

NAIST-IS-MT0651084

修士論文

画像特徴点の特異度を考慮したランドマーク照合
による一枚の画像からのカメラ位置・姿勢推定の
高速化とロバスト性の向上

西海 嘉志

2008年3月17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。

西海 嘉志

審査委員：

横矢 直和 教授 (主指導教員)

小笠原 司 教授 (副指導教員)

山澤 一誠 准教授 (副指導教員)

佐藤 智和 助教 (副指導教員)

画像特徴点の特異度を考慮したランドマーク照合 による一枚の画像からのカメラ位置・姿勢推定の 高速化とロバスト性の向上*

西海 嘉志

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は難しく，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．これに対して，現実環境を撮影した画像を用い，画像撮影時のカメラ位置・姿勢に基づいて CG によるナビゲーション情報を重畳合成することで，ユーザに直感的な案内情報を提供する拡張現実感 (Augmented Reality; AR) ナビゲーションに関する研究が盛んに行われている．このような分野の研究は，従来比較的計算能力の高いウェアラブルコンピュータや車載機器上での映像の実時間処理を前提に行われてきた．これに対して，計算能力の低い携帯端末上での AR ナビゲーションを想定し，扱いが容易な一枚の画像からカメラの位置・姿勢推定を行う手法が提案されている．この手法では，三次元位置と見え方が既知の複数のランドマークを事前にデータベース化した上で，それらのランドマークを入力画像上で探索し，照合したランドマークの情報を用いてカメラ位置・姿勢を 6 自由度で推定する．しかし，従来手法では，環境中に類似したランドマークが多く存在する場合に誤対応が発生し推定が失敗するという問題や，データベースに登録されたすべてのランドマーク

* 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-MT0651084, 2008 年 3 月 17 日.

と入力画像上の特徴点の対応付けを総当たりで行うため計算コストが大きいという問題が残されている．このような問題を解決するために，本研究では，環境内に類似したランドマークが少ない特異なランドマークでは誤対応の発生率が低いこと注目し，信頼度の指標としてランドマークの特異度を用いたランドマークと自然特徴点の対応付けを行う．具体的には，特異度の高いランドマークを優先的に対応点探索に用い，対応付けに成功した特異度の高いランドマークから空間距離を考慮してランドマークを選択することで，特異度の低いランドマークに対しても誤対応の低減を図るとともに，ランドマークの探索範囲を空間中で限定にすることで処理の高速化を図る．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周辺で撮影した入力静止画像からのカメラ位置・姿勢推定を行い，処理時間，精度，推定成功率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，画像特徴点の特異度

Fast and Robust Camera Parameter Estimation from a Still Image Based on Landmark Database Considering Uniqueness of Image Features*

Yoshiyuki Nishiumi

Abstract

Human navigation systems on cellular phones can navigate users showing his/her position and navigation information on 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 navigation system, a novel technique called Augmented Reality (AR) has been proposed and investigated. In AR applications, guiding information are drawn on an image of real environment at geometrically correct position based on the absolute position and posture of the camera. Most of researches in this field are investigated for wearable computers or car-mounted systems because they have capability of real-time processing of video frames. On the other hand, cellular phones do not have them. Thus, to realize the AR navigation system on cellular phones, the method has been proposed that estimate camera parameters from a still image based on feature landmark database. However, this approach has some problems. First, camera parameter estimation fails when similar landmarks exist in target environment because they cause incorrect matching between landmarks in the database and feature points in an image. Secondly, calculation cost is expensive as the result of feature point matching using all landmarks in the database. In this thesis, to solve these

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0651084, March 17, 2008.

problems, we focus on the uniqueness of image features in landmark database. Because unique features do not have similar features in landmark database, they result in a high correspondence ratio in the feature points matching. Therefore, we employ uniqueness of image features as a measure for the confidence of feature point matching. Additionally, to make correspondence of un-unique features effectively, searching range for un-unique features are also limited using spatial distance to unique features. In experiments, the effectiveness of the proposed method by comparing estimation accuracy, processing time and success rate with the previous work.

Keywords:

camera parameter estimation , still image, feature landmark database, uniqueness of image features

目次

1. はじめに	1
2. カメラ位置・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	5
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 本研究の位置付けと方針	12
3. ランドマークデータベースの構築	17
3.1 ランドマークデータベースの構成要素	17
3.2 全方位動画像による環境の三次元復元	19
3.3 特異度を考慮したランドマーク情報の抽出	20
3.3.1 特徴点の固有スケールの算出	20
3.3.2 多重スケール特徴ベクトルの生成	22
3.3.3 特異度の算出	23
4. 特異度を用いた静止画像からのカメラ位置・姿勢推定	25
4.1 データベースの選択	26
4.2 特異度を用いたランドマークの照合	26
4.2.1 特異度を用いた基準ランドマークの選択	26
4.2.2 空間距離を考慮したランドマークの選択	28
4.3 観測地点の整合性によるランドマークの選択	29
4.4 選択されたランドマークを用いたカメラ位置・姿勢の推定	31
4.4.1 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・姿勢の推定	32

4.4.2	暫定カメラ位置・姿勢に基づくカメラ位置・姿勢の再推定	33
5.	実験	35
5.1	実験条件	35
5.2	従来手法との比較によるロバスト性向上効果の確認	37
5.3	従来手法との比較による処理速度向上効果の確認	47
6.	まとめ	48
	謝辞	49
	参考文献	50

目 次

1	Möhring ら [16] の携帯端末上での AR システム	5
2	興相ら [22] の映像上への注釈表示システム	7
3	Yang ら [24] の手法における三次元モデルと入力画像の比較	8
4	各カメラ位置・姿勢推定手法の位置付け	13
5	画像を用いてカメラ位置・姿勢推定する各手法の位置付け	14
6	提案手法の処理の流れ	15
7	本研究で想定するサーバ・クライアント型システム	16
8	ランドマークデータベースの構成要素	18
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	特徴ベクトルの生成手順	23
12	特異なランドマークの概念図	24
13	ランドマークと画像上特徴点の比較例	27
14	ランドマークの固有スケールと撮影距離の関係	28
15	空間距離を考慮した対応候補ランドマークの選択	29
16	1つのランドマークからの投票例	30
17	再投影誤差の最小化によるカメラ位置・姿勢推定	33
18	全方位型マルチカメラシステム Ladybug と入力画像の一部	37
19	データベースに登録された自然特徴点と全方位カメラの位置	38
20	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	39
21	各撮影地点におけるカメラ位置・姿勢の推定成功・失敗の判定結果 (方向 1)	40
22	各撮影地点におけるカメラ位置・姿勢の推定成功・失敗の判定結果 (方向 2)	41
23	A 地点における従来手法と提案手法の投票, 推定結果の比較	44
24	B 地点における従来手法と提案手法の対応付けの比較	45
25	C 地点における従来手法と提案手法の対応付けの比較	46

表 目 次

1	カメラ位置・姿勢推定に用いたパラメータ	36
2	提案手法と従来手法の推定成功率・精度の比較 (方向 1)	43
3	提案手法と従来手法の推定成功率・精度の比較 (方向 2)	43
4	提案手法と従来手法の処理時間の比較 [秒]	47

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，これらのシステムを用いた場合でも複雑な交差点などでは地図と現実空間の関係を正しく把握することは難しく，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，最近では一部の複雑な交差点などで三次元的な CG を表示することにより直感的な道案内をするサービスが提供され始めている．しかしながら，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狭く，利用できる場所が限定されている．これらの問題を解決するために，現実空間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，ユーザ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は，撮影しているカメラの正確な絶対位置・姿勢情報が必要となる．

従来，カメラの位置・姿勢を推定することを目的とした研究が数多く行われているが，これらの研究は GPS やジャイロなどのセンサを組み合わせる手法，画像を用いる手法，それらのハイブリッドに分類できる．センサを用いる手法は，一般に複数のセンサを組み合わせるカメラの位置・姿勢を推定する．この手法は，広域環境での利用が可能であり，広く研究されているが，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) 人工的なランドマークを用いる手法，(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，誤差の蓄

積や計算レート，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，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，すでに広く普及しているカメラ付き携帯端末を用いて広域環境で利用することが可能なカメラ位置・姿勢推定の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携帯端末でも容易に取得可能な一枚の静止画像からカメラ位置・姿勢の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・姿勢推定が可能な薄らの手法 [1] を基礎とし，データベース中の各ランドマークの特異度を考慮することにより，カメラ位置・姿勢推定の高速化とロバスト性の向上を図る．本研究では，環境内に見え方の類似したランドマークが少ない特異なランドマークには，特徴点の対応付けにおいて誤対応の発生率が低いという特長があることに注目し，信頼度の指標としてランドマークの特異度を用いた自然特徴点の対応付けを行う．具体的には，まず特異度の高いランドマークから優先的に対応点探索を行い，次に対応付けに成功した特異度の高いランドマークと他のランドマークの空間距離を考慮してランドマークを対応付けることで，環境内に類似したランドマークを持つ特異度の低いランドマークに対しても誤対応の低減を図るとともに，対応点探索の範囲を限定することで処理の高速化を図る．

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，特異度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に基づき，各ランドマークの特異度を信頼度の指標として用いるカメラ位置・姿勢推定の手法について述べる．5章では，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章で本研究のまとめと今後の課題について述べる．

2. カメラ位置・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処理可能なものを、(1)GPSやジャイロなどのセンサを用いる手法 [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]、(2)画像を用いる手法 [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 1]、(3)それらのハイブリッド手法 [33, 34, 35, 36, 37, 38, 39] 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提案手法の概要を述べる。

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・姿勢推定手法は、超音波センサなどのインフラを環境内に人工的に配置し、利用する手法 [2, 3, 4] 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 [5, 6, 7, 8, 9, 10, 11, 12] 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、特別なインフラは用いないが、絶対方位情報が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 [2, 3, 4] では、超音波センサや赤外線ビーコンなど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長がある。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・姿勢を計測するセンサを組み合わせる手法 [5, 6, 7, 8, 9, 10, 11, 12] では、インフラによるカメラ位置・姿勢の取得が不可能な状況に至っても、センサを用いてインフラから最後に取得した絶対位置・姿勢からの相対的なカメラの運動を計測し、カメラ位置・姿勢の推定を継続できる。このアプローチでは、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・姿勢推定を実現できる [8]。また、インフラの設置密度を下げることでインフラの設置・維持コストの削減を図るために、センサによる自立計測をより高精度に行う研究も行われている [11, 12]。しかし、このようなインフラとセンサの組み合わせによる手法を利用するためには、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ず、既存の携帯端末上で手法を利用することは難しい。

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のみを用いて、環境に対する絶対的なカメラの位置・姿勢を推定するためには、一般に環境の事前知識が必要となる。環境の事前知識としては、以下のようなものが用いられる。

- 三次元位置関係が既知の人工的なマーカ [13, 14, 15, 16, 17, 18, 19, 20]
- 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 [21, 22, 23]
- 事前に作成した環境の三次元モデル [24, 25, 27, 28, 29]
- 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 [30, 31, 32, 1]

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、各手法の詳細を述べる。



図 1 Möhring ら [16] の携帯端末上での AR システム

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・姿勢の推定手法としては，ARToolkit[13]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用いる手法 [14, 15, 16, 17, 18]，人工的なマーカと自然特徴点の追跡を併用する手法 [19, 20] 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．Wagner ら [19] や Möhring ら [16] 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・姿勢が推定できることを実証している．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，画像マーカを利用する手法で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環境の景観を損ねてしまうという問題がある．

これに対して，中里ら [17] は天井に半透明の再帰性反射材を利用し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，景観を損ねることなくカメラ位置・姿勢を推定する手法を提案している．また，斎藤ら [18] は部屋の内装となる模様位置情報を埋め込むことで，人工的なマーカを周囲と調和させ，

目立ちにくくす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を写らなければカメラの位置・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削減のために，Neumann ら [19] や Davison ら [20] は，人工的なマーカと自然特徴点の追跡を併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．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，人工的なマーカが写らない状態では，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 [21, 22] 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 [23] などがある．

岩佐らの手法 [21] 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度を評価し，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で，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 [22] は図 2 に示すような映像上への注釈情報の重畳合成を実現するために，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ド・アノテーション手法を提案している．この手法は，環境中の複数地点で事前に撮影



図 2 興相ら [22] の映像上への注釈表示システム

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択することで入力画像の大まかな位置を推定している．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．

Cipolla ら [23] は，画像処理を用いてデータベース内の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，まず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する．次に，入力画像とデータベース画像の直線の対応関係を用いることで，データベース画像に対する相対的なカメラの位置・姿勢を推定する．この手法では，静止画像 1 枚を入力として扱い，サーバ・クライアント型システムの枠組みによりサーバ上で推定を行うため，携帯端末でも利用可能である．しかし，並行直線が画像内に複数存在していることを前提として推定を行うため，利

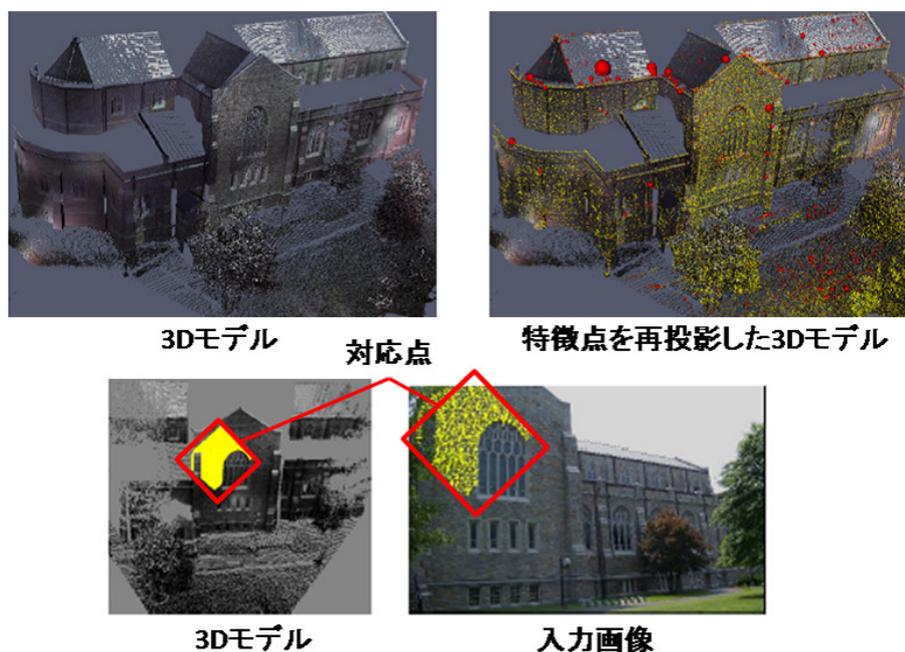


図 3 Yang ら [24] の手法における三次元モデルと入力画像の比較

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 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る手法 [24, 25] や三次元モデルのワイヤフレームを用いた手法 [26]，自然特徴点とワイヤフレームモデルを用いた手法 [27, 28] などがある．Yang ら [24] は，図 3 に示すような環境の三次元モデルのテクスチャ上の特徴量を入力画像上の特徴量と照合することで，三次元モデルと入力画像の対応関係を求め，1 枚の静止画像からカメラの位置・姿勢を推定する手法を提案している．また，Vacchetti ら [25] は，三次元モデルと複数のキーフレーム画像を同様に対応付け，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 らは，自然特徴点とワイヤフレームモデルを併用したカメラ位置・姿

勢推定手法 [27] も提案している．一般的に，ワイヤフレームモデルのみを用いたカメラ位置・姿勢推定手法 [26] 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取ることで位置・姿勢の推定を行うが，環境中にエッジが多い場合，ワイヤフレームモデルと入力画像中のエッジの誤対応が発生しやすくカメラ位置・姿勢推定が不安定になりやすい．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 [27] 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関わらずロバストなカメラ位置・姿勢推定を実現している．また Rosten ら [28] は，Vacchetti らの手法 [27] と同様に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作成を行い，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 FAST オペレータを用いることで，カメラの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・姿勢推定を実現している．

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 での利用が可能である．しかし，これらの手法ではモデル作成のために多くの人的コストがかかり，ナビゲーションなどを前提に比較的規模の大きなシーンをモデル化することは容易ではない．

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 [30]，大江ら [31]，Snavely ら [32]，薄ら [1] 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，及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位置・姿勢を 6 自由度で推定する．

Skrypnyk らの手法 [30] 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 特徴点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．

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・姿勢を行う手法 [31] を提案している．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・姿勢推定を実現する．この手法は，structure-from-motion による三次元復元において，基準となる測量済みの三次元位置を与えることで，誤差の累積を防いでいる．しかし，この手法は動画像を入力とし，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，カメラ位置・姿勢の推定を実現しており，カメラ位置・姿勢のよい初期値を与えることができない 1 枚の画像からのカメラ位置・姿勢推定に利用することはできない．

薄ら [1] 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 [31] と同様に全方位動画像から structure-from-motion によって自然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の位置・姿勢推定を行う．この手法は，静止画像1枚のみからカメラ位置・姿勢推定を行うことが可能であり，市販のカメラ付き携帯端末での使用が可能である．しかし，環境中のすべてのランドマークと入力画像中の自然特徴点を照合するため，計算コストが大きいという問題や，環境中に類似した特徴をもつランドマークが複数存在する場合に，特徴点のマッチングにおける誤対応が発生し，推定が失敗するといった問題が残されている．

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・姿勢手法 [33, 34, 35, 36, 37, 38, 39] は，2.1 節，2.2 節で述べた各々の手法の欠点をお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とっている．

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 [33, 34, 35] では，画像マーカから絶対位置を取得できるため，定期的に画像中にマーカを捉えることでセンサの蓄積誤差を抑制することが可能である．しかし，定期的にマーカを画像内に捉える必要から屋外などの広域環境で利用する場合はマーカの設置に人的コストがかかり，環境の景観を害する可能性がある．

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 [36] 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が高いデータベース画像の撮影位置・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．従って，広範囲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であり，同等の範囲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，機器構成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 [37] は，モデルと入力

画像の対応付けの際に加速度センサの情報も用いることで、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位置・姿勢推定を継続可能である。しかし、広域範囲に適用した場合には、やはり三次元モデルの作成に多くの人的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 [38, 39, 40] 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。この手法では、景観を損なうことなく、高精度にカメラ位置・姿勢推定をすることが可能であり、データベースの構築コストも低い。しかし、文献 [38] の手法は、カメラ位置が固定されており、カメラの移動を考慮していない。また、文献 [39] 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、建築物などの GPS の計測精度が悪い環境では利用が難しい。文献 [40] の手法は、大江ら [31] の手法をベースとしており、1 枚の静止画像からのカメラ位置・姿勢推定には利用できない。

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とっており、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 本研究の位置付けと方針

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、カメラ位置・姿勢推定の従来手法の特徴を示す。前節までに概観したように、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・姿勢推定手法はすで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や、使用する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

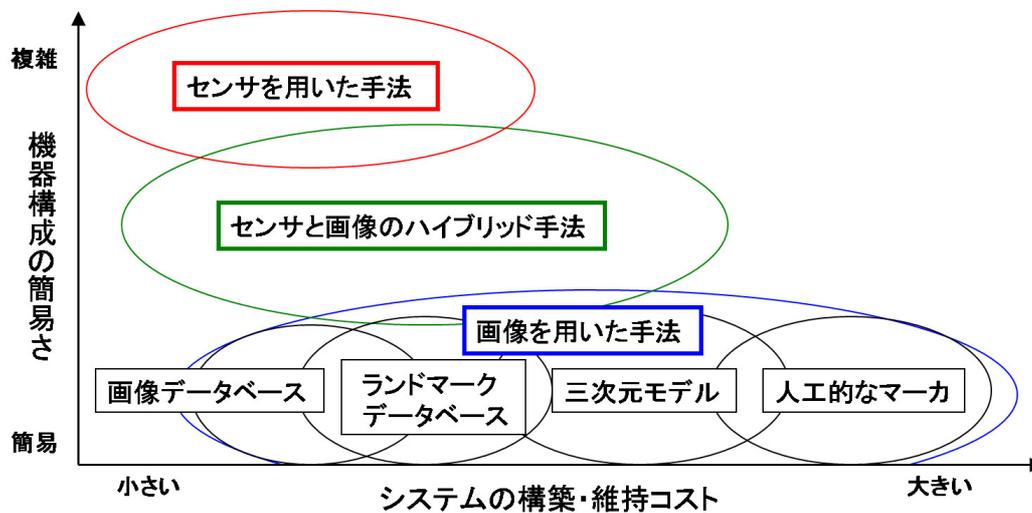


図 4 各カメラ位置・姿勢推定手法の位置付け

う問題があり，現在すでに広く普及した携帯電話や PDA などを利用することは難しい．一方，画像を用いる手法は，携帯電話や PDA に内蔵されているカメラのみを利用することで，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で利用可能である．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・維持コストを抑える必要がある．従来手法において，薄らの手法 [1] 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に推定を行うことができる．薄らの手法では，環境中に類似したランドマークが多く存在する場合に誤対応が発生し推定が失敗するという問題や，データベースに登録されたすべてのランドマークと入力画像上の特徴点の対応付けを総当たりで行うため計算コストが大きいという問題が残されている．このような問題を解決するために，本研究では，環境内に類似したランドマークが少ない特異なランドマークでは誤対応の発生率が低いことに注目し，信頼度の指標としてランドマークの特異度を用いたランドマーク

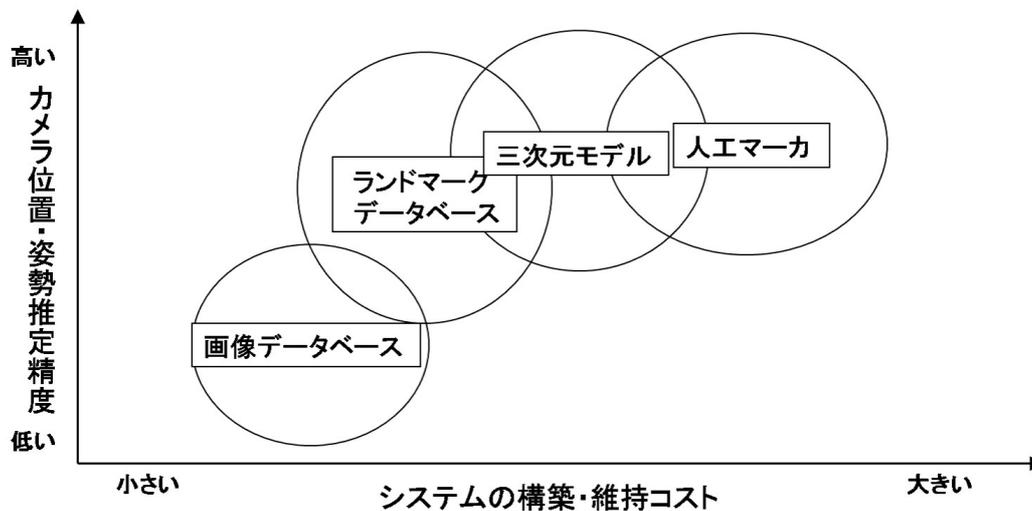


図 5 画像を用いてカメラ位置・姿勢推定する各手法の位置付け

と自然特徴点の対応付けを行う．具体的には，特異度の高いランドマークを優先的に対応点探索に用い，対応付けに成功した特異度の高い基準となるランドマーク(以降，基準ランドマーク)からの空間距離を考慮してランドマークを選択することで，特異度の低いランドマークに対しても誤対応の低減を図るとともに，ランドマークの探索範囲を空間中で限定にすることで処理の高速化を図る．

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(オフライン処理)と，ユーザがシステムを利用する際に行うカメラ位置・姿勢推定処理(オンライン処理)の二段階から構成される．全体の処理の流れを図6に示す．本手法では，文献[1]の手法を基礎とし，特異度を考慮したランドマーク情報の獲得(A-2)を追加し，特異度を考慮した基準ランドマークの選択(B-3)，基準ランドマークからの空間距離を考慮した類似度評価によるランドマークの選択(B-4)の処理を新たに行う．ただし，基本的な処理の流れは手法[1]と同様である．

なお，本研究では，薄らの手法と同様に，図7に示すようなサーバ・クライアント型システムを想定し，サーバ側でカメラの位置・姿勢推定を行うものとする．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

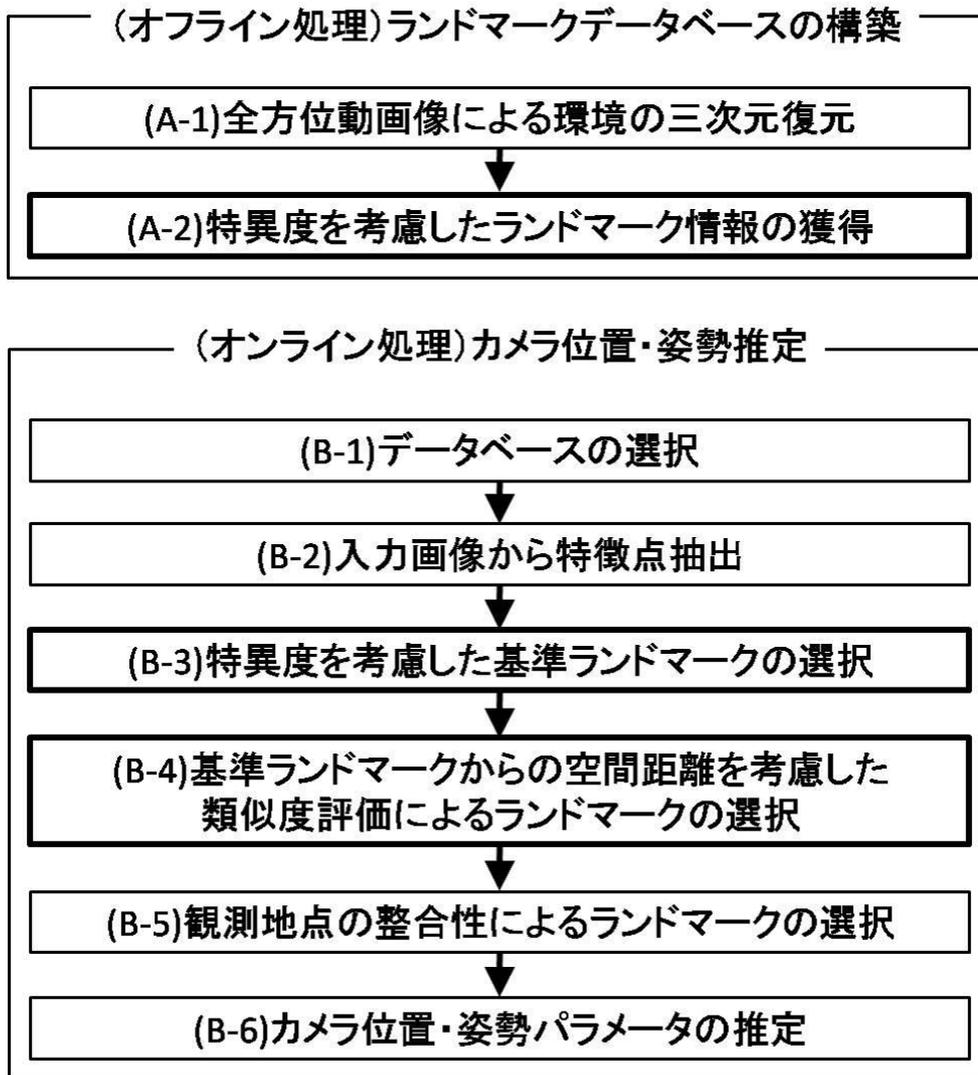


図 6 提案手法の処理の流れ

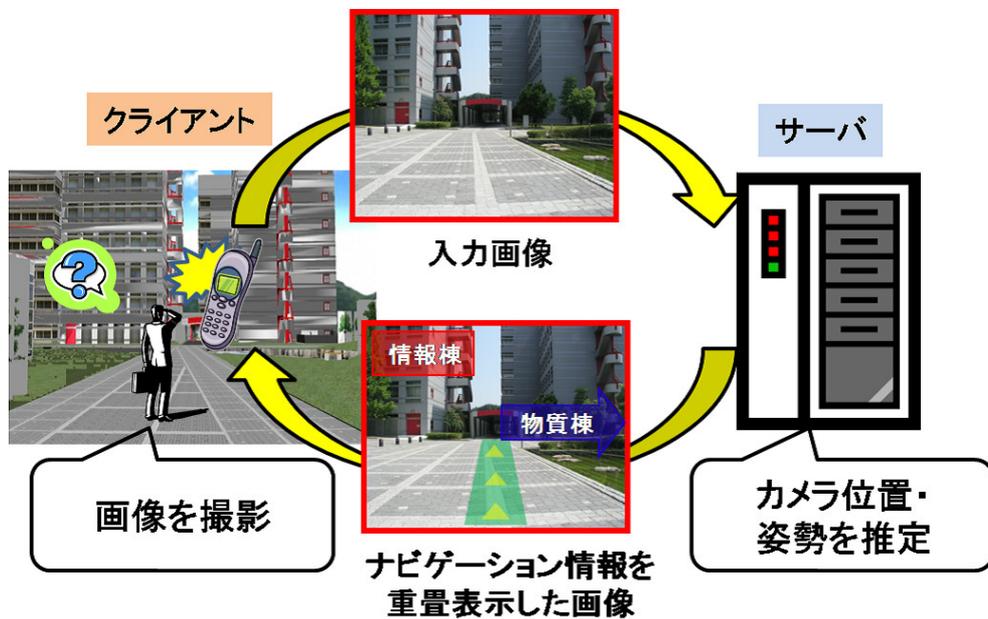


図 7 本研究で想定するサーバ・クライアント型システム

と GPS 情報をネットワークを介してサーバに送信する．サーバは受信した画像・GPS 情報とサーバ上の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 でナビゲーションを重畳表示し，結果をクライアントに送信する．このようにサーバで推定処理を行うことで，計算リソースの乏しい携帯電話・PDA 上においても，画像からのカメラ位置・姿勢推定結果を利用することができる．

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は、図6に示すように、全方位動画像による三次元復元(A-1)、特異度を考慮したランドマーク情報の獲得(A-2)の2つの処理により構成される。全方位動画像による三次元復元処理では、まず対象となる広域環境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を行う。次に、撮影された全方位動画像上の自然特徴点追跡による環境の三次元復元を行い、自然特徴点の三次元位置と同画像撮影時のフレームごとの全方位カメラ位置・姿勢を推定する(A-1)。ランドマーク情報の獲得処理では、環境の三次元復元により推定された自然特徴点の三次元位置、画像情報、特異度を算出し、それらをランドマークとしてデータベースに登録する(A-2)。

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処理(A-1),(A-2)について順に述べる。

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[1]で用いられているデータベースを基礎として、ランドマーク探索における高速化とロバスト性の向上を実現するために各ランドマークに特異度情報を付加し、作成する。ランドマークデータベースの構成要素を図8に示す。ランドマークは、1つの(a)ランドマークの三次元位置、複数の(b)撮影位置ごとの情報、(c)ランドマークの特異度、から成る。撮影地点情報は、(b-1)ランドマーク撮影時の全方位カメラの位置、(b-2)輝度勾配によるランドマークの特徴ベクトル、(b-3)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記述する。

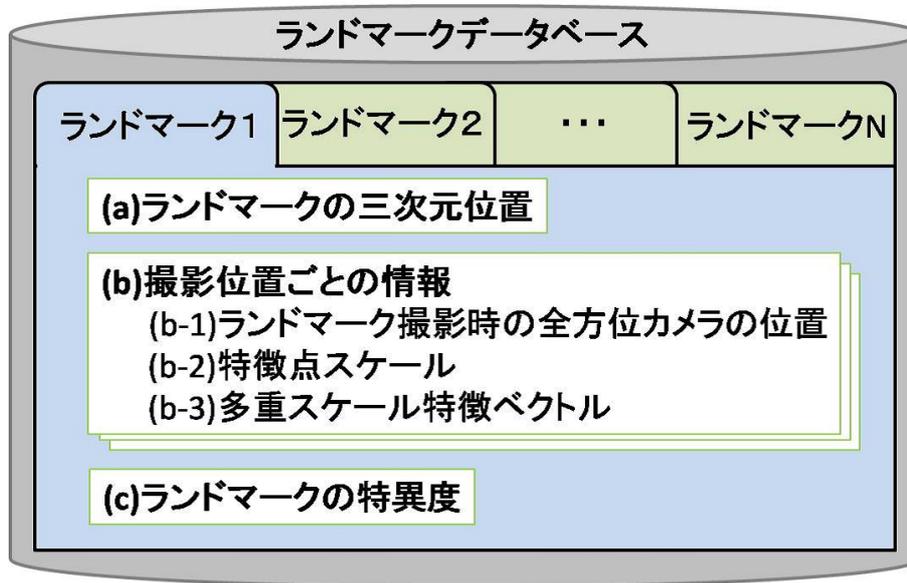


図 8 ランドマークデータベースの構成要素

(a) ランドマークの三次元位置

ランドマークの三次元位置は，対応付けられた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．自然特徴点の三次元位置は，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

ランドマークは撮影されたカメラ位置によって見え方が異なるため，1つのランドマークに対して，データベース構築時の全方位カメラの撮影位置ごとの情報を登録する．撮影地点情報は以下の3つの要素から構成される．

(b-1) ランドマーク撮影時の全方位カメラの位置

4.2 節，4.3 節で述べるランドマークの探索範囲の限定，カメラ位置・姿勢候補の投票に用いる．自然特徴点の三次元位置と同様に 3.2 節で述べる環境の三次元復元によって得られる．

(b-2) 特徴点スケール

撮影画像上の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であり，観測地点の整合性によるランドマークの選択 (B-5) に用いる．この特徴点スケールは，3.3.1 項で述べる Harris-Laplacian による特徴点スケールの算出処理によって得られる．

(b-3) 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[41] による多次元ベクトルを多重スケールで保持し，類似度評価による対応点探索に用いる．これらの特徴ベクトルは3.3.2 項で述べる多重スケール特徴ベクトルの生成によって得られる．

(c) ランドマークの特異度

自然特徴点の特異度は，カメラ位置・姿勢推定時のランドマーク探索における対応付けの信頼度の指標である．特異度の高いランドマークほど対応付けの信頼度が高くなることに注目し，ランドマーク探索の優先度の決定に用いる．特異度は各ランドマークに1つ設定する．特異度の算出方法は3.3.3 節で述べる．

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に登録するランドマークの情報は，利用環境内を全方位カメラで撮影し，structure-from-motion に基づく三次元復元を行うことによって作成する．本研究では佐藤らの手法 [42] を用いて，移動を伴って撮影した全方位動画像上において，三次元位置が既知の少数の基準特徴点と Harris オペレータ [43] 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（カメラの三次元的な軌跡）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，この手法により得られた三次元復元結果より自然特徴点の三次元位置 (図 8(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 ((図 8(b-1)) を取得する．

3.3 特異度を考慮したランドマーク情報の抽出

上述の手法によって三次元位置が推定されたすべての自然特徴点に対して，データベースの構築に必要な画像情報を全方位動画像から取得する．以下，特徴点の固有スケール，多重特徴ベクトル，特異度の算出手法を述べる．

3.3.1 特徴点の固有スケールの算出

本手法では，次章で述べるオンライン処理において，データベースに格納されたランドマークと入力画像中の自然特徴点の対応付けを行うことでカメラの位置・姿勢を推定する．ランドマークと入力画像上の自然特徴点对応付けを正しく行うためには，ランドマークの周辺画像と入力画像の自然特徴点の周辺画像の領域を同範囲とする必要がある．本研究では，自然特徴点の周辺画像の切り出し領域を空間的に一意に定めるために特徴点の固有スケールを算出する．

まず，カメラの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，レンズ歪みを補正した上で，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[44] 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 [44] 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

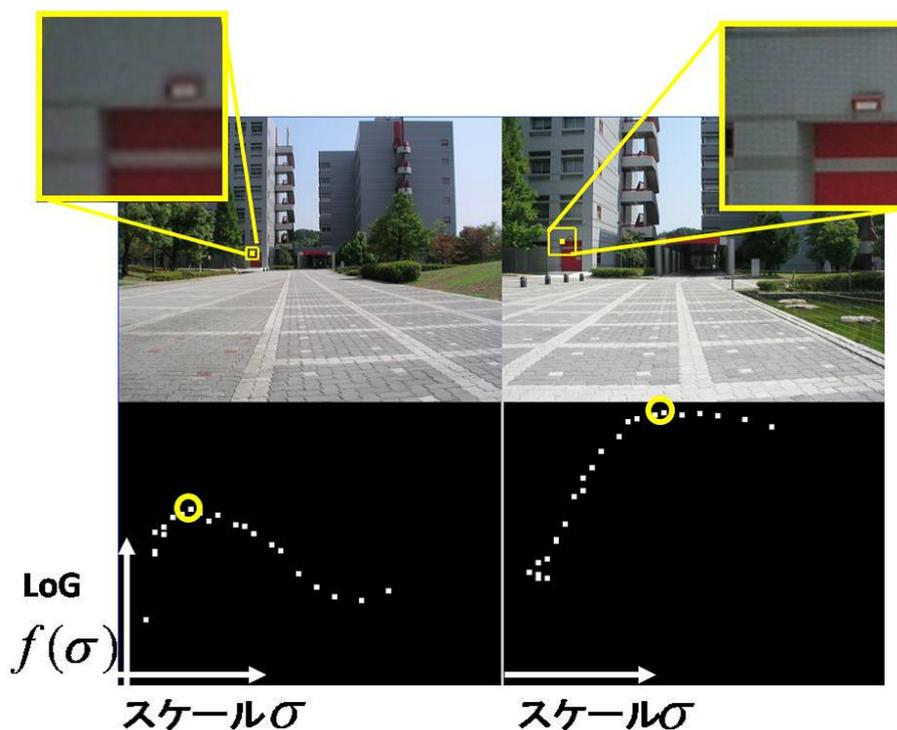


図 9 Harris-Laplacian 適用例

とで行う。

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

ここで、スケール σ は Gaussian フィルタの半径を表しており、 s 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、特徴点までの距離の違いにより画像スケールが異なる場合でも図 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 によって一意に決定することができる ((図 8(b-2)))。

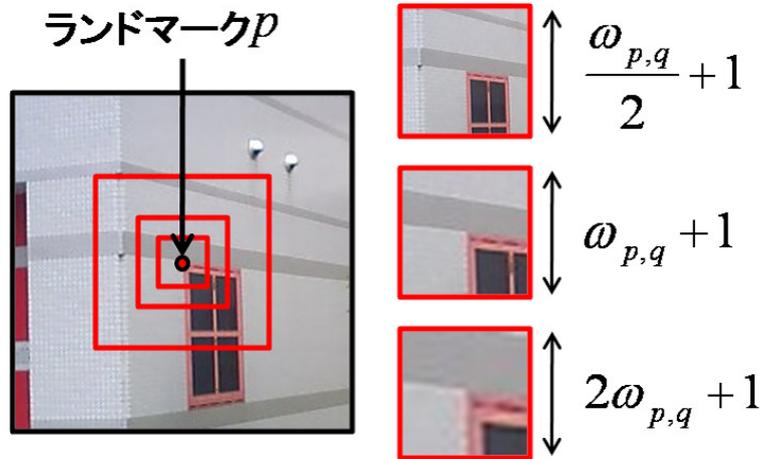


図 10 多重スケール特徴ベクトル生成領域の例

3.3.2 多重スケール特徴ベクトルの生成

ここでは，前項で算出された特徴点の固有スケールを用いて，SIFT[41] による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ごとに生成する．

1. 生成領域の決定

ランダムマーク p の q 番目の撮影地点情報における特徴点の固有スケールを $\omega_{p,q}$ とすると，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の生成領域とする．

2. 特徴ベクトルの生成

SIFT-descriptor[41] 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．生成手順を図 11 に示す．SIFT-descriptor では，回転不変なマッチングを可能にするために，まず手順 1 で求めた輝度勾配ベクトルの生成領域に対して 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のピークに対応する輝度勾配方向を求め，特徴ベクトルの生成領域に対応する画像パターンを輝度勾配のピーク方向に

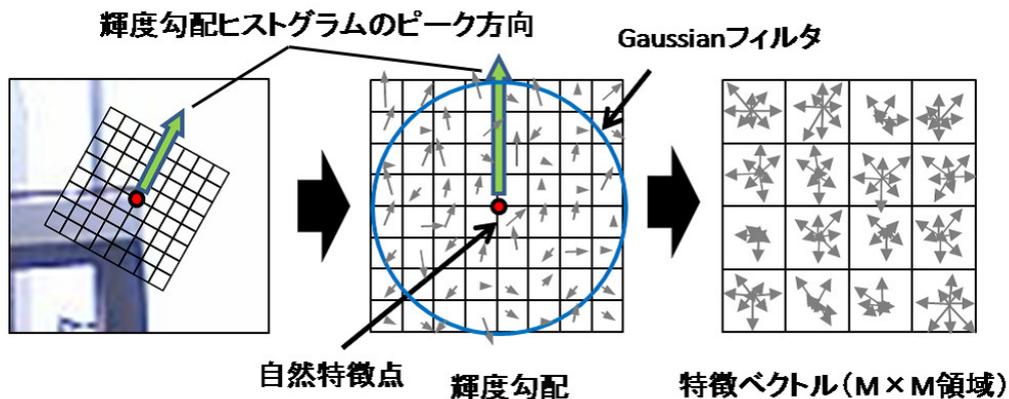


図 11 特徴ベクトルの生成手順

回転させる．さらに，画像を $M \times M$ の領域に分割し，分割領域ごとに輝度勾配を 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し，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 $D = M \times M \times 8$ 次元の特徴ベクトル $f = (v_1, \dots, v_D)$ を生成する．

3.3.3 特異度の算出

4章で述べるカメラの位置・姿勢推定において，効率的にランドマークの照合を行うため，構築するデータベースのランドマークに対応付けの信頼度の指標である特異度を設定する．オンライン処理におけるランドマークの照合では，データベース中のランドマークと入力画像中の自然特徴点の特徴量を比較することで照合を行うが，類似したランドマークが環境に複数存在する場合，それらに対応する特徴量も類似したものとなり，誤対応が発生しやすい．逆に，図 12 に示すような，環境内に類似したランドマークを持たない特異なランドマークは誤対応が発生しにくい．そこで，次章で述べるオンラインでのランドマークの照合においては，特異度の高いランドマークを優先的に照合に用いて誤対応の少ない対応付けを行う．また，対応づいた特異度の高いランドマークを，特異度の低いランドマークの絞り込みの基準としても利用する．

本研究では，ランドマーク p に関する特異度 E_p を，ランドマーク p と他のす

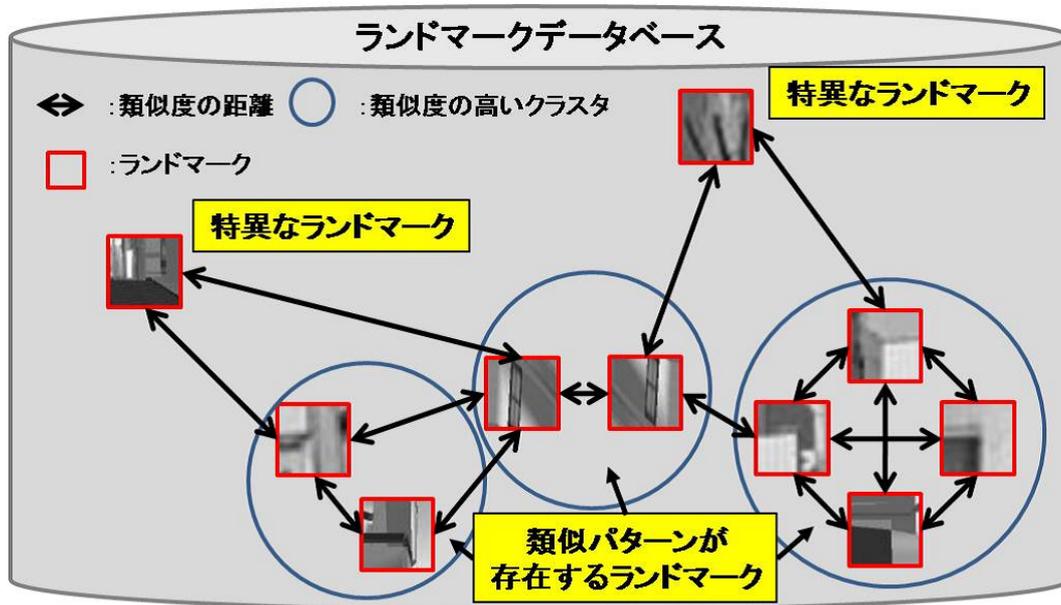


図 12 特異なランドマークの概念図

すべてのランドマークの特徴空間上での距離が短い上位 N 点の平均距離として以下の式で定義する．

$$E_p = \frac{1}{N} \sum_{k=1}^N |f_p - f_{h_{pk}}| \quad (2)$$

ここで、 f_p はランドマーク p に関する D 次元特徴ベクトルを表し、 h_{pk} は p 以外のすべてのランドマークを特徴空間上でランドマーク p に近い順に並べたインデックスリストにおける k 番目の要素を表す．式 (2) に用いるランドマークのスケールはランドマークの中間スケールとする．また、ランドマーク p に関する特徴ベクトルは、撮影地点ごとに存在するため、本研究では、ランドマーク p が撮影された複数フレームに対する中央フレームにおけるランドマークの見え方をランドマーク p の代表的な見え方とし、中央フレームに対する特徴ベクトルを特徴空間上での距離の算出に用いる．なお、特異度 E_p は、値が高いほど環境内で特異なランドマークであることを表す．

4. 特異度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、一枚の静止画像からカメラの位置・姿勢を推定する方法について述べる。なお、カメラ位置・姿勢推定処理の手順は文献 [1] と同様であるが、本研究では、対応点探索処理 (図 6(B-3),(B-4)) において 3.3.3 項で述べた特異度を考慮したランドマークの照合を新たに行う。また、観測地点の整合性によるランドマークの選択処理 (図 6(B-5)) において、推定精度の向上のためにランドマークの再投影処理を追加する。

以下では、前章で述べた手法により広域環境に対するランドマークデータベースが構築されていることを前提に、オンラインでのカメラ位置・姿勢推定の手順を説明する。提案手法では、まず GPS の情報を用いて使用するランドマークデータベースの範囲を限定する (B-1)。次に、入力された静止画像から検出された画像上の自然特徴点の固有スケールを算出し、算出された特徴点の固有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)。特異度が高いランドマークから順に、入力画像上の自然特徴点と照合を行い、正しく対応付く可能性の高い、基準ランドマークを選択する (B-3)。次に、処理 (B-3) で得られた基準ランドマークから三次元空間上で一定の空間距離内に存在のランドマークをデータベースから取り出し、新たに取り出したランドマークと入力画像の自然特徴点の対応付けを行う (B-4)。そして、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・姿勢候補を決定し、推定したカメラ位置・姿勢候補を用いて、撮影可能なランドマークを入力画像に再投影し、カメラ位置・姿勢推定に用いるランドマークを最終的に決定する (B-5)。最後に、処理 (B-5) で対応付けられたランドマークと入力画像上の特徴点の対応点の組を複数用いてを用いてカメラ位置・姿勢推定を行う (B-6)。以下では、各処理について詳しく述べる。

なお、本研究では利用者の携帯端末に取り付けられている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 データベースの選択

携帯機器に内蔵された GPS から位置情報を取得し、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 (B-1)。この処理では、まず GPS 情報とランドマークの撮影地点を基に数十 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 であると仮定し、GPS による計測位置と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 特異度を用いたランドマークの照合

本節では、3.3.3 項で定義し、算出した特異度を用いてランドマークの照合を行う手法について述べる。まず、特異度の高いランドマークを優先的に対応付けに用いることで、対応付けの信頼度の高い少数のランドマークと入力画像上の特徴点の組を決定する。この処理で対応付けられた少数のランドマークを基準ランドマークとし、基準ランドマークに対して三次元空間中で空間距離が近いランドマークをデータベースから取り出し、再度ランドマークの照合を行う。特異度、空間距離を考慮したランドマークの絞り込みにより、従来研究 [1] で行われていた総当たりによるランドマークの照合における誤対応の低減、および処理速度の向上を図る。以下、4.2.1 項で特異度の高いランドマークに対する対応点 (基準ランドマーク) の決定手法について述べ、4.2.2 項で、4.2.1 項で決定した基準ランドマークを用いた空間距離を考慮したランドマークの探索手法について述べる。

4.2.1 特異度を用いた基準ランドマークの選択

ここでは、3.3.3 項で定義したランドマークの特異度を用いてランドマークの照合を行い、後述の処理 (B-4) において、ランドマークの探索範囲を空間的に限定

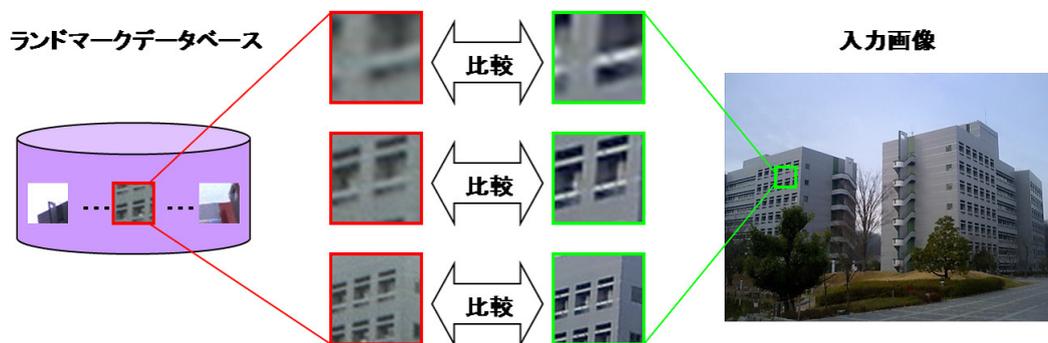


図 13 ランドマークと画像上特徴点の比較例

する際に指標として用いる基準ランドマークの決定手法 (B-2) について述べる . 特異度を用いた基準ランドマークの選択は以下の手順で行う .

- (1) 現在までに選択されていない最も特異度の高いランドマーク p をデータベースから選択する .
- (2) 処理 (1) で選択されたランドマーク p と入力画像上の全ての特徴点を照合する .

ランドマークの照合には , 選択されたランドマーク p の i 番目の撮影地点の特徴ベクトル $\mathbf{f}_{\text{LM}i} = (v_{\text{LM}i1}, \dots, v_{\text{LM}iD})$ と入力画像上の自然特徴点 j から得られた特徴ベクトル $\mathbf{f}_{\text{IN}j} = (v_{\text{IN}j1}, \dots, v_{\text{IN}jD})$ の特徴空間上での二乗誤差 $S(i, j)$ を用いる .

$$S(i, j) = |\mathbf{f}_{\text{LM}i} - \mathbf{f}_{\text{IN}j}|^2 = \sum_{d=1}^D (v_{\text{LM}id} - v_{\text{IN}jd})^2 \quad (3)$$

ここでは全ての (i, j) の組に対して $S(i, j)$ を最小とする組 (i_m, j_m) を選択する . 次に , 二乗誤差 $S(i_m, j_m)$ が閾値以下であればランドマーク p と自然特徴点 j_m を対応点とし , そうでなければ処理 (1) に戻る .

- (3) 処理 (2) で対応付いた特異なランドマーク p を撮影可能なカメラ位置 V_p を算出する .

データベースに登録されているランドマーク p の三次元位置 Q_p , 撮影位置 C_p , 固有スケール ω_p , 自然特徴点 j_m の固有スケール ω_{shoot} から , 図 14 に示

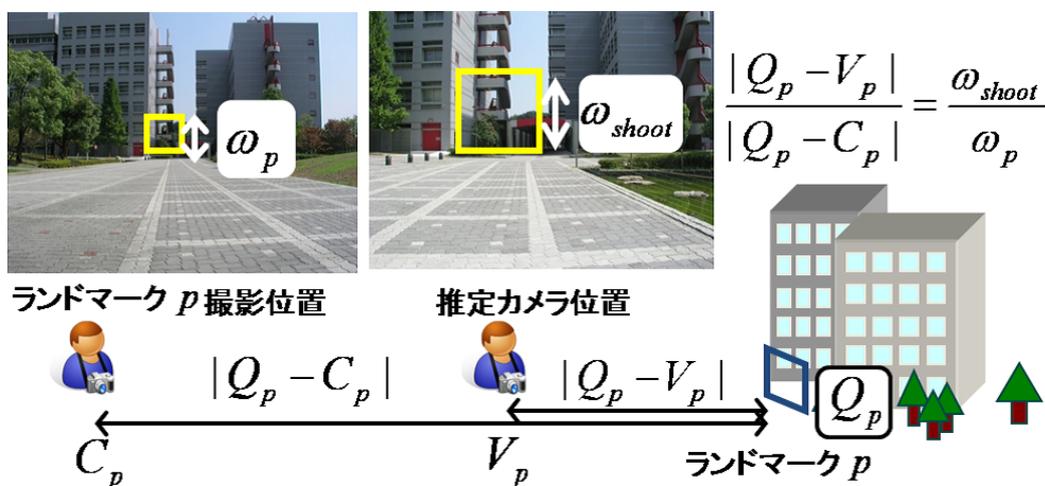


図 14 ランドマークの固有スケールと撮影距離の関係

すランドマークの固有スケールと撮影距離の関係を用いて V_p を算出する .

$$V_p = Q_p - \frac{\omega_p}{\omega_{shoot}}(Q_p - C_p) \quad (4)$$

- (4) 現在までに処理 (3) で算出された p 以外のランドマークに対するすべての $V_n(n \neq p)$ について, $|V_p - V_n| < K$ (K は閾値) を満たす V_n の個数をカウントする .
- (5) 処理 (4) でカウントした個数が閾値 J 以上であれば, $|V_p - V_n| < K$ を満たす全てのランドマークを基準ランドマークとし, 処理を終了する . それ以外の場合には処理 (1) に戻る . ただし, 一定個数の特異なランドマークを用いても, カウントが J に満たない場合には本処理を中断し, 従来手法と同様に総当たりでランドマークと自然特徴点の対応付け処理を行う .

4.2.2 空間距離を考慮したランドマークの選択

ここでは, 4.2.1 節の手順で産出された基準ランドマークからの空間距離を考慮して少数のランドマークを取り出し, 入力画像上の特徴点と対応付ける . 具体的

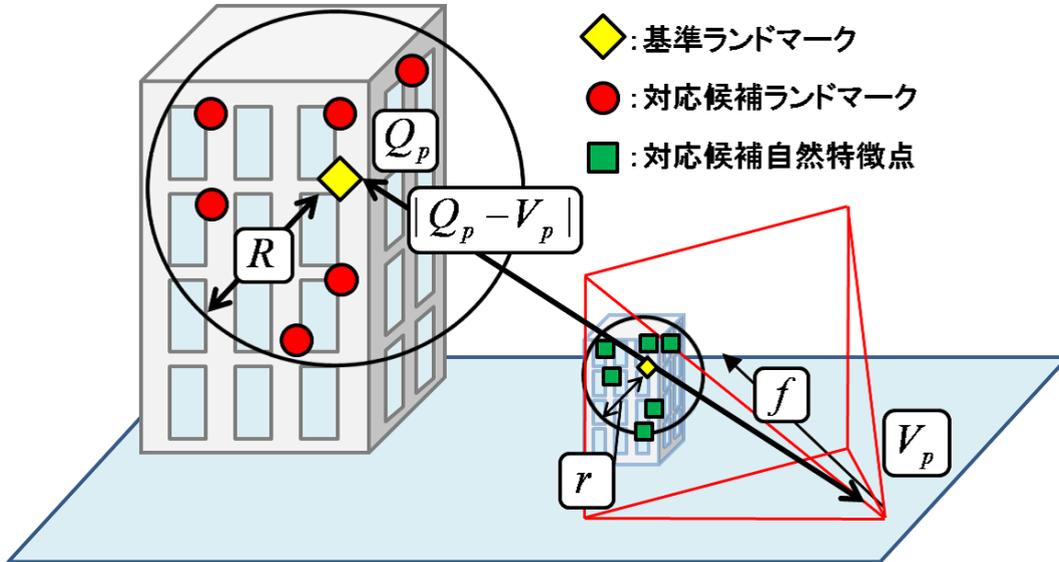


図 15 空間距離を考慮した対応候補ランドマークの選択

には，図 4.2.2 に示すような，基準ランドマーク p の三次元位置 Q_p を中心とする半径 R の球状範囲内に存在するランドマーク群と，基準ランドマーク p と対応づいた画像上の自然特徴点を中心とする半径 r の炎上範囲内に存在する自然特徴点群を取り出し，取り出したランドマークと入力画像上の自然特徴点を 4.2.1 項で述べた処理 (2) と同様の方法で照合する．これにより，画像上の各特徴点との類似度が高い複数のランドマークをデータベース中から選択する．ただし，ここで半径 r は経験的に与えるものとし，空間範囲 R は次式によって決定する．

$$R = \frac{r}{f} |Q_p - V_p| \quad (5)$$

ただし， f はカメラの焦点距離を表す．

4.3 観測地点の整合性によるランドマークの選択

4.2.1 節では基準ランドマークを用いた投票を行ったが，用いた対応付けの組み合わせは数点であり，高精度にカメラ位置・姿勢は求められない．そこで本節では，4.2.2 節で空間距離を考慮して対応付けられた入力特徴点とランドマークの組

- 入力画像の特徴点に類似したランドマーク
- × ランドマーク撮影時の全方位カメラ位置

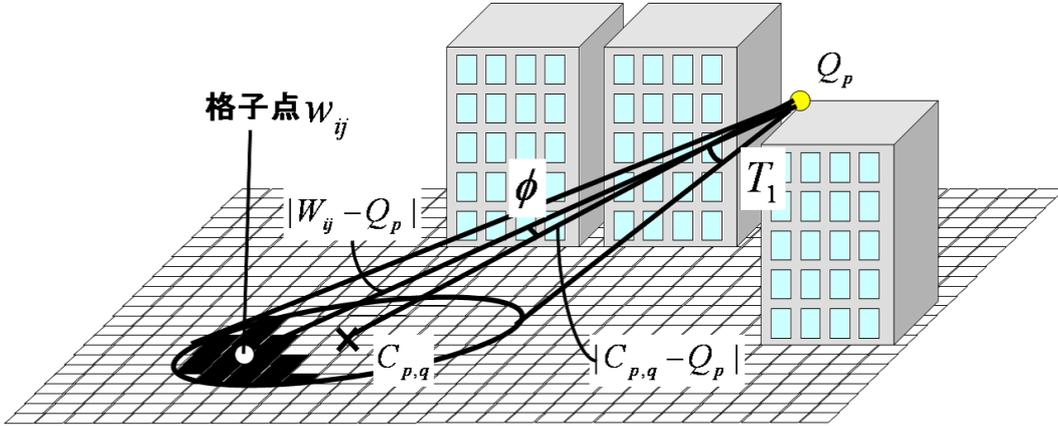


図 16 1つのランドマークからの投票例

み合わせを用いてカメラの位置・姿勢投票を行うことで高精度なカメラ位置・姿勢候補を決定する．また，多数の対応付けの組み合わせを用いることで，さらなる誤対応の低減を図る．

具体的には，図 16 に示すように，GPS 情報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，それぞれの行毎に水平方向の姿勢の回転する l 個（360 度を $360/l$ 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 4.2 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．次に，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 $w_{ij} (-n \leq i \leq n, -n \leq j \leq n)$ を以下のように定義する．

$$w_{ij} = \begin{pmatrix} w_i \\ w_j \end{pmatrix} = \begin{pmatrix} g_x + L \times i \\ g_y + L \times j \end{pmatrix} \quad (6)$$

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 4.2 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 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，

$$(1 - \alpha) \frac{\omega_{LM}}{\omega_{IN}} \leq \frac{|\mathbf{W}_{ij} - \mathbf{Q}_p|}{|\mathbf{C}_{p,q} - \mathbf{Q}_p|} \leq (1 + \alpha) \frac{\omega_{LM}}{\omega_{IN}} .$$

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．

$$m = \left\lceil \frac{\theta l}{2\pi} \right\rceil \quad (7)$$

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．以上の処理により得られた投票結果に対して，投票数が最大となる位置・姿勢に投票したランドマークを抽出し，次節で述べるカメラ位置・姿勢の推定処理に用いる．

4.4 選択されたランドマークを用いたカメラ位置・姿勢の推定

前節で述べた手法により算出されたカメラの位置・姿勢候補および投票結果を用いてカメラの位置・姿勢推定を行う．ここでは，ランドマークの三次元位置と入力特徴点の画像上での二次元位置の組を複数用いて PnP 問題 [45] を解くことでカメラの位置・姿勢を推定する．提案手法では，まず前節で述べた 3 自由度のカメラ位置・姿勢候補に対して，投票に用いられたランドマークと画像上の特徴点の組から 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の位置・姿勢を求める。以下では各ステップの処理について述べる。

4.4.1 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・姿勢の推定

カメラの位置・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の位置・姿勢を求める手法について述べる。まず、4.3節で述べた手法により得られたカメラの位置・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の位置・姿勢を推定する。ただし、投票されるランドマーク中には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[46]を用いて誤対応を排除する。その後、図17に示すように、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 e が最少となるカメラ位置・姿勢を求める。

$$e = \frac{1}{g} \sum_{i=1}^g \{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2\} \quad (8)$$

ただし、 g 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 e が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。

- 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

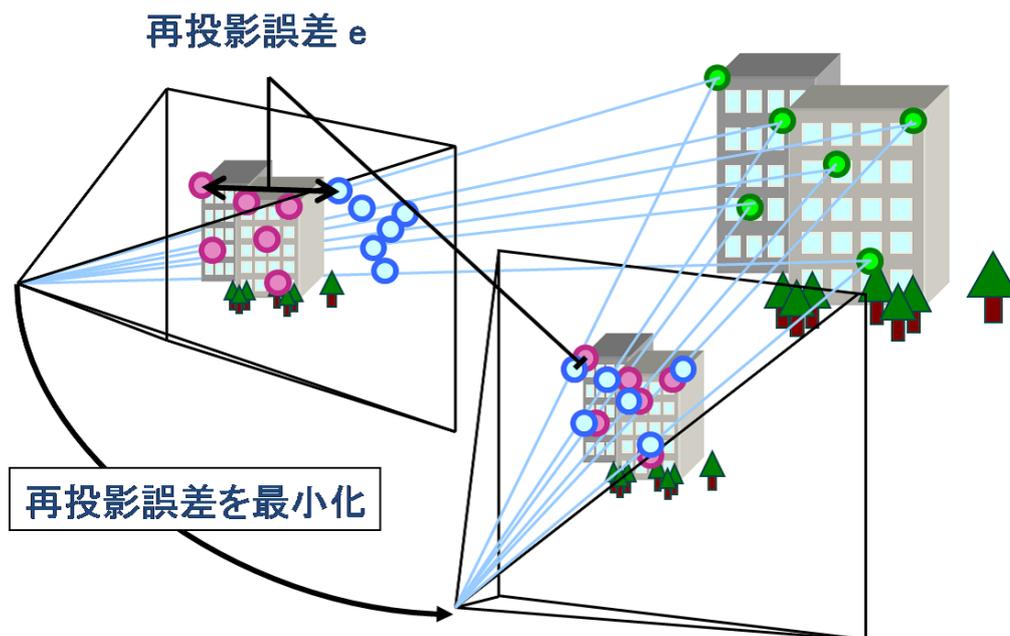


図 17 再投影誤差の最小化によるカメラ位置・姿勢推定

4.4.2 暫定カメラ位置・姿勢に基づくカメラ位置・姿勢の再推定

前項の処理で求められたカメラの位置・姿勢に基づいて、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・姿勢の推定の高精度化を図る。以下に、本研究における再推定処理の手順を示す。

(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。

以上の手順によって最終的なカメラ位置・姿勢を算出する。

5. 実験

屋外環境においてランドマークデータベースを構築し，提案手法によるカメラ位置・姿勢の推定の精度，推定成功率，処理速度を従来手法 [1] と比較することで，ロバスト性と処理速度の向上効果について検証した．

5.1 実験条件

本実験では，文献 [1] で入力として用いられた全方位動画像を用いて特異度を付加したランドマークデータベースを新たに構築し用いた．この全方位動画像は，図 18 に示す全方位型マルチカメラシステム (Point Grey Research 社 Ladybug) を用いて屋外環境を移動しながら取得した．図 18 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，図 19 にオフライン処理で復元されたランドマークの三次元位置と全方位カメラの位置の関係を，地表に対する上面図として示す．同図中の太い実線は全方位カメラのカメラパスを，点群はランドマークの位置を表している．本実験では，データベース構築時にカメラパスが交差するように撮影を行い，そのパスの全長は約 80m であった．また，データベースには約 12,500 個のランドマークが登録され，各ランドマークに対して平均 7.5 地点の撮影地点ごとの情報が登録された．

構築したランドマークデータベースに基づき，表 1 に示すパラメータを用いて，カメラの位置・姿勢を推定した．本実験では，図 20 に示すようなカメラパス周辺における 5m の格子点上 ($6 \times 6 = 36$ 地点) で，矢印 1,2 の方向を向いてそれぞれ撮影した静止画像 (解像度 640×480 画素) を入力画像として与えた．また，推定精度の検証を行うため，入力画像内の特徴点に対応するランドマークの三次元位置をトータルステーションで測定し，その画像上での座標を手動で与えた上で，それらを用いてカメラ位置・姿勢を推定することで真値を作成した．ただし，72 枚中 6 枚の画像についてはトータルステーションによって十分な数の三次元位置を計測できず，真値が得られないため，本実験では真値を取得可能な 66 枚の入力画像に対してカメラの位置・姿勢を推定し評価に用いた．なお，本実験ではサーバ・クライアントシステムの構築は行わず，カメラ内部パラメータをあらかじめ，Tsai

表 1 カメラ位置・姿勢推定に用いたパラメータ

再推定に用いる再投影誤差の閾値 d	5
輝度勾配ベクトルの次元数 D	128
尤もらしさを判定する再投影誤差の閾値 e [画素]	6
基準ランドマークの選択における閾値 J	6
基準ランドマークの選択における閾値 K [m]	5.0
投票処理における各格子のサイズ [m]	0.5
位置投票用の格子の分割数 n	160
姿勢投票用の格子の分割数 l	72
処理 (A-2) における特徴点の抽出数 N	10
処理 (B-4) における二次元距離 r [画素]	210
投票領域の絞り込みにおける閾値 T_1 [度]	10
カメラ位置・姿勢再推定における閾値 T_2 [度]	20
投票領域の絞り込みにおける閾値 α	0.2



図 18 全方位型マルチカメラシステム Ladybug と入力画像の一部

の手法 [47] によって校正し，撮影時はカメラの内部パラメータを固定した．さらに，今回は用いたランドマークデータベースの規模が比較的小さいため，GPS の位置情報を用いた絞り込みは行わず，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．

5.2 従来手法との比較によるロバスト性向上効果の確認

本節では，従来手法との比較によって，提案手法による推定成功率，推定位置・姿勢の精度の向上効果について検証する．図 21，図 22 に，4.4 節で述べた基準に基づきシステムが判定した，従来手法 [1] および提案手法に対する撮影地点ごとの推定成功・失敗の結果を示す．図中の青の曲線はデータベース構築時のカメラパスを，赤枠の丸はカメラの位置・姿勢の推定が成功した位置を，×印はカメラの位置・姿勢の推定が失敗した位置を表す．

従来手法と提案手法の推定成功率，推定精度および再投影誤差の平均と標準偏

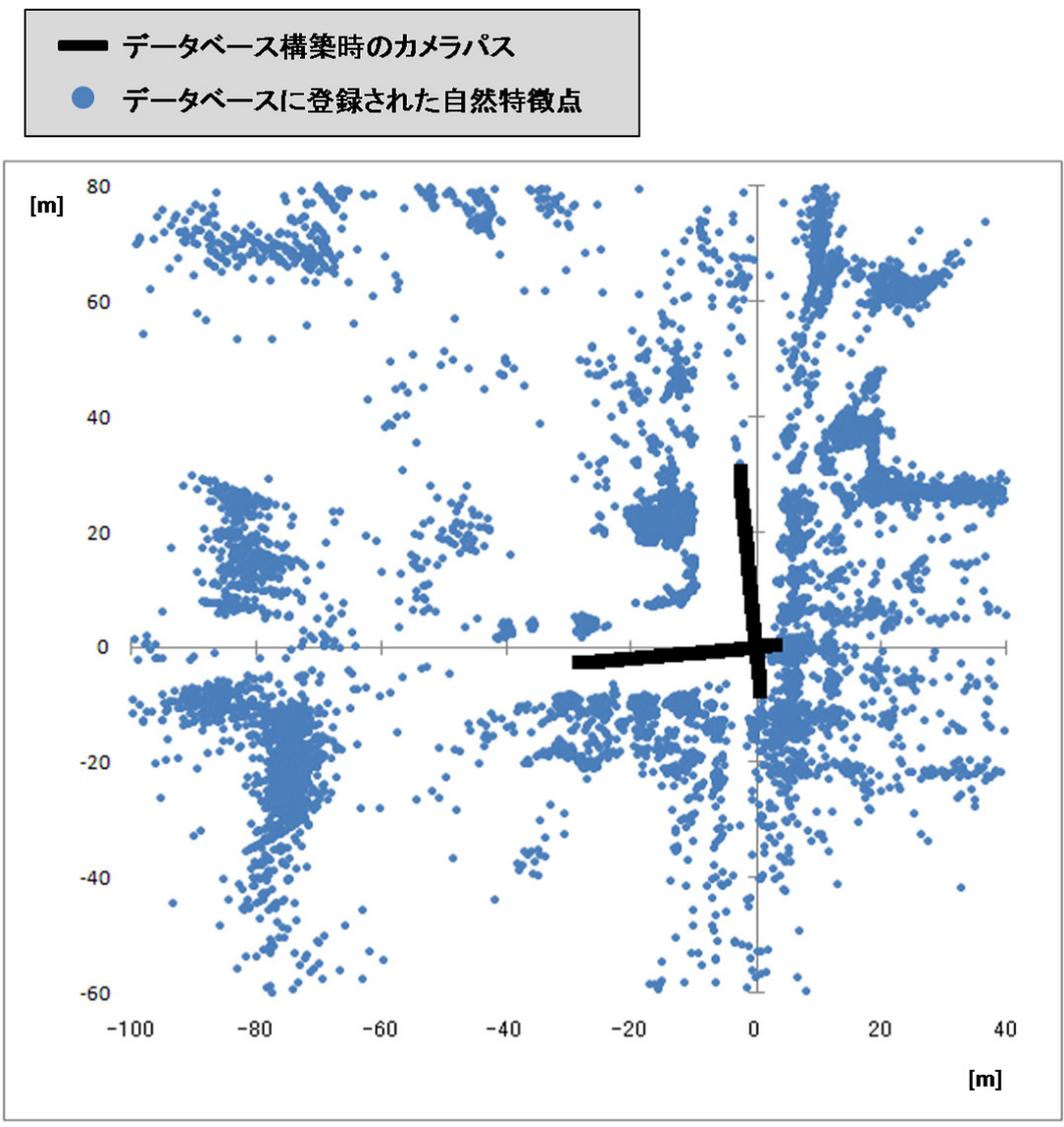


図 19 データベースに登録された自然特徴点と全方位カメラの位置

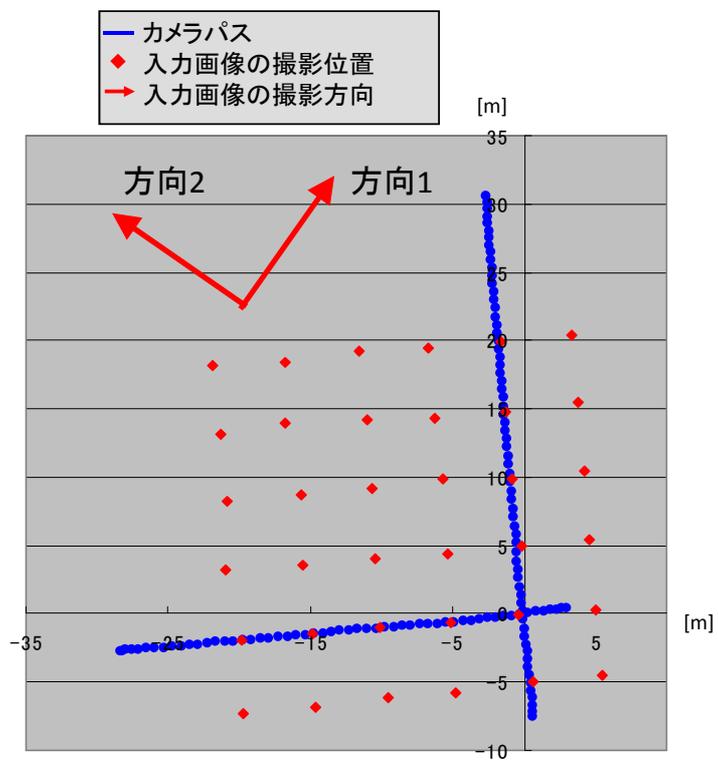
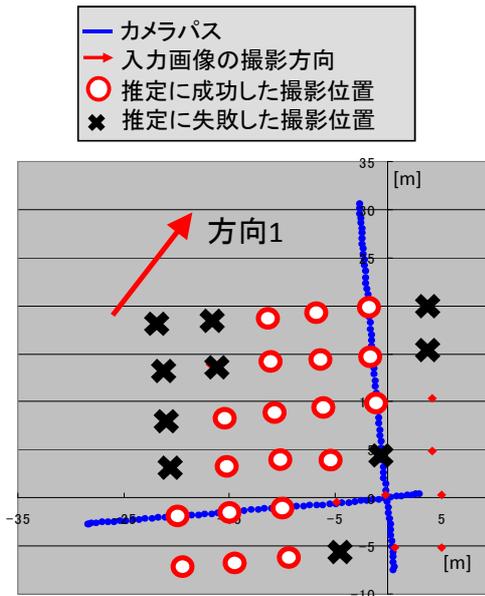
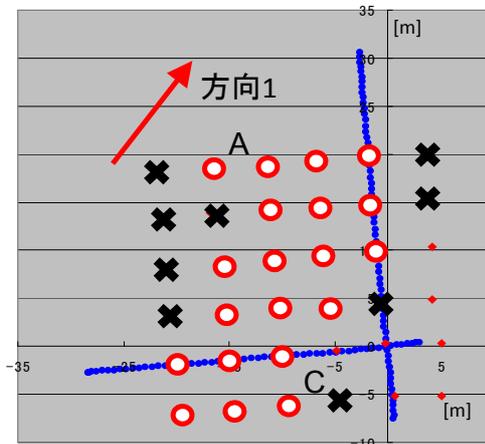


図 20 データベース構築時のカメラパスと入力画像の撮影位置・方向



(a)従来手法の結果



(b)提案手法の結果

図 21 各撮影地点におけるカメラ位置・姿勢の推定成功・失敗の判定結果(方向1)

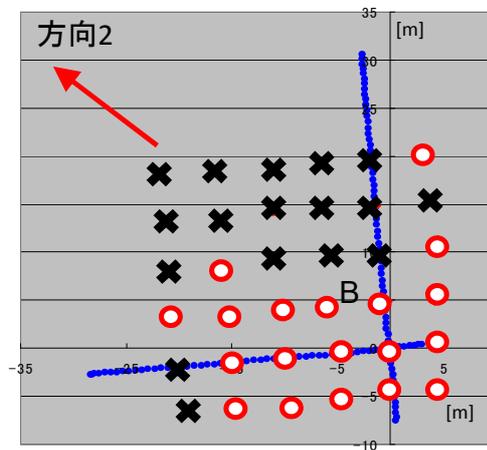
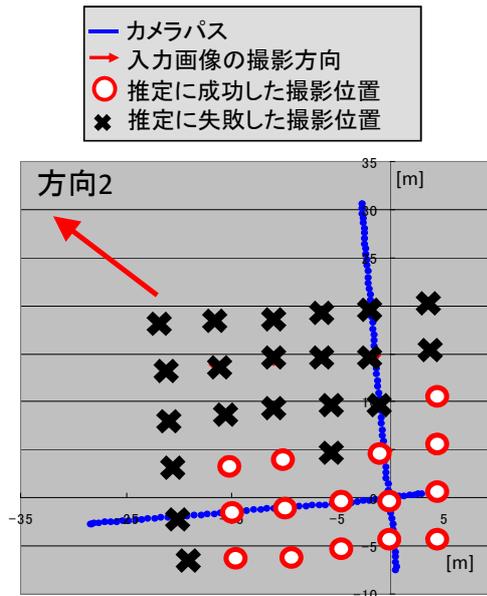


図 22 各撮影地点におけるカメラ位置・姿勢の推定成功・失敗の判定結果(方向2)

差を表 2, 表 3 に示す。同表より, 従来手法と比較して, 提案手法の推定成功率が向上していることがわかる。提案手法では, ランドマークの探索範囲を絞り込むことで誤対応の発生率が低減し, カメラ位置・姿勢を推定可能な地点を増やすことができた。位置・姿勢の平均誤差・標準偏差に関しては, 方向 1 では大きな差は見られなかったが, 方向 2 で改善が見られた。

従来手法で推定が失敗し, 提案手法により推定が成功した地点について考察する。図 21 の A 地点における従来手法および提案手法の投票, 推定結果を図 23 に示す。同図のように従来手法においては, 投票段階において誤った対応付けやスケール算出により, 投票位置が拡散し, 投票結果が真値から大きく離れていることがわかる。提案手法では, 従来手法に比べ投票位置の拡散が抑制され, 真値に近いところへ投票の最大値が表れていることがわかる。また, 図 24 に図 22 の B 地点における対応付けの結果を示す。同図中, 四角は推定に用いられた入力画像上の特徴点を, 丸は各特徴点に対応付けられたランドマークを表す。従来手法では自然物や, 窓枠のような類似したパターンにおいて誤対応が発生している。一方, 提案手法では類似パターンによる誤対応は発生していない。これらのことから, 提案手法によって対応付けの精度が向上したことを確認した。

しかし, 図 25 に表す図 21 の C 地点における対応付けの結果のように, 従来手法と比較して提案手法の対応付けの個数が減少していること入力画像が存在する。これは, 基準ランドマークの位置が検出されなかった画像上の一部の領域において, 空間を考慮した対応付けを行った際に入力画像上の特徴点もしくはランドマークが探索範囲から漏れてしまったためであると考えられる。今後, 入力画像上の基準ランドマークの位置関係を用いて探索範囲のパラメータ r を動的に変更するといった手法の改善が必要である。

表 2 提案手法と従来手法の推定成功率・精度の比較 (方向 1)

手法	従来	提案
推定成功率 (%)	65.5	68.9
平均位置誤差 (m)	0.7	0.7
位置誤差の標準偏差 (m)	0.4	0.5
平均姿勢誤差 (度)	0.9	0.8
姿勢誤差の標準偏差 (度)	0.5	0.5
平均再投影誤差 (画素)	2.3	2.1
再投影誤差の標準偏差 (画素)	0.9	1.0

表 3 提案手法と従来手法の推定成功率・精度の比較 (方向 2)

手法	従来	提案
推定成功率 (%)	41.7	52.7
平均位置誤差 (m)	6.8	2.8
位置誤差の標準偏差 (mm)	9.1	4.3
平均姿勢誤差 (度)	3.9	1.2
姿勢誤差の標準偏差 (度)	4.5	1.9
平均再投影誤差 (画素)	2.0	2.2
再投影誤差の標準偏差 (画素)	1.1	3.9

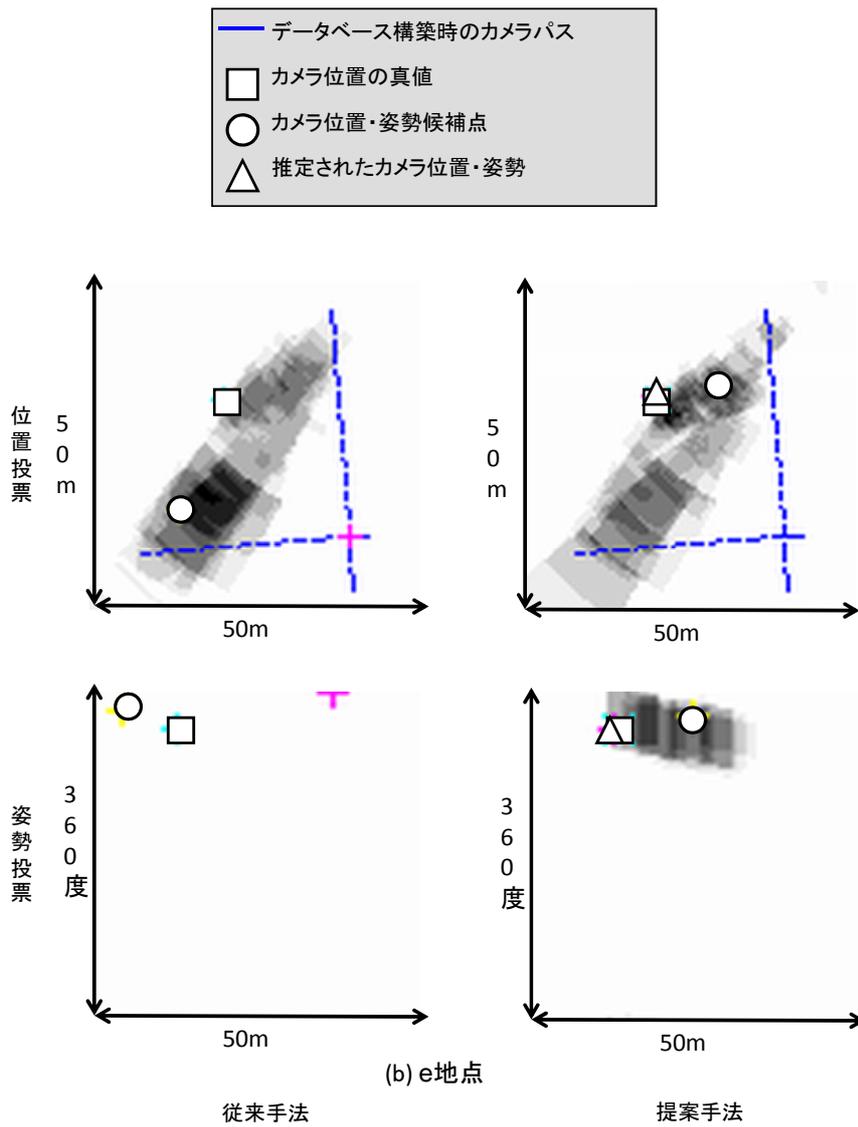
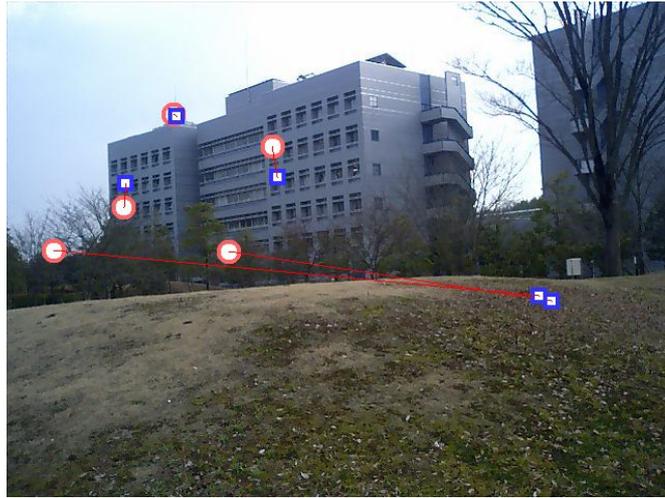


図 23 A 地点における従来手法と提案手法の投票，推定結果の比較



(a)B地点における従来手法による特徴点の対応付けの結果



(b)B地点における提案手法による特徴点の対応付けの結果

図 24 B 地点における従来手法と提案手法の対応付けの比較



(a)C地点における従来手法による特徴点の対応付けの結果



(b)C地点における提案手法による特徴点の対応付けの結果

図 25 C 地点における従来手法と提案手法の対応付けの比較

5.3 従来手法との比較による処理速度向上効果の確認

本節では、従来手法との比較によって提案手法の処理時間の向上効果について確認する。表4に、PC(Xeon 3.0GHz, メモリ 16GB) を用いた場合の1枚の入力画像に対するオンラインでのカメラ位置・姿勢推定の各処理に対する処理時間を示す。本実験ではを用いて検証を行った。表から、従来手法と比較してランドマークの選択および投票に要する時間が短縮されていることが確認できる。提案手法では前処理としての特異度を用いたランドマークの探索範囲を絞り込みに3.8秒かかるが、ランドマークの選択および投票に要する時間が35.7秒となり、ランドマークの対応付け処理を従来手法の約1/3に短縮することができた。全体の処理時間は従来手法が136.5秒、提案手法が57.0秒であり、約2/5の処理時間となっている。以上の結果から、提案手法による大幅な高速化が実現できることが確認できる。

表4 提案手法と従来手法の処理時間の比較 [秒]

手法	従来	提案
入力画像からの特徴量の抽出 (B-1)	14.5	14.7
特異度による基準ランドマークの選択 (B-2)	-	3.8
ランドマークの選択 (B-3) と投票 (B-4)	111.6	35.7
カメラ位置・姿勢推定 (B-5)	10.3	2.8
合計	136.5	57.0

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、対応付けの信頼度としての特異度を考慮して、ランドマークと画像上の特徴点の対応付けにおける探索範囲を絞り込むことで、一枚の静止画像からのカメラ位置・姿勢推定の処理速度を高速化し、ロバスト性を向上させる手法を提案した。提案手法では、まず、オフライン処理として、ランドマークの特異度を考慮したランドマークデータベースを、環境を撮影した全方位動画画像から半自動で作成する。そして、オンライン処理としてカメラの位置・姿勢の推定を行う。推定処理では、まず作成したランドマークデータベース中のランドマークを入力画像上の特徴点と特異度を考慮して対応付け、対応付けの基準ランドマークを選択する。次に、基準ランドマークの三次元座標を中心として、空間距離を考慮して基準ランドマーク周辺に存在するランドマークを入力画像上の特徴点と対応付け、投票を行うことで尤もらしい組み合わせを選択する。最後に、このように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の位置・姿勢推定を行う。提案手法では、画像上の特徴点と対応付けるランドマークを絞り込むことで、誤対応を低減し、投票における精度向上・処理の高速化を行う。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特異度を考慮し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。入力画像に自然物が写り込んでいても従来手法と比較し、投票の精度を向上することに成功し、推定成功率が向上することを確認した。ただし、最終的なカメラ位置・姿勢の推定結果は従来手法と同等であった。また、ランドマークと自然特徴点の対応付けの処理時間が約 $1/3$ に、全体処理を約 $2/5$ に短縮され、高速向上効果を確認した。

今後の課題として、さらなる精度の向上・高速化が挙げられる。手法の改善案としては、空間距離を考慮したランドマークの選択において、空間距離の考慮を入力画像上の各特徴点に合わせて行うことが考えられる。高速化により、実際のサーバ・クライアント型システムにおける応答時間を短縮し、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の実現が可能となる。

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，御鞭撻頂いたロボティクス講座小笠原 司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准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教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教に心より感謝致します．特に，佐藤智和助教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．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高橋 美央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] 薄充孝, 佐藤智和, 横矢直和: “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撮影時のカメラ位置・姿勢推定”, 画像の認識・理解シンポジウム講演論文集, pp. 660–665, 2007.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] 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.
- [11] E. Foxlin: “Pedestrian tracking with dhow-mounted inertial sensors,” IEEE Computer Graphics and Applications, Vo. 25, No. 6, pp. 38–46, 2005.
- [12] 山中 一樹, 神原 誠之, 横矢 直和: ”装着した 3次元磁気センサによる歩き・走り状態に対応したユーザの自己位置推定”, 計測自動制御学会 第 72 回パターン計測部会研究会, pp. 15–20, 2007
- [13] 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.
- [14] 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.
- [15] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. IEEE Int. Symp. on Wearable Computers, pp. 21–23, 2003.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] U. Neumann and S. You: “Natural feature tracking for augmented-reality,” IEEE Transactions on Multimedia, Vo. 1, No. 1, pp. 53–64, 1999.
- [20] A. Davison, Y. G. Cid and N. Kita: “Real-time 3D slam with wide-angle vision,” Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [21] 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.
- [22] 興梠正克, 蔵田武志, 坂上勝彦, 武村治雄: “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. 2, pp. 310–320, 2001
- [23] R. Cipolla, D. Robertson and Tordoff: “Image-based localization”, Proc. Int. Conf. Virtual Systems and Multimedia, pp. 2–29, 2004
- [24] G. Yang J. Becker C. V. Stewart: “Estimating the Location of a Camera with Respect to a 3D Model”, 3DIM, pp. 159–166, 2007
- [25] 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.
- [26] A.I .Comport, E.Maechand, M.Pressigout, F.Chaumette: “Real-Time Markerless Tracking for Augmented Reality: The Virtual Visual Servoing Framework”, IEEE Transactions on visualization and computer graphics, pp. 615–628, 2006

- [27] 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.
- [28] 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.
- [29] L.Wang, S.You, U.Neumann: “Single View Camera Calibration for Augmented Virtual Environments”, IEEE Virtual Reality Conference, pp 255–258, 2007
- [30] I. Skrypnik 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.
- [31] 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
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.

- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 松田 幸大, 池田 聖, 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と姿勢センサに基づく高速な回転にロバストなカメラ位置・姿勢推定”, 画像の認識・理解シンポジウム講演論文集, pp. 1558–1563, 2007
- [41] D.G.Lowe: “Distinctive image features from scale-invariant keypoints”, Int. Journal of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004
- [42] 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.
- [43] C. Harris and M .Stephens: “A combined corner and edge detector”, Proc. Alvey Vision Conf., pp. 147–151, 1998
- [44] K.Mikolajczyk and C.Schmid: “Scale & affine invariant interest point detectors”, Int. J. of Computer Vision, Vol. 60, No. 1, pp. 63–86, 2004
- [45] R. Klette, K. Schluns and A. koschan Eds: “Computer vision: Three-dimensional data from image,” Springer, 1998.

- [46] R. J. Rousseeuw: “Least median of squares regression,” J. American Stat. Assoc, Vol. 79, pp. 871–880, 1984.
- [47] 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.