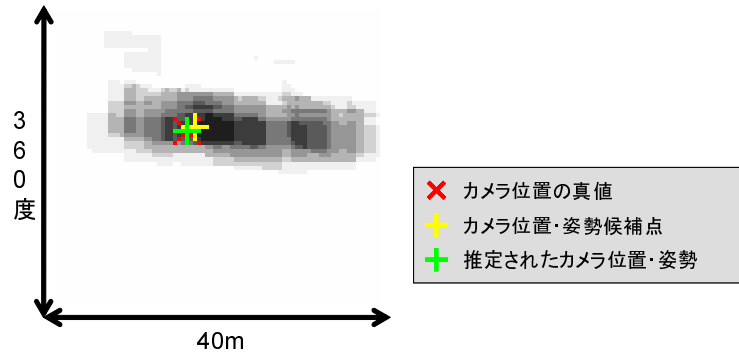


(b) k 地点

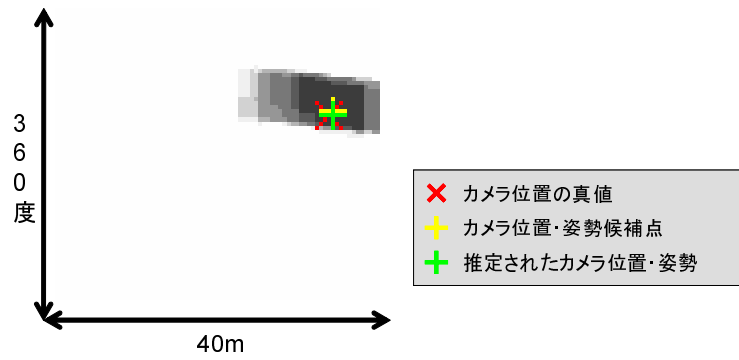


(c) l 地点

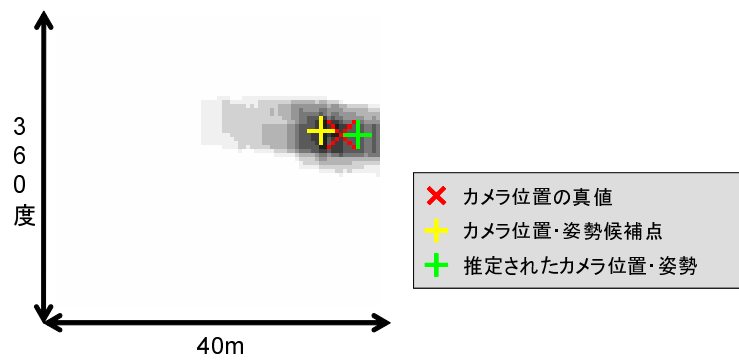
図 32 方向 2 における推定に成功した地点での位置投票例



(a) i 地点



(b) k 地点



(c) l 地点

図 33 方向 2 における推定に成功した地点での姿勢投票例



(a) a 地点



(b) d 地点



(c) j 地点

図 34 方向 2 における推定に失敗した入力画像例



図 35 類似構造を持つ建物 (入力画像)



図 36 類似構造を持つ建物 (データベース構築時撮影画像)

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、画像スケールに依存しない特徴量による特徴点の対応付けとスケール情報を利用したカメラ位置・姿勢候補の絞り込みを行うことで、データベース構築時のカメラパスから大きく離れた地点においても利用可能な、静止画像からのカメラ位置・姿勢推定法を提案した。提案手法では、まず、スケール情報を考慮した自然特徴点ランドマークデータベースを、環境を撮影した全方位画像から半自動で作成する。次に、作成したランドマークデータベース中のランドマークと、入力画像上の特徴点をスケール不変特徴量を用いて対応付け、スケール情報を考慮した投票を行うことで尤もらしい組み合わせを絞り込む。最後に、このようにして選択されたランドマークと入力画像上の特徴点の組み合わせを用いてカメラ位置・姿勢推定を行う。提案手法は、スケール不変特徴量を用いることで、撮影距離が異なっても同一特徴点を正しく対応付けることが可能であり、スケール情報を考慮した投票を行うことで、データベース構築時のカメラ位置から離れた場所で撮影された画像を入力とした場合にもカメラの位置・姿勢を正しく推定する。実験では、実際の屋外環境を全方位カメラで撮影することで構築された自然特徴点ランドマークデータベースを用いて、複数の静止画像からのカメラの位置・姿勢を推定した。また、スケール変化への対応評価実験とロバスト性の検証実験の2つを行い、それぞれの実験で、推定された位置・姿勢と手動で作成した正解データとを比較した。その結果、人工物を入力画像中に十分捉えている場合は、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、ARによるヒューマンナビゲーションを行うのに十分な精度でロバストに推定を行えることを確認できた。ただし、自然物が多く存在し、安定に利用可能なランドマークが少ない場合には十分な精度でカメラ位置・姿勢の推定を行うことができなかった。

今後の課題として、処理の高速化やサーバ・クライアント型システムの構築、システムとユーザのインタラクションなどが挙げられる。処理の高速化では、SIFT-descriptorの主成分分析を行い、次元数を削減することが考えられる。また、初めに得られた入力画像からカメラ位置・姿勢推定に失敗した場合、その入力画像から得られた情報からある程度の位置・姿勢を絞り込み、再度ユーザに位置・姿

勢を指示しながら画像の撮り直しを要求するなどが考えられる。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端末を用いた AR によるヒューマンナビゲーションシステムが実現可能となる。

謝辞

本研究の全過程を通して，懇切なる御指導，御鞭撻を賜った視覚情報メディア講座横矢直和教授に心より感謝致します．また，本研究の遂行にあたり，有益な御助言，御鞭撻頂いた知能情報処理学講座木戸出正継教授に厚く御礼申し上げます．そして，本研究を通じ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致します．さらに，本研究の全過程において，的確な御助言・御指導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．また，本研究を行うにあたり，多大なる御助言，御鞭撻を賜っ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．さらに，本研究の遂行にあたり，細やかな御助言・御鞭撻を頂いた視覚情報メディア講座池田聖特任助手に心より御礼申し上げます．また，研究室での生活を支えていただいた視覚情報メディア講座事務補佐員守屋智代女史に深く感謝致します．最後に，研究活動だけでなく日々の生活においても大変お世話になった視覚情報メディア講座の諸氏に心より感謝致します．

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. Conf. 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] N. B. Priyantha, A. Chakraborty and H. Balakrishnan: “The cricket location-support system,” Proc. ACM/IEEE Int. Conf. on Mobile Computing and Networking, pp. 32–43, 2000.
- [4] B. Thomas, B. Close, J. Donoghue, J. Squires, P. Bondi, M. Morris and W. Piekarski: “Arquake: An outdoor/indoor augmented reality first person application,” Proc. Int. Symp. on Wearable Computers, pp. 139–146, 2000.
- [5] P. Dähne and J. Karigiannis: “Archeoguide: System architecture of a mobile outdoor augmented reality system,” Proc. Int. Conf. on Mixed and Augmented Reality, pp. 263–264, 2002.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] R. Tenmoku, M. Kanbara and N. Yokoya.: “A positioning method combining specification of users absolute 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.
- [10] E. Foxlin: “Pedestrian tracking with dhow-mounted inertial sensors,” IEEE Computer Graphics and Applications, Vo. 25, No. 6, pp. 38–46, 2005.
- [11] 濱口明宏, 神原誠之, 横矢直和: “装着した3次元位置姿勢センサによるウェアラブルユーザの自己位置推定”, 電子情報通信学会 技術研究報告, MVE2005-59, , 2005.
- [12] 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.
- [13] 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, IE2002-53, pp. 65–70, 2002.
- [14] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. IEEE Int. Symp. on Wearable Computers, pp. 21–23, 2003.
- [15] M. Möhring, C. Lessig and O. Bimber: “Video see-through ar on consumer cell-phones,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 252–253, 2004.
- [16] 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.
- [17] 齊藤滋, 谷川智洋, 廣瀬通孝: “コード化した模様を内装に施した室内における位置同定システム”, 電子情報通信学会 技術研究報告, MVE2006-1, , 2006.

- [18] U. Neumann and S. You: “Natural feature tracking for augmented-reality,” IEEE Transactions on Multimedia, Vo. 1, No. 1, pp. 53–64, 1999.
- [19] A. Davison, Y. G. Cid and N. Kita: “Real-time 3D slam with wide-angle vision,” Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [20] 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.
- [21] 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.
- [22] R. Cipolla, D. Robertson and B. Tordoff: “Image-based localization,” Proc. Int. Conf. Vertual Systems and Multimedia, pp. 22–29, 2004.
- [23] L. Vacchetti, V. Lepetit and P. Fua: “Stable real-time 3D tracking using online and offline information,” IEEE Transaction on Pattern Analysis and Machine Intelligence, Vo. 26, No. 10, pp. 1385–1391, 2004.
- [24] L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3D camera tracking,” Proc. IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp. 48–57, 2004.
- [25] E. Rosten and T. Drummond: “Fusing points and lines for high performance tracking,” Proc. IEEE Int. Conf. on Computer Vision, Vo. 2, No. 10, pp. 1508–1515, 2005.
- [26] 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.

- [27] 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.
- [28] N. Snavely, S. M. Seitz and R. Szeliski: “Photo tourism: Exploring photo collections in 3D,” Proc. ACM SIGGRAPH 2006, Vo. 25, pp. 835–846, 2006.
- [29] 中川智香, 佐藤智和, 横矢直和: “自然特徴点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”, 画像の認識・理解シンポジウム (MIRU2006) 講演論文集, pp. 249–254, 2006.
- [30] 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.
- [31] S. You and U. Neumann: “Fusion of vision and gyro tracking for robust augmented reality registration,” Proc. IEEE on Virtual Reality Conf., pp. 71–78, 2005.
- [32] 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.
- [33] M. Kouroggi 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.
- [34] G. Reitmayr and T. W. Drummond: “Going out: Robust model-based tracking for outdoor augmented reality,” Proc. IEEE on Virtual Reality Conf., pp. 71–78, 2005.
- [35] 佐藤清秀, 穴吹まほろ, 山本裕之, 田村秀行: “屋外装着型複合現実感のためのハイブリッド位置合わせ手法”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 129–137, 2002.

- [36] 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.
- [37] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int. Journal of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.
- [38] 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.
- [39] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.
- [40] K. Mikolajczyk and C. Schmid: “Scale & affine invariant interest point detectors,” Int. J. of Computer Vision, Vol. 60, No. 1, pp. 63–86, 2004.
- [41] R. Klette, K. Schluns and A. Koschan Eds: “Computer vision: Three-dimensional data from image,” Springer, 1998.
- [42] R. J. Rousseeuw: “Least median of squares regression,” J. American Stat. Assoc, Vol. 79, pp. 871–880, 1984.
- [43] 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.