

修士論文

三次元点群への局所平面当てはめによる
法線情報付き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

川戸 一希

2011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。

川戸 一希

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)

三次元点群への局所平面当てはめによる 法線情報付き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*

川戸 一希

内容梗概

現実環境を撮影した画像に対してCGなどの仮想物体を重畳することで情報を付加するビデオシーズルー型拡張現実感においては、現実環境と仮想物体との幾何学的な位置合わせのためにカメラの位置・姿勢を推定する必要があり、従来から様々なカメラ位置・姿勢推定法が提案されている。中でもランドマークデータベースを用いた手法はStructure from Motion(SfM)法を用いて半自動でデータベースを構築できるため、低い人的コストで広域環境に適用できるという特徴を持つ。ただし、従来手法ではランドマークと自然特徴点の対応付け処理において、カメラの姿勢の変化のみを考慮した画像の補正を行っているため、ランドマーク取得時のカメラ位置から離れた地点において、見え方が大きく変化するランドマークを正しく対応付けることができず、推定処理が失敗するという問題があった。この問題に対して、本研究ではSfM法により得られた三次元自然特徴点群に局所平面を当てはめることで、新たにその法線方向を算出する。この法線情報を用いて、平面上に存在すると判定されたランドマークの見え方を入力画像の撮影位置・姿勢に応じて幾何学的に補正し、対応付けの精度を向上させることで、視点位置の変化に対するカメラ位置・姿勢推定の頑健性を向上させる。実験では提案手法の有効性を確認するために、ランドマーク取得時のカメラ位置から離れた地点にお

*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-MT0951033, 2011年3月17日.

いて，従来手法とカメラ位置・姿勢の推定精度を比較し，視点位置の変化に対する頑健性を検証する．

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

Extrinsic Camera Parameter Estimation Using Landmark Database with Normal of Local Plane Fitted to 3D Point Cloud*

Kazuki Kawato

Abstract

In the field of augmented reality which efficiently provides location-based information by overlaying virtual objects to the real world, it is important to solve a geometric registration problem between real and virtual worlds. To solve this problem, vision-based registration methods which result in estimating extrinsic camera parameters are widely used because they can achieve pixel-level geometric registration. As one of these methods, a feature landmark-based method has been proposed. In this method, a feature landmark database which is automatically constructed by the structure-from-motion (SfM) is used to estimate extrinsic camera parameters. Although this method can be easily applied to large-scale environments, an extrinsic camera parameter estimation process will fail where the camera position is far from the position of acquiring landmark information. This is due to the visual aspect change of landmark whose image templates of landmarks are compensated for by considering only captured camera pose. In this study, to achieve robust landmark matching against viewpoint change, planes are locally fitted to 3D point cloud acquired by the SfM. Viewpoint independent image templates of landmarks which are judged existing on local planes are then generated by projecting the input image to estimated local planes. By using

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0951033, March 17, 2011.

these image templates, user's movable areas are expanded more widely than that of the conventional method. To demonstrate the effectiveness of the proposed method, the accuracy of estimated camera parameters where the camera position is far from the position of acquiring landmark information is compared to the conventional method.

Keywords:

extrinsic camera parameter estimation, landmark database, natural feature, augmented reality

目次

1. はじめに	1
1.1 拡張現実感技術の概観	1
1.2 カメラ位置・姿勢推定手法の分類	4
1.2.1 カメラ以外のセンサに基づく手法	5
1.2.2 画像に基づく手法	6
1.2.3 カメラ以外のセンサと画像を併用した手法	12
1.3 視点位置の変化に頑健な自然特徴点の対応付け手法	14
1.4 本研究の位置づけと方針	16
2. 局所平面当てはめによる法線情報付きランドマークデータベースの構築	20
2.1 ランドマークデータベースの構成要素	20
2.2 全方位カメラを用いた環境の三次元復元	22
2.3 主成分分析による三次元点群への局所平面の当てはめ	22
2.4 局所平面を考慮したランドマーク情報の取得	23
2.4.1 多重スケールの画像テンプレートの生成	23
2.4.2 多重スケールの視点位置に依存しない画像テンプレートの生成	25
3. 法線情報付き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	28
3.1 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	28
3.2 法線情報に基づくランドマークの選択	29
3.3 自然特徴点とランドマークとの対応付けによるカメラ位置・姿勢推定	31
3.3.1 正規化相互相関による自然特徴点とランドマークとの対応付け	31
3.3.2 自然特徴点とランドマークとの対応関係を用いたカメラ位置・姿勢推定	32
4. 実験	34
4.1 ランドマークデータベースの構築	34

4.2	カメラ位置・姿勢の推定精度の比較	37
4.2.1	自然特徴点とランドマークとの対応付け精度の比較	38
4.2.2	仮想物体の重畳誤差の比較	42
5.	まとめ	45
	謝辞	46
	参考文献	47

目 次

1	拡張現実感の応用例	2
2	入力画像の画質を考慮した仮想物体の重畳 [9]	3
3	カメラ位置・姿勢推定法の分類	4
4	超音波センサを用いた位置・姿勢推定 [11]	5
5	Naimark らの円形マーカ [21]	8
6	景観を損ねないマーカ	8
7	航空写真を用いた車載カメラの位置・姿勢の推定 [15]	10
8	文字の重心から算出可能な特徴量を用いた紙面の入力画像の対応 付け [27]	10
9	自然特徴点の対応付け手法の分類	14
10	局所平面パッチを用いた三次元モデルの対応付け [41]	16
11	画像に基づく手法の位置づけ	17
12	提案手法の処理の流れ	19
13	ランドマークデータベースの構成要素	21
14	主成分分析の範囲	23
15	従来の画像テンプレートの作成方法	24
16	視点位置に依存しない画像テンプレートの生成方法	26
17	生成された視点位置に依存しない画像テンプレートの例(左 スケー ル 1 倍, スケール中央 5 倍, スケール右 9 倍)	27
18	カメラに対する画像テンプレートの傾き	30
19	ランドマークと自然特徴点との再投影誤差	33
20	全方位型マルチカメラシステムと全方位画像の 1 フレーム	35
21	推定された自然特徴点の三次元位置と全方位カメラの位置・姿勢	35
22	再投影されたランドマークとその法線情報 (1/2)	36
23	再投影されたランドマークとその法線情報 (2/2)	37
24	カメラの初期位置と推定されたカメラパス	38
25	各フレームにおけるランドマーク数の推移	40

26	選択されたランドマークの内，カメラ位置姿勢推定に用いられる割合の推移	40
27	提案手法が有効なフレームでのカメラ位置・姿勢推定に用いられたランドマークの比較：(赤) 自然特徴点と対応付けられなかったランドマーク (緑) 対応づいたが誤対応として排除されたランドマーク (青) 最終的なカメラ位置・姿勢推定に用いられたランドマーク	41
28	再投影誤差の計算に用いた真値の例	43
29	各フレームにおける再投影誤差の推移	43
30	提案手法によるカメラ位置・姿勢推定結果を用いた仮想物体の重畳表示	44

表 目 次

1	カメラ位置・姿勢推定に用いた機器	38
2	実験に用いたパラメータ	39
3	1 フレームあたりのランドマークの平均数	40

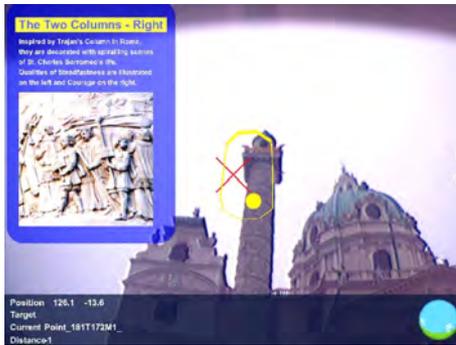
1. はじめに

近年，現実環境を撮影した画像中にCGなどによって表現される仮想環境を位置合わせして描画することで情報を付加するビデオシースルー型の拡張現実感技術に関する研究が盛んに行われている．このような拡張現実感において，現実環境と仮想環境との位置合わせを実現するためには，一般にカメラの位置・姿勢を推定することが必要となり，現在までに様々な手法が提案されている．中でも，ランドマークデータベースを用いた手法 [1] は Structure-from-Motion 法 (SfM 法) を用いることで広域な環境においても比較的容易にデータベースを構築できるという特長を持つ．しかし，従来手法はオンラインでのランドマークと自然特徴点の対応付け処理において，カメラの姿勢の変化のみを考慮した画像の補正を行っているため，ランドマーク取得時のカメラ位置から離れた地点において，見え方が大きく変化するランドマークを正しく対応付けることができず，推定処理が失敗するという問題が残されていた．そこで本研究では武富らの手法 [1] を基礎とし，新たに SfM 法により得られた三次元自然特徴点群に局所平面を当てはめることでその法線方向を算出する．この法線情報を用いて，平面上に存在すると判定されたランドマークに対して，入力画像を算出された局所平面上へ投影することで撮影位置・姿勢に依存しない画像パターンを生成する．これにより，対応付けの成功率を向上させ，視点位置の変化に対して頑健なカメラ位置・姿勢推定を実現する．

以下では，まず本研究の背景となる拡張現実感技術について概観する．次に拡張現実感におけるカメラ位置・姿勢推定に関する従来研究，および視点位置の変化に頑健な自然特徴点の対応付け手法について述べる．最後に，本研究の位置づけと方針について明確にする．

1.1 拡張現実感技術の概観

拡張現実感技術は，図 1 に示すようにナビゲーション [2,3]，文化遺産の仮想復元 [4-6] などを含む様々な分野への応用が期待されている．このような拡張現実感を実現するためには，以下に示す現実環境と仮想環境の幾何学的整合性，光学



(a) ナビゲーションへの応用 [2]



(b) 文化遺跡の復元への応用 [5]

図 1 拡張現実感の応用例

的整合性，時間的整合性の問題を解決する必要がある．

- 幾何学的整合性

現実環境と仮想環境との位置合わせや現実物体と仮想物体の隠蔽関係などの幾何学的な整合性に関する問題である．中でも，現実環境と仮想環境の位置合わせに関する問題の解決は拡張現実感では必要不可欠である．位置合わせの問題を解決するためには，一般に，ユーザ視点の位置・姿勢や合成対象となる画像を撮影したカメラの位置・姿勢を推定する必要がある．

- 光学的整合性

現実環境と仮想環境の間の画質のずれや現実環境と仮想環境の間の照明条件の不一致によって生じる陰影や写り込みなどに関する不整合の問題である．この問題は，文化遺産の復元 [6] やインテリアデザイン [7] などの写実的な画像の合成を必要とするアプリケーションで重要である．光学的整合性の問題を解決するために，利用環境中にカメラや鏡面球を設置することで実世界の光源環境を推定する手法 [6,8] や，図 2 に示すような，撮影画像に生じるモーションブラーやローリングシャッターによるノイズなどを仮想物体に反映することで重畳される仮想物体と撮影画像間の画質を一致させ



(a) 画質の考慮なし



(b) 画質の考慮あり

図 2 入力画像の画質を考慮した仮想物体の重畳 [9]

る手法 [9, 10] が提案されている。

- 時間的整合性

現実環境と仮想環境の間の時間ずれに関する問題である。時間的整合性問題は、計算機による仮想空間の合成に必要な処理時間に起因するため、完全な解決は困難な問題である。特に光学シースルー型拡張現実感では現実環境は直接ユーザに提示されるのに対して、仮想環境は計算機で処理された後にユーザに提示されるため、現実環境と仮想環境に時間のずれが知覚されやすいという問題がある。一方、ビデオシースルー型拡張現実感では現実環境と仮想環境の両方が計算機で処理されるため、合成画像の提示には遅延が生じるが、提示される画像上においては現実環境と仮想物体の間に時間のずれが生じないという特長がある。

本研究は、広域な屋外環境における頑健なカメラ位置・姿勢推定を目指したものであり、これは上記の幾何学的整合性問題の一部を解決することを目的としている。次節では、これ関連するカメラ位置・姿勢推定の従来研究について詳しく述べる。

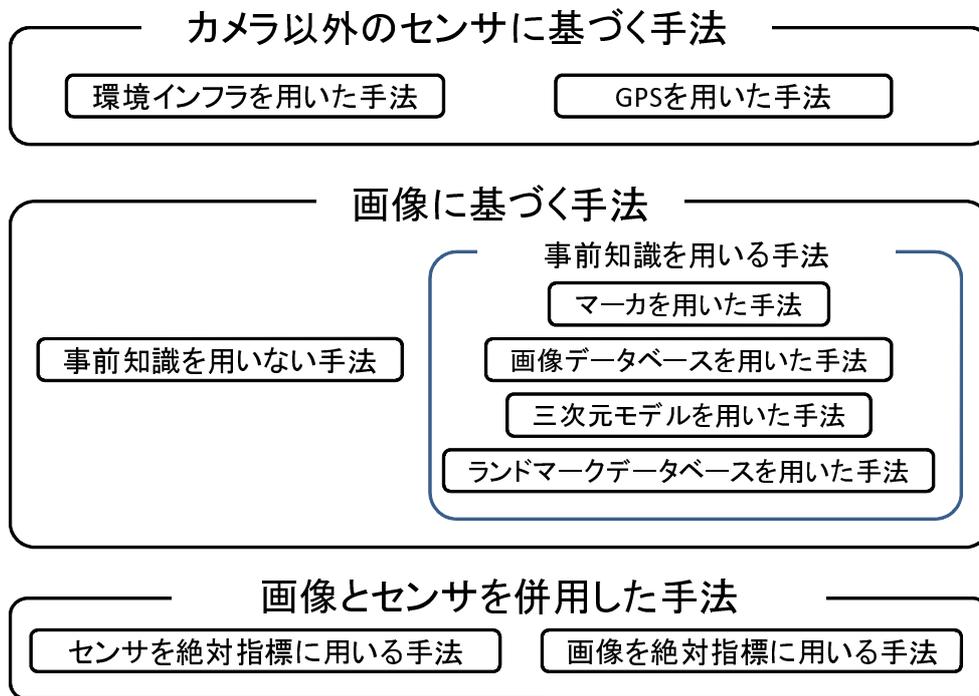
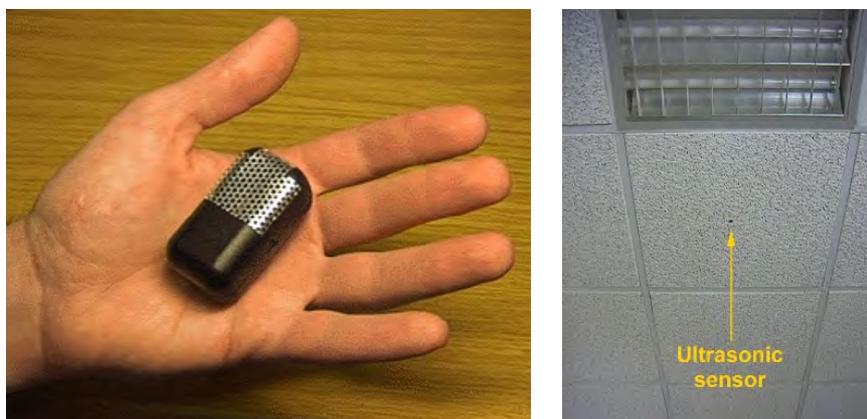


図 3 カメラ位置・姿勢推定法の分類

1.2 カメラ位置・姿勢推定手法の分類

拡張現実感における合成画像の提示方法は、光学シースルー型とビデオシースルー型に大別できる。光学シースルー型拡張現実感は、機器を通さずに現実環境を直接ユーザに提示できるため、現実環境の提示に時間の遅れは生じないが、高精度な位置合わせや時間的整合性問題の解決が難しいという問題がある。一方で、ビデオシースルー型拡張現実感は、整合性問題（幾何学的整合性、光学的整合性、時間的整合性）を比較的容易に解決することができるため、近年の拡張現実感で主流となっている。そこで、本節では、ビデオシースルー型拡張現実感における位置合わせの従来手法について概説する。ビデオシースルー型拡張現実感では、現実環境と仮想環境との位置合わせを実現するために、合成対象となる画像を撮影したカメラの位置・姿勢を推定する必要がある。カメラの位置・姿勢を推定する手法は、図 3 に示すように、GPS やジャイロなどのカメラ以外のセンサに基づ



(a) 超音波発信器

(b) 超音波センサ

図 4 超音波センサを用いた位置・姿勢推定 [11]

く手法 [2, 3, 11, 12] , カメラで撮影された画像に基づく手法 [13–17] , GPS やジャイロなどのカメラ以外のセンサと画像を併用した手法 [18, 19] に分類できる。以下では、それぞれの手法の特徴について述べる。

1.2.1 カメラ以外のセンサに基づく手法

カメラ以外のセンサに基づく手法は、環境に設置された絶対位置・姿勢が推定可能なインフラを用いた手法と GPS と相対的な姿勢が取得可能なセンサを併用した手法に大別することができる。

インフラを用いた手法は、環境中に超音波センサ、無線 LAN、磁気センサ、RFID などの機器を設置することによって絶対位置・姿勢を推定する。Harter ら [11] は図 4 のように、ユーザが超音波発信器を持ち、そこから発信される超音波の強度を環境中に設置した複数の超音波センサによって計測することで位置・姿勢を推定する手法を提案している。この手法は頑健にカメラの位置・姿勢を推定することができるが、広域な環境に適用するためには、多数のセンサを設置する必要があり、機器の設置や維持にかかる人的コストが膨大になるという問題がある。これに対して、天目ら [20] はこれらのインフラに加えて、歩数計を用いてインフラ

間でのユーザの相対的な動きを推定することで、疎にインフラが配置された環境においても位置計測が可能な手法を提案している。しかし、広域な環境に適用する場合には、依然として機器の設置や維持にかかる人的コストが大きくなるという問題が残る。

一方、Feinerら [2] は GPS とジャイロセンサを併用し、広域な屋外環境においてカメラの位置・姿勢を推定する手法を提案している。この手法は位置・姿勢を取得するためのインフラを環境中に設置する必要がないが、一般的な GPS の計測レートは低いため、カメラのすべてのフレームにおいて位置情報を算出することは難しい。これに対して、神原ら [12] は計測レートの高い慣性航法装置を併用することで GPS から位置情報を取得できない場合においてもカメラの位置・姿勢を推定できる手法を提案している。しかし、GPS を用いる手法 [2, 12] は、周辺に GPS の電波を遮るような遮蔽物が存在する場合、カメラ位置・姿勢の推定精度が低下するという問題がある。

カメラ以外のセンサに基づく手法は、頑健なカメラ位置・姿勢の推定が可能であるが、カメラとセンサとのキャリブレーション誤差に起因して高精度な位置合わせが難しいという問題がある。

1.2.2 画像に基づく手法

画像に基づく手法は、事前知識を用いない手法と事前知識を用いた手法に大別することができる。また、事前知識を用いた手法は、マーカを用いた手法 [21–23]、画像データベースを用いた手法 [14, 15]、三次元モデルを用いた手法 [16, 17]、ランドマークデータベースを用いた手法に分類することができる。以下では、それぞれの手法について詳細に述べる。

事前知識を用いない手法

事前知識を用いない手法 [13, 24] は、一般に、visual-SLAM (Simultaneous Localization and Mapping) と呼ばれ、入力画像中の自然特徴点を追跡することで、利用環境の三次元復元とカメラ位置・姿勢の推定を実時間で同時に行う。visual-SLAM による手法は、手軽に拡張現実感環境を構築することができるが、三次元

復元のための最適化処理の計算コストが大きく、推定処理において、多くの自然特徴点を扱うことができないという問題がある。この問題に対して、Klein らの PTAM(Parallel Tracking and Mapping) [24] はカメラ位置・姿勢推定処理と三次元復元のための最適化処理を並列化し、多くの自然特徴点を利用可能とすることで、安定したカメラ位置・姿勢推定を実現している。しかし、これらの手法は、それ単体では基準となる座標系を事前に定義することができないため、ナビゲーションなど現実環境の位置情報に依存した情報を提示するアプリケーションへ適用することは難しいという問題がある。

この問題を解決するために、Lee ら [25] は人の手をインタフェースとし、ユーザがインタラクティブに仮想物体を配置する HandyAR を提案している。しかし、広域な環境において、インタラクティブに仮想物体を配置することは、ユーザにとって大きな負担になるという問題がある。

マーカを用いた手法

マーカを用いた手法は、ARToolKit [26] で用いられる正方形マーカなど、形状と三次元位置が既知のマーカを利用環境中に設置し、これらのマーカをカメラで撮影することでカメラの位置・姿勢を推定する。マーカを用いた手法は、カメラで認識しやすい人工的なマーカを用いることで、頑健なカメラ位置・姿勢推定を実現している。Naimark ら [21] は、図 5 に示すように、検出が容易な円形のマーカを利用環境中に配置することでカメラ位置・姿勢の推定を実現している。しかし、この手法では利用環境中に多数のマーカを配置する必要があり、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。景観を損ねるという問題に対して、齊藤ら [22] や中里ら [23] は図 6 に示すように、利用環境に調和するようなマーカや再帰性反射材を用いた不可視マーカを用いることで、景観を損ねないマーカを用いたカメラ位置・姿勢推定の手法を提案している。

これらの手法を用いることで、景観を損ねるという問題は緩和できるが、広域環境に適用する場合には、依然として、マーカの設置に多くの人的コストが必要になるという問題は残されている。



(a) マーカを配置した環境

(b) 配置された円形マーカ

図 5 Naimark らの円形マーカ [21]



(a) 部屋の内装に適したマーカ [22]



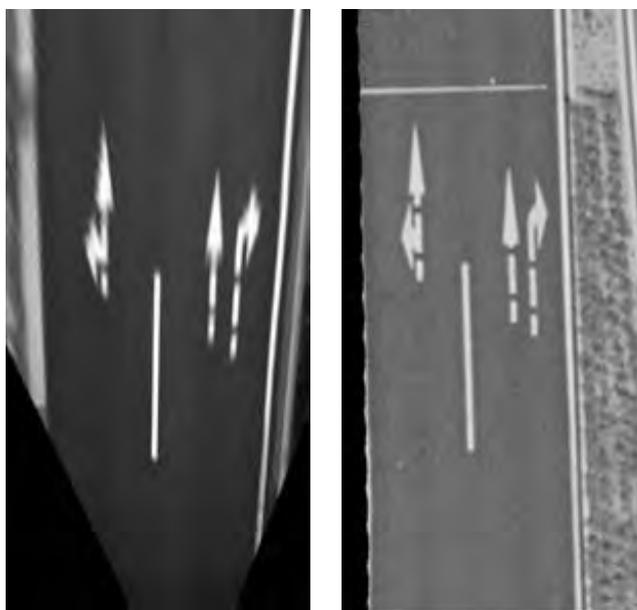
(b) 再帰性反射材からなる不可視マーカ [23]

図 6 景観を損ねないマーカ

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た手法は、事前に撮影時の位置・姿勢情報を付加した画像をデータベースへ登録しておき、入力画像とデータベース中に存在する画像を対応付けることでカメラ位置・姿勢を推定する。Cipollaら [14] は、入力画像とデータベース画像とを対応付け、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を、正対した位置から見ているように画像の補正を行った後、データベース画像と比較をすることで、入力画像とデータベース画像を対応付けている。この手法では、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、複雑な環境への適用は難しい。また、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ないという問題がある。野田ら [15] は航空写真をデータベースとして用い、航空写真に写っている路面領域と車載カメラで撮影した入力画像中の路面領域とを対応付けることで、車載カメラの位置・姿勢を推定する手法を提案している。この手法では、図7に示すように、車載カメラによって撮影された路面領域のテクスチャを射影変換することで俯瞰画像を生成し、航空写真中の路面領域と入力画像中の路面領域の対応付けを行っている。この手法は、広域な環境を効率的に撮影できる航空写真を事前知識としているため、データベース構築の人的コストが低いという特長を持つ。しかし、路面領域と車載カメラの相対関係は一定であるという仮定の元で射影変換を行っているため、ヒューマンナビゲーションのようにカメラの姿勢が大きく変化する拡張現実感への応用は難しい。

一方で、Uchiyamaら [27] は、テーブルトップ上での、文書を対象とした拡張現実感を実現するために、文書の画像をデータベースとして用いた手法を提案している。この手法は、図8に示すように、文書画像中の文字の重心位置の幾何学的な配置から算出可能な特徴量を用いて、データベース画像と入力画像との対応付けを行うことで、カメラ位置・姿勢を推定する。この手法は、拡張現実感を用いた絵本、紙面上への広告・注釈等の紙面に特化した拡張現実感アプリケーションへの応用が考えられる。



(a) 射影変換によって変形した車載カメラ画像

(b) 航空写真

図 7 航空写真を用いた車載カメラの位置・姿勢の推定 [15]



図 8 文字の重心から算出可能な特徴量を用いた紙面の入力画像の対応付け [27]

三次元モデルを用いた手法

三次元モデルを用いた手法 [16, 17] は、事前に利用環境の三次元モデルを作成しておき、入力画像中の自然特徴点やエッジを三次元モデルと対応付けることでカメラの位置・姿勢を推定している。Drummondら [28] は事前に環境のワイヤーフレームモデルを作成し、ワイヤーフレームモデルと入力画像中のエッジを対応付けることでカメラの位置・姿勢を推定している。しかし、入力画像中にエッジが多く存在する場合、誤対応が起こりやすく、カメラ位置・姿勢の推定精度が低下するという問題がある。これに対して、Vacchettiら [16] はワイヤーフレームモデルを用いたカメラ位置・姿勢推定の手法に、自然特徴点の追跡処理を組み合わせることで、入力画像中にエッジが多く存在する場合においても、頑健にカメラの位置・姿勢を推定する手法を提案している。しかし、これらの三次元モデルを用いた手法は、事前に利用環境の三次元モデルを手で作成する必要があるため広域な環境に適用する場合、人的コストが膨大になるという問題がある。

三次元モデルの作成コストを低減するために、Bleserら [29] は visual-SLAM を併用することで、三次元モデルが写らない場合においても自然特徴点を追跡し、カメラの位置・姿勢を推定する手法を提案している。この手法は、visual-SLAM による相対的なカメラ位置・姿勢推定処理を組み合わせることで、利用環境全体のモデルを作成するのではなく、部分的なモデルを作成することで、事前知識獲得のためのコストを低減できる。しかし、長時間モデルが写らない場合には推定誤差が蓄積するため、広域環境へ適用するには、やはり多くの人的コストが必要となる。

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は、利用環境中に存在する自然特徴点の三次元位置とそれらの周辺の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、入力画像中の自然特徴点とデータベース内のランドマークを対応付けることでカメラの位置・姿勢を推定する。Skrypnikら [30] や Arthら [31] は、複数枚の画像から半自動で三次元復元を行うことでデータベースの構築を行っている。しかし、画像情報のみを用いてデータベースの構築を行っているため、広域な環境へ適用する場合には、三次元復元の復元結果に誤差が蓄積するという問

題がある．

一方で，データベース構築時に，三次元位置が既知の基準点や GPS の測位情報を用いて，蓄積誤差の生じない三次元復元を行うことで，広域な環境に適用可能なカメラ位置・姿勢推定の手法が提案されている [1]．しかし，この手法ではランドマークと自然特徴点の対応付け処理において，カメラの姿勢の変化のみを考慮し，ランドマーク周辺の画像情報の補正を行っているため，ランドマーク取得時のカメラ位置から離れた地点において，見え方が大きく変化するランドマークを正しく対応付けることができず，推定処理が破綻するという問題がある．

この問題に対して，武富ら [32] はレンジセンサで取得した密な奥行き情報を併用し，幾何学的にランドマーク周辺の画像情報を補正することで対応付け精度の向上を図っている．しかし，密な奥行き情報の取得には多くの人的コストが必要となるため，広域環境への適用は困難である．また，Wu ら [33] は MVS (Multi Views Stereo) によって得られた三次元形状を用いて，自然特徴点周辺の局所形状に対して平面を当てはめた後，その平面に対して入力画像を投影し，視点位置に依存しない画像パターンを生成することで，視点位置の変化に対して頑健なカメラ位置・姿勢推定の手法を提案している．しかし，法線を求めることができないような環境では，ランドマークは自然特徴点と正しく対応付けることができず，カメラ位置・姿勢の推定精度が低下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た手法では，頑健なカメラ位置・姿勢推定を実現するためには，入力画像中の自然特徴点とデータベース中のランドマークを頑健に対応付ける必要がある．そこで視点位置の変化に頑健な自然特徴点の対応付け手法について 1.3 節で述べる．

1.2.3 カメラ以外のセンサと画像を併用した手法

カメラ以外のセンサと画像を併用した手法 [18] は，カメラ以外のセンサに基づく手法と画像に基づく手法の問題点を補い，頑健で高精度なカメラ位置・姿勢推定を実現している．これら手法は，カメラ以外のセンサを絶対指標として用いる手法と画像を絶対指標として用いる手法に分類することができる．

カメラ以外のセンサを絶対指標として用いる手法は画像を補助的に利用するこ

とで、カメラ位置・姿勢推定の精度の向上を図っている。Kimら [18] はGPSとジャイロを用いて大局的な位置・姿勢を推定し、航空写真から推定されるワイヤフレームモデルと入力画像中の建物のエッジとを対応付けることで、広域な環境に容易に適用可能なカメラ位置・姿勢推定の手法を提案している。しかし、航空写真から簡易な三次元モデルを推定しているため、利用環境の周囲に複雑な形状の建物が多く存在する場合、カメラ位置・姿勢の推定精度が低下するという問題がある。糸ら [34] はGPSを併用することで、広域環境に適用可能なSfM法を提案している。この手法では三次元復元の最適化処理においてGPSの測位情報を用いることで蓄積誤差を含まないカメラ位置・姿勢推定を実現している。しかし、撮影された動画全体での最適化処理を必要とするため、実時間処理が必須の拡張現実感のようなアプリケーションには適さない。

画像を絶対指標として用いる手法は、カメラ以外のセンサを補助的に用いることで、カメラ位置・姿勢推定の頑健性を向上させている。小竹ら [19] はジャイロから取得できる重力方向に対するカメラの傾きの情報を用いることで、推定すべきパラメータの数を削減し、線形計算により低い計算コストでカメラ位置・姿勢を推定する手法を提案している。しかし、この手法は自然特徴点の対応関係は正しく取得されていることを前提としているため、マーカの角などの認識精度の高い自然特徴点を用いる必要がある。一方で、Kleinら [35] はジャイロを併用し、入力画像中のマーカに発生したモーションブラーを推定することで、高速なカメラの移動に対して頑健な手法を提案している。しかし、広域な環境に適用する場合、マーカ設置の人的コストが大きくなるという問題がある。また、Youら [36] はジャイロから得られる角加速度情報によってランドマークの出現位置を予測し、ランドマーク探索範囲を限定する手法を提案している。しかし、ランドマークデータベースを用いた手法での問題と同様に、ランドマーク構築時のカメラ位置から離れた場合など、環境によって自然特徴点とランドマークとを正しく対応付けることができず、カメラ位置・姿勢の推定精度が低下するという問題が残されている。

カメラ以外のセンサを絶対指標とする手法は、画像を併用することで位置合わせの精度を高めることができる。しかし、カメラ以外のセンサを絶対指標とする場合、センサによって利用環境に制限があるため、様々な環境に適用するために

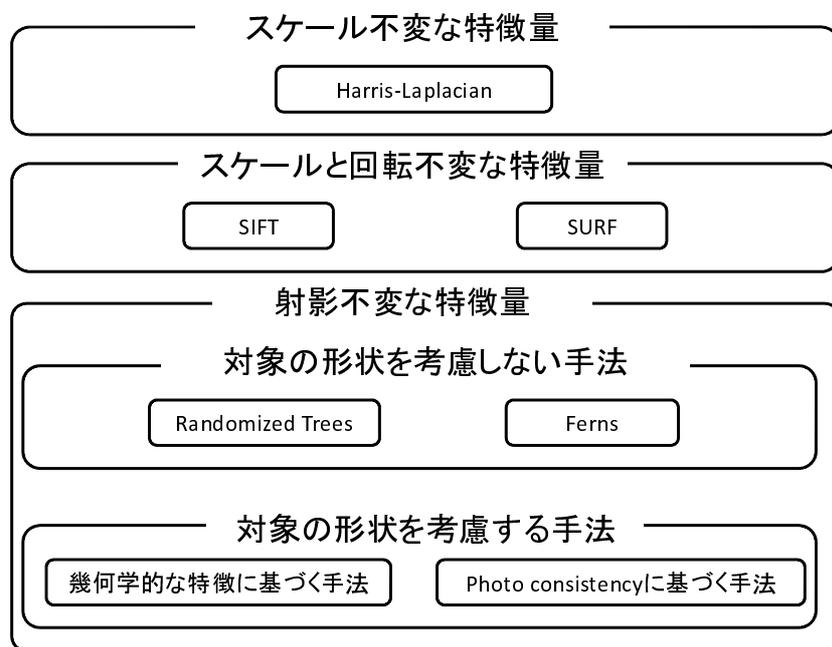


図 9 自然特徴点の対応付け手法の分類

は多種類のセンサを併用する必要がある。一方、画像を絶対指標とする手法は、カメラ以外のセンサを併用することで、頑健なカメラ位置・姿勢推定を実現している。しかし、利用環境の構築にかかる人的コストの問題や、環境によってカメラ位置・姿勢の推定精度が低下する問題が残されている。

1.3 視点位置の変化に頑健な自然特徴点の対応付け手法

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を頑健に対応付ける必要がある。そこで、本節では、視点位置が大きく異なる画像間で頑健な自然特徴点の対応付けを実現可能な従来手法について述べる。これらの手法の分類を図9に示す。

Mikolajczyk ら [37] は二次元画像上でのスケール変化に対して頑健な Harris-Laplace 検出器を提案している。この手法では、DoG(Difference of Gaussians) フィルタを用いて自然特徴点に固有なスケール情報を算出することで、スケー

ルに対して不変な特徴量を抽出し，自然特徴点の対応付けを行っている．Loweら [38] は，さらにこの手法を発展させ，二次元画像上でのスケール変化と回転に対して不変な SIFT 特徴量を提案している．この手法は，自然特徴点周辺の輝度勾配情報から算出される特徴量を用いて自然特徴点の対応付けを行っている．しかし，SIFT 特徴は自然特徴点の抽出処理に多くの計算コストが必要となるという問題があった．これに対して，Bayらは，自然特徴点の抽出に用いられる DoG フィルタをヘッセ行列で近似することで，高速に自然特徴点を抽出する SURF 特徴 [39] を提案している．しかし，これらの視点位置の変化によって射影歪みが発生した場合には，正しく対応付けることができない．

一方で，射影歪みが生じた場合にも，自然特徴点の頑健な対応付けを実現する手法が提案されている．Lepetitら [40] は視点位置の異なる地点の画像を仮想的に生成しておき，Randomized Trees によってそれらを学習することで，視点位置の変化に対して頑健な対応付けを実現している．また，Özuysalら [40] は仮想的に生成した視点位置の画像群から不変特徴量を抽出することで，より頑健に対応付ける手法 (Ferns) を提案している．しかし，これらの手法では，実際の幾何形状を考慮していないため，一部の視点位置では，自然特徴点を対応付けることができないという問題がある．

一方，Wuら [41] は図 10 のように MVS (Multi Views Stereo) によって得られた三次元形状を用いて，自然特徴点周辺の局所形状に対して平面を当てはめた後，その平面に対して入力画像を投影し，視点位置に依存しない画像パターンを生成することで，得られる自然特徴点の対応関係を求める手法 (VIP) を提案している．しかし，この手法では，平面上に存在しない自然特徴点は正しく対応付けることができないため，複雑な環境では自然特徴点の対応付けの精度が低下するという問題がある．また，Furukawaら [42] は MVS の対応点探索処理において，物体の局所形状に当てはまる平面を，観測画像から見積もられる Photo Consistency に基づいて決定することで頑健な対応付けを実現している．しかし，Wuらの手法と同様に，平面上に存在しない自然特徴点は正しく対応付けることができないため，法線を求めることができないような場所では自然特徴点の対応付けの精度が低下するという問題がある．また，再帰的に仮定する平面の方向を変え，当ては



図 10 局所平面パッチを用いた三次元モデルの対応付け [41]

まる平面の方向を探索するため，計算コストが大きく，実時間処理が必要な拡張現実感には適していない．

このように，自然特徴点の対応関係を求める手法は多く提案されているが，複雑な環境においても，自然特徴点の対応関係を正しく求められる手法は存在しない．

1.4 本研究の位置づけと方針

本節では，複雑な広域屋外環境下での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．広域な屋外環境において頑健なカメラ位置・姿勢を推定するためには，拡張現実感環境の構築に必要な人的コストと位置合わせの精度の問題を解決する必要がある．1.2節で概観したように，これまでにカメラの位置・姿勢を推定する手法は数多く提案されているが，これら2つの問題を同時に解決することは難しい．これら2つの問題を同時に解決するためには，カメラ以外のセンサ，および，画像に基づく手法の双方を改善する必要がある．本研究では画像に基づく手法に着目し，広域屋外環境へ適用可能な手法の実現を目指す．

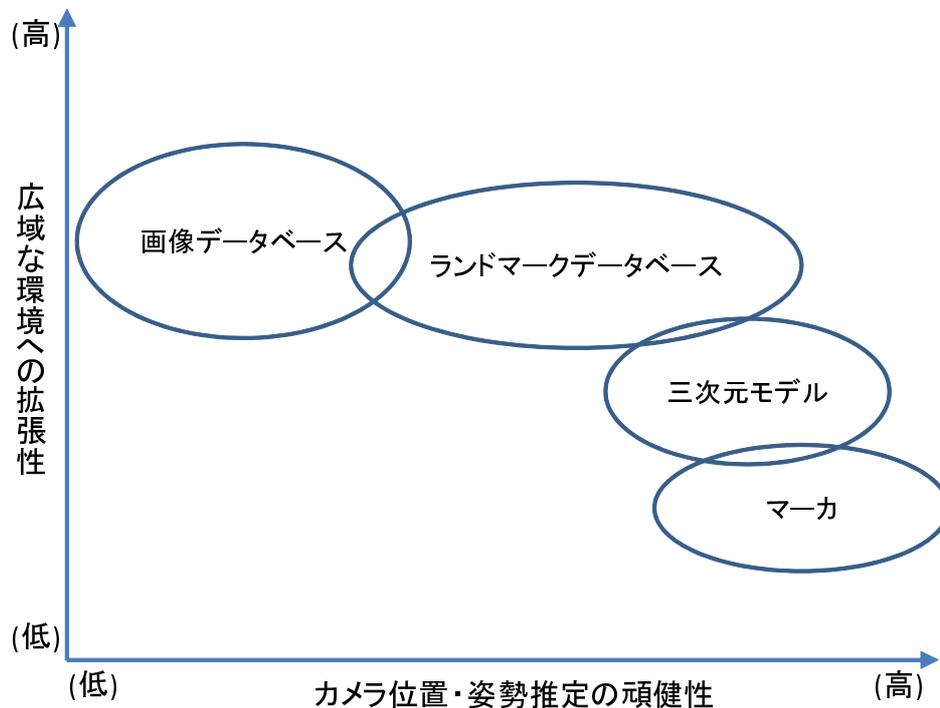


図 11 画像に基づく手法の位置づけ

画像に基づく手法のカメラ位置・姿勢推定の頑健性と広域な環境への拡張性の関係を図 11 に示す。マーカを用いた手法、および、三次元モデルを用いた手法は、拡張現実感環境を構築するための人的コストが大きく、広域な環境には適していない。また、画像データベースを用いた手法は、カメラ位置・姿勢推定の精度が低く、拡張現実感へ適用することは難しい。一方、ランドマークデータベースを用いた手法は、比較的低い人的コストでデータベースを構築でき、特定の環境では高精度にカメラの位置・姿勢を推定することができる。そのため本研究では、ランドマークデータベースを用いた手法 [1] に着目する。

武富らの手法 [1] では、事前に多くの地点からランドマークを撮影しておくことで、複雑な環境においてもカメラ位置・姿勢を推定することができる。しかし、データベース構築時のカメラパスから離れた地点では、自然特徴点とランドマークを正しく対応付けることができず、カメラ位置・姿勢推定の精度が低下する。

これに対して，Wuらの手法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，頑健に自然特徴点とランドマークを対応付けることができる．しかし，複雑な環境では，自然特徴点とランドマークを正しく対応付けることができず，カメラ位置・姿勢推定の精度が低下する．これに対して，本研究では，武富らの手法を基礎とし，Wuらの手法を適切に融合することで，広域な環境においても頑健なカメラ位置・姿勢の推定を実現する．また，法線が推定できないような環境に対しては，武富らの手法と同様に，複数の地点から撮影した見え方を持つことで対応する．

武富らの手法では，自然特徴点とランドマークとの対応付けの処理において，カメラに正対した平面を仮定し，そこに入力画像を投影することで，視線方向に依存しない画像パターンを生成し，自然特徴点とランドマークとの対応付けを行っている．そのため，ランドマーク取得時のカメラ位置から離れた地点において，見え方が大きく変化するランドマークを正しく対応付けることができず，推定処理が失敗する．視点位置の変化に対して頑健に自然特徴点を対応付けるには，周辺の幾何形状を考慮した変形が必要であるが，広域な環境の三次元モデルの作成には膨大な人的コストが必要となるという問題がある．

そこで，本研究では，SfM法により得られた自然特徴点群に局所平面を当てはめることで，幾何形状を考慮した画像テンプレートを生成する．提案手法の処理の流れを図12に示す．図中の太枠は本研究で改良した処理を表す．提案手法では，まず，オフラインでのデータベース構築処理において，主成分分析に基づく三次元点群への局所平面当てはめによって各自然特徴点に対応する法線情報を算出する．次に，この法線情報を用いて，平面上に存在すると判定されたランドマークに対して，算出された局所平面上へ入力画像を投影することで，撮影位置に依存しない画像テンプレートを生成する．その後，オンライン処理でのカメラ位置・姿勢推定処理において，生成した撮影位置に依存しない画像テンプレートを用いることで，自然特徴点とランドマークとの対応付けの精度を向上させ，視点位置の変化に対して頑健なカメラ位置・姿勢推定を実現する．また，平面上に存在しないランドマークには従来手法を適用し，対応付けを行うことで複雑な形状を持つ環境に対応する．

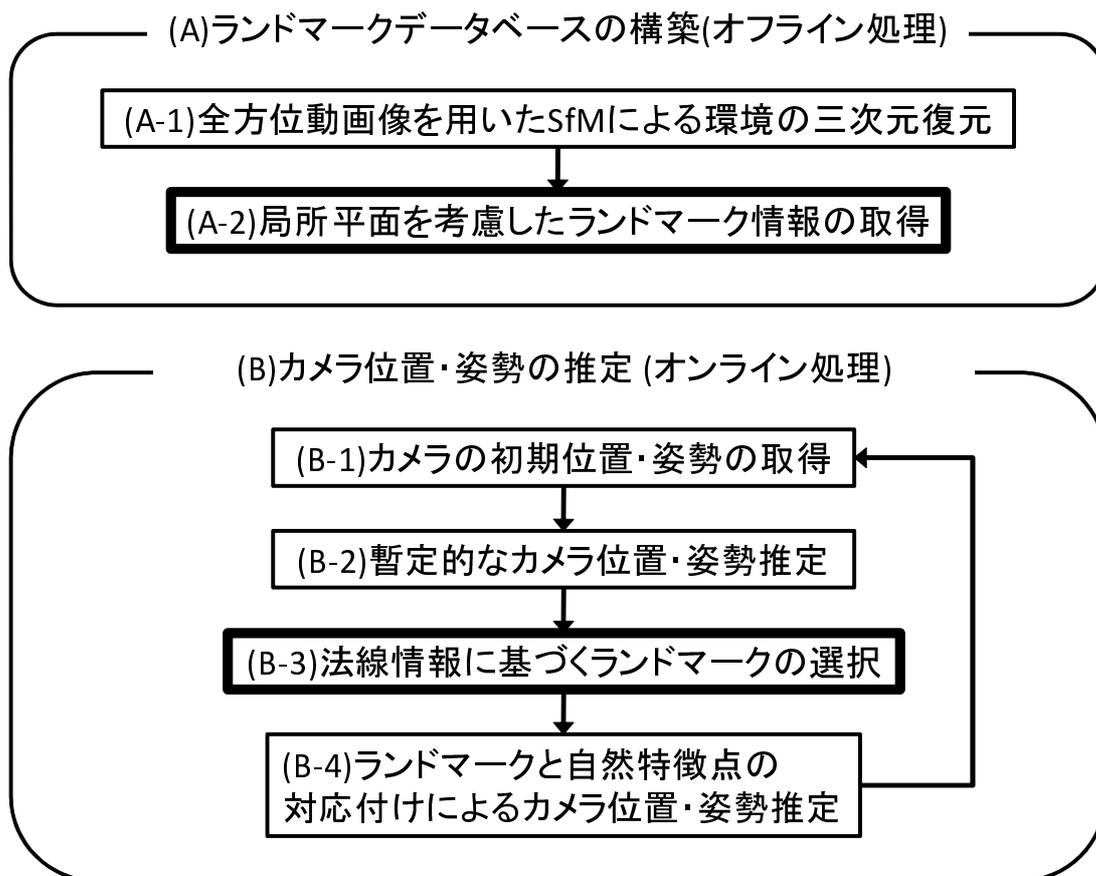


図 12 提案手法の処理の流れ

2. 局所平面当てはめによる法線情報付きランドマークデータベースの構築

本研究では，ランドマークデータベース構築時のカメラパスから離れた場合においても頑健なカメラ位置・姿勢推定を実現するために，新たにランドマークの法線情報を付加したデータベースを構築する．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，図 12 で示したランドマークデータベース構築処理である，(A-1) 全方位動画像を用いた環境の三次元復元，(A-2) 局所平面を考慮したランドマーク情報の取得，について述べる．

2.1 ランドマークデータベースの構成要素

本研究では，従来手法 [1] で用いられているデータベースに新たにランドマークの法線情報を付加したデータベースを用いる．図 13 に示すように，各ランドマークは (A) ランドマークの三次元位置，(B) ランドマークの法線情報，(C) 撮影地点ごとの情報によって構成される．また，撮影地点ごとの情報は，(i) 撮影時の全方位カメラの位置，(ii) ランドマークを中心とした多重スケールの画像テンプレート，(iii) 多重スケールの視点位置に依存しない画像テンプレート，(iv) 画像テンプレートの基本スケールから成る．以下にそれぞれの要素のについて詳述する．

(A) ランドマークの三次元位置

オンラインでのカメラ位置・姿勢推定時において，入力画像中の自然特徴点とランドマークデータベース内のランドマークを対応付け，それらの二次元座標と三次元位置の関係からカメラ位置・姿勢を推定する際に用いられる．ランドマークの三次元位置の情報は次節で述べる三次元復元によって得られ，環境に固定された世界座標系で保持される．



図 13 ランドマークデータベースの構成要素

(B) ランドマークの法線情報

三次元復元によって得られた自然特徴点の三次元点群に対して、後述する局所平面当てはめを用いることで得られる、ランドマークが平面上に存在するかどうかの真偽値、および、真である場合には平面の法線方向を保持する。法線情報は視点位置に依存しない画像テンプレートの生成に用いられる。

(C) 撮影地点ごとの情報

入力画像中の自然特徴点とランドマークデータベース内のランドマークを対応付けるために、ランドマークを中心とした画像テンプレートを保持する。ランドマークが平面上に存在しない場合、位置に依存しない画像テンプレートを生成することは難しいため、従来手法と同様に、複数の撮影地点での画像テンプレートとその撮影位置を保持することで撮影地点の変化による見え方の変化に対応する。一方、平面上に存在すると判定されたランドマークに対しては、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正しくランドマークを対応付けるために、後述する手法を用いて視点位置に依存

しない画像テンプレートを生成し，保持しておく．

2.2 全方位カメラを用いた環境の三次元復元

本研究では，佐藤らの手法 [43] または Ikeda らの手法 [44] を用いることにより，広域で複雑な環境においても蓄積誤差の生じない三次元復元を行う．これらの手法では，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，SfM 法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外部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 [43]，または，GPS から得られる位置情報 [44] を利用して，動画像全体での最適化処理を行い，カメラの外部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とで，蓄積誤差の生じない三次元復元を行う．

2.3 主成分分析による三次元点群への局所平面の当てはめ

本研究では，視点位置に依存しない画像テンプレートを生成するために，SfM 法によって取得された三次元点群に対して，主成分分析を用いて局所平面を当てはめ，平面上に存在するランドマークの判定とその法線方向の算出を行う．具体的には，まず，図 14 に示すように，ランドマーク $L_i(x_i, y_i, z_i)$ を中心とする半径 r 内に存在するランドマーク群 Ω_i の三次元座標を主成分分析することで，第一，第二，第三主成分に対応する固有値 $\lambda_1, \lambda_2, \lambda_3$ と固有ベクトル u_1, u_2, u_3 を算出する．ランドマーク群 Ω_i が同一平面上に分布している場合，第三主成分の方向の単位ベクトル u_3 は当てはめるべき平面の法線方向を表し，全ての点が平面上に存在する場合，固有値 λ_3 は λ_2 に比べ，非常に小さな値を取る．そこで，本研究では，以下の条件を用いて平面上に存在するランドマークを検出する．

$$|\lambda_3| < |\lambda_2|/m \quad (1)$$

ここで， m は正の値を取る定数である．

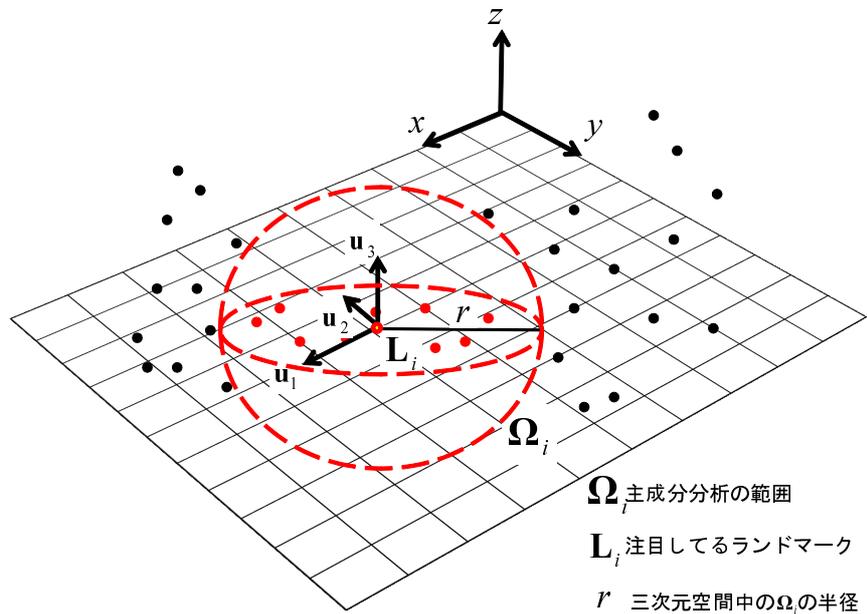


図 14 主成分分析の範囲

2.4 局所平面を考慮したランドマーク情報の取得

2.3節で述べた手法を用いることで、平面上に存在すると判定されたランドマークに対して、視点位置に依存しない画像テンプレートを生成する。ただし、オンライン処理において、カメラ位置から遠いランドマークに対して視点位置に依存しない画像テンプレートを生成すると、空間分解能の低下により、大きく歪んだ画像テンプレートが生成される。そのため、本研究では平面上に存在すると判定されるランドマークを含む全てのランドマークに対して、従来手法と同様の方法により多重スケールの画像テンプレートを生成する。以下では、具体的な画像テンプレートの生成方法について述べる。

2.4.1 多重スケールの画像テンプレートの生成

視線方向の変化による見え方の変化に対応するために、従来手法と同様の方法で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、図 15 に示

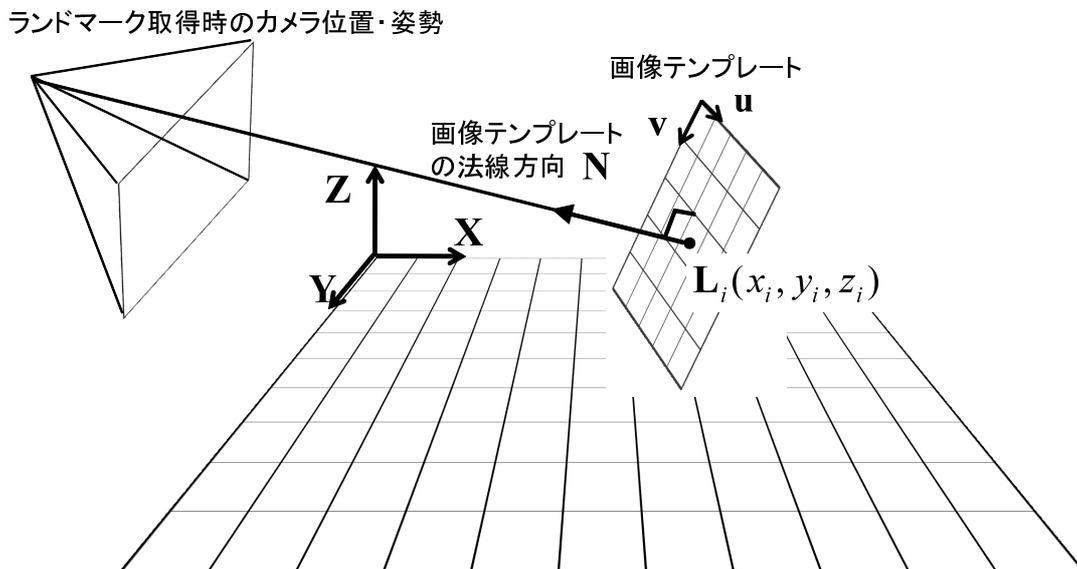


図 15 従来の画像テンプレートの作成方法

すように，世界座標系において，以下の条件を満たす正方形の平面を仮定し，この面上に入力画像を投影することで画像テンプレートを作成する．

- ランドマークの三次元位置を中心とする
- ランドマークの三次元位置とランドマーク撮影時のカメラ位置とを結ぶ直線に対して垂直である．
- 世界座標系での大きさが $S \times S$ である．

ここで，画像テンプレートの空間中でのサイズ S は，入力画像の 1 画素と画像テンプレートの 1 画素が対応するように以下の式により決定する．

$$S = \frac{D}{f} 2^{(p-1)} d \quad (2)$$

ただし， d はカメラの CCD における画素間の距離， f はカメラの焦点距離， D はランドマークの奥行き距離， p は多重スケールテンプレートのスケール値 (1, 2, 3) を表す．また，画像テンプレートの基底 \mathbf{u} ， \mathbf{v} は，以下のように決定する

$$\mathbf{v} = \frac{\hat{\mathbf{u}} \times \mathbf{N}}{\|\hat{\mathbf{u}} \times \mathbf{N}\|} \quad (3)$$

$$\mathbf{u} = \frac{\mathbf{N} \times \mathbf{v}}{\|\mathbf{N} \times \mathbf{v}\|} \quad (4)$$

$$\hat{\mathbf{u}} = \mathbf{Z} \times \mathbf{N} \quad (5)$$

ただし， N は画像テンプレートの法線方向を表す．ここで，画像テンプレートの法線ベクトル N が Z 軸に対して平行である場合，画像テンプレートの座標 u ， v を一意に決定することができない．そのような場合には，以下の式で u ， v を決定する．

$$\mathbf{v} = \frac{\hat{\mathbf{u}} \times \mathbf{N}}{\|\hat{\mathbf{u}} \times \mathbf{N}\|} \quad (6)$$

$$\mathbf{u} = \frac{\mathbf{N} \times \mathbf{v}}{\|\mathbf{N} \times \mathbf{v}\|} \quad (7)$$

$$\hat{\mathbf{u}} = \mathbf{X} \times \mathbf{N} \quad (8)$$

最後に，仮定した平面上に撮影画像を投影することで，撮影時のカメラ姿勢やレンズ歪みの影響を排除した画像テンプレートを作成する．画像テンプレートの画素数を $N \times N$ とした場合，この投影面を等間隔な格子で $N \times N$ に分割し，各格子点の三次元座標を撮影画像面上に投影した上で，その投影座標の画素値をその点の画素値とする．このとき，画像テンプレートの1画素が入力画像上の1画素と対応するように画像テンプレートの大きさを決定し，これを基本スケールの画像テンプレートとする．

このような手法により，多重スケール ($p = 1, 2, 3$) の画像テンプレートを作成し，これらを撮影地点ごとに保持する．ただし，このような手法でスケールが2倍，4倍の画像テンプレートを作成した場合，投影された点の間に存在する画素の色情報は用いられずエイリアシングが生じる．本研究では，多重スケールの画像テンプレートを作成する際には，撮影画像を $1/2$ ， $1/4$ に縮小した画像を投影することで周辺の色情報を取得し，この問題を回避する．

2.4.2 多重スケールの視点位置に依存しない画像テンプレートの生成

主成分分析を用いた平面当てはめによって平面上に存在すると判定されたランドマークに対して，視点位置に依存しない画像テンプレートを生成する．ここでは，図16に示すように，主成分分析により算出したランドマークの法線ベクト

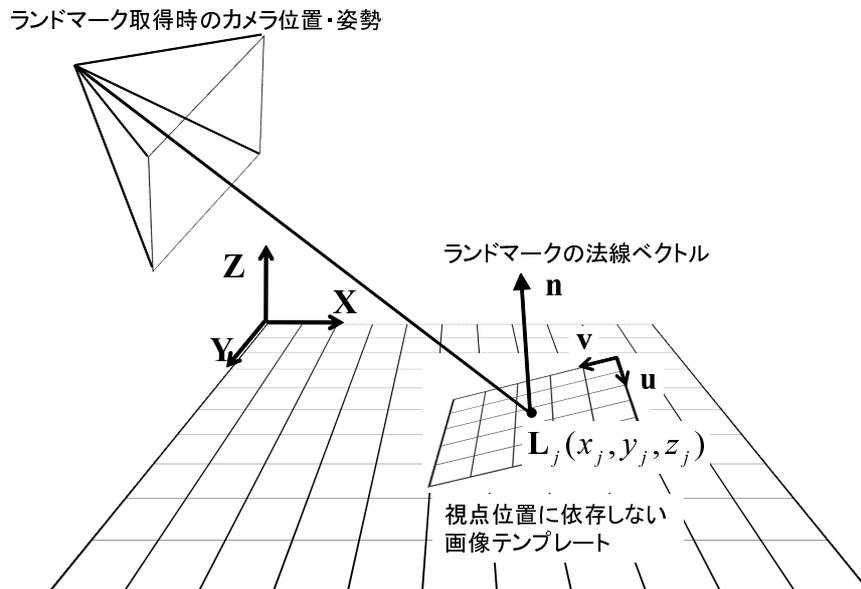
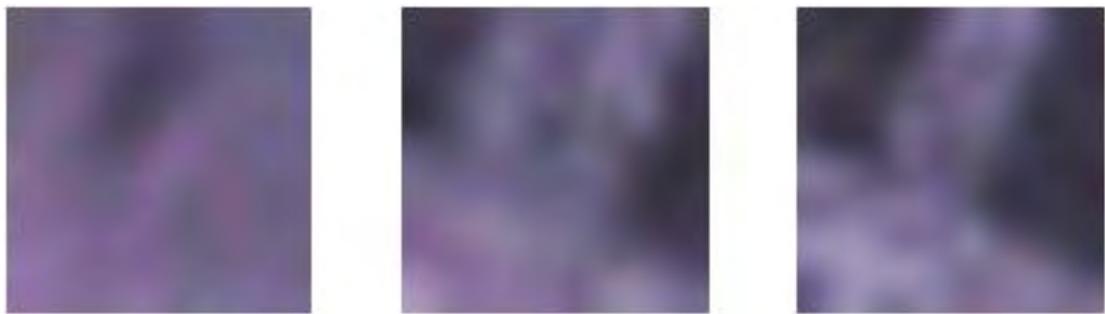
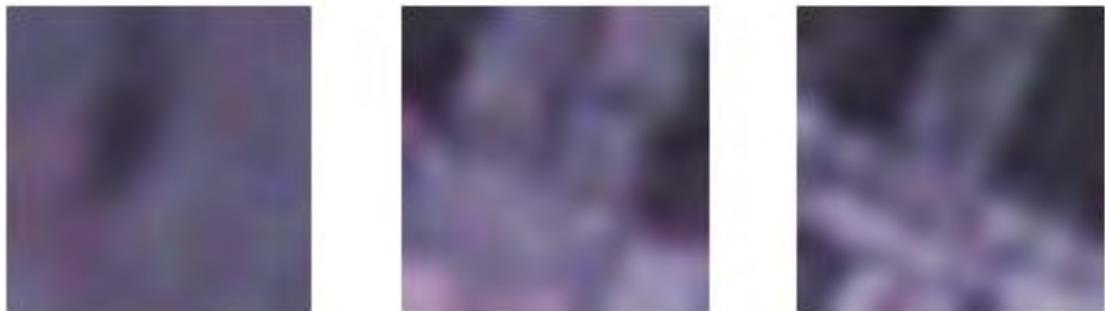


図 16 視点位置に依存しない画像テンプレートの生成方法

ル n に対して垂直な正方形の平面を仮定し，2.4.1 節と同様の方法で u, v を算出する．ここで，視点位置に依存しない画像テンプレートは，カメラの投影中心とランドマークを結ぶ直線に対して大きく傾くことがある．そのため，図 17 に示すように，従来と同様の手法によって，スケールの大きい画像テンプレートを生成した場合，画像テンプレートの一部の解像度が低下する．そこで，本研究では，高解像度な画像テンプレートを作成し，その画像テンプレートから必要な範囲を切り出し，縮小することで視点位置に依存しない画像テンプレートを生成する．これにより，画像テンプレート生成時の解像度の低下を抑える．



(a) 従来の手法で作成した画像テンプレート



(b) 高解像度な画像テンプレートから切り出した画像テンプレート

図 17 生成された視点位置に依存しない画像テンプレートの例 (左 スケール1倍 , スケール中央 5 倍 , スケール右 9 倍)

3. 法線情報付き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

本章では，図 12 に示した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．オンラインでのカメラ位置姿勢推定処理は，従来手法と同様に，まず，初期フレームでのカメラ位置・姿勢を何らかの方法により取得する (B-1)．次に，暫定的なカメラ位置・姿勢推定 (B-2)，ランドマークの選択 (B-3)，ランドマークと自然特徴点との対応付けによるカメラ位置・姿勢 (B-4) を繰り返す．ただし本研究では，平面上に存在するランドマークは 2 種類の画像テンプレートを保持するため，ランドマークの選択処理において，画像テンプレートの法線情報，および，ランドマークとカメラの距離に基づき，どちらの画像テンプレートを用いるかを適応的に選択する．

3.1 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

連続フレーム間でランドマークを追跡することで，現フレームの暫定的なカメラ位置・姿勢を推定する．暫定的なカメラ位置・姿勢推定では，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が微小であることから，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内に存在する自然特徴点を対象として 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．

次に，追跡した自然特徴点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，まず，LMedS 基準を用いて誤対応を排除する．次に，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位置・姿勢を推定する．

3.2 法線情報に基づくランドマークの選択

ここでは、対応付けの候補となる N_{max} 個のランドマークとその画像テンプレートをデータベースより選択する。提案手法では、平面上に存在すると判定されたランドマークは、従来の画像テンプレートに加え、視点位置に依存しない画像テンプレートの画像テンプレートを保持している。そのため、それらの画像テンプレートの法線情報に基づき、対応付けに用いる画像テンプレートの種類を適応的に決定する。

ランドマークの選択では、以下の条件に基づき、ランドマークとその画像テンプレートを選択する。

- (条件 1) ランドマークが暫定カメラの画角内に存在する：ランドマークの三次元位置を暫定的なカメラ位置・姿勢を用いて撮像面に投影する。これにより、入力画像上に投影されたランドマークのみを用いる。
- (条件 2) ランドマーク撮影時のカメラ位置とランドマークとを結ぶ直線と暫定カメラ位置とランドマークとを結ぶ直線のなす角が閾値以下であるか、または、ランドマークが平面上に存在し、ランドマークの三次元位置と暫定カメラ位置との距離が閾値以下である：図 18 に示すように、ランドマークの三次元位置とランドマークが撮影されたカメラ位置とを結ぶ直線とランドマークの三次元位置と暫定カメラ位置を結ぶ直線とのなす角 θ_c が閾値以上であれば、そのランドマークを用いない。しかし、 θ_c が閾値以上であっても、そのランドマークが平面上に存在し、ランドマークの三次元位置と暫定カメラ位置との距離 d が閾値 D_{max} 以下の場合、このランドマークを用いる。
- (条件 3) 既に選択されたランドマークと入力画像上で密接していない：まず、入力画像を格子状に分割する。ランドマークの三次元位置を暫定的なカメラ位置・姿勢を用いて入力画像上に投影したときに、同じ格子内に投影されるランドマークが既に選択されていた場合、そのランドマークを用いない。

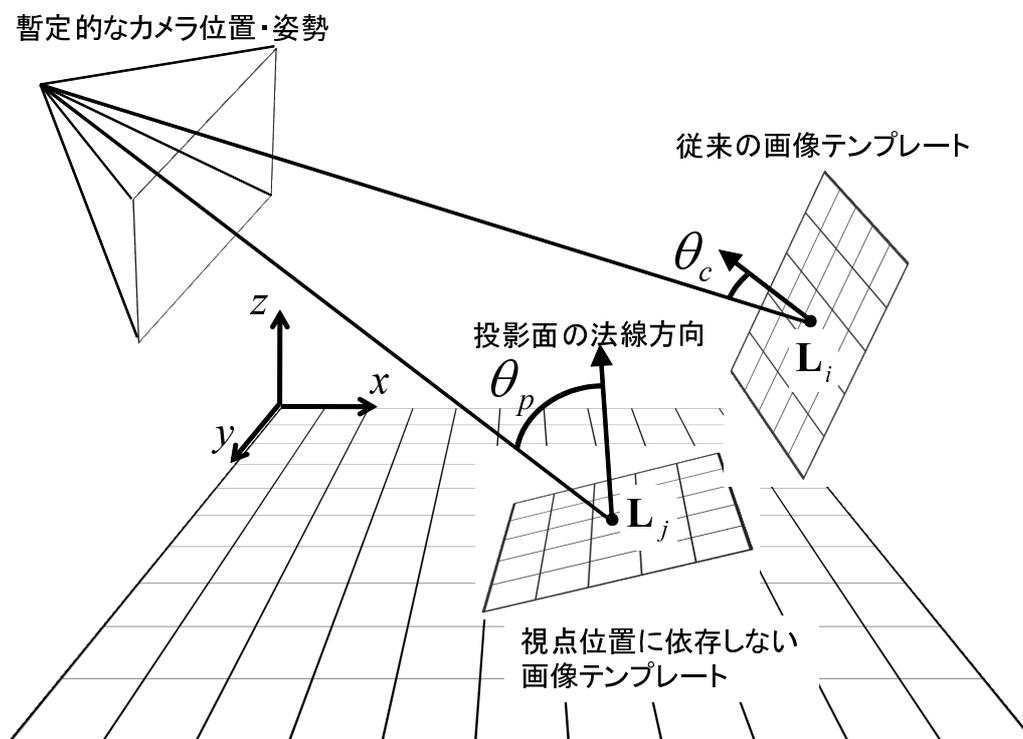


図 18 カメラに対する画像テンプレートの傾き

ここで，図 18 に示すように，多くの場合，ランドマークの三次元位置と暫定カメラ位置を結ぶ直線とランドマークの法線ベクトルとの成す角 θ_p はランドマークの三次元位置とランドマークが撮影されたカメラ位置とを結ぶ直線とランドマークの三次元位置と暫定カメラ位置を結ぶ直線とのなす角 θ_c に対して大きくなる．そのため，従来手法 [1] と同様に θ_c ， θ_p の大きさを基準に画像テンプレートの選択を行うと，従来の画像テンプレートを用いるランドマークが優先的に用いられるという問題が生じる．そこで，本研究では以下の式で定義する優先度 C を新たに定義し，この優先度に基づいてテンプレートの選択を行う．具体的には，まず (条件 1)，(条件 2)，(条件 3) を全て満たすランドマークを，ランドマーク取得時のカメラ位置と暫定カメラ位置の距離 d_{view} でソートし， d_{view} が小さい順に N'_{max} 個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を以下に定義する

優先度 C を用いてソートし，優先度の高い N_{max} 個のランドマークを選択する．

$$C = \begin{cases} (w\theta_c + 1)^{-1} & \text{(従来の画像テンプレートを選択した場合)} \\ (\theta_p + 1)^{-1} & \text{(視点位置に依存しない画像テンプレートを選択した場合)} \end{cases} \quad (9)$$

ここで， w は θ_c と θ_p のバランスを取るための重みである．

3.3 自然特徴点とランドマークとの対応付けによるカメラ位置・姿勢推定

自然特徴点とランドマークとの対応付けによるカメラ位置・姿勢推定処理では正規化相互相関を用いて選択されたランドマークと入力画像中の自然特徴点とを対応付け，それらの組に含まれる誤対応を LMedS 基準により排除した後に，再投影誤差を最小化することによってカメラ位置・姿勢推定を行う．本手法では平面上に存在するランドマークと存在しないランドマークが存在するが，これらのランドマークに対して同様の処理を行う．以下では，これらの処理について詳述する．

3.3.1 正規化相互相関による自然特徴点とランドマークとの対応付け

正規化相互相関による自然特徴点とランドマークとの対応付けでは，まず，暫定的なカメラ位置・姿勢を用いて，選択されたランドマークを入力画像上に投影し，ランドマークが投影された画像上での位置を中心とする一定ウィンドウ内の自然特徴点を検出する．検出されたすべての自然特徴点に対して，データベース構築時と同様にして，画像テンプレート，および，視点位置に依存しない画像テンプレートを生成し，ランドマークの画像テンプレート，および，視点位置に依存しない画像テンプレートと照合する．これによって相関値が最大となる自然特徴点が閾値以上の相関値を持つ場合，その自然特徴点をランドマークと対応付ける．

3.3.2 自然特徴点とランドマークとの対応関係を用いたカメラ位置・姿勢推定

自然特徴点とランドマークとの対応関係を用いたカメラ位置・姿勢推定処理では，自然特徴点とランドマークの組に含まれる誤対応を LMedS 基準を用いて排除し，再投影誤差を最小化することによってカメラ位置・姿勢を推定する．以下ではこれらの処理の流れについて述べる．

- (1) F 個の対応付けられたランドマークと自然特徴点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) 図 19 に示すように全てのランドマーク L_i に対して，暫定的なカメラパラメータを用いて画像上に投影した二次元座標 \hat{L}_i と対応する自然特徴点の二次元座標 x_i との距離の二乗誤差 R_i (以下，再投影誤差) の中間値 RM を以下のように算出する．

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

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

- (3) 上記の処理 (1), (2) を t 回繰り返し，中間値 RM が最小となる暫定的なカメラパラメータを用いて各ランドマークの再投影誤差を評価し，閾値を超える組を誤対応として排除する．
- (4) 誤対応を排除した後の F' 組のランドマークと自然特徴点を用いてカメラ位置・姿勢を推定する．

手順 (2), (4) におけるカメラ位置・姿勢推定は，自然特徴点とランドマークとの組が l 個である場合，以下の式で定義される再投影誤差の和 E を非線形最小化することで推定する．

$$E = \sum_{i=1}^l (\hat{L}_i - x_i) \quad (12)$$

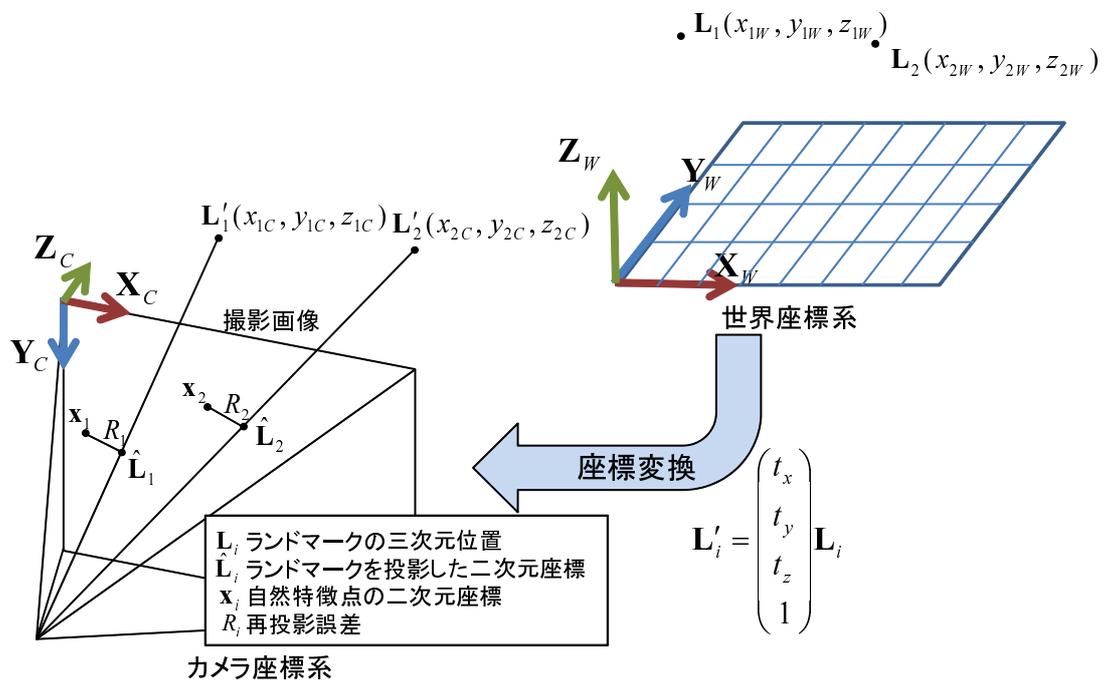


図 19 ランドマークと自然特徴点との再投影誤差

4. 実験

提案手法の有効性を示すために，従来手法 [1] との比較により，カメラパスから離れた場合におけるカメラ位置・姿勢推定の頑健性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動画像を用いた実験について述べる．

4.1 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 20 左に示すような全天球動画像を撮影できる全方位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を用いて屋外環境を移動撮影することにより，図 20 右に示すような画像系列を取得した．その後，佐藤らの手法により，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した結果に対して，局所平面を当てはめ，法線情報を算出することでランドマークデータベースを構築した．三次元復元の結果を図 21 に示す．本実験において，データベースに登録されたランドマークの数は約 6500 個であり，1 つのランドマークあたりの撮影地点数は平均 6 個所であった．推定された三次元点群に対して，主成分分析に用いる点群の範囲 r を 1m, 3m, 5m と変化させ，法線方向の算出を行った結果を図 22 に示す．図中の青色の点は平面以外と判定されたランドマーク，赤色の点は平面と判定されたランドマーク，赤色の線はそのランドマークの法線方向を表す．同図は，4.2 節以降で用いる入力画像の 1 フレームに対して，後述するカメラパラメータの真値を用いてランドマークを重畳合成している．これらの図から，半径 r を 1m として平面当てはめを行った場合には，3m, 5m と比較して，平面ではないと判定されたランドマークが多く，特に，遠方の地面のランドマークの大半は平面ではないと判定されている．これは，主成分分析の範囲内に十分な数のランドマークが含まれず，安定して法線を推定することができなかつたためであると考えられる．半径 r を 5m として平面当てはめを行った場合には，地面の左端の一部と右端に存在するランドマークが平面ではないと判定されている．こ



図 20 全方位型マルチカメラシステムと全方位画像の1フレーム

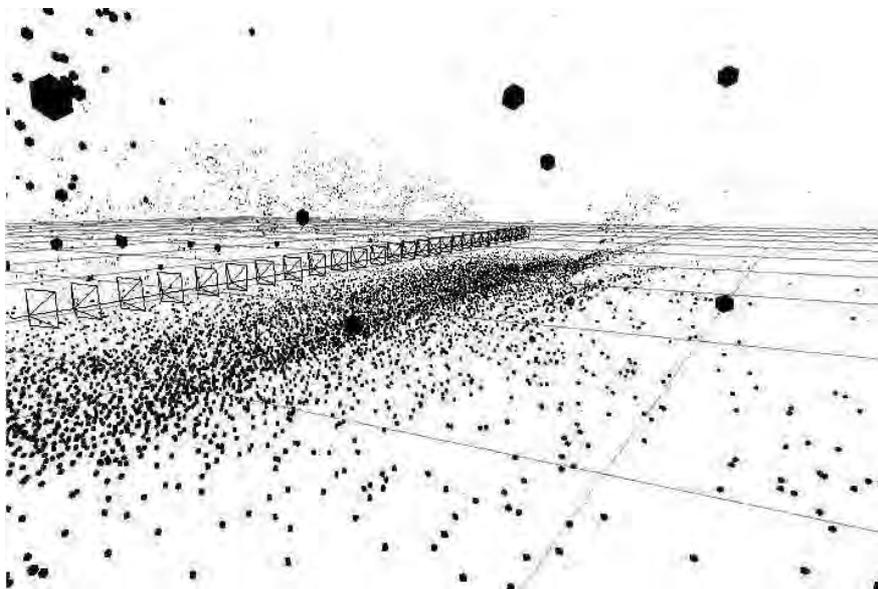
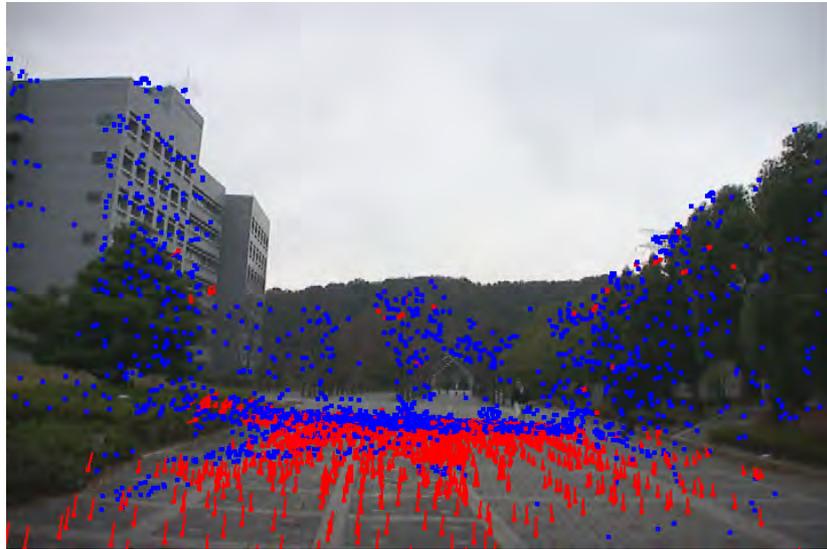
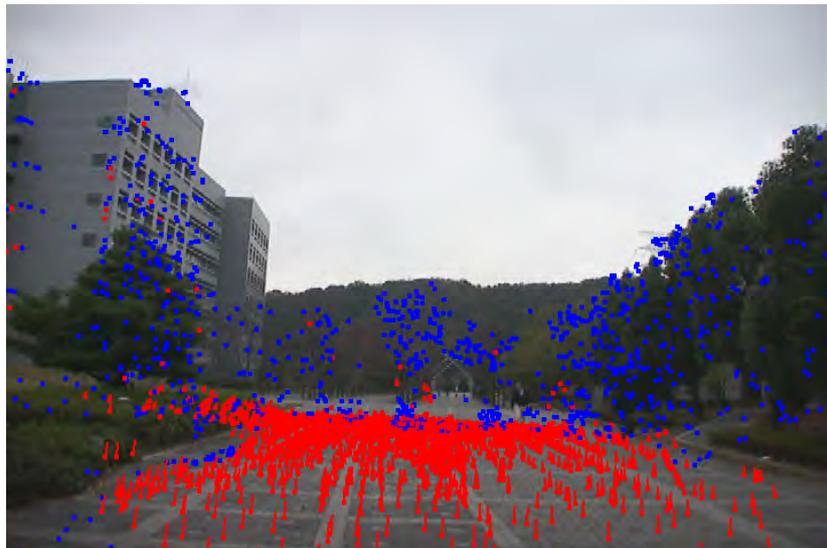


図 21 推定された自然特徴点の三次元位置と全方位カメラの位置・姿勢

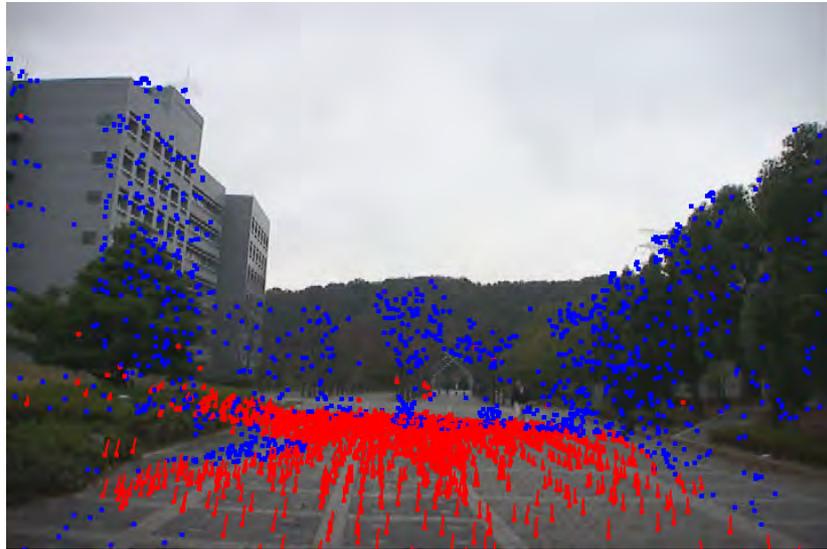


(a) 半径 r を 1m として主成分分析を行った結果



(b) 半径 r を 3m として主成分分析を行った結果

図 22 再投影されたランドマークとその法線情報 (1/2)



(a) 半径 r を 5m として主成分分析を行った結果

図 23 再投影されたランドマークとその法線情報 (2/2)

これは、主成分分析の範囲となる半径が大きすぎるため、周辺に存在する木などに存在するランドマークの影響を受けたためであると考えられる。また、半径 r を 1m, 3m, 5m で行った全ての場合において、図中の左に存在する建物の壁は平面として判定されなかった。これらのランドマークは地面のランドマークと比較して遠方に存在するため、主成分分析で用いる範囲内に十分な数のランドマークが含まれず、安定して法線を推定できなかったためであると考えられる。これらの結果より、次節で述べる推定精度の比較実験では、半径 r を 3m とした場合に推定された法線情報を用いて構築したデータベースを利用した。

4.2 カメラ位置・姿勢の推定精度の比較

本実験では、表 1 に示すような機器を用い、図 24 に示すように、ランドマークデータベース構築時のカメラパスから約 10m 離れた地点からランドマークデータベース構築時のカメラパスに近づくような経路で、動画像 (170 フレーム) を撮影

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((解像度 720 × 480 画素, プログレッシブスキャン, 15fps)
広角レンズ	SONY VCL-HG0758
計算機	CPU:Xeon(TM) 2.67GHz(4 コア)×2, Memory:12.0GB

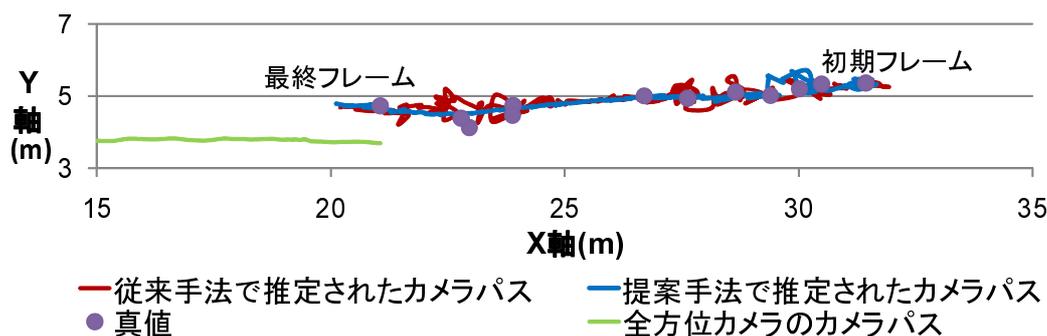


図 24 カメラの初期位置と推定されたカメラパス

し, 推定されたカメラ位置・姿勢を従来手法 [1] と比較した. ただし, 本実験では, 提案手法, 従来手法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・姿勢は環境内の自然特徴点の三次元位置をトータルステーションにより計測し, 画像上での自然特徴点と手動で対応付け, PnP 問題を解くことで算出した. 本実験で用いたパラメータを表 2 に示す. 世界座標系は, X 軸, Y 軸が現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で, 上向きの座標系とした. 以下では, 自然特徴点とランドマークとの対応付け精度, 推定されたカメラパス, および, 仮想物体重畳時の位置ずれの比較について述べる.

4.2.1 自然特徴点とランドマークとの対応付け精度の比較

まず, 自然特徴点とランドマークの対応付け精度を検証するために, データベースから選択されたランドマーク, 自然特徴点と対応付けられたランドマーク, 最終的なカメラ位置・姿勢推定に用いられたランドマークの数を従来手法と提案手法で比較した. 得られたランドマークの数の推移を図 25, 各フレームにおける

表 2 実験に用いたパラメータ

	従来手法	提案手法
自然特徴点追跡時の画像テンプレートサイズ	15 × 15 画素	15 × 15 画素
自然特徴点追跡時の探索範囲	120 × 60 画素	120 × 60 画素
ランドマークの画像テンプレートサイズ	15 × 15 画素	15 × 15 画素
ランドマークの 視点位置に依存しない画像テンプレートサイズ	-	15 × 15 画素
視点位置に依存しない画像テンプレートのスケール	-	1 倍 5 倍 9 倍
ランドマークの対応点候補の探索範囲	20 × 20 画素	20 × 20 画素
ランドマークの最大選択数	100 個	100 個
ランドマークとランドマーク取得時のカメラ位置と を結ぶ直線とランドマーク暫定カメラ位置とを結ぶ 直線のなす角 θ_c の閾値	15°	15°
従来の画像テンプレートと視点位置に依存しない 画像テンプレート間における角度の重み w	-	9
LMedS の繰り返し回数	600 回	600 回

データベースから選択されたランドマークの内，カメラ位置・姿勢推定に用いられる割合を図 26，1 フレームあたりの平均数を表 3 に示す．各フレームにおいて，データベースから選択されたランドマークの内，カメラ位置・姿勢推定に用いられる割合に対して有意水準 1%とした t 検定を行い，提案手法と従来手法との比較を行った．その結果，有意差が認められたことから，平均的に提案手法がより多くのランドマークを頑健に対応付けることができると考えられる．特に 97 フレームから 148 フレームでは，従来手法よりも多くのランドマークが推定に利用されていることが確認できる．図 27 に，これらのフレームでカメラ位置・姿勢推定に用いられたランドマークの画像上の位置の例を示す．丸印は，3.2 節で述べた (条件 1) から (条件 3) を満たすランドマークを表しており，図中の赤の丸印は正規化相互相関を用いた照合の結果，入力画像の自然特徴点と対応付けられなかったランドマーク，緑の丸印は処理 (B-4) で LMedS 基準により誤対応と判定されたランドマーク，青の丸印は最終的なカメラ位置・姿勢推定に用いられたランドマークを表す．同図 (a)，(c) より従来手法ではユーザに近い地面上に存在するランドマークはデータベースから選択されておらず，遠方のランドマークのみをカメラ位置・姿勢推定に利用している．それに対して，提案手法では地面上に存在

表 3 1 フレームあたりのランドマークの平均数

	提案手法 [1]	従来手法
データベースから選択されたランドマーク数	100	98.5
自然特徴点と対応付けられたランドマーク数	79.6	73.3
最終的なカメラ位置・姿勢推定に用いられたランドマーク数	51.7	46.1
選択されたランドマークの内、カメラ位置・姿勢推定に用いられる割合	0.80	0.76

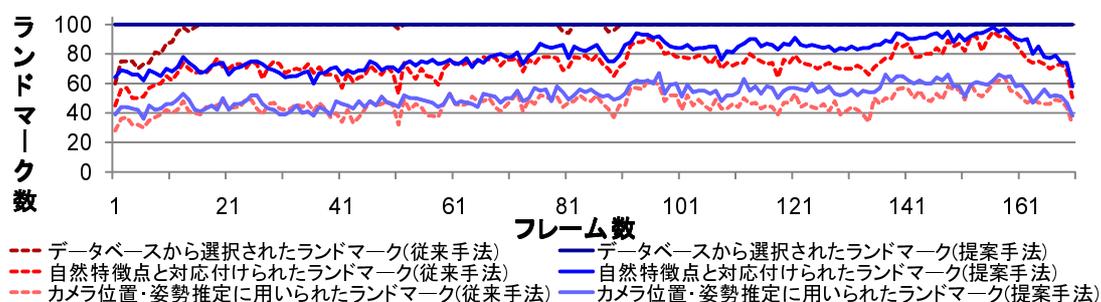


図 25 各フレームにおけるランドマーク数の推移

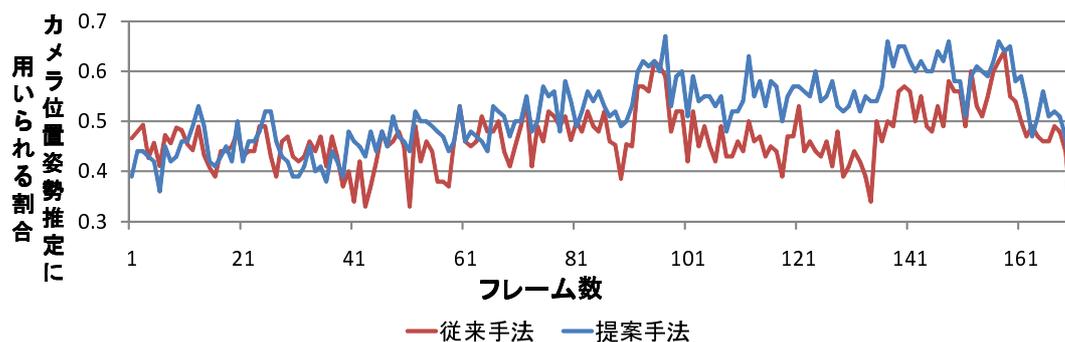


図 26 選択されたランドマークの内、カメラ位置姿勢推定に用いられる割合の推移

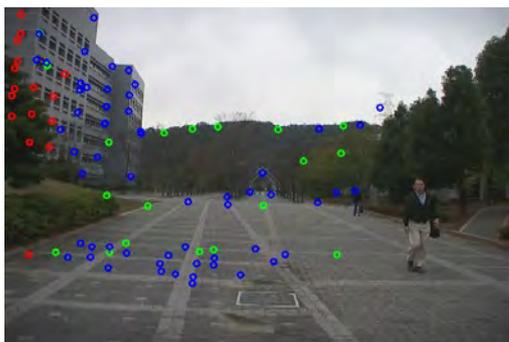


100フレーム

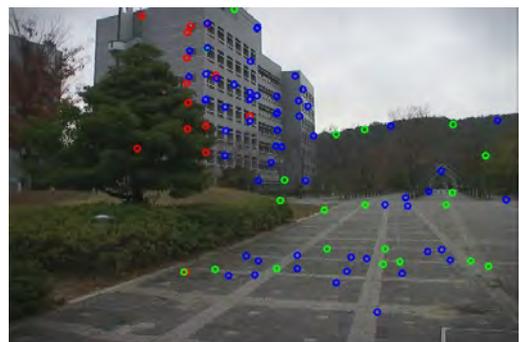


120フレーム

(a) 従来手法



100フレーム



120フレーム

(b) 提案手法

図 27 提案手法が有効なフレームでのカメラ位置・姿勢推定に用いられたランドマークの比較：(赤) 自然特徴点と対応付けられなかったランドマーク (緑) 対応づいたが誤対応として排除されたランドマーク (青) 最終的なカメラ位置・姿勢推定に用いられたランドマーク

するランドマークもデータベースから選択し，遠方のランドマークと地面のランドマークの双方を対応付けている．以上のことから，提案手法では，ユーザの近くに存在するランドマークについても有効に利用できていることが分かる．

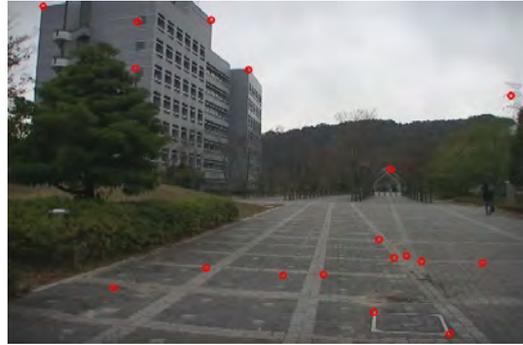
次に，従来手法と提案手法により推定されたカメラパスの比較を行った．推定されたカメラパスと真値，データベース構築時の全方位カメラのカメラパスを図 24 に示す．真値は，初期フレームにおけるカメラの位置・姿勢と同様に手動で対応点を与え，PnP 問題を解くことで算出した．この結果から，提案手法により推定されたカメラパスは真値から大きく離れることなく，おおむね正しく推定できていることが確認できる．それに対して，従来手法により推定されたカメラパスでは一部のフレームにおいて，真値から大きく外れるような誤差が生じていることが分かる．これは提案手法では，カメラの位置を決定するために重要なユーザの近くに存在するランドマークを正しく対応づけることで，カメラ位置の推定精度が向上したためであると考えられる．

4.2.2 仮想物体の重畳誤差の比較

仮想物体を重畳した場合における画像上での仮想物体の重畳の誤差を確認するために，手動で対応関係を求めた各フレームにおけるランドマークの画像上での二次元座標と推定されたカメラ位置・姿勢を用いてランドマークを再投影した二次元座標との誤差（以下，再投影誤差）を計測した．真値として用いたランドマークの画像上の位置を図 28 に，各フレームにおける再投影誤差の推移を図 29 に示す．また，提案手法によって推定されたカメラ位置・姿勢を用いて仮想物体の重畳を行った結果を図 30 に示す．このとき，平均再投影誤差は従来手法で 10.8 画素，提案手法で 7.4 画素であった．図 29 から，提案手法の 107 フレームから 142 フレームにおいて精度向上がみられる．また，各フレームにおける平均再投影誤差に対して，有意水準を 1%とした t 検定により，有意差を確認した．また，図 30 より仮想物体が位置ずれなく幾何学的に正しい位置に重畳されていることが分かった．



(a) 10 フレーム



(b) 110 フレーム

図 28 再投影誤差の計算に用いた真値の例

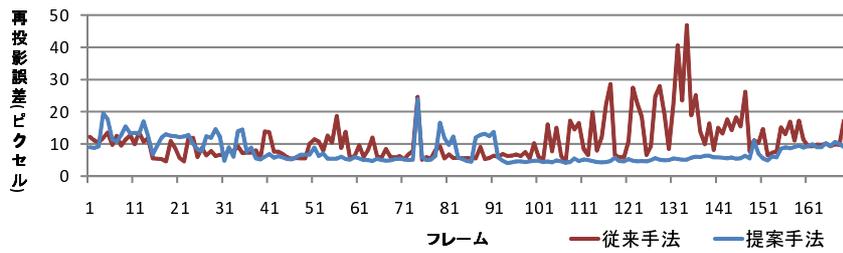


図 29 各フレームにおける再投影誤差の推移



(a) 110 フレーム



(b) 115 フレーム



(c) 120 フレーム



(d) 125 フレーム



(e) 130 フレーム



(f) 135 フレーム

図 30 提案手法によるカメラ位置・姿勢推定結果を用いた仮想物体の重畳表示

5. 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、ランドマークの三次元位置に局所平面を当てはめ、これを用いて幾何学的に補正した画像テンプレート利用することで、ランドマークデータベース構築時の全方位カメラの経路から離れた地点で撮影した入力画像に対しても頑健なカメラ位置・姿勢推定を実現する手法を提案した。入力画像とランドマークデータベースの対応付けにおいて、平面上に存在するランドマークに対して視点位置に依存しない画像テンプレートを用いることで、視点位置に近いランドマークの頑健な対応付けを実現し、これにより、カメラ位置・姿勢の推定精度の向上を実現させた。

実験では、データベース構築時のカメラパスから離れた地点において、自然特徴点と対応づいたランドマークの数、および、誤対応として排除されたランドマークの数を比較することで、自然特徴点とランドマークとの対応付けの頑健性について検証した。実験の結果、従来手法と比較して、平面と判定されたランドマークを頑健に対応付けられていることを確認した。また、拡張現実感画像の比較により、仮想物体の位置合わせの精度が向上することを確認した。

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やテクスチャの少ない環境での頑健性の向上が挙げられる。データベースを更新することにより得られる、異なる季節や時間に構成された、複数のランドマーク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間を補間することも考えられる。また、カメラ以外のセンサを併用することで、より頑健なランドマークの対応付け、および、カメラ位置・姿勢の推定が可能であると考えられる。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

本研究の全過程を通して、懇切なる御指導、御鞭撻を賜りまし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、御鞭撻を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、本研究を進めるにあたり、始終暖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、本研究を通じて、的確な御助言、御鞭撻を頂いた環境知能学講座 神原誠之 准教授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、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、貴重な御助言や御指摘を頂きました視覚情報メディア講座 武富貴史氏に深く御礼申し上げます。そして、本研究を進めるにあたり、適切な御助言や御指摘を頂きました視覚情報メディア講座 糸秀行氏に心より感謝致します。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 中村美奈 女史に厚く御礼申し上げます。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] 武富貴史, 佐藤智和, 横矢直和. 拡張現実感のための優先度情報を付加した自然特徴点ランドマークデータベースを用いた実時間カメラ位置・姿勢推定. 電子情報通信学会論文誌 (D), Vol. J92-D, No. 8, pp. 1440–1451, 2009.
- [2] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. A Touring Machine : Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Personal and Ubiquitous Computing*, Vol. 1, No. 4, pp. 208–217, 1997.
- [3] Tobias Höllerer, Steven Feiner, and John Pavlik. Situated Documentaries: Embedding Multimedia Presentations in the Real World. In *Proceedings of 3rd International Symposium on Wearable Computers*, pp. 79–86, 1999.
- [4] Patrick Dähne and John N. Karigiannis. ARCHEOGUIDE : System Architecture of a Mobile Outdoor Augmented Reality System. In *Proceedings of 1st International Symposium on Mixed and Augmented Realit*, pp. 263–264, 2002.
- [5] Ryuhei Tenmoku, Yusuke Nakazato, Atsushi Anabuki, Masayuki Kanbara, and Naokazu Yokoya. Nara Palace Site Navigator: Device-Independent Human Navigation using a Networked Shared Database. In *Proceedings of 10th International Conference on Virtual Systems and Multimedia*, pp. 1234–1242, 2004.
- [6] Tetsuya Kakuta, Takeshi Oishi, and Katsushi Ikeuchi. Development and Evaluation of Asuka-Kyo MR Contents with Fast Shading and Shadowing. In *Proceedings of 13th International Conference on Virtual Systems and Multimedia*, pp. 254–260, 2007.
- [7] Michele Fiorentino, Raffaele de Amicis, Giuseppe Monno, and Andre Stork. Spacedesign: A Mixed Reality Workspace for Aesthetic Industrial Design.

In *Proceedings of 1st International Symposium on Mixed and Augmented Reality*, pp. 86–94, 2002.

- [8] Peter Supan, Ines Stuppacher, and Michael Haller. Image Based Shadowing in Real-Time Augmented Reality. *International Journal of Virtual Reality*, Vol. 5, No. 3, pp. 1–6, 2006.
- [9] Georg Klein and David Murray. Compositing for Small Cameras. In *Proceedings of 7th International Symposium on Mixed and Augmented Reality*, pp. 57–60, 2008.
- [10] 奥村文洋, 神原誠之, 横矢直和. 焦点外れによるぼけとモーションブラーの推定に基づく拡張現実感における光学的整合性の実現. 電子情報通信学会論文誌 (D), Vol. J90-D, No. 8, pp. 2126–2136, 2007.
- [11] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The Anatomy of a Context-Aware Application. *Wireless Networks*, Vol. 8, No. 2, pp. 187–197, 2002.
- [12] 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. 電子情報通信学会 技術研究報告, PRMU104–572, pp. 933–938, 2005.
- [13] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 9, pp. 1052–1067, 2007.
- [14] Roberto Cipolla, Duncan Robertson, and Ben Tordoff. Image-Based Localisation. In *Proceedings of 10th International Conference on Virtual Systems and Multimedia*, pp. 22–29, 2004.
- [15] 野田雅文, 高橋友和, 出口大輔, 井手一郎, 村瀬洋, 小島祥子, 内藤貴志. 空撮画像と時系列車載カメラ画像との照合による自車位置推定. 電子情報通信学会 技術研究報告, PRMU306–109, pp. 177–182, 2009.

- [16] Luca Vacchetti., Vincent Lepetit, and Pascal Fua. Stable Real-Time 3D Tracking using Online and Offline Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1385–1391, 2004.
- [17] Muriel Pressigout and Eric Marchand. Real-Time Hybrid Tracking using Edge and Texture Information. *International Journal of Robotics Research*, Vol. 26, No. 7, pp. 689–713, 2007.
- [18] Sehwan Kim, Stephen DiVerdi, Jae S. Chang, Taehyuk Kang, Ronald Iltis, and Tobias Höllerer. Implicit 3D Modeling and Tracking for Anywhere Augmentation. In *Proceedings of 15th Symposium on Virtual Reality Software and Technology*, pp. 19–28, 2007.
- [19] 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.
- [20] Ryuhei Tenmoku, Masayuki Kanbara, and Naokazu Yokoya. A Wearable Augmented Reality System Using Positioning Infrastructures and a Pedometer. In *Proceedings of 2nd International Symposium on Wearable Computers*, pp. 110–117, 2003.
- [21] Leonid Naimark and Eric Foxlin. Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-Inertial Self-Tracker. In *Proceedings of 1st International Symposium on Mixed and Augmented Reality*, pp. 27–36, 2002.
- [22] 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, MVE106–91, pp. 1–6, 2006.
- [23] 中里祐介, 神原誠之, 横矢直和. 不可視マーカを用いた位置・姿勢推定のための環境構築とユーザ位置・姿勢推定システム. 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 257–266, 2008.

- [24] G. Klein and D. Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *Proceedings of 6th International Symposium on Mixed and Augmented Reality*, pp. 225–234, 2007.
- [25] Taehee Lee and Tobias Höllerer. Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 3, pp. 355–368, 2008.
- [26] Hirokazu Kato and Mark Billinghurst. Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. In *Proceedings of International Workshop on Augmented Reality*, pp. 85–94, 1999.
- [27] Hideyuki Uchiyama and Hideo Saito. Augmenting Text Document by On-Line Learning of Local Arrangement of Keypoints. In *Proceedings of 8th International Symposium on Mixed and Augmented Reality*, pp. 95–98, 2009.
- [28] Tom Drummond and Roberto Cipolla. Real-Time Visual Tracking of Complex Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 932–946, 2002.
- [29] Gabriele Bleser, Harald Wuest, and Didier Stricker. Online Camera Pose Estimation in Partially Known and Dynamic Scenes. In *Proceedings of 5th International Symposium on Mixed and Augmented Reality*, pp. 56–65, 2006.
- [30] Iryna Skrypnyk and David G. Lowe. Scene Modelling, Recognition and Tracking with Invariant Image Features. In *Proceedings of 3rd International Symposium on Mixed and Augmented Reality*, pp. 110–119, 2004.
- [31] Clemens Arth, Daniel Wagner, Manfred Klopschitz, Arnold Irschara, and Dieter Schmalstieg. Wide Area Localization on Mobile Phones. In *Proceedings of 8th International Symposium on Mixed and Augmented Reality*, pp. 73–82, 2009.

- [32] Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya. AR Cultural Heritage Reconstruction Based on Feature Landmark Database Constructed by using Omnidirectional Range Sensor. In *Proceedings of 10th Asian Conference on Computer Vision Workshop on e-Heritage*, 2010.
- [33] Changchang Wu, Friedrich Fraundorfer, Jan M. Frahm, and Marc Pollefeys. 3D Model Search and Pose Estimation from Single Images using VIP Features. In *Proceedings of 21st Conference on Computer Vision and Pattern Recognition Workshop on Search in 3D*, pp. 1–8, 2008.
- [34] 糸秀行, 武富貴史, 佐藤智和, 横矢直和. GPS 測位の信頼度と外れ値を考慮した動画像と GPS からのカメラ位置・姿勢推定. 情報処理学会 研究報告, CVIM173–36, pp. 229–236, 2010.
- [35] Georg Klein and Tom Drummond. Tightly Integrated Sensor Fusion for Robust Visual Tracking. *Image and Vision Computing*, Vol. 22, No. 10, pp. 769–776, 2004.
- [36] Suyu You and Ulrich Neumann. Fusion of Vision and Gyro Tracking for Robust Augmented Reality Registration. In *Proceedings of Virtual Reality 2001 Conference*, pp. 71–78, 2001.
- [37] Krystian Mikolajczyk and Cordelia Schmid. Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, Vol. 60, No. 1, pp. 63–86, 2004.
- [38] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [39] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF : Speeded Up Robust Features. In *Proceedings of 9th European Conference on Computer Vision*, pp. 404–417, 2006.

- [40] Vincent Lepetit and Pascal Fua. Keypoint Recognition using Randomized Trees. *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, pp. 1465–1479, 2006.
- [41] Changchang Wu, Brian Clipp, Xiaowei Li, Jan M. Frahm, and Marc Pollefeys. 3D Model Matching with Viewpoint-Invariant Patches (VIP). In *Proceedings of 21st Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [42] Yasutaka Furukawa and Jean Ponce. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376, 2010.
- [43] 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.
- [44] Sei Ikeda, Tomokazu Sato, Koichiro Yamaguchi, and Naokazu Yokoya. Construction of Feature Landmark Database Using Omnidirectional Videos and GPS Positions. In *Proceedings of 6th International Conference on 3-D Digital Imaging and Modeling*, pp. 249–256, 2007.