

NAIST-IS-MT0651063

修士論文

優先度情報を付加したランドマークデータベース
に基づく実時間でのカメラ位置・姿勢推定

武富 貴史

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。

武富 貴史

審査委員：

横矢 直和 教授 (主指導教員)

加藤 博一 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)

優先度情報を付加したランドマークデータベース に基づく実時間でのカメラ位置・姿勢推定*

武富 貴史

内容梗概

近年、CGなどを現実環境を撮影した画像中に位置合わせして描画することで情報を付加する拡張現実感技術に関する研究が盛んである。拡張現実感において、現実環境と仮想環境の位置合わせを行うためには、一般にカメラの位置・姿勢を推定することが必要となり、現在までに様々なカメラ位置・姿勢推定手法が提案されている。このような分野では、一般に、現実環境を撮影した画像をCGの合成対象として用いるだけでなく、カメラ位置・姿勢推定にも用いることで、精度の高い位置合わせを実現している。これらの手法の多くは、環境に対する事前知識を用いることで、現実環境に対する仮想物体の合成位置を決定している。中でもランドマークデータベースを用いたカメラ位置・姿勢推定手法は、データベース構築の人的コストが低く、広域な環境を対象とした場合にもカメラ位置・姿勢推定の誤差が累積しないという特長を持つ。しかし、従来手法ではデータベースに登録されている多数のランドマークと入力画像上の自然特徴点を正しく対応付けるために多くの処理時間を必要とし、実時間でのカメラ位置・姿勢の推定が困難である。また、全てのランドマークを対等に扱っているため、自然特徴点との対応付けに失敗しやすいランドマークを何度も選択し、推定処理が失敗するという問題がある。そこで、本研究では連続フレーム間でのランドマークの追跡とランドマークへの優先度情報の付加により、照合すべき自然特徴点数とデータベースから選択するランドマーク数を削減することで、実時間でのカメラ位置・姿勢

* 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-MT0651063, 2008年3月17日.

推定処理を実現する。また，ランドマークに優先度情報を付加することにより，自然物が多く存在し，推定に有効なランドマークが少ない場所において位置・姿勢推定のロバスト性の向上を図る。ランドマークの優先度は，過去に同じ場所でユーザによって撮影された入力に対する推定結果をデータベースへフィードバックすることにより，各ランドマークが利用される確率を算出することで決定する。実験では提案手法の有効性を確認するため，従来手法との比較によって処理速度とロバスト性の向上効果について検証を行った。

キーワード

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

Real-time camera parameter estimation using a feature landmark database with priorities of landmarks*

Takafumi Taketomi

Abstract

In the field of augmented reality, many kinds of image based camera parameter estimation methods have been proposed to achieve accurate geometric registration. To achieve geometric registration between real and virtual worlds, some kinds of databases that contain knowledge of the target environment are usually used. As one of such approaches, a landmark based camera parameter estimation method has been proposed. In this method, database construction cost is lower than other approaches and the method can be used in various environments. However, computational cost is expensive and thus it cannot work in real-time. Furthermore, ineffective landmarks are often selected from the database and it prevents the robust camera parameter estimation. This thesis describes a real-time camera parameter estimation method using a feature landmark database with priorities of landmarks. To achieve real-time camera parameter estimation, priorities are associated with landmarks by using a previously captured video sequences. According to these priorities, unreliable landmarks are efficiently discarded to realize fast and robust camera parameter estimation. To demonstrate the effectiveness of the proposed method, the robustness and the computational cost is compared between the proposed and conventional methods.

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0651063, March 17, 2008.

Keywords:

camera parameter estimation, landmark database, natural features, augmented reality

目次

1. はじめに	1
2. カメラ位置・姿勢推定の従来研究と本研究の位置づけ	4
2.1 センサを用いたカメラ位置・姿勢推定	4
2.2 画像を用いたカメラ位置・姿勢推定	5
2.2.1 SLAMによるカメラ位置・姿勢推定手法	6
2.2.2 マーカを用いた手法	6
2.2.3 画像データベースを用いた手法	8
2.2.4 環境の三次元モデルを用いた手法	8
2.2.5 ランドマークデータベースを用いた手法	10
2.3 センサと画像を用いたハイブリッドなカメラ位置・姿勢推定	10
2.3.1 絶対指標としてセンサを用いる手法	11
2.3.2 絶対指標として画像を用いる手法	12
2.4 本研究の位置づけと方針	13
3. ランドマークデータベースの構築	19
3.1 ランドマークデータベースの構成要素	19
3.2 全方位動画画像を用いた環境の三次元復元	20
3.3 ランドマーク情報の取得	21
3.4 ランドマークの優先度の定義	22
4. カメラ位置・姿勢の推定	23
4.1 フレーム間でのランドマークの追跡による暫定的なカメラ位置・姿勢の推定	23
4.2 優先度を用いたランドマークの選択	24
4.3 ランドマークと自然特徴点の対応付けによるカメラ位置・姿勢推定	25
4.4 推定結果のフィードバックによる優先度の更新	27
5. 実験	28

5.1	ランドマークデータベースの構築	28
5.2	ロバスト性と処理速度の向上効果の検証	29
5.2.1	ロバスト性向上効果の検証	31
5.2.2	処理速度向上効果の検証	35
6.	まとめ	38
	謝辞	39
	参考文献	40
	付録	45
A.	LMedS 基準による誤対応の排除とカメラ位置・姿勢推定	45

目 次

1	超音波センサを用いた位置・姿勢推定 [1]	5
2	景観への影響を抑えたマーカ	7
3	画像データベースを用いたカメラ位置・姿勢推定 [21]	9
4	SIFT 特徴量を用いた三次元復元 [28]	11
5	Reitmayr らの手法 [38] で用いられる三次元モデル	13
6	カメラ位置・姿勢推定手法の位置づけ	14
7	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（オンライン処理）	15
8	撮影姿勢に依存しない画像パターンの作成	16
9	提案手法の処理の流れ	18
10	ランドマークデータベースの構成要素	20
11	多重スケールの画像テンプレート	21
12	全方位型マルチカメラシステムと全方位画像	28
13	環境の三次元復元結果	29
14	推定されたカメラパスと真値	31
15	自然物が多く映り込んでいるフレーム	33
16	ランドマーク数とカメラの推定位置誤差の関係	34
17	ランドマーク数と推定に失敗したフレーム数の関係	35
18	カメラ位置・姿勢推定結果を用いた仮想物体の重畳表示	37

表 目 次

1	カメラ位置・姿勢推定に用いた機器	30
2	実験で用いたパラメータ	30
3	カメラ位置推定精度の比較	32
4	カメラ姿勢推定精度の比較	32
5	処理時間の比較	36

1. はじめに

近年、実世界を撮影した画像中にCGなどによって表現される仮想世界を位置合わせして描画することで情報を付加する拡張現実感技術に関する研究が盛んに行われている。拡張現実感において、CGなどを所定の位置に描画するためには、一般にカメラの位置・姿勢を推定することが必要となる。従来から、カメラの位置・姿勢を推定する手法としてGPSやジャイロなどのセンサを用いた手法[1, 2, 3, 4, 5, 6, 7, 8, 9], カメラで撮影された画像を用いる手法[10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29], GPSやジャイロなどのセンサとカメラ画像を併用したハイブリッド手法[30, 31, 32, 33, 34, 35, 36, 37, 38]が提案されている。

センサを用いた手法[1, 2, 3, 4, 5, 6, 7, 8, 9]は、GPSなどの絶対位置を取得するセンサと加速度センサやジャイロなどの相対位置を取得するセンサを組み合わせることで、屋内や広域な屋外環境において高速でロバストなカメラ位置・姿勢推定を実現している。しかし、センサによって利用可能な環境に制約があり、様々な環境で用いるためには多数のセンサを組み合わせる必要がある。また、センサとカメラのキャリブレーション誤差や計測誤差に起因して、一般に画像を用いた手法に対して高精度な位置合わせが難しいという問題がある。

画像からのカメラ位置・姿勢推定手法[10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]は、カメラから得られる画像をCGなどを重畳表示するための背景として用いるだけでなく、カメラ位置・姿勢の推定にも利用することで実世界と仮想世界の高精度な位置合わせを実現している。画像からのカメラ位置・姿勢推定手法には、環境の事前知識を用いない手法[10, 11, 12]と事前知識を用いる手法[13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]がある。環境の事前知識を用いない手法では、一般に入力画像中の自然特徴点を追跡することにより、カメラ位置・姿勢の推定と環境のモデル化を同時に行うSLAM(Simultaneous Localization and Mapping)と呼ばれるアプローチが用いられる。このアプローチでは、事前にデータベースを構築する必要がなく、手軽に拡張現実感環境を構築できるという特長があるが、広域な環境を対象としたカメラの位置・姿勢推定では、カメラ位置・姿勢の推定誤差が蓄積するという問題が

あり、利用できる範囲が限定される。環境の事前知識を用いる手法では、事前知識として、マーカを用いる手法 [13, 14, 15, 16, 17, 18, 19, 20], 画像データベースを用いる手法 [21, 22, 23], 環境の 3 次元モデルを用いる手法 [24, 25, 26, 27], ランドマークデータベースを用いる手法 [28, 29] がある。マーカを用いる手法 [13, 14, 15, 16, 17, 18, 19, 20] では、あらかじめ環境内に三次元位置が既知のマーカを配置しておき、入力画像中のマーカを識別することによってカメラの位置・姿勢を推定する。マーカを用いる手法は比較的高速かつ高精度にカメラ位置・姿勢の推定を行うことができるが、広域な環境ではマーカを設置するコストが膨大となるという問題がある。画像データベースを用いる手法 [21, 22, 23] では、入力画像とデータベース中の画像を対応付けることでカメラの位置・姿勢を推定しているが、6 自由度で正確なカメラ位置・姿勢を推定することが難しい。環境の三次元モデルを用いる手法 [24, 25, 26, 27] では、入力画像中のエッジや自然特徴点などをあらかじめ作成した三次元モデルと対応付けることによって、実時間でのカメラの位置・姿勢推定を実現している。しかし、一般に広域で複雑な屋外環境の三次元モデルを作成することは難しく、データベースの構築に多くの人的コストを必要とするという問題がある。ランドマークデータベースを用いる手法 [28, 29] では、自然特徴点の三次元位置と自然特徴点周辺の局所的な画像情報をランドマークとしてデータベースへ登録しておき、ランドマークと入力画像中の自然特徴点を対応付けることでカメラ位置・姿勢の推定を行う。自然特徴点の三次元位置は structure-from-motion によって自動で推定されるため、データベース構築の人的コストが低く、複雑な環境を対象とした場合にも比較的容易にデータベースを作成することができる。しかし、データベースに登録されている多数のランドマークと画像上の特徴点を正しく対応付けるために多くの処理時間を必要とし、実時間でのカメラ位置・姿勢の推定が困難であるという問題が残されている。

GPS やジャイロなどのセンサとカメラ画像を併用したハイブリッド手法 [30, 31, 32, 33, 34, 35, 36, 37, 38] は、絶対指標として GPS やコンパス、赤外線ビーコンなどのセンサの計測結果を用い、画像を補助的に利用する手法 [30, 31, 32] と、画像から推定されるカメラ位置・姿勢を絶対指標とし、相対位置・姿勢を計測する加速度センサやジャイロなどのセンサを補助的に利用する手法 [33, 34, 35, 36, 37, 38]

に大別できる。前者は、比較的ロバスト性の高いカメラ位置・姿勢推定を実現できるが、センサによって利用可能な環境に制約があり、様々な環境で用いるためには多数のセンサを組み合わせることが必要となる。後者は、実世界と仮想世界の位置合わせの精度を高めやすいという利点があるが、画像からの絶対位置・姿勢推定の従来手法には、広域で複雑な環境への適用が困難であるという問題やカメラ位置・姿勢推定の処理コストの問題が残されており、解決が必要である。

本研究では、拡張現実感において位置合わせの精度を高めやすい画像からのカメラ位置・姿勢推定手法の中でも、広域で複雑な環境においても比較的 low コストにデータベースを構築することができるランドマークデータベースを用いる手法 [29] に着目し、従来手法 [29] で問題となっていた対応付けの処理コストの問題を解決する手法を提案する。具体的には、連続フレーム間でのランドマークの追跡による暫定的なカメラパラメータの推定とランドマークへの優先度情報の付加により、ランドマークと自然特徴点の照合回数を削減することで高速化を図り、屋内外の広域環境において蓄積誤差を含まない実時間でのカメラ位置・姿勢推定を実現する。

以下、2章では、カメラ位置・姿勢推定の従来研究と本研究の位置づけについて述べる。3章では、優先度情報を付加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たカメラ位置・姿勢推定手法について述べる。5章では、処理速度とロバスト性の向上効果について検証し、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・姿勢推定の従来研究と本研究の位置づけ

本章では、まずカメラ位置・姿勢推定の従来研究を、センサを用いる手法 [1, 2, 3, 4, 5, 6, 7, 8, 9], 画像を用いる手法 [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29], センサと画像を併用したハイブリッド手法 [30, 31, 32, 33, 34, 35, 36, 37, 38] に大別し、それぞれの特徴と問題点について述べる。次に、本研究の位置づけと方針について述べる。

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・姿勢推定手法には、絶対位置・姿勢の計測が可能なインフラを用いる手法 [1, 2, 3] と、絶対位置・姿勢を取得可能なセンサと相対位置・姿勢を取得可能なセンサを組み合わせる手法 [4, 5, 6, 7, 8, 9] がある。

インフラを用いる手法 [1, 2, 3] では、図1のように環境内に超音波センサや赤外線ビーコンなどを設置し、それらを計測することでユーザの絶対位置・姿勢を推定する。これらの手法では、比較的高精度に位置・姿勢を推定することが可能である。しかし、広域な環境でとぎれなく位置・姿勢を推定するためには、インフラを密に設置する必要があり、インフラの設置・維持コストが問題となる。

絶対位置・姿勢を取得可能なセンサと相対位置・姿勢を取得可能なセンサを組み合わせた手法 [4, 5, 6, 7, 8, 9] では、GPS や赤外線ビーコンなどから間欠的に取得される絶対位置・姿勢からの相対位置・姿勢を加速度センサやジャイロなどを用いて計測し、カメラ位置・姿勢推定を行う。神原ら [7] は、高精度に絶対位置を取得できるが計測周期が長い RTK-GPS と計測周期の短い慣性航法センサを組み合わせることで、誤差の累積しない高レートでのカメラ位置・姿勢推定を実現している。また、Foxlin ら [8] や山中ら [9] は、先に述べたインフラを用いる手法の問題点である設置コストを低減するために、インフラを疎に配置し、自律計測によってインフラ間における相対的な運動を推定する手法を提案している。しかし、絶対位置の計測に用いられる GPS や赤外線ビーコンなどのインフラには、それぞれ屋内・屋外での利用が難しいという問題がある。また、拡張現実感での利用を前提とした場合にはセンサとカメラのキャリブレーション誤差や計測誤差

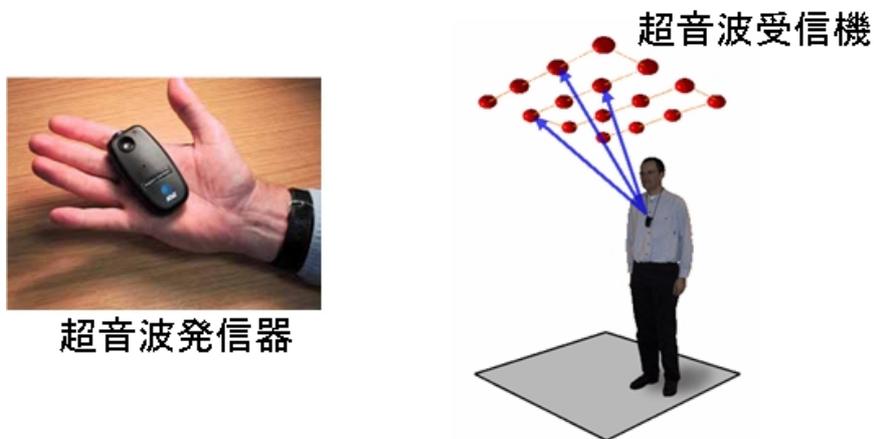


図 1 超音波センサを用いた位置・姿勢推定 [1]

に起因して、画素単位での位置合わせが難しいという問題がある。

2.2 画像を用いたカメラ位置・姿勢推定

画像からのカメラ位置・姿勢推定手法は、センサを用いた手法に比べ、拡張現実感における実世界と仮想世界の位置合わせの精度を高めやすいという利点がある。画像からのカメラ位置・姿勢推定手法は、環境の事前知識を用いずにカメラ位置・姿勢の推定と環境のモデリングを同時に行う SLAM(Simultaneous Localization And Mapping) と呼ばれる手法 [10, 11, 12] と、事前知識としてあらかじめ構築された環境の情報を用いる手法 [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29] に大別できる。以下では、まず SLAM によるカメラ位置・姿勢推定手法について述べる。次に環境の事前知識を用いる手法を、画像データベースを用いる手法 [21, 22, 23]、環境の三次元モデルを用いる手法 [24, 25, 26, 27]、ランドマークデータベースを用いる手法 [28, 29]、に分類しそれぞれの手法の特徴と問題点について述べる。

2.2.1 SLAMによるカメラ位置・姿勢推定手法

SLAMによるカメラ位置・姿勢推定手法 [10, 11, 12] では、入力画像中の自然特徴点を追跡することにより、カメラ位置・姿勢の推定と環境の三次元モデリングを同時に行い、事前知識なしに拡張現実感環境を構築する。

このアプローチでは、一般にカメラ位置・姿勢の推定処理と環境の三次元復元処理を交互に実行するため、計算コストの問題により、実時間処理で多くの自然特徴点を利用することが困難であった [10, 11]。これに対して、Klein ら [12] は、カメラ位置・姿勢推定処理と環境の三次元復元処理を並列化し、非同期に実行することで、環境の三次元復元の最適化処理に割かれる計算時間が特徴点追跡処理のフレームレートに影響することを防いでいる。これにより、多くの自然特徴点を安定に追跡することができ、安定性の高いカメラ位置・姿勢の推定を実現している。しかし、広域な環境を対象としたカメラの位置・姿勢推定では、カメラ位置・姿勢の推定誤差が蓄積するという問題があり、利用できる範囲が限定される。また、SLAMによるカメラ位置・姿勢の推定は、それ単体では現実環境に対する絶対的なカメラ位置・姿勢やスケール情報を取得できないため、拡張現実感技術を用いたナビゲーションのような実環境の位置に依存した情報の提示を必要とするアプリケーションにそのまま用いることはできない。

2.2.2 マーカを用いた手法

マーカを用いた手法 [13, 14, 15, 16, 17, 18, 19, 20] は、パターンや形状などが既知の画像マーカを撮影画像中から抽出することでカメラ位置・姿勢を推定する。画像中のマーカの識別は比較的容易であり、実時間で蓄積誤差を含まないカメラ位置・姿勢の推定が可能である。このような画像マーカを事前に環境内に設置し、その三次元位置をデータベースとして保持しておくことでカメラの位置・姿勢を推定することが可能である [14, 15, 16, 17]。しかし、広域で複雑な環境で利用することを想定した場合には、これらの手法では環境内に多数のマーカを設置する必要があり、マーカの設置に多くの人的コストがかかるという問題や景観を損ねるという問題がある。

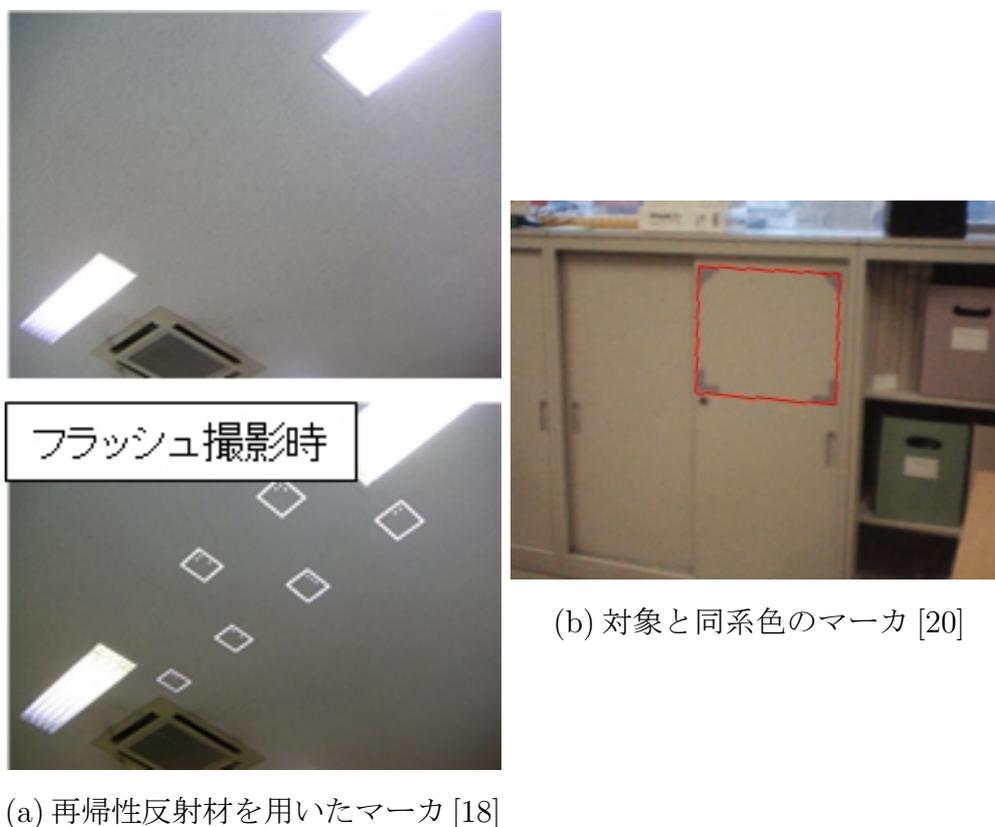


図 2 景観への影響を抑えたマーカ

環境内に多数のマーカを設置することで景観を損ねるという問題に対し，再帰性反射材を用いた不可視マーカ（図 2(a)）を設置し，それを赤外線カメラで撮影することでカメラ位置・姿勢を推定する手法 [18]，壁紙や床面に周りの環境と調和するような位置情報を付加したパターンを用いる手法 [19]，マーカを設置する対象と同系色のマーカ（図 2(b)）を利用する手法 [20] が提案されている．これらの手法を用いることで，景観を損ねるという問題を緩和できるが，広域で複雑な環境ではマーカの設置に多くの人的コストが必要となるという問題が残されている．

2.2.3 画像データベースを用いた手法

画像データベースを用いる手法 [21, 22, 23] は、事前に撮影した画像とその撮影位置・姿勢の情報を保持したデータベースを構築しておき、入力画像と類似した画像をデータベース中から検索することでカメラの位置・姿勢を推定している。岩佐ら [22] や興梠ら [23] は、入力画像と最も類似した画像をデータベース中より選び出し、その画像の撮影位置を入力画像の撮影位置として推定している。これらの手法では、大まかなカメラ位置の推定は可能であるが、正確なカメラ位置を推定することは難しい。

一方、Cipolla ら [21] は、図 3 のように、データベース中の画像を撮影したカメラ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。この手法では、まず画像中のエッジを検出し、それらを統合することで線分を抽出する。次に、抽出された線分より消失点を求めることで画像の垂直化を行い入力画像とデータベース中の画像を対応付ける。最後に、得られた対応関係よりデータベース中の画像を撮影したカメラ位置からの相対的なカメラ位置・姿勢を推定することで、入力画像を撮影したカメラの位置・姿勢を推定している。この手法では撮影地点の高さやカメラのピッチ成分を推定することができず、6 自由度で正確なカメラ位置・姿勢を推定することは難しい。

2.2.4 環境の三次元モデルを用いた手法

環境の三次元モデルを用いる手法 [24, 25, 26, 27] では、入力画像中のエッジや自然特徴点などを、あらかじめ作成した三次元モデルと対応付けることによって、実時間でのカメラの位置・姿勢推定を実現している。Drummond ら [24] は、環境の三次元モデルとしてワイヤースケルトンモデルを用い、入力画像中のエッジとワイヤースケルトンモデルを対応付けることでカメラ位置・姿勢の推定を行っている。この手法では、実時間でのカメラ位置・姿勢推定を実現しているが、入力画像中に多くのエッジが存在する場合に誤対応が起こりやすく、ロバストなカメラ位置・姿勢の推定が困難であるという問題がある。このような問題に対し、Vacchetti ら [26] は、ワイヤースケルトンモデルを用いたカメラ位置・姿勢推定手法と自然特徴点の追跡処理を組み合わせることによって、入力画像中に多くのエッジが存在す

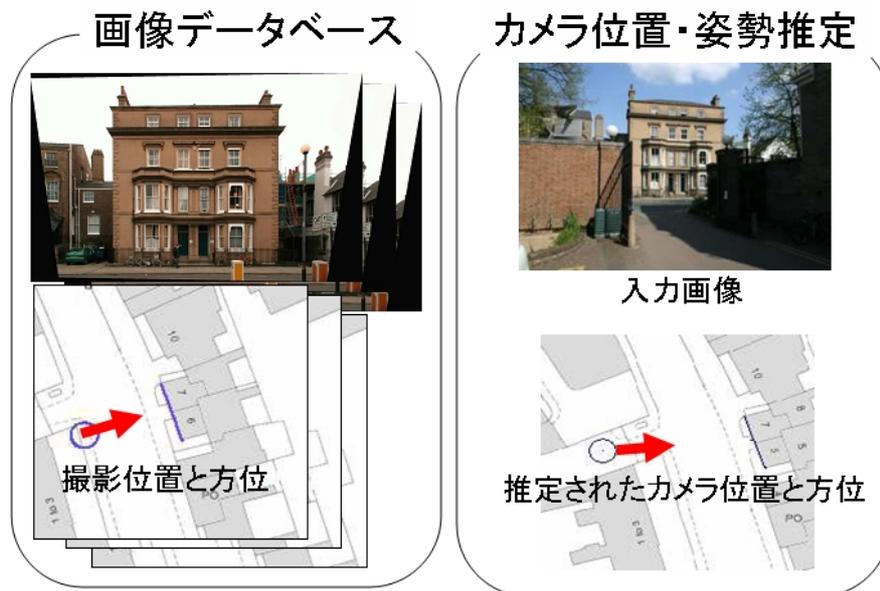


図 3 画像データベースを用いたカメラ位置・姿勢推定 [21]

る場合にも、ロバストなカメラ位置・姿勢推定を実現している。この手法では、対応するエッジの探索範囲を入力画像上へ投影したモデルの輪郭と垂直な直線上に限定することで、高速な対応付け処理を行っている。三次元モデルを用いた手法では、実時間でカメラの位置・姿勢推定を実現できるが、一般に広域で複雑な屋外環境の三次元モデルを作成することは難しく、データベースの構築に多くの人的コストを必要とするという問題がある。

このようなモデル作成のコストを低減可能なアプローチとして、環境の一部のワイヤフレームモデルと SLAM を併用した手法が提案されている [27]。この手法では、三次元モデルが写らない場合にも、自然特徴点の追跡によってカメラ位置・姿勢の推定を継続する。この手法では、環境全体のモデルを作成するのではなく、環境の一部のモデルを作成することで、データベース構築のコストを低減できるが、長い間モデルが写らないと推定誤差が累積するため、広域な環境での利用にはやはりモデル作成のための多くの人的コストが必要となる。

2.2.5 ランドマークデータベースを用いた手法

ランドマークデータベースを用いた手法 [28, 29] は、自然特徴点の三次元位置と自然特徴点周辺の局所的な画像情報を事前にランドマークとしてデータベースへ登録しておき、オンラインでランドマークと入力画像中の自然特徴点を対応付けることでカメラ位置・姿勢の推定を行う。

Skrypnyk ら [28] は、環境内の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、入力画像中の SIFT 特徴量と対応付けることでカメラ位置・姿勢を推定する手法を提案している。この手法では、複数枚の画像から SIFT 特徴点を算出し、画像間での対応付けを行うことで図 4 に示すような特徴点の三次元復元を行っている。このような特徴点の対応付けによる三次元復元により、自動でデータベースを構築することが可能である。しかし、広域な環境へ適用する場合には特徴点の三次元復元に誤差が蓄積するという問題や入力画像中からの SIFT 特徴量算出に多くの処理時間を必要とし、ビデオレートでの推定が困難であるという問題がある。

大江ら [29] は、三次元位置が既知の少数の基準点を用いる手法 [39] により蓄積誤差の生じない環境の三次元復元を行うことで、ランドマークデータベースを構築している。ランドマークデータベースには自然特徴点の三次元位置と自然特徴点周辺の局所的な画像テンプレートの情報が保持されており、入力画像中の自然特徴点の二次元座標とランドマークの三次元座標を対応付けることで、カメラ位置・姿勢を推定する。この手法では、全方位カメラを用いてランドマーク情報を収集するため、広域で複雑な環境に対しても比較的 low コストにデータベースを構築できる。しかし、Skrypnyk ら [28] の手法と同様に、計算コストが高く、実時間でのカメラ位置・姿勢の推定が困難であるという問題が残されている。

2.3 センサと画像を用いた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た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, 37, 38] は、センサを用いたカメラ位置・姿勢推定手法と画像を用いたカメラ位置・姿勢推定手法のそれぞれの問題点を互いに補うことによって、高精



図 4 SIFT 特徴量を用いた三次元復元 [28]

度なカメラ位置・姿勢推定を実現している。以下では、センサと画像を用いたハイブリッドなカメラ位置・姿勢推定手法を、絶対指標としてセンサの計測結果を用い、画像を補助的に利用する手法 [30, 31, 32] と、画像による計測結果を絶対指標とし、相対位置・姿勢を計測するセンサと組み合わせる手法 [33, 34, 35, 36, 37, 38] に大別しそれぞれの手法の特徴と問題点について述べる。

2.3.1 絶対指標としてセンサを用いる手法

絶対指標としてセンサを用いる手法 [30, 31, 32] では、GPS や赤外線ビーコンなどの計測結果を絶対指標として用い、画像を補助的に利用することで比較的ロバストなカメラ位置・姿勢の推定を実現している。山本ら [30] は、RTK-GPS (Real Time Kinematic GPS) と慣性センサの計測結果を用いて絶対的なカメラ位置・姿勢を推定し、入力画像中の特徴点とデータベース中のランドマークを用いて、ICP (Iterative Closest Point) アルゴリズムによってカメラ位置・姿勢の補正を行うことで実世界と仮想世界の位置合わせの精度を高めている。Hu ら [31] は、GPS

とジャイロから得られる計測結果に加えて、三次元位置情報、車線数などが記録されている道路地図データを用い、道路形状と入力画像中の特徴を対応付けることで最終的なカメラ位置・姿勢を推定している。横地ら [32] は、入力画像中の自然特徴点を追跡することによって推定される相対的な位置・姿勢と GPS から得られる絶対位置を組み合わせることで蓄積誤差の生じないカメラ位置・姿勢推定手法を提案している。これらのセンサからの計測結果を絶対指標として用いる手法 [30, 31, 32] では、比較的ロバスト性の高いカメラ位置・姿勢推定を実現できるが、センサを用いた手法と同様にセンサによって利用可能な環境に制限があるという問題が残る。

2.3.2 絶対指標として画像を用いる手法

絶対指標として画像を用いる手法 [33, 34, 35, 36, 37, 38] では、入力画像中の特徴を三次元モデルやランドマークと対応付ける際に、加速度センサやジャイロなどによる相対的な位置や姿勢の計測結果を補助的に利用することで、カメラ位置・姿勢推定のロバスト性の向上を図っている。このような手法として、姿勢センサより得られる加速度情報や角速度情報を利用することで、検出すべき自然特徴点の位置を予測する手法 [35, 36] や、センサ情報によりモーショントラjectoryを考慮することでロバストに対応付け処理を行う手法 [33, 34, 35] が提案されている。これらの手法では、カメラが高速に並進・回転した場合にもロバストにカメラ位置・姿勢を推定できるが、事前知識として三次元モデルを用いる手法にはモデル作成コストの問題が、ランドマークデータベースを用いる手法にはカメラ位置・姿勢推定の処理コストの問題が残されている。

小竹ら [37] は、画像上の自然特徴点と三次元座標との対応に加えて、姿勢センサから得られる重力軸の傾きの情報を利用することによって、線形計算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。この手法では、カメラ位置・姿勢を推定する際に非線形最適化を必要とせず、低い計算コストでの位置・姿勢の推定を実現している。しかし、この手法では特徴点の画像上の座標と三次元座標との対応が既知であることを前提としており、マーカの頂点などの比較的識別性の高い自然特徴点を必要としている。



図 5 Reitmayr らの手法 [38] で用いられる三次元モデル

Reitmayr らの手法 [38] では，複数枚の画像と地図から取得したスケール情報を用いて，PhotoBuilder[40] によって人手で図 5 に示すような環境のテクスチャ付のポリゴンモデルを作成し，カメラ位置・姿勢の推定に利用している．この手法ではテクスチャ情報を用いるため詳細なモデルを必要とせず，建物などの形状の大部分を平面として近似することができモデル構築の人的コストを抑えることができる．しかし，自然物が多く存在するような複雑な環境への適用は難しい．

2.4 本研究の位置づけと方針

本節では，拡張現実感への応用を前提として，従来手法に対する本研究の位置づけを述べる．前節までに概観したように，これまでにカメラの位置・姿勢を推定する手法が数多く提案されている．図 6 に各手法のロバスト性と位置合わせの精度の関係を示す．センサを用いる手法は比較的ロバスト性の高いカメラ位置・姿勢の推定を実現できるが，拡張現実感での利用を前提とした場合にはセンサとカメラのキャリブレーション誤差や計測誤差に起因して，画素単位での位置合わせが難しいという問題がある．画像からのカメラ位置・姿勢の推定は，センサを用いた手法に比べ，拡張現実感における実世界と仮想世界の位置合わせの精度を高めやすいという利点があるが，ロバストなカメラ位置・姿勢の推定を実現するためには，カメラ以外のセンサから得られる情報を統合する必要がある．センサ

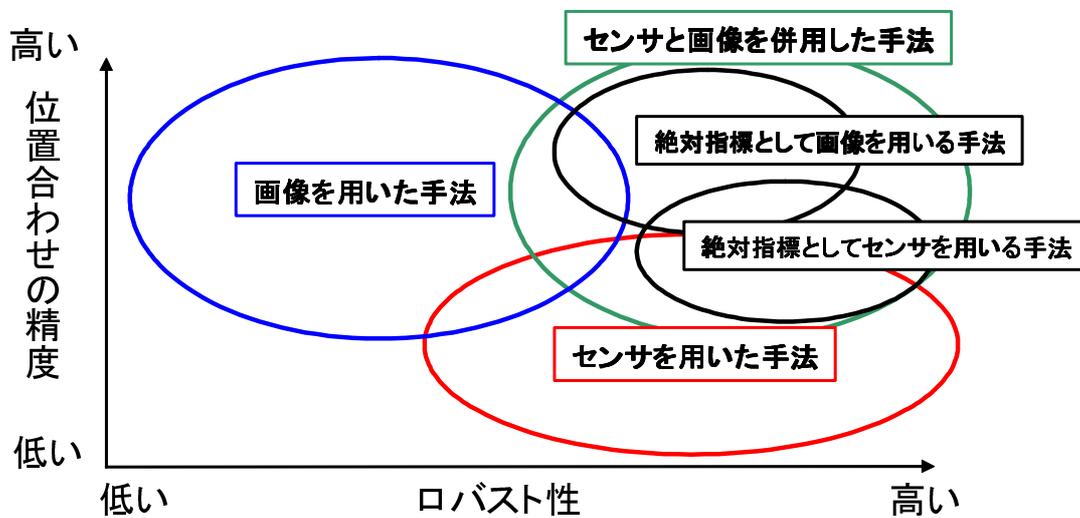


図 6 カメラ位置・姿勢推定手法の位置づけ

と画像を用いたハイブリッドなカメラ位置・姿勢推定手法は、センサを用いることでロバストなカメラ位置・姿勢推定を実現し、また画像を用いることで実世界と仮想世界の位置合わせの精度を高めている。ハイブリッドなカメラ位置・姿勢推定において、絶対指標としてセンサの計測結果を用いた場合には、比較的ロバスト性の高いカメラ位置・姿勢の推定が可能であるが、センサによって利用可能な環境に制限があり、様々な環境で利用するためには多数のセンサを組み合わせる必要がある。そのため、多くの手法では画像から推定される位置・姿勢を絶対指標とし、加速度センサやジャイロなどの計測結果を補助的に用いることによりロバスト性を向上させるアプローチがとられる。画像から推定される位置・姿勢を絶対指標とする手法では、センサからの計測結果を統合することによりロバスト性を向上させているが、画像からの位置・姿勢の推定には利用環境の構築コストやカメラ位置・姿勢推定の処理コストの問題が残されており、解決が必要である。

本研究では、拡張現実感において実世界と仮想世界の位置合わせの精度を高めやすい画像からの位置・姿勢推定の中でも、広域で複雑な環境においてもデータベースの構築コストが比較的安く、蓄積誤差を含まないカメラ位置・姿勢を推定することができるランドマークデータベースを用いる手法 [29] に着目し、従来手



図7 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（オンライン処理）

法 [29] で問題となっていたカメラ位置・姿勢推定処理の計算コストを低減することで実時間でのカメラ位置・姿勢推定を実現する。以下では、まず従来手法 [29] について概説し、次に本研究の方針について述べる。

従来手法 [29] では、オンラインでのカメラ位置・姿勢推定処理において、図7に示すように、データベース中より選択した N 個のランドマークと入力画像中より抽出した平均 F 個の自然特徴点とを照合することで対応付け、カメラ位置・姿勢の推定を行っている。従来手法 [29] では、全体の処理時間の大半をこの対応付け処理が占めており、実時間での処理を達成するためには対応付け処理の高速化が必要である。

従来手法の対応付け処理における計算コスト C_{prev} は、ランドマークと自然特

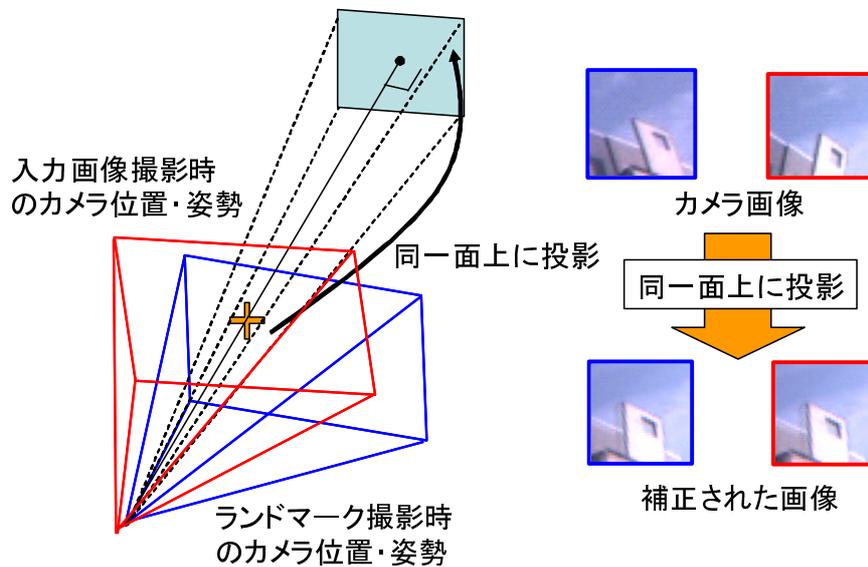


図 8 撮影姿勢に依存しない画像パターンの作成

徴点 1 個当たりの照合コストを A とすると

$$C_{prev} = NFA \quad (1)$$

となる。照合コスト A には、図 8 に示すような撮影姿勢に依存しない画像パターンの作成コストと正規化相互相関によるテンプレートマッチングのコストが含まれる。本研究において、ランドマークと自然特徴点の照合コスト A を低減することは困難であり、対応付け処理の計算コストを低減するためには、投影するランドマーク数 N と照合される自然特徴点数 F を削減する必要がある。従来から、対応付け処理の計算コストを低減するために、あらかじめ探索空間を KD-Tree によって分割しておくことで、高速に照合する手法 [41] が提案されているが、大量のデータや高次元の特徴空間に対して適用する場合には、計算コストが増大するという問題がある。

そこで、本研究では、対応付け処理に用いられるランドマーク数 N と照合される自然特徴点数 F を削減するため、以下の 2 つの改良を加える。

- (1) 連続フレーム間でランドマークを追跡し暫定的なカメラの位置・姿勢を推定することで照合すべき自然特徴点数 F を削減
- (2) ランドマークへ優先度情報を付加することで、正しく対応づく可能性の高い少数のランドマークを選択し、選択するランドマーク数 N を削減

これにより、屋内外の広域環境において蓄積誤差を含まない実時間でのカメラ位置・姿勢推定を実現する。

提案手法の処理の流れを図9に示す。図中の太枠は、本研究で追加または改良した処理を表す。本手法は、オフラインでのデータベースの構築処理 (A) とオンラインでのカメラ位置・姿勢の推定処理 (B) からなる。オフラインでのデータベース構築処理 (A) では、まず全方位カメラを用いて利用環境内を移動しながら撮影し、全方位動画像中の自然特徴点を自動で追跡することにより環境の三次元復元を行う (A-1)。次に、三次元復元の結果よりランドマーク情報を取得しデータベースへ登録する (A-2)。オンラインでのカメラ位置・姿勢の推定処理では、まず初期フレームでのカメラ位置・姿勢を何らかの手法により取得し (B-1)、続いて逐次的なカメラ位置・姿勢推定処理 (B-2)~(B-4) を繰り返す。逐次推定では、まず現フレームの暫定的なカメラ位置・姿勢を推定する (B-2)。暫定的なカメラ位置・姿勢を推定しておくことで、ランドマークの対応点候補の探索範囲を従来よりも狭くし、照合すべき自然特徴点数を削減する。次に、優先度に基づいてランドマークを選択することで、入力画像中の自然特徴点と正しく対応づく可能性の高い推定に有効な少数のランドマークを選択する (B-3)。最後に、選択したランドマークと入力画像中の自然特徴点を対応付け、現フレームのカメラ位置・姿勢を推定する (B-4)。逐次推定の完了後に推定結果をデータベースへフィードバックすることで、ランドマークの優先度を更新する (B-5)。

以上のように、本手法ではオンライン処理において、暫定的なカメラ位置・姿勢推定処理 (B-2) を追加することで照合すべき自然特徴点の削減し、またランドマーク選択処理 (B-3) において、新たにランドマークの優先度情報を用いることで、データベース中より選択するランドマーク数の削減する。これらの改良により、自然特徴点とランドマークの対応付け処理 (B-4) の処理コストを低減し、実時間でのカメラ位置・姿勢推定を実現する。

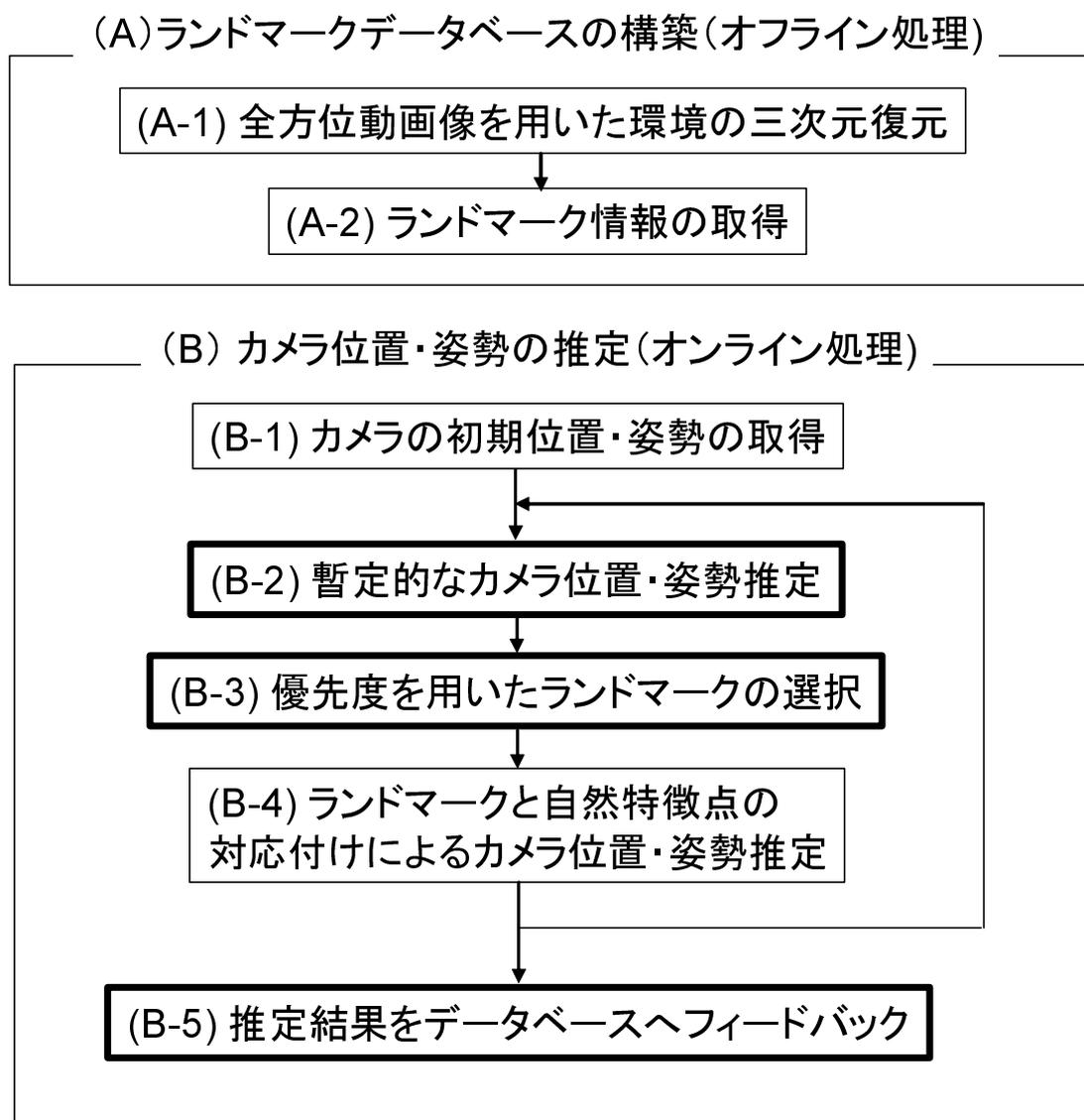


図 9 提案手法の処理の流れ

3. ランドマークデータベースの構築

本章では、オフラインでのランドマークデータベース構築処理について述べる。本研究では、4章で述べるオンラインでのカメラ位置・姿勢の推定において、推定に有効なランドマークを効率的に選択するために、従来手法 [29] で用いられているデータベースに優先度情報を付加した新たなランドマークデータベースを用いる。ここでは、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る。次に、実際のデータベース構築処理である、全方位動画像を用いた環境の三次元復元 (A-1)、ランドマーク情報の取得 (A-2) について述べる。最後に、本研究で用いるランドマークの優先度の定義について述べる。

3.1 ランドマークデータベースの構成要素

本研究では、推定に有効なランドマークを効率的に選択するために、従来手法 [29] で用いられているデータベースに優先度情報を付加した新たなランドマークデータベースを用いる。本研究で用いるランドマークデータベースには図 10 に示すように、(a) ランドマークの三次元位置、(b) 撮影地点ごとの情報、(c) ランドマークの優先度の情報、が保持される。撮影地点ごとの情報は、(b-1) 多重スケールの画像テンプレート、(b-2) ランドマーク撮影時の全方位カメラの位置、から成る。以下では、それぞれの要素について述べる。

(a) ランドマークの三次元位置

次章で述べるオンラインでのカメラ位置・姿勢推定では、ランドマークの三次元位置と画像上の二次元座標を対応付けることによりカメラの位置・姿勢を推定するため、ランドマークの三次元位置情報が必要となる。

(b) 撮影地点ごとの情報の取得

ランドマークを中心とする画像テンプレートを保持し、この画像テンプレートと入力画像中の自然特徴点周辺の画像テンプレートを照合することで、ランドマークと自然特徴点とを対応付ける。撮影地点によるランドマークの見え方の違いに対応するため、撮影地点ごとに図 11 に示すような多重スケール

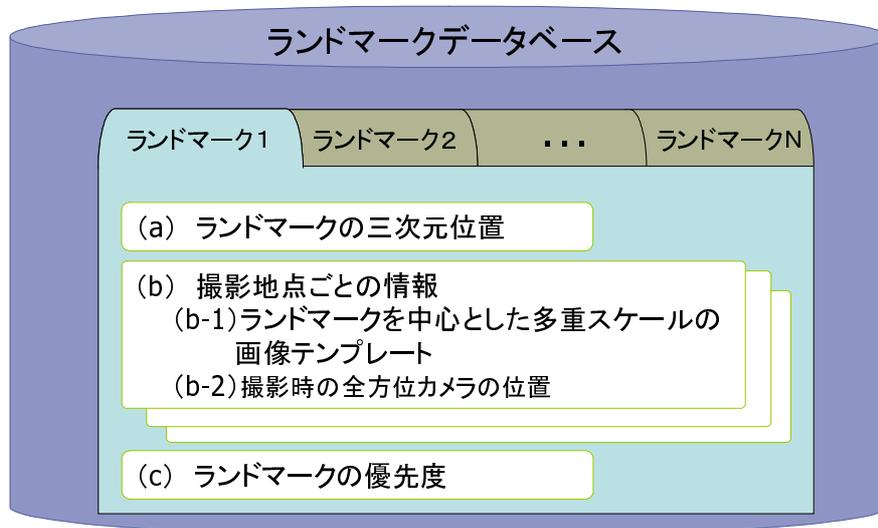


図 10 ランドマークデータベースの構成要素

の画像テンプレートを作成し、データベースへ登録する (b-1). また、オンラインでのカメラ位置・姿勢推定において、データベース中から推定に用いるランドマークを選択するために、ランドマーク撮影時の全方位カメラの位置をデータベースへ登録する (b-2).

(c) ランドマークの優先度

オンラインでのカメラ位置・姿勢の推定において、正しく対応づく可能性の高いランドマークをデータベース中から効率的に選択するために、ランドマークに優先度を設定する. 優先度は、過去に同じ場所でユーザによって撮影された入力に対する推定結果を利用して、各ランドマークがカメラ位置・姿勢推定に利用される確率を算出することで決定する.

3.2 全方位動画像を用いた環境の三次元復元

本研究では、佐藤らの手法 [39] または Ikeda らの手法 [42] を用いることで、広域で複雑な環境においても蓄積誤差の生じない環境の三次元復元を行う. これら



図 11 多重スケールの画像テンプレート

の手法では、まず、対象となる環境を移動しながら全方位カメラを用いて撮影する。次に、Harris オペレータ [43] によって検出された動画像中の自然特徴点を自動追跡することで、structure-from-motion によって自然特徴点の三次元位置と画像上の座標、および全方位カメラのカメラパラメータを推定する。最後に、三次元位置が既知の少数の基準点 [39] または GPS から得られる絶対位置 [42] を利用して、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。このような手法を用いることで、広域で複雑な環境においても比較的容易に環境の三次元復元を行うことができる。

3.3 ランドマーク情報の取得

環境の三次元復元の結果より、ランドマークの三次元位置 (a) と撮影地点ごとの情報 (b) を取得しデータベースへ登録する。以下ではそれぞれの情報の取得方法について述べる。

(a) ランドマークの三次元位置

次章で述べるオンラインでのカメラ位置・姿勢推定では、ランドマークの三次元位置と画像上の二次元座標を対応付けることによりカメラの位置・姿勢を推定するため、ランドマークの三次元位置情報が必要となる。ランドマークの三次元位置は 3.2 項で述べた三次元復元処理によって得られるものであ

り，環境に固定された世界座標系で保持されている。

(b) 撮影地点ごとの情報の取得

撮影地点によるランドマークの見え方の違いに対応するため，撮影地点ごとに多重スケールの画像テンプレートを作成し，データベースへ登録する．ここでは，従来手法 [29] と同様に，世界座標系において，カメラの投影中心とランドマークの三次元位置を結ぶ直線に対して垂直な面に画像上での自然特徴点周辺のパターンを投影することで，カメラの撮影姿勢に依存しない画像テンプレートを作成し登録する (b-1)．また，オンラインでのカメラ位置・姿勢推定において，データベース中から推定に用いるランドマークを選択するために，ランドマーク撮影時の全方位カメラの位置をデータベースへ登録する (b-2)．

3.4 ランドマークの優先度の定義

次章で述べるオンラインでのカメラ位置・姿勢の推定において，正しく対応づく可能性の高いランドマークをデータベース中から効率的に選択するために，ランドマークに優先度を付加する．優先度は，過去に同じ場所でユーザによって撮影された入力に対する推定結果を利用して，各ランドマークがカメラ位置・姿勢推定に利用される確率を算出することで決定する．ランドマーク i の優先度 P_i は，オンライン処理においてランドマーク i がデータベースから選択された回数 D_i とカメラ位置・姿勢推定の際に LMedS 基準 (付録 A 参照) によって誤対応として排除されずに推定に用いられた回数 E_i を用いて，以下の式で定義する．

$$P_i = \frac{E_i}{D_i} \quad (2)$$

本研究では，オンライン処理の完了時に，ランドマークの利用頻度 D_i と E_i をデータベースにフィードバックすることで，ランドマーク i の優先度 P_i を更新する．ただし，本研究では P_i の初期値として全てのランドマークに同じ優先度を設定しておく．この優先度の更新処理については次章で詳しく述べる．

4. カメラ位置・姿勢の推定

本章ではオンラインでのカメラ位置・姿勢推定処理について述べる．本研究では，図9に示したように，まず初期フレームでのカメラ位置・姿勢を何らかの手法により取得し (B-1)，続いて逐次的なカメラ位置・姿勢推定処理 (B-2)～(B-4) を繰り返す．逐次推定では，カメラ位置・姿勢推定の高速化とロバスト性の向上のために，まず連続フレーム間でランドマークを追跡し，暫定的なカメラの位置・姿勢を推定することで，ランドマークの対応点探索範囲を限定する (B-2)．次に，優先度に基づいてランドマークを選択することで，正しく対応づく可能性の高いランドマークを効率的にデータベース中から取り出す (B-3)．最後に，選択したランドマークと入力画像中の自然特徴点を対応付け，ロバスト推定により誤対応を排除した後，現フレームのカメラ位置・姿勢を推定する (B-4)．逐次推定の完了後にランドマークの利用頻度情報をデータベースへフィードバックすることで，ランドマークの優先度を更新する (B-5)．以下では，処理 (B-2)～(B-5) について詳しく述べる．

4.1 フレーム間でのランドマークの追跡による暫定的なカメラ位置・姿勢の推定

本研究では，連続フレーム間でランドマークを追跡し暫定的なカメラの位置・姿勢を推定することで，照合すべき自然特徴点数を削減する．暫定的なカメラ位置・姿勢の推定では，まず前のフレームで自然特徴点と対応付けられ，誤対応として排除されずにカメラ位置・姿勢の推定に用いられたランドマークを選択し，現フレームにおける対応点を探索する．ランドマークの対応点の探索において，連続フレーム間ではパターンの変形や明るさの変化は微小であることから，前フレームにおけるランドマークの位置を中心とする $T \times T$ 画素の画像をテンプレートとして用い，ランドマークの前フレームの画像上の位置を中心とする一定のウィンドウ W_1 内で SSD (Sum of Squared Differences) を評価尺度としたテンプレートマッチングによりランドマークの追跡を行う．ただし，ランドマークの追跡処理はウィンドウ W_1 内で Harris オペレータによって検出された自然特徴点のみを

対象として行う。次に，追跡したランドマークを用いて現フレームの暫定的なカメラ位置・姿勢を推定する。カメラ位置・姿勢の推定では，LMedS 基準を用いて誤対応を排除した後に，正しい対応関係を用いてカメラ位置・姿勢の推定を行い，再投影誤差を最小化することによって現フレームの暫定的なカメラパラメータ \hat{M} を推定する。

ここで，ランドマークの追跡処理に要する計算コスト C_{track} は，ウィンドウ W_1 内で抽出される自然特徴点数を平均 F 個，追跡するランドマーク数を N_{track} 個とすると，

$$C_{track} = N_{track}FB + E_{LMedS} \quad (3)$$

となる。ただし， E_{LMedS} は誤対応の排除とカメラ位置・姿勢の推定にかかる計算コストであり， B は追跡処理においてランドマークと自然特徴点の1組のペアに対して SSD を算出する計算コストである。このコストは2.4節で述べた，幾何学的な変形や輝度値の正規化を必要とするランドマークの照合コスト A に対してはるかに小さい。

4.2 優先度を用いたランドマークの選択

本研究では，ランドマークの優先度情報に基づいてランドマークを選択することで，正しく対応づく可能性の高いランドマークを効率的にデータベース中から取り出す。ランドマークの選択処理 (B-3) では，まず処理 (B-2) で推定された暫定的なカメラパラメータ \hat{M} とランドマークが撮影された全方位カメラの位置を用いて，以下の幾何学的な条件を満たす画像テンプレートをデータベース中より選択する。

- ランドマークが入力画像の画角内に存在する
- ランドマーク撮影時のカメラ位置とランドマークを結ぶ直線と，暫定的なカメラ位置とランドマークを結ぶ直線の成す角が閾値以下
- ランドマーク撮影時のカメラ位置と現フレームの暫定的なカメラ位置の間の距離が閾値以下

- 入力画像上での他のランドマークとの距離が閾値以上

次に、選択されたランドマークの候補から優先度 P の高い順にランドマークを N_{prior} 個選択する。ただし、本手法では現フレームのカメラ位置・姿勢を推定するために、フレーム間で追跡したランドマークも用いるため、 N_{prior} は最大 N_{max} 個とし、 $N_{prior} = N_{max} - N'_{track}$ とする。ここで、 N'_{track} は前節で述べた暫定的なカメラ位置・姿勢の推定に用いた自然特徴点数である。

4.3 ランドマークと自然特徴点の対応付けによるカメラ位置・姿勢推定

ここでは、データベースより選択した優先度の高いランドマークと入力画像中の自然特徴点を対応付け、カメラ位置・姿勢の推定を行う。ランドマークと自然特徴点の対応付け処理では、まずランドマークの追跡により推定した暫定的なカメラパラメータ \hat{M} を用いて、データベースより選択されたランドマーク $j(j = 1, 2, \dots, N_{prior})$ を以下の式により入力画像上へ投影する。

$$\begin{bmatrix} a_j u_j \\ a_j v_j \\ a_j \end{bmatrix} = \hat{M} \begin{bmatrix} x_j \\ y_j \\ z_j \\ 1 \end{bmatrix} \quad (4)$$

ただし、 (x_j, y_j, z_j) はランドマーク j の世界座標系での三次元位置、 (u_j, v_j) はランドマーク j の画像上への投影位置、 a_j はカメラ座標系でのランドマーク j の奥行きを表す。次に、投影したランドマーク j の画像上の位置 (u_j, v_j) を中心とする一定ウィンドウ W_2 内の自然特徴点を検出する。ただし、ここではウィンドウ W_2 のサイズを暫定的なカメラ位置・姿勢を推定する際のウィンドウ W_1 のサイズよりも小さく設定することで、ランドマークの画像上での探索範囲を小領域に限定する。最後にウィンドウ内で検出された全ての自然特徴点について、データベース構築時と同様に、カメラの撮影姿勢に依存しない画像テンプレートを作成し、作成した画像テンプレートとランドマークの画像テンプレートを正規化相互相関

法によって照合する．本手法では相関値が最大となる自然特徴点をランドマークと対応付ける．

次に，ランドマークと対応付けられた自然特徴点と，処理 (B-2) で暫定的なカメラ位置・姿勢の推定に用いた自然特徴点を用いて，現フレームのカメラの位置・姿勢を推定する．現フレームのカメラの位置・姿勢は，処理 (B-2) と同様に，LMedS 基準を用いた誤対応の排除後に，再投影誤差を最小化することによって推定する．

なお，本処理 (B-4) では，暫定的なカメラ位置・姿勢を用いて画像上の探索範囲を限定することで，ウィンドウ内で抽出される対応点候補として用いられる自然特徴点の平均数を $\frac{S_2}{S_1}F$ (S_1, S_2 はそれぞれウィンドウ W_1, W_2 の面積を表す) に削減でき，また，手順 (B-2) で追跡されたランドマークを本処理で対応付けるランドマークから除外することで，ランドマークの数を $(N_{max} - N'_{track})$ に削減できる．よって，本処理におけるランドマークと自然特徴点の対応付け処理の計算コスト C_{proj} は，

$$C_{proj} = \frac{S_2}{S_1} N_{prior} F A \quad (5)$$

$$= \frac{S_2}{S_1} (N_{max} - N'_{track}) F A \quad (6)$$

となる．また，処理 (B-2) と処理 (B-4) によるランドマーク対応付けの総コスト C_{new} は，

$$C_{new} = C_{track} + C_{proj} \quad (7)$$

$$= N_{track} F B + E_{LMedS} + \frac{S_2}{S_1} (N_{max} - N'_{track}) F A \quad (8)$$

となる．ここで，暫定的なカメラ位置・姿勢の推定コスト C_{track} は，従来の対応付け処理の計算コスト $C_{prev} = N_{max} F A$ に対して，はるかに小さいため，本処理におけるランドマークと照合すべき自然特徴点数 F とデータベースより選択するランドマーク数 N の削減による処理コストの低減効果は大きいと考えられる．

4.4 推定結果のフィードバックによる優先度の更新

カメラの位置・姿勢推定完了後、ランドマークが利用された頻度の情報をデータベースへフィードバックすることで、データベース中の各ランドマークの優先度を更新する (B-5)。ここでは頻度の情報として、カメラ位置・姿勢推定の逐次処理において、ランドマーク i が処理 (B-3) でデータベース中から選択された回数 D_{inew} と処理 (B-4) において LMedS 基準により誤対応として排除されずにカメラ位置・姿勢の推定に用いられた回数 E_{inew} を利用することでデータベース中のランドマークの優先度を更新する。優先度の更新は、ユーザより新たに取得した頻度情報と、過去に同じ場所でカメラ位置・姿勢の推定を行ったユーザより得られている頻度情報を用いて、優先度 P_i を次式で更新する。

$$P_i = \frac{E_{iold} + E_{inew}}{D_{iold} + D_{inew}} \quad (9)$$

ここで、 D_{iold} 、 E_{iold} は、それぞれランドマーク i が過去にデータベース中から選択された回数、過去に誤対応として排除されずにカメラ位置・姿勢の推定に用いられた回数を表す。

5. 実験

提案手法の有効性を示すために，従来手法 [29] との比較によりロバスト性と処理速度の向上効果について検証した．以下では，まずランドマークデータベースの構築について述べ，次に単眼カメラで撮影した実動画像を用いた実験について述べる．

5.1 ランドマークデータベースの構築

本実験で用い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．データベースの構築には，図 12 左に示すような全天球の動画像を取得できる全方位型マルチカメラシステム (Point Grey Research 社 Ladybug, 解像度 $768 \times 1024 \times 6$ 台) を用いた．このカメラシステムにより屋外環境 (約 75m) を図 12 右に示すような画像系列として移動撮影し，佐藤らの手法 [39] により，自然特徴点の三次元位置と全方位カメラのカメラパラメータを推定することによりランドマークデータ



図 12 全方位型マルチカメラシステムと全方位画像

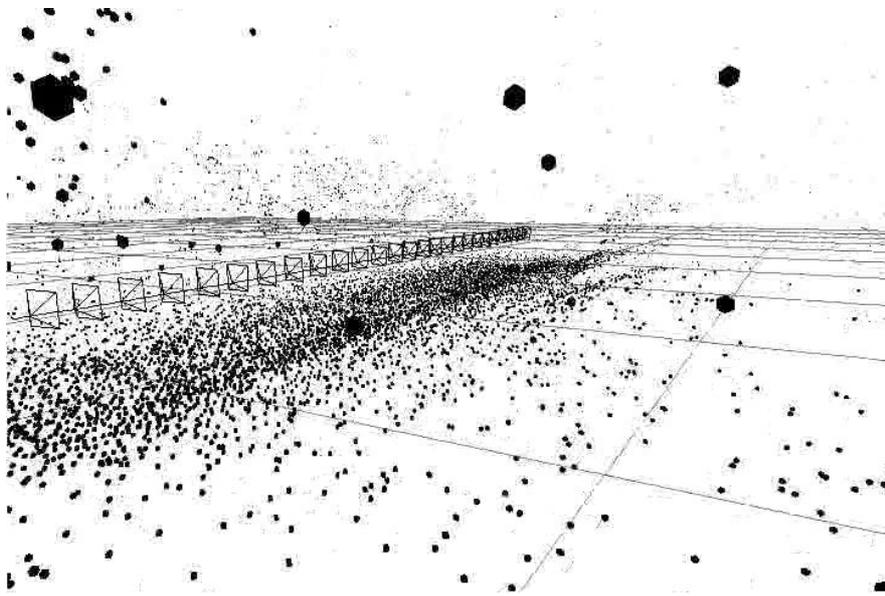


図 13 環境の三次元復元結果

ベースを構築した。環境の三次元復元の結果を図 13 に示す。ランドマークデータベース中のランドマーク数は約 12400 個であり、一つのランドマークにつき平均 8ヶ所の異なる地点で撮影された画像テンプレートが登録された。

5.2 ロバスト性と処理速度の向上効果の検証

提案手法の有効性を示すために、従来手法 [29] との比較によりロバスト性と処理速度の向上効果について検証した。まず、ロバスト性向上効果を示すために、位置・姿勢推定精度及び推定失敗率について、従来手法と同一条件での比較を行った。次に、処理速度向上効果について、オンラインでのカメラ位置・姿勢推定の各処理にかかる処理時間を比較した。

本実験では、表 1 に示すような機器を用いてカメラ位置・姿勢の推定を行った。カメラ位置・姿勢の推定は、あらかじめ撮影した動画像 (解像度 720×480 画素, プログレッシブスキャン, 15fps, 1000 フレーム) に対して行った。本実験では、提案手法、従来手法 [29] とともにカメラの内部パラメータは既知とし、初期フレー

ムのカメラ位置・姿勢は手動で与えた。また、世界座標系はX軸，Y軸が実環境における地面に対して水平，Z軸が地面に対して垂直な座標系とした。本実験で用いたカメラ位置・姿勢推定処理における各パラメータを表2に示す。本実験では、提案手法におけるランドマーク追跡処理の対応点探索範囲を120×60画素とし、データベース中より選択したランドマークの対応点候補の探索範囲を従来手法[29]では120×60画素、提案手法では20×20画素とした。また、提案手法では、あらかじめ同じ場所で撮影した3本の動画像の推定結果を優先度を設定し、また、ランドマークの優先度の初期値として0.5を用いた。

表1 カメラ位置・姿勢推定に用いた機器

入力撮影カメラ	SONY DSR-PD-150
広角レンズ	SONY VCL-HG0758
マシンスペック	CPU:Core2Extreme 2.93GHz, Memory:2.0GB

表2 実験で用いたパラメータ

	従来手法 [29]	提案手法
ランドマーク追跡処理の際の 画像テンプレートサイズ	-	15×15画素
ランドマークの 画像テンプレートサイズ	15×15画素	
自然特徴点追跡の際の 探索範囲	-	120×60画素
ランドマークの対応点候補の 探索範囲	120×60画素	20×20画素
学習データ	-	動画3本分の推定結果
優先度の初期値	-	0.5

5.2.1 ロバスト性向上効果の検証

ロバスト性向上効果の検証のため、カメラ位置・姿勢の推定精度と推定に必要なランドマーク数について従来手法 [29] との比較を行った。位置・姿勢推定精度の定量的な評価に用いるカメラ位置・姿勢の真値は、環境内の特徴点の三次元位置をトータルステーションを用いて測定し、10 フレームおきに各フレームの画像上でそれらの特徴点を手動で対応付け、PnP 問題を解くことで算出した。ただし、再投影誤差が 1.5 画素以上であるフレームや自然物が入力画像中の大半を占めるなどの理由で、トータルステーションによって計測できる特徴点が少ないフレームについては評価対象にしなかった。本実験では、従来手法 [29]、提案手法ともにランドマークの選択処理において、データベース中より選択するランドマークの最大数を $N_{max} = 100$ とした。

図 14 に初期フレームから 1000 フレーム目までの入力画像に対して推定されたカメラパスと真値、データベース構築時の全方位カメラのカメラパスを示す。提案手法で推定されたカメラの平均位置推定誤差、およびカメラの光軸向きの平均

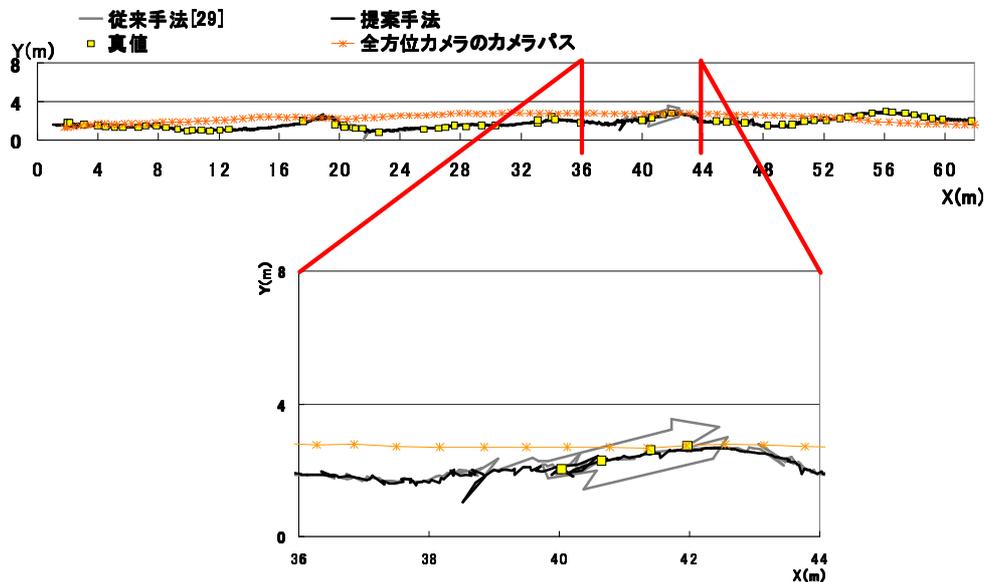


図 14 推定されたカメラパスと真値

推定誤差を表3, 表4に示す. 有意水準を5%と設定したt検定により, 提案手法と従来手法の位置・姿勢推定精度を比較した結果, 有意な差は認められなかったことから, ほぼ同等の精度で推定できていることが確認できる. ただし, 図15に示すように, 自然物が多く映り込んでいるようなシーンに対する推定結果(図14, X=38~44m付近)については, 従来手法[29]では推定されたカメラ位置がばらついているのに対し, 提案手法ではカメラ位置が比較的滑らかに推定されており, 本手法を用いることにより推定に利用できるランドマークが少ないシーンでもロバストな推定が実現できていることが確認できる. なお, このようなロバスト性の向上については表3, 表4に示した定量評価値からは確認できないが, これは本実験において, 全フレームに対する真値を作成しておらず, カメラ位置がばらついているフレームに対しては, 上記の精度評価に含まれていないためである. 次に, データベースから選択するランドマークの最大数 N_{max} を変化させ, 以下の4つの手法に対するカメラ位置・姿勢推定の精度と推定失敗率を比較した.

(手法A) 本手法の基礎とした大江らの手法[29]

(手法B) 提案手法において, 暫定的なカメラ位置・姿勢の推定による探索範囲の限定のみを行った場合

表3 カメラ位置推定精度の比較

	従来手法 [29]	提案手法
平均位置誤差 (mm)	220	226
標準偏差 (mm)	141	141

表4 カメラ姿勢推定精度の比較

	従来手法 [29]	提案手法
カメラの光軸向きの平均推定誤差 (度)	0.37	0.41
標準偏差 (度)	0.35	0.42



(a)640 フレーム



(b)660 フレーム

図 15 自然物が多く映り込んでいるフレーム

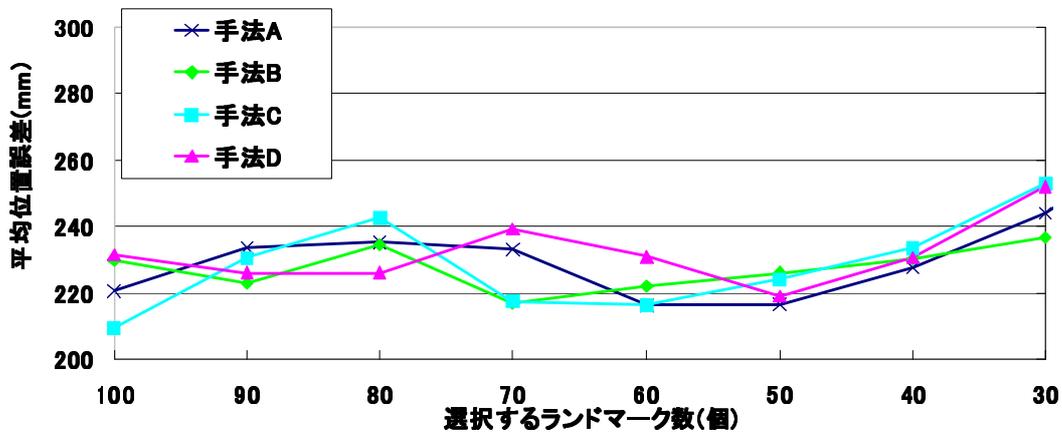


図 16 ランドマーク数とカメラの推定位置誤差の関係

(手法 C) 提案手法において、優先度を用いたランドマークの選択のみを用いた場合

(手法 D) 暫定的なカメラ位置・姿勢の推定による探索範囲の限定と優先度を用いたランドマークの選択を両方行った場合

なお、本実験では、カメラ位置・姿勢の逐次推定に途中で失敗した場合には、手動で失敗したフレームのカメラ位置・姿勢を与えることにより、カメラ位置・姿勢の逐次推定処理を再開し、最終フレーム（1000 フレーム目）まで推定を完了させた上で推定が成功した全てのフレームに対する推定誤差の平均値と推定失敗の回数を算出した。

データベースから選択するランドマーク数とカメラ位置の推定誤差の関係を図 16 に示す。全ての手法において、選択するランドマーク数 N_{max} が 40 以下の場合に位置誤差が増加する傾向があるが、手法ごとの明確な差は見られない。また、推定されたカメラの位置誤差は、全ての手法でほぼ同等の精度となっていた。図 17 にデータベースから選択するランドマーク数と推定に失敗したフレーム数の関係を示す。手法 A と手法 B では、データベース中より選択するランドマーク数の減少に伴い、カメラ位置・姿勢の推定に失敗するフレームが発生するが、優先度を用いてデータベース中よりランドマークを選択する手法 C、手法 D では最終フ

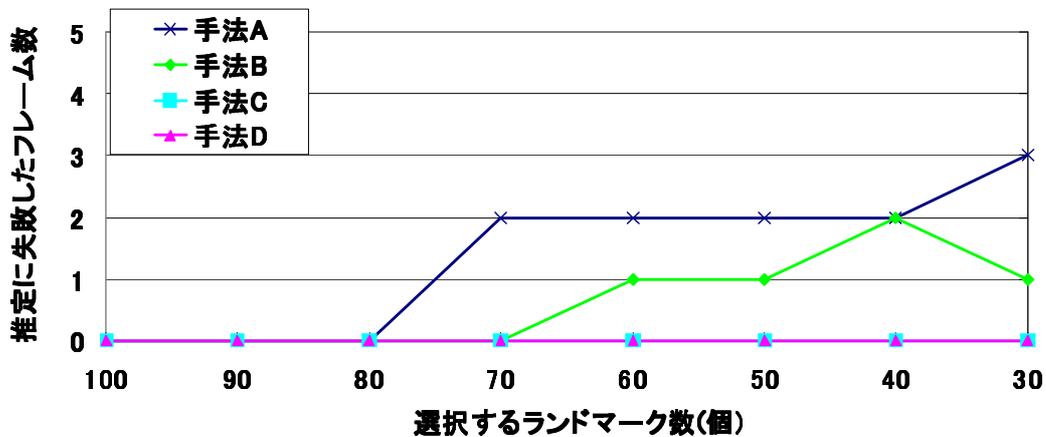


図 17 ランドマーク数と推定に失敗したフレーム数の関係

フレームまで失敗せずにカメラ位置・姿勢を推定できている。このことから、優先度を用いることで、正しく対応づく可能性の高いランドマークを優先的に選択できていることが確認できる。

5.2.2 処理速度向上効果の検証

ここでは、最終フレームまで推定に失敗せず、かつ $N_{max} = 100$ とした場合とほぼ同等のカメラパスを推定可能な、従来手法では $N_{max} = 80$ 個、提案手法では $N_{max} = 60$ 個、の場合の処理時間の比較を行った。このときの提案手法および従来手法で推定されたカメラの平均位置誤差、およびカメラの光軸向きの平均誤差はそれぞれ 231mm, 0.42 度, 235mm, 0.39 度であった。カメラ位置・姿勢推定の各処理の処理時間と全体の処理時間を表 5 に示す。提案手法では、ランドマークの追跡処理による暫定的なカメラ位置・姿勢の推定処理 (B-2) に 28ms を要するが、対応点探索範囲を小領域に限定することで照合すべき自然特徴点数が削減され、ランドマークと入力画像中の自然特徴点の対応付けの処理時間が 15ms(従来手法 [29] の約 1/20) となった。また、誤対応の排除とカメラ位置・姿勢推定の処理時間が 17ms(従来手法 [29] の約 1/4) となった。これは、提案手法では現フレームの暫定的なカメラ位置・姿勢を推定することによって、対応点探索範囲を小領

域に限定したことと、優先度に基づいてランドマークを選択することで、正しく対応づく可能性の高いランドマークを優先的にデータベース中より選択したことで、誤対応の割合が減少したためである。全体の処理時間は平均66ms(従来手法[29]の約1/6)となり、1秒間に約15フレームの実時間でカメラ位置・姿勢推定を実現した。

提案手法で推定されたカメラ位置・姿勢($N_{max} = 60$ のとき)を用いて入力動画像上に仮想物体を重畳表示した結果を図18に示す。生成された動画像から仮想物体が位置ずれなく幾何学的に正しい位置に描画されていることを確認した。

表 5 処理時間の比較

	処理時間 (ms)		処理時間の比
	従来手法[29] (ランドマーク数80個)	提案手法 (ランドマーク数60個)	
暫定的な カメラ位置・姿勢の推定	-	28	-
ランドマークの選択	12	1	0.08
ランドマークの対応付け	316	15	0.05
誤対応の排除と カメラ位置・姿勢推定	61	17	0.28
画像の取得など	4	5	1.25
全体	393(2.5fps)	66(15.1fps)	0.16



(a)200 フレーム目



(b)400 フレーム目



(c)600 フレーム目



(d)800 フレーム目

図 18 カメラ位置・姿勢推定結果を用いた仮想物体の重畳表示

6. まとめ

本論文では、ランドマークデータベースを用いた動画像からのカメラ位置・姿勢推定において、ランドマークと自然特徴点の照合回数を削減することで計算コストを低減し、実時間でのカメラ位置・姿勢推定処理を実現した。提案手法では、連続フレーム間でランドマークを追跡し暫定的なカメラの位置・姿勢を推定することで、ランドマークの対応点探索範囲を小領域に限定し、照合すべき自然特徴点数を削減した。また、過去に同じ場所で推定を行ったユーザの推定結果を利用して、ランドマークへ優先度情報を付加することで、正しく対応づく可能性の高い少数のランドマークを選択し、対応付け処理に用いるランドマーク数を削減した。これにより、従来、多くの計算コストを必要としていたランドマークと自然特徴点の対応付け処理の計算コストが低減され、実時間での処理が可能となった。

実験では、提案手法で推定されたカメラ位置の推定精度と推定失敗率および処理時間を従来手法と比較することで、ロバスト性と処理速度の向上効果について検証した。実験の結果、従来手法とほぼ同等の精度で、実時間でのカメラ位置・姿勢推定処理が実現できることを確認した。また、定量的な評価には現れなかったが、目視によるカメラパスの比較により、推定に利用できるランドマークが少ないシーンでもロバストな推定が実現できることを確認した。

今後の課題として、データベースの更新による利用環境の変化への対応、テクスチャの少ないシーンでのロバスト性の向上などがあげられる。データベースの更新では、異なる季節や時間に構築されたいくつかのデータベースを用いることで、データベースが構築されていない季節や時刻のデータベースを補間することが考えられる。また、テクスチャが少ないシーンでは、カメラ以外のセンサからの計測結果を統合することが考えられる。このような課題を解決することによって、拡張現実感技術を用いたナビゲーションなどの様々な分野への応用が期待できる。

謝辞

本研究を進めるにあたり，終始暖かく御助言，御鞭撻を頂いた視覚情報メディア講座 横矢直和 教授に心より感謝致します。また，本研究の遂行にあたり，有益な御助言，御鞭撻頂いたインタラクティブメディア設計学講座 加藤博一 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，御鞭撻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。さらに，本研究の遂行にあたり，細やかな御助言，御鞭撻を頂いた像情報処理学講座 池田聖 助教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 高橋美央 女史に深く感謝致します。最後に，忙しい中，本研究の実験をサポートして頂いた水戸博之氏をはじめ，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. 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. Int. Conf. on Mobile Computing and Networking, pp. 32–43, 2000.
- [4] T. Höllerer, S. Feiner and J. Pavlik: “Situated documentaries: Embedding multimedia presentations in the real world,” Proc. Int. Symp. on Wearable Computers, pp. 79–86, 1999.
- [5] 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.
- [6] R. Azuma, B. Hoff, H. Neely and R. Sarfaty: “A motion-stabilized outdoor augmented reality system,” Proc. IEEE Virtual Reality, pp. 252–259, 1999.
- [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] E. Foxlin: “Pedestrian tracking with shoe-mounted inertial sensors,” IEEE Computer Graphics and Applications, Vo. 25, No. 6, pp. 38–46, 2005.

- [9] 山中一樹, 神原誠之, 横矢直和: “装着した3次元磁気センサを用いた歩き・走り状態に対応したユーザの自己位置推定”, 電子情報通信学会 技術研究報告 MVE2006-62, 2006.
- [10] A. Davison, Y. G. Cid and N. Kita: “Real-time 3D slam with wide-angle vision,” Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [11] B. Williams, G. Klein and I. Reid: “Real-time SLAM relocalisation,” Proc. Int. Conf. on Computer Vision, 2007.
- [12] G. Klein and D. Murray: “Parallel tracking and mapping for small AR workspaces,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 225–234, 2007.
- [13] H. Kato and H. Billinghurst: “Marker tracking and hmd calibration for a video-based augmented reality conferencing system,” Proc. 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, 2002.
- [15] M. Kalkusch, T. Lidy, M. Knapp, G. Reitmayr, H. Kaufmann and D. Schmalstieg: “Structured visual markers for indoor pathfinding,” Proc. Int. Augmented Reality Toolkit Workshop, 2002.
- [16] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. Int. Symp. on Wearable Computers, pp. 21–23, 2003.
- [17] L. Naimark and E. Foxlin: “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 27–36, 2002.
- [18] 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.

- [19] S. Saito, A. Hiyama, T. Tanikawa and M. Hirose: “Indoor marker-based localization using coded seamless pattern for interior decoration,” Proc. IEEE Virtual Reality, pp. 67–74, 2007.
- [20] 吉田友祐, 天目隆平, 柴田史久, 木村朝子, 田村: “美観と頑健性を両立させた複合現実感用半人為的幾何位置合わせマーカの研究 (第1報)”, 電子情報通信学会 技術研究報告 PRMU2006-195, 2007.
- [21] R. Cipolla, D. Robertson and B. Tordoff: “Image-based localization,” Proc. Int. Conf. Virtual Systems and Multimedia, pp. 22–29, 2004.
- [22] 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.
- [23] M. Kouroggi and T. Kurata: “Personal positioning based on walking locomotion analysis with self-contained sensors and wearable camera,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 103–112, 2003.
- [24] T. Drummond and R. Cipolla: “Real-time visual tracking of complex structures,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 27, pp. 932–946, 2002.
- [25] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: “Fully automated and stable registration for augmented reality applications,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 93–102, 2003.
- [26] L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3D camera tracking,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 48–57, 2004.
- [27] G. Bleser, H. Wuest and D. Stricker: “Online camera pose estimation in partially known and dynamic scenes,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 56–65, 2006.

- [28] 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.
- [29] 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.
- [30] 山本治由, 津田崇博, 塚田正人, 北原格, 亀田能成, 大田友一: “屋外における複合現実型情報提示の広域化 - 特性の異なるカメラ群の利用 -”, 電子情報通信学会 技術研究報告 MVE2006-67, 2006.
- [31] Z. Hu, K. Uchimura, H. Lu and F. Lamosa: “Fusion of vision, 3D gyro and GPS for camera dynamic registration,” Proc. Int. Conf. on Pattern Recognition, pp. 351–354, 2004.
- [32] 横地裕次, 池田聖, 佐藤智和, 横矢直和: “動画像と GPS による位置情報を用いたカメラ外部パラメータの推定”, 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 650–657, 2005.
- [33] B. Jiang, U. Neumann and S. You: “A robust hybrid tracking system for outdoor augmented reality,” Proc. IEEE Virtual Reality, pp. 3–10, 2004.
- [34] G. S. W. Klein and T. W. Drummond: “Tightly integrated sensor fusion for robust visual tracking,” Image and Vision Computing, Vol.22, pp. 769–776, 2004.
- [35] 松田幸大, 池田聖, 佐藤智和, 横矢直和: “ランドマークデータベースと姿勢センサを用いた高速な回転にロバストなカメラの位置・姿勢推定”, 電子情報通信学会 技術研究報告 ITS2006-76, 2007.
- [36] S. You and U. Neumann: “Fusion of vision and gyro tracking for robust augmented reality registration,” Proc. IEEE Virtual Reality, pp. 71–78, 2001.

- [37] 小竹大輔, 佐藤清秀, 内山晋二, 山本裕之: “傾斜角拘束を利用したハイブリッド位置合せ手法”, 電子情報通信学会論文誌 (D), Vol.J90-D, No. 8, pp. 2070–2080, 2007.
- [38] G. Reitmayr and T. Drummond: “Going Out: Robust model-based tracking for outdoor augmented reality,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 109–118, 2006.
- [39] 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.
- [40] D. P. Robertson and R. Cipolla: “Building architectural models from many views using map constraints,” Proc. European Conference on Computer Vision, pp. 155–169, 2002.
- [41] J. S.Beis and D.G.Lowe: “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces,” Proc. Conf. on Computer Vision and Pattern Recognition, pp. 1000–1006, 1997.
- [42] S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: “Construction of feature landmark database using omnidirectional videos and GPS positions,” Proc. Int. Conf. on 3-D Digital Imaging and Modeling, pp. 249–256, 2007.
- [43] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.

付録

A. LMedS 基準による誤対応の排除とカメラ位置・姿勢推定

LMedS 基準による誤対応の排除では、以下の処理を行う。

- (1) 対応付けられたランドマークの三次元座標と二次元座標の組からランダムに 6 組以上を選択し、暫定的なカメラパラメータを推定する。
- (2) 全ての組について、ランドマークの二次元座標と三次元位置を暫定的なカメラパラメータによって画像上に投影した座標との距離の二乗誤差を再投影誤差として求める。

ランドマーク i の二次元座標を x_i 、ランドマーク i の三次元座標を暫定的なカメラパラメータで画像上に投影した座標を \hat{x}_i とすると、ランドマーク i の再投影誤差 R_i と再投影誤差の中間値 RM は以下の式によって表される。

$$R_i = |x_i - \hat{x}_i|^2 \quad (10)$$

$$RM = \text{median}(R_1, R_2, \dots, R_n) \quad (11)$$

ただし、 n は対応付けられたランドマークの数とする。

- (3) 上記の (1), (2) の処理を複数回繰り返す、再投影誤差の中間値 RM が最小となる暫定カメラパラメータを選択する。
- (4) 選択されたカメラパラメータによって各ランドマークの投影誤差を評価し、閾値を超える結果を誤対応として削除する。

最後に、得られた正しいと判定された対応関係のみを用いて、再投影誤差が最小となるカメラパラメータを求める。