

博士論文

拡張現実感のためのランドマークデータベースに基づく
カメラ位置・姿勢推定の高速化と高精度化に関する研究

武富 貴史

2011年3月10日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

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

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)

拡張現実感のためのランドマークデータベースに基づく カメラ位置・姿勢推定の高速化と高精度化に関する研究

*

武富 貴史

内容梗概

現実環境を撮影した画像中にCGなどを重畳表示することで情報を付加する拡張現実感，特にビデオシーズルー型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実現するためには，カメラの位置・姿勢を高精度に推定する必要がある．現在までに様々な手法が提案されている．これらの手法では，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．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，structure-from-motion法を用いて半自動でデータベースを構築できることから，低い人的コストで広域な環境に適用できるという特長を持つ．しかし，従来手法ではデータベースに登録されているランドマークと入力画像上の自然特徴点との照合に多くの処理時間を必要とし，拡張現実感に必須の実時間でのカメラ位置・姿勢の推定が実現されていなかった．また，カメラの近くに存在するランドマークは撮影位置の違いにより見え方が大きく変化するため，入力画像上の自然特徴点と正しく対応付けることが難しく，ユーザに近い位置にCGを高精度に位置合わせすることが難しいという問題があった．本研究では，実時間でのカメラ位置・姿勢推定を実現するために，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

* 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 博士論文, NAIST-IS-DD0861010, 2011年3月10日.

の付加により，対応付け処理において用いられる対応点候補数を削減し，計算コストを低減する．また，高精度な位置合わせを実現するために，厳密な位置合わせが要求される場所におけるスポット的な位置合わせ精度の向上を目的として，全周レンジファインダにより取得した密な奥行き情報を用いてランドマークの見え方を補正することで，ユーザの位置の近くに存在するランドマークの対応付けの成功率を向上させる．本論文では，1章で拡張現実感における現実環境と仮想環境の位置合わせを実現するための技術的な課題と従来研究を概観し，本研究の位置付けと方針を明確にする．2章では，従来のランドマークデータベースに基づくカメラ位置・姿勢推定手法について概観する．3章では，対応点候補数の削減によるランドマークデータベースに基づくカメラ位置・姿勢推定の高速化手法について述べ，4章では密な奥行情報を用いたランドマークデータベースの構築によるカメラ位置・姿勢推定精度の向上について述べる．5章では提案手法を用いた実アプリケーションの例を示す．最後に，6章で本研究について総括し，今後の展望について述べる．

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

Fast and Accurate Camera Parameter Estimation Based on Feature Landmark Database for Augmented Reality*

Takafumi Taketomi

Abstract

Augmented reality (AR) is a technique which enhances the real world by overlaying virtual objects. In this field, extrinsic camera parameter estimation is required for geometric registration of real and virtual worlds. To achieve geometric registration, especially for video see-through augmented reality, vision-based registration methods are extensively investigated because they can achieve pixel-level geometric registration. As one of these methods, a feature landmark-based camera parameter estimation method has been studied to achieve geometric registration in a large-scale environment. In this method, a feature landmark database is automatically constructed by using a structure-from-motion technique before the online camera parameter estimation process is carried out. One of significant problems in this method is a computational cost for matching process which is executed to find correspondences between landmarks in a database and natural features in an input image. Most of conventional methods cannot work in real-time due to high computational cost for testing a large number of matching candidates. Additionally, the accuracy of estimated camera parameters is insufficient for specific applications, which needs to overlay CG objects at the position close to the user's viewpoint. This is due to the difficulty in compensation of

* Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0861010, March 10, 2011.

visual pattern change of close landmarks only from the sparse 3-D information obtained by structure-from-motion. The aim of this research is achievement of fast and accurate geometric registration using the feature landmark database. In this thesis, Chapter 1 gives a perspective of the study on geometric registration between real and virtual worlds. The conventional feature landmark based camera parameter estimation methods are then reviewed in Chapter 2. In Chapter 3, in order to achieve fast camera parameter estimation, priorities are associated with landmarks by using previously captured video sequences to discard unreliable matching candidates. In Chapter 4, to improve the accuracy of geometric registration at the spot where CG objects must be placed at the position close to the user, we newly compensate for visual patterns of landmarks using a dense depth map obtained by an omnidirectional laser range sensor. To demonstrate the effectiveness of the proposed method, some AR applications using the proposed method are described in Chapter 5. Finally, Chapter 6 summaries the presented study.

Keywords:

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

目次

1. 序論	1
1.1 拡張現実感の技術課題	1
1.2 拡張現実感技術におけるカメラ位置・姿勢推定手法の分類	3
1.2.1 利用環境に関する事前知識を用いない手法	5
1.2.2 利用環境に関する事前知識を用いる手法	7
1.3 自然特徴点に基づくカメラ位置・姿勢推定の高速化手法	15
1.3.1 特徴点抽出の高速化手法	15
1.3.2 対応点探索の高速化手法	16
1.4 本研究の位置付けと方針	17
1.5 本論文の構成	19
2. 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・姿勢推定の従来研究	20
2.1 ランドマークデータベースの構築	20
2.1.1 ランドマークデータベースの構成要素	20
2.1.2 structure-from-motion 法による環境の三次元復元	21
2.1.3 ランドマーク情報の取得	24
2.2 カメラ位置・姿勢の推定	24
2.2.1 ランドマークの選択	25
2.2.2 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	26
3. ランドマークへの優先度情報の付加と自然特徴点の追跡によるカメラ位置・姿勢推定の高速化	28
3.1 序言	28
3.2 ランドマークの優先度の定義	30
3.3 優先度情報を用いた実時間でのカメラ位置・姿勢推定	31
3.3.1 初期カメラ位置・姿勢の推定	31

3.3.2	自然特徴点の追跡による暫定的なカメラ位置・姿勢の推定	33
3.3.3	優先度情報に基づくランドマークの選択	34
3.3.4	ランドマークと自然特徴点の照合によるカメラ位置・姿勢の推定	35
3.3.5	推定結果のフィードバックによる優先度情報の更新	35
3.4	計算コストの比較	36
3.5	実験と考察	38
3.5.1	推定に用いるランドマーク数の決定	41
3.5.2	処理速度の比較	44
3.5.3	推定精度の比較	45
3.6	結言	46
4.	局所的な三次元形状を考慮したランドマークデータベースの構築によるカメラ位置・姿勢推定精度の向上	47
4.1	序言	47
4.2	全周レンジファインダを用いたランドマークデータベースの構築	48
4.2.1	奥行情報とテクスチャ情報の取得	49
4.2.2	局所形状を考慮したランドマーク情報の取得	52
4.3	実験と考察	55
4.3.1	画像テンプレート補正効果の検証	56
4.3.2	カメラ位置・姿勢推定精度の評価	57
4.4	結言	60
5.	提案手法を用いた実アプリケーション	63
5.1	ヒューマンナビゲーションへの応用	64
5.2	カーナビゲーションへの応用	67
5.3	MR-PreViz への応用	73
5.4	バーチャル歴史体験への応用	79
6.	結論	85
6.1	まとめと考察	85

6.2 今後の展望	86
謝辞	89
参考文献	90
付録	106
A. 再投影誤差最小化によるカメラ位置・姿勢推定	106
B. LMedS 基準による誤対応の排除とカメラ位置・姿勢推定	108

目 次

1	拡張現実感技術を用いたアプリケーション	2
2	カメラ位置・姿勢推定手法の分類	4
3	平面領域の追跡によるカメラ位置・姿勢推定 [SB02]	6
4	Visual-SLAM による拡張現実感環境の構築 [KM07]	7
5	カメラ座標系と世界座標系の関係	8
6	景観への影響を抑えたマーカ	9
7	画像データベースを用いたカメラ位置・姿勢推定 [CRT04]	11
8	ワイヤフレームモデルを用いたカメラ位置・姿勢推定手法 [DC02]	12
9	Reitmayr らの手法で用いられる三次元モデル [RD06a]	13
10	マーカを用いた自然特徴点データベースの構築	14
11	利用環境の規模と画像に基づくカメラ位置・姿勢推定手法の関係	18
12	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 [大江 05]	21
13	多重スケールの画像テンプレート	22
14	座標系と再投影誤差の定義	23
15	画像テンプレートの生成	25
16	提案手法のオンライン処理の流れ	29
17	対応点候補の絞り込み方法の違い	32
18	ランドマークデータベースの構成要素	32
19	ランドマークの対応点の探索範囲	36
20	全方位型マルチカメラシステムとデータベースの構築に用いた全 方位動画像の一部	40
21	三次元復元の結果	40
22	ランドマーク数と推定に失敗したフレーム数の関係	41
23	従来手法 [大江 05] で推定に失敗したフレーム	43
24	拡張現実感による文化遺産の復元	48
25	計測システムと各センサの座標系の関係	49
26	全周レンジデータと全方位画像	50
27	全方位奥行き画像	51

28	検出された空領域	51
29	提案手法によるランドマーク情報の取得	53
30	画像テンプレートの補正例	54
31	局所形状を考慮した画像テンプレートの生成	55
32	対応づいたランドマークの画像上の位置 (局所形状を考慮しない場合)	58
33	対応づいたランドマークの画像上の位置 (局所形状を考慮した場合)	59
34	各フレーム毎の位置誤差	61
35	各フレーム毎の光軸方向の角度誤差	61
36	登録されているランドマークの位置と 90 フレーム目のカメラ位置の関係	62
37	利用環境と撮影時のカメラパス	64
38	ヒューマンナビゲーションへの応用例	66
39	車両の動きのモデル化	68
40	データ取得時のカメラ配置	69
41	走行環境のデータベースの構築	70
42	推定されたカメラパス (予測無しの場合)	71
43	推定されたカメラパス (予測有りの場合)	71
44	カーナビゲーション情報の提示結果	72
45	MR-PreViz 環境のデータベースの構築	74
46	試作システムの機器構成	75
47	CG キャラクタの合成結果 (CG モデルは立命館大学 田村研究室より提供)	77
48	ランドマークと対応づいた自然特徴点の位置	78
49	復元された川原寺の CG モデル (東京大学 池内研究室より提供)	81
50	川原寺跡地	81
51	バーチャル歴史体験システムの機器構成	82
52	ユーザの外観	82
53	入力画像 (川原寺)	83

54	川原寺のCGモデルの合成結果	84
55	センサを用いた移動量の計測 [神原 08]	87
56	撮影日時の違いによる環境 (照明条件) の変化の例	88

表 目 次

1	実験で用いたパラメータ	39
2	1 フレーム当たりの処理時間の比較 (ミリ秒)	44
3	推定精度・誤対応の発生率の比較	45
4	誤対応の発生率と推定に用いられたランドマーク数の比較	46
5	相関値の比較	56
6	推定精度の比較	57
7	各アプリケーションにおける三次元情報の取得方法	63

1. 序論

現実環境を撮影した画像中にCGなどの仮想物体を描画することで情報を付加するビデオスルー型拡張現実感を実現するためには、一般に、カメラの位置・姿勢を推定する必要があり、現在までに様々な手法が提案されている。中でも、画像に基づくカメラ位置・姿勢推定手法は、画素単位での高精度な位置合わせを実現することが可能であり、特にランドマークデータベースに基づく手法 [大江 05] は広域で複雑な環境においても比較的容易に適用できるという特長を持つ。しかし、従来のランドマークデータベースに基づく手法は、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照合コストが高く実時間での処理が困難であり、また、カメラの近くに存在するランドマークは撮影位置の違いにより見え方が大きく変化するため、入力画像中の自然特徴点と正しく対応付けることが難しく、ユーザに近い位置にCGを重畳表示する場合に十分な推定精度が得られずジッタが生じるという問題が残っていた。そこで、本研究では、従来のランドマークデータベースに基づく手法 [大江 05] を基礎とし、まず、対応点候補となるランドマークおよび自然特徴点数を削減することで照合コストを低減し高速なカメラ位置・姿勢推定を実現する。また、全周レンジファインダにより取得された密な奥行き情報を用いてランドマークの見え方を補正することで、ユーザに近い位置に存在するランドマークを効果的に対応付けることにより厳密な位置合わせが要求される場所におけるスポット的な位置合わせ精度の向上を実現する。

以下では、まず、背景となる拡張現実感を実現するために解決すべき技術的課題について述べ、拡張現実感におけるカメラ位置・姿勢推定手法の従来研究を概観する。次に、自然特徴点を用いたカメラ位置・姿勢推定処理の高速化手法について述べる。最後に、本研究の位置づけと方針を明確にする。

1.1 拡張現実感の技術課題

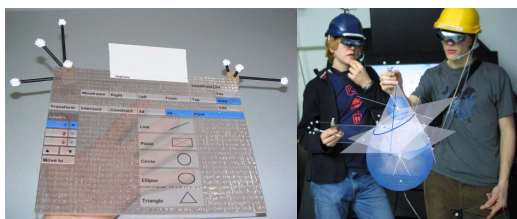
現実環境に仮想環境を実時間で重畳表示する拡張現実感技術は、図1に示すように、ナビゲーション [WS03, MMT⁺08]、景観シミュレーション [DK02, TNA⁺04, 角田 08]、教育支援 [KD07, LMW⁺04]、エンターテインメント [TCD⁺00, HSHS05]



(a) ナビゲーションへの応用 [WS03]



(b) 景観シミュレーションへの応用 [DK02]



(c) 教育への応用 [KD07]



(d) エンターテインメントへの応用 [TCD+00]

図 1 拡張現実感技術を用いたアプリケーション

などの様々な分野への応用が期待されている。これらの拡張現実感技術を用いたアプリケーションを開発するためには、仮想物体を違和感なく合成するために、以下に示すような現実環境と仮想環境の間の整合性問題を解決する必要がある [TYK01]。

幾何学的整合性

現実環境と仮想環境との位置合わせ、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。

光学的整合性

実物体と仮想物体に生じる表面の陰影や影，現実環境と仮想環境との画質のずれ，写り込みなどの光学的な整合性に関する問題．

時間的整合性

ユーザの移動，視線の変化，実環境の変化などによって生じる現実環境に対する仮想物体の追従の時間遅れの問題．

光学的整合性の実現は，インテリアデザインや製品設計のような写実的な拡張現実感を必要とするアプリケーション [安室 03] で重要である．光学的整合性を実現するために，鏡面球や魚眼カメラを用いて光源環境を推定し仮想環境に反映させる手法 [神原 04, SSH06] や合成対象となる画像と仮想物体の間の画質を一致させる手法 [KM08, OKY06] などが提案されている．時間的な整合性は，特に光学シースルー型の拡張現実感システムで問題となる．光学シースルー型の拡張現実感とは，現実環境にハーフミラーなどを用いて仮想環境を重畳することで，拡張現実感を実現する．そのため，現実環境を時間遅れなく提示できるが，仮想環境を提示する際の遅延により現実環境と仮想環境の間に位置ずれが生じるという問題がある．一方，ビデオシースルー型の拡張現実感システムでは，提示される環境全体が，合成画像を生成するための計算時間などによって遅れて提示される．しかし，現実環境と仮想環境の遅延は等しいため，遅延による両者の位置ずれは生じにくい．幾何学的整合性の問題は，拡張現実感を実現する上で最も基本的な問題であり，特に，現実環境と仮想環境との位置合わせ問題の解決は必要不可欠であり現在までに様々な手法が提案されている．本研究も現実環境と仮想環境との位置合わせを実現するための研究であり，次節ではこれに関する従来アプローチについて概観する．

1.2 拡張現実感技術におけるカメラ位置・姿勢推定手法の分類

初期の拡張現実感システムでは，ジャイロなどのセンサを用いてユーザの位置・姿勢を計測し，光学シースルー型ディスプレイを用いて仮想環境を提示するようなシステムが数多く開発されていた [Azu97, 佐藤 03]．しかし，光学シースルー

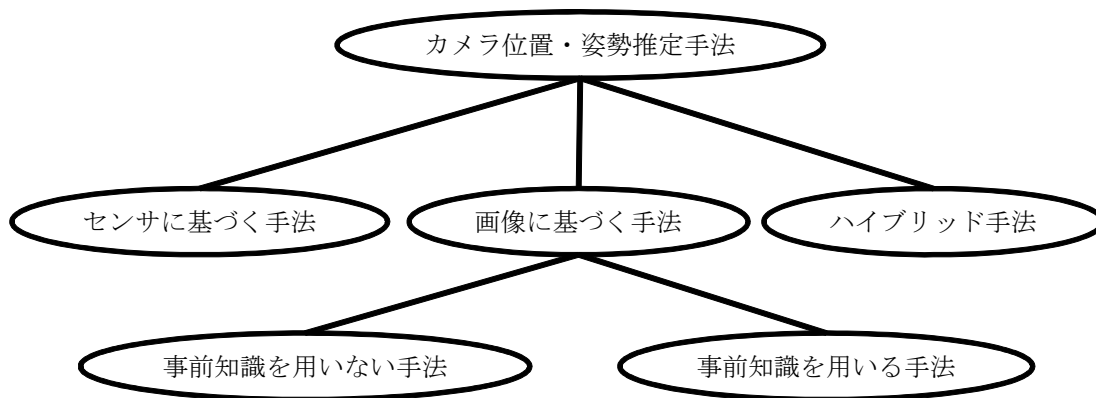


図 2 カメラ位置・姿勢推定手法の分類

型の拡張現実感システムは，時間的整合性の問題や実物体と仮想物体の隠蔽関係を再現することが難しいという問題がある．一方で，ビデオスルー型の拡張現実感システムは，カメラで撮影した画像中に仮想物体を重畳表示することにより，時間的整合性の問題や実物体と仮想物体の隠蔽関係の再現を比較的容易に実現できることから，近年の拡張現実感システムの主流となっている．

ビデオスルー型拡張現実感では，一般に，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 2 に示すようにカメラ以外のセンサを用いる手法，カメラで撮影された画像を用いる手法，それらのハイブリッド手法に大別できる．カメラ以外のセンサを用いる手法は，利用環境中にインフラを設置する手法 [HHS⁺02, PCB00] 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 [HFP99, TKY04, AHNS99, 神原 05] がある．このような，センサに基づくカメラ位置・姿勢推定手法は，ロバストなカメラ位置・姿勢の推定が可能であるが，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．一方で，近年，計算機の性能の向上によって，画像に基づくカメラ位置・姿勢推定手法が盛んに研究されている．画像に基づくカメラ位置・姿勢推定手法では，撮影した画像を CG などの仮想物体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画素単位での高精度な位置合わせを実現している。このような、画像に基づくカメラ位置・姿勢推定手法は、利用環境に関する事前知識を用いる手法と利用環境に関する事前知識を用いない手法に大別できる。以下では、それぞれの手法の特徴と問題点について述べる。

1.2.1 利用環境に関する事前知識を用いない手法

事前知識を用いない手法として、Simon らは図 3 のように多くの人工物は平面で構成されていることを仮定し、初期フレームにおいて手動で指定した平面領域を追跡することによりカメラの相対的な位置・姿勢を推定する手法を提案している [SFZ00, SB02]。これに対して、任意のシーンにおいて事前知識を用いずにカメラの位置・姿勢を推定する手法が提案されている [DMM03, BWS06, ED06, CGCMC07, KM07]。これらの手法は 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により、図 4 に示すようなカメラ位置・姿勢の推定と環境の三次元構造の取得を同時に行い、事前知識なしに拡張現実感環境を構築することができる。

このアプローチでは、一般に、自然特徴点の追跡、カメラ位置・姿勢の推定、自然特徴点の三次元位置の推定・更新、を繰り返すことでシーンの三次元構造を随時更新していくため、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。これに対して、Klein らはカメラ位置・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 [KM07]。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。ただし、Visual-SLAM によるカメラ位置・姿勢推定は、広域な環境を対象とした場合には、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、現実環境と仮想環境の幾何学的関係が得られない。このような問題に対して、インタラクティブな操作によって、随時、仮想物体を配置する手法 [RED07, LH09]

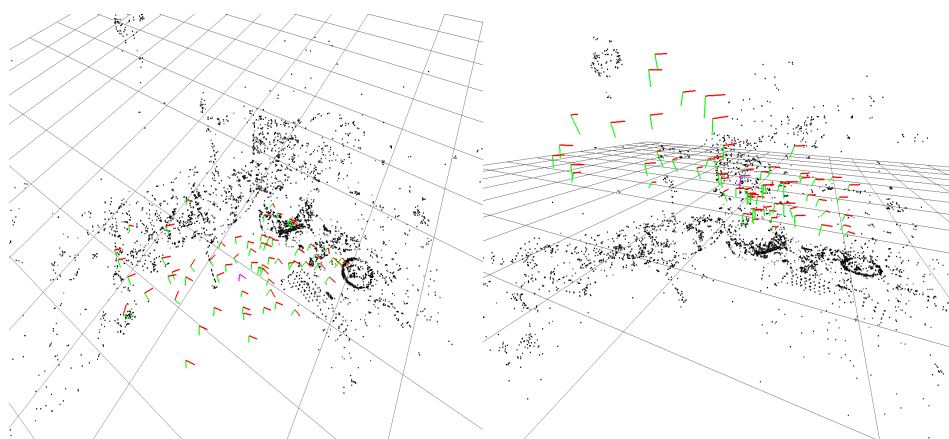


図 3 平面領域の追跡によるカメラ位置・姿勢推定 [SB02]

や環境中の一部の領域に対するワイヤフレームモデルを併用することで現実環境と仮想環境との位置合わせを実現する手法 [BWS06] が提案されているが、ナビゲーションのような現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への応用は難しい。



(a) 入力画像



(b) 推定されたカメラ位置と自然特徴点の三次元位置

図 4 Visual-SLAM による拡張現実感環境の構築 [KM07]

1.2.2 利用環境に関する事前知識を用いる手法

事前知識を用いる手法では，マーカ，画像データベース，三次元モデル，自然特徴点データベースなどを事前知識として用い，オンラインで撮影された画像中に存在するこれらの特徴を対応付けることでカメラの位置・姿勢を推定する．これらの手法では，多くの場合，図5に示すように，特徴点の三次元位置 F_p と画像上の検出座標 x_p の n 組の対応から PnP 問題を解くことによってカメラの位置・姿勢 M_{wc} を推定する．PnP 問題の解法としては，様々な手法が提案されているが，一般には，高精度にカメラ位置・姿勢の推定が可能な再投影誤差最小化によるアプローチが採用されている(付録 A 参照)．以下では，利用する事前知識の種類によって手法を分類し，それぞれの手法の特徴と問題点について述べる．

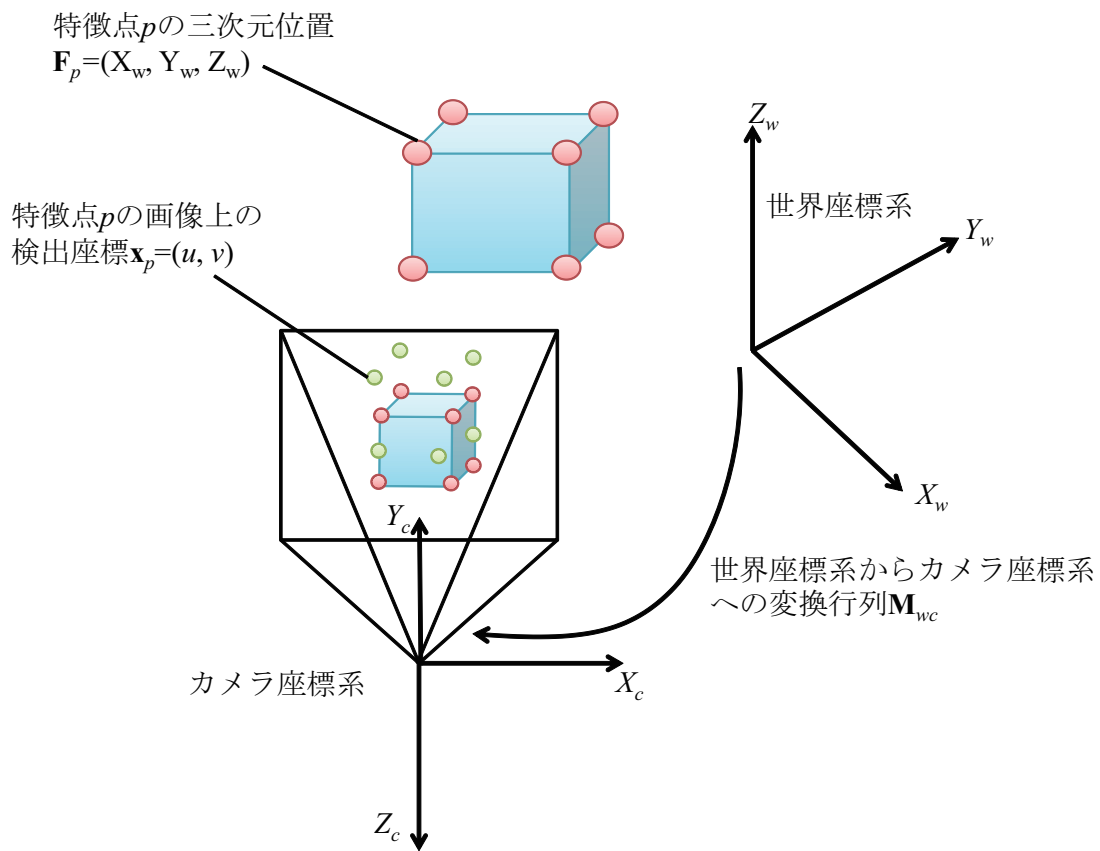


図 5 カメラ座標系と世界座標系の関係

マーカを用いる手法

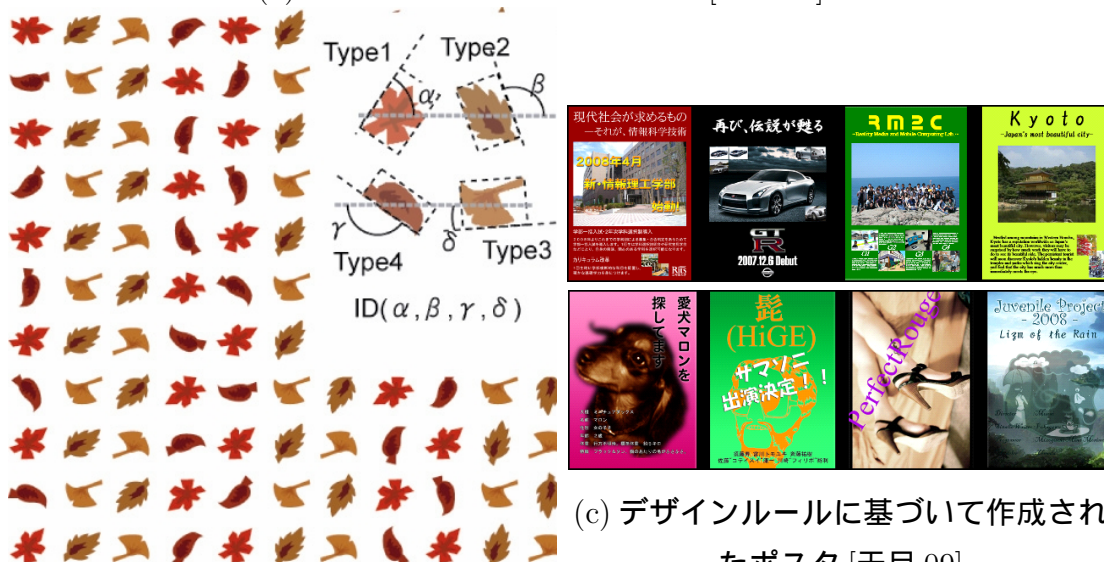
マーカを用いる手法では、パターンや形状などが既知の画像マーカを撮影画像中から抽出することでカメラ位置・姿勢を推定する [KB99, 羽原 02, KLK⁺02, WS03, NF02, 中里 05, SHTH07, 天目 09]。画像中のマーカの識別は比較的容易であり、実時間で蓄積誤差を含まないカメラ位置・姿勢の推定が可能である。このような画像マーカを事前に環境内に設置し、その三次元位置をデータベースとして保持しておくことでカメラの位置・姿勢を推定することが可能である。しかし、広域で複雑な環境で利用することを想定した場合には、これらの手法では環境内に多数のマーカを設置する必要があり、マーカの設置に多くの人的コストがかかるとい



フラッシュ無しで撮影

フラッシュ有りで撮影

(a) 再帰性反射材を用いたマーカ [中里 05]



(b) 位置情報を付加したパターン

[SHTH07]

(c) デザインルールに基づいて作成されたポスタ [天目 09]

図 6 景観への影響を抑えたマーカ

う問題や景観を損ねるといった問題がある [KB99, 羽原 02, KLK+02, WS03, NF02] .

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った問題に対し, 再帰性反射材を用いた不可視マーカを設置し, 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 [中里 05] (図 6(a)), 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 [SHTH07] (図 6(b)),

マーカとして利用可能なデザインルールに基づいて作成されたポスタをマーカとして利用する手法 [天目 09] (図 6(c)) などが提案されている。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。

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る手法は、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 [岩佐 01] や興相ら [興相 01] 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。Cipolla ら [CRT04] は、図 7 に示すように、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。

一方で、テーブルトップ上で拡張現実感を実現することを目的として、平面状の紙面を追跡の対象とし、データベース中の紙面の画像特徴と対応付けることでカメラの位置・姿勢を推定する手法が提案されている。これらの手法は Paper AR と呼ばれ、平面画像の特徴量としてテクスチャを用いる手法 [KBG03]、SIFT 特徴量 [Low04] を用いる手法 [WRM+08]、文書画像中の文字領域の重心位置から算出される特徴量を用いる手法 [USSM09] が提案されている。これらの手法は、デジタルサイネージや教育などへの応用が期待される。

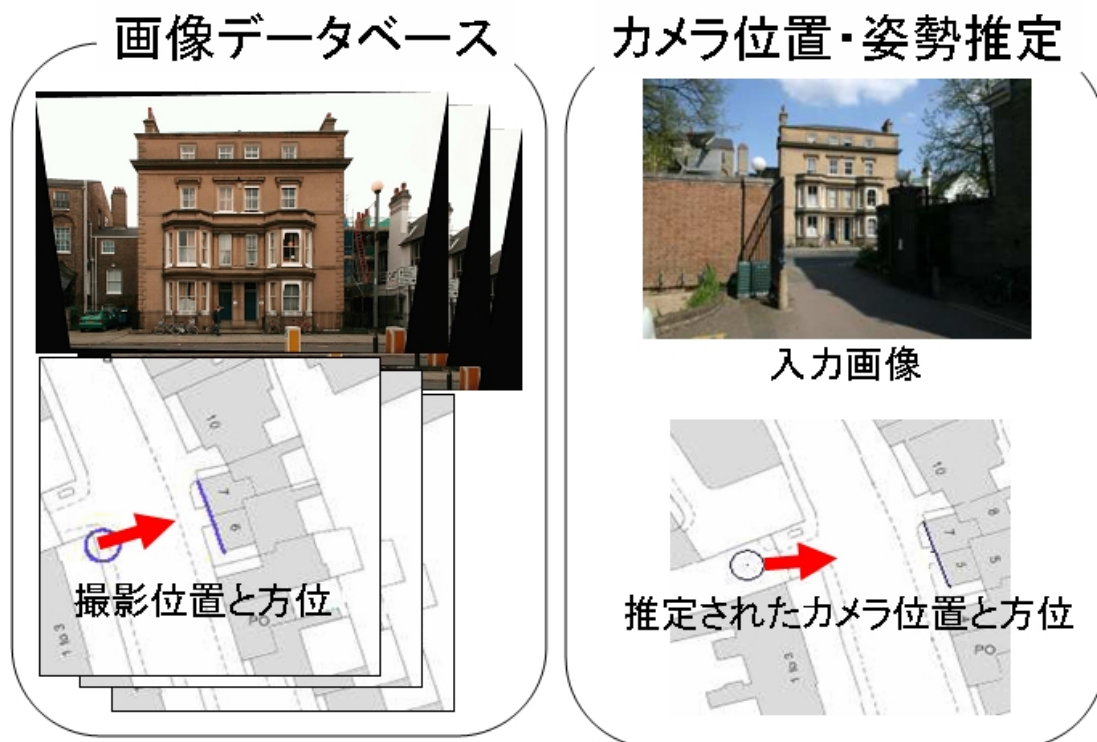


図 7 画像データベースを用いたカメラ位置・姿勢推定 [CRT04]

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 [DC02] や Comport ら [CMPC06] は，環境の三次元モデルとしてワイヤーフレームモデルを用い，図 8 に示すように入力画像中のエッジとワイヤーフレーム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，Lepetit ら [LVTF04] は，ワイヤーフレームモデルと複数枚のキーフレーム画像を用いることでキーフレーム画像中の自然特徴点と入

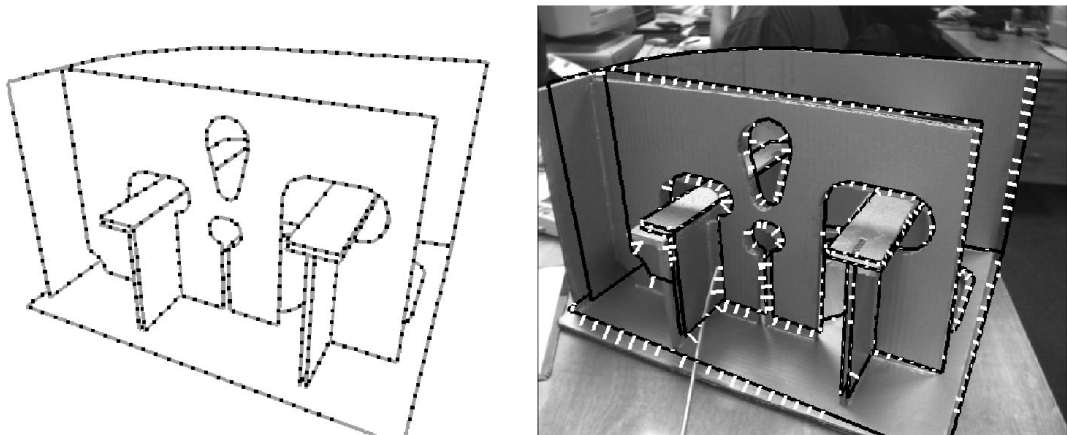


図 8 ワイヤーフレームモデルを用いたカメラ位置・姿勢推定手法 [DC02]

力画像中の自然特徴点を物体形状を考慮しながら対応づけることによりカメラの位置・姿勢を推定する手法を提案している。Vacchettiら [VLF04] 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。一方で、図9に示すようなテクスチャ付き三次元モデルを用いる手法も提案されている [RD06a]。この手法では、形状特徴だけでなくテクスチャから得られる輝度情報も用いることで屋外環境における頑健なモデルの追跡を実現している。

これらの三次元モデルを用いた手法では、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。これを解決するために自動または半自動でモデルを作成する手法についても研究が行われている [Beh99, RC02, PL07, NPD07]。これらの手法は、特定の環境においてモデル作成の人的コストを低減することができるが、複雑なシーンに対してこれらの手法を適用することは難しい。一方で、環境の一部のワイヤーフレームモデルと Visual-SLAM を併用することでモデル作成のコストを低減する手法が提案されている [BWS06]。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



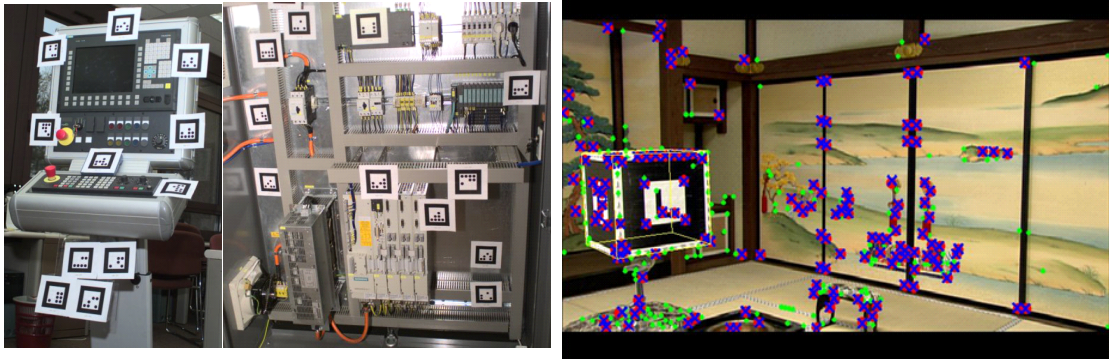
図 9 Reitmayr らの手法で用いられる三次元モデル [RD06a]

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，事前知識獲得のため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．

自然特徴点データベースを用いる手法

自然特徴点データベースを用いた手法は，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．

初期の自然特徴点データベースを用いた手法として，Ravela ら [RDLW95] や Uenohara ら [UK91] は事前に手動で自然特徴点の三次元位置とその見え方をデータベースに登録しておき，入力画像中からこれらの点を検出することでカメラの位置・姿勢を推定する手法を提案している．これらに対して，近年，Genc ら [GRS⁺02] や樋下ら [樋下 09] は，学習用の映像を撮影する際に，図 10 に示すように人工マーカをシーン中に配置しておき，これらのマーカに基づいて映像中の自然特徴点の三次元位置を推定し自動でデータベースを構築している．オンライン



(a)Genc ら [GRS⁺02]

(b)樋下ら [樋下 09]

図 10 マーカを用いた自然特徴点データベースの構築

での推定時にはマーカを撤去した上で自然特徴点ランドマークを検出することで、マーカを用いないカメラ位置・姿勢推定を実現している。しかし、広域な環境に適用する場合には、多数のマーカを設置する必要があり、多くの人的コストが必要となる。

Skrypnik ら [SL04] や Arth ら [AWK⁺09] では、structure-from-motion 法によって推定した自然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。これらの手法では、マーカを用いずに自動または半自動でデータベースを構築することが可能であるが、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がある。大江ら [大江 05] は、三次元位置が既知の少数の基準点を用いる手法により蓄積誤差の生じない環境の三次元復元 [佐藤 05] 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、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。Ikeda ら [ISYY07] は、データベース構築時に GPS を併用することによって、より広域な環境にお

いても効率よくランドマークデータベースを構築する手法を提案している。しかし，Skrypnyk らの手法 [SL04] と同様に，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 [大江 05, ISYY07] はオンライン処理において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。また，オンラインでのカメラ位置・姿勢推定時に，カメラの近くに存在するランドマークは撮影位置の違いにより見え方が大きく変化するため，入力画像中の自然特徴点と正しく対応付けることが難しく，ユーザに近い位置に CG を重畳表示する場合に十分な推定精度が得られずジッタが生じるという問題がある。

1.3 自然特徴点に基づくカメラ位置・姿勢推定の高速化手法

自然特徴点に基づく手法は，複雑な環境においても比較的容易にデータベースを構築することができるという特長を持つが，一般に，自然特徴点の抽出や対応点の探索に多くの計算コストを必要とするため，実時間でのカメラ位置・姿勢の推定が難しいという問題がある。しかし，近年，自然特徴点の抽出および対応点の探索を高速に実現可能なアルゴリズムが多数提案されている。以下では，自然特徴の抽出および対応点探索の高速化手法について特徴と問題点について述べる。

1.3.1 特徴点抽出の高速化手法

従来，拡張現実感の分野において画像中の特徴点を抽出する手法として KLT [ST94]，Harris [HS88] などのオペレータが用いられてきた。中でも，Harris オペレータは回転，拡大などの画像の変形操作を行った場合にも同一の点に特徴点が検出される再現率が高いオペレータである [SMB98] ことから多くの手法で用いられている。これらのオペレータに対して，近年，高速に特徴点を抽出することが可能な FAST と呼ばれるオペレータが提案されている [RD05, RD06b, RPD10]。FAST は，KLT や Harris などのオペレータと比較し 10 倍以上の処理速度で特徴点を抽出することができ，高周波のテクスチャが少ない画像において KLT や Harris と同等以上の再現率で特徴点を抽出することができるという特長を持つ。また，KLT

による特徴点抽出を高速化するために，多重解像度処理によって特徴点抽出を行う領域を限定する手法 [KM09] が提案されている．

一方，画像のスケールが変化した場合にも同一の位置に特徴点を抽出可能な手法が提案されている [Lin98, MS01, MS04, Low04]．これらの手法では，特徴点の抽出と同時に特徴点周辺の画像パターンから特徴点固有のスケール情報を算出しており，算出されたスケール情報に基づいて画像パターンの特徴量を抽出することで，画像スケールの変化に依存しないパターンマッチングを実現することができる．中でも，SIFT 記述子 [Low04] はスケール不変，回転不変の特徴量であることから多くの手法で用いられている．しかし，SIFT は特徴点の抽出処理の計算コストが高く，実時間での処理が必要なアプリケーションに適用することは難しいという問題があった．これに対して，高速に SIFT 特徴量を算出する手法として，FAST によって抽出された特徴点に対して SIFT 記述子を算出する手法 [WRM⁺08]，GPU を用いる手法 [SFPG06]，SURF と呼ばれる特徴点およびスケール情報を近似計算によって抽出する手法 [BETG08] などが提案されている．

1.3.2 対応点探索の高速化手法

データベースにランドマークとして登録されている自然特徴点と画像上の自然特徴との照合処理において，データベース中のランドマーク情報が増大すると，対応点の照合処理に必要な計算コストが膨大となり実時間での処理が困難となる．このような問題に対して，照合処理の高速化を行う手法として，KD Tree を用いた近似最近傍探索による手法 [SL04]，Randomized Tree を用いた分類木を用いる手法 [SLF05, WKR07] などが提案されている．KD Tree を用いた近似最近傍探索は処理時間を削減することができるが，最近傍が距離計算の対象から外れてしまい求まらない可能性がある．また，Randomize Tree を用いた手法では多くの見えの変化を一つの決定木で表現するのは困難であるという問題がある．これに対して，Özuysal らは，事前に特徴点周辺のパターンに対してアフィン変換や雑音を付与することによって，見え方を統計的にモデル化しスケール不変，回転不変の特徴量を抽出している [OCLF10]．この手法は，高速な照合処理を実現することが可能であるが，クラス分類による対応付け処理に多くのメモリ容量を必要と

するという問題がある．Keら [KS04] は，SIFT で検出した勾配情報に対して主成分分析を適用することで識別性能を向上させた PCA-SIFT を提案している．この手法では，主成分分析を行うことによって，特徴空間の次元数を削減することができるため，照合のための計算コストを削減することができる．

一方で，データベース中から対応付けの候補となるランドマークを高速に選択する手法が提案されている．Irscharaら [IZFH09] は，データベース構築時の画像中に現れる SIFT 特徴ベクトル群をクラスタリングし，Visual Words として辞書を作成しておくことで，Vocabulary Tree[NS06] を用いて対応付けの候補となるランドマークを高速にデータベース中より選択している．Arthら [AWK⁺09] は，事前にデータベース中のランドマークの可視判定を行い，コンピュータグラフィックスの分野で用いられていた Potential Visible Set というアプローチ [ARFPB90] を利用することで対応付けの候補となるランドマークを高速に選択する手法を提案している．

1.4 本研究の位置付けと方針

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付けを述べる．1.2 節で概観したように，これまでにカメラで撮影された画像からカメラの位置・姿勢を推定する手法が数多く提案されている．画像に基づくカメラ位置・姿勢推定手法には，Visual-SLAM による手法，マーカを用いる手法，画像データベースを用いる手法，三次元モデルを用いる手法，自然特徴点データベースを用いる手法などがある．図 11 に，利用環境の規模と各手法の適用範囲を示す．Visual-SLAM を用いた手法は，利用環境に対する事前知識を必要としないため，手軽に拡張現実感環境を構築することができるが，広域な環境ではカメラ位置・姿勢推定の誤差が蓄積する．また，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，仮想世界と実世界の幾何学的関係が得られない．一方で，事前知識を用いる手法では，事前に定義した絶対座標系でのカメラの位置・姿勢を推定することができるため，ナビゲーションなどの位置に依存したアプリケーションへの応用が可能である．しかし，マーカを用いる手法，画像データベースを用いる手法，三次元モデルを用いる手法では，広

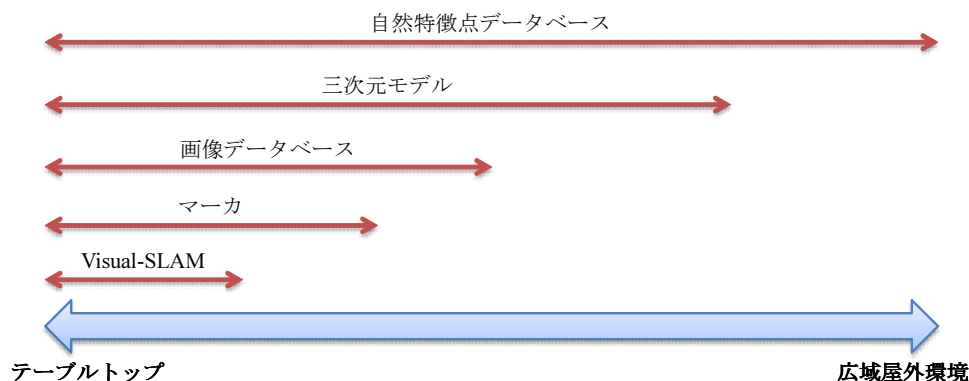


図 11 利用環境の規模と画像に基づくカメラ位置・姿勢推定手法の関係

域で複雑な環境において事前知識を獲得するための人的コストが膨大となるという問題がある．このような問題に対して，自然特徴点データベースを用いた手法では，structure-from-motion 法を用いることによって，比較的低い人的コストでデータベースを構築することができる．特に，大江らの手法 [大江 05] では，データベースの構築に蓄積誤差の生じない環境の三次元復元手法 [佐藤 05, ISYY07] を用いることで，広域で複雑な環境においても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という特長を持つ．ただし，大江らの手法では，オンラインでのカメラ位置・姿勢の推定処理において，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対応付け処理に多くの計算コストを必要とし，実時間での処理が困難である．また，オンラインでのカメラ位置・姿勢推定時に，カメラの近くに存在するランドマークは撮影位置の違いにより見え方が大きく変化するため，入力画像中の自然特徴点と正しく対応付けることが難しく，ユーザに近い位置に CG を重畳表示する場合に十分な推定精度が得られずジッタが生じるという問題がある．

そこで，本研究では，大江らの手法 [大江 05] を基礎とし，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，対応付けの候補となるランドマークと自然特徴点の数を削減することで対応付け処理の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する．また，厳密な位

置合わせが要求される場所において，全周レンジファインダにより取得した密な奥行き情報を用いてランドマークの見え方を補正することで，スポット的な位置合わせ精度の向上を実現する．

1.5 本論文の構成

本論文の構成は以下の通りである．2章では本研究の基礎となるランドマークデータベースの従来研究 [大江 05] について概観する．3章では，自然特徴点の追跡処理およびランドマークへの優先度情報の付加によりオンラインでのカメラ位置・姿勢推定処理の計算コストを削減する手法について述べる．4章では，密な奥行き情報を用いることでランドマーク周辺の局所形状を考慮した画像テンプレートの補正を行うことで，カメラ位置・姿勢推定の精度を向上させる手法について述べる．5章では，提案手法を用いた実アプリケーションの例を示す．最後に，6章で本論文のまとめと今後の展望について述べる．

2. 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・姿勢推定の従来研究

本章では、本研究の基礎となるランドマークデータベースを用いたカメラ位置・姿勢推定手法[大江05]について概観する。図12に示すように、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、オフラインでのランドマークデータベース構築処理とオンラインでのカメラ位置・姿勢推定処理に分かれている。以下では、それぞれの処理について詳しく述べる。

2.1 ランドマークデータベースの構築

ここでは、オフラインでのデータベース構築処理について述べる。大江らの手法[大江05]では、まず、利用環境の三次元情報を取得するために、利用環境中を移動しながら撮影した全方位動画像に対して structure-from-motion 法を用いた三次元復元を行う。次に、structure-from-motion 法の結果からランドマーク情報を取得し、データベースに登録する。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、structure-from-motion 法による環境の三次元復元およびランドマーク情報の取得について述べる。

2.1.1 ランドマークデータベースの構成要素

大江らの手法では、ランドマーク情報として (a) 自然特徴点の三次元位置、(b) 撮影地点ごとの情報がデータベースに保持される。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) テンプレート作成時の視点位置から成る。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

2.2 節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。

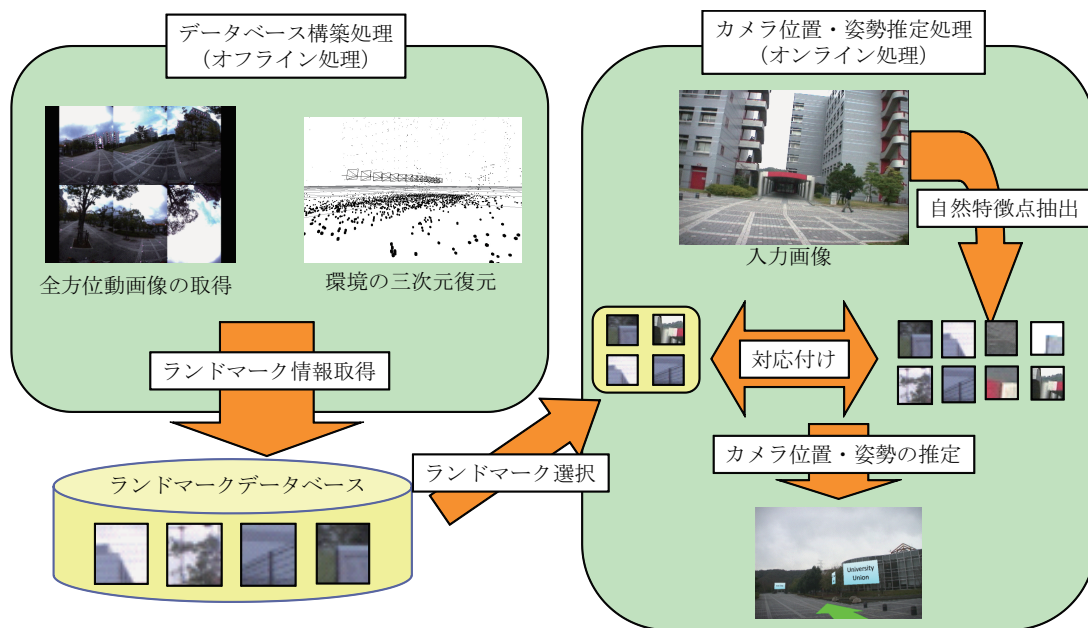


図 12 ランドマークデータベースを用いたカメラ位置・姿勢推定 [大江 05]

(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 13 に示すような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。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。

2.1.2 structure-from-motion 法による環境の三次元復元

利用環境の三次元情報を取得するために structure-from-motion 法による三次元復元を行う。ここでは、佐藤らの手法 [佐藤 05] または Ikeda らの手法 [ISYY07]



図 13 多重スケールの画像テンプレート

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 [HS88] 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 [佐藤 05] または GPS から得られる絶対位置 [ISYY07] 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．

具体的には，図 14 に示すような再投影誤差に基づく誤差関数 E を特徴点の三次元位置 p_j および全方位カメラの位置・姿勢 R_i, t_i をパラメータとして最小化する．

- 少数の基準点を用いる場合 [佐藤 05]

$$E = \sum_{i \in \mathbf{F}} A_i \sum_{c=1}^n \sum_{j \in \mathbf{P}_{ic}} W_j \Phi_{ijc}^2 \quad (1)$$

- GPS を用いる場合 [ISYY07]

$$E = \frac{\omega}{|\mathbf{F}_{gps}|} \sum_{i \in \mathbf{F}_{gps}} \Psi_i + \frac{1 - \omega}{\sum_{i \in \mathbf{F}} \sum_{c=1}^n |\mathbf{P}_{ic}|} \sum_{i \in \mathbf{F}} \sum_{c=1}^n \sum_{j \in \mathbf{P}_{ic}} W_j \Phi_{ijc}^2 \quad (2)$$

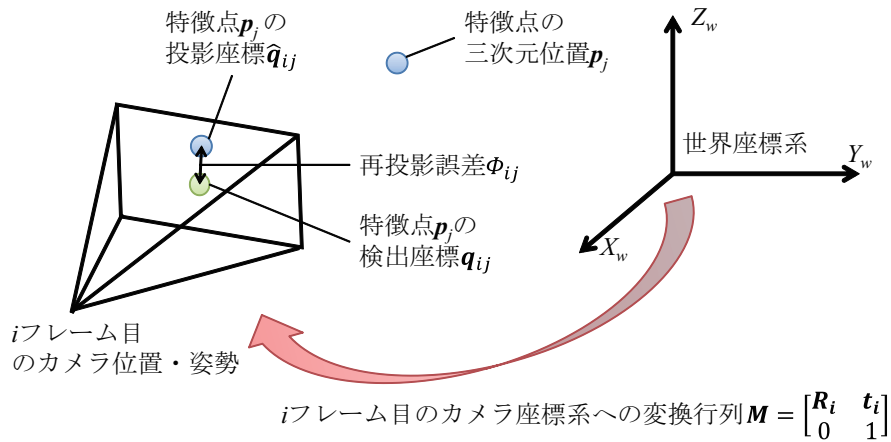


図 14 座標系と再投影誤差の定義

ここで、 F 、 F_{gps} はそれぞれ動画全体フレームの集合と GPS 測位情報が得られたフレームの集合、 P_{ic} は第 i フレームにおいて要素カメラ c の画像上で検出される特徴点の集合、 Ψ_i は推定されたカメラ位置と GPS から得られる測位位置との誤差、 ω は Ψ_i および Φ_i に対する重み係数、 A_i は各フレームに対する重み係数、 Φ_{ijc} は全方位カメラの要素カメラ c に写る特徴点 j の第 i フレームにおける再投影誤差を表す。重み係数 A_i は基準点を与えられていないフレームでは 1、基準点を与えられているフレームでは 1 より十分に大きな値を与える。また、 W_j は特徴点 j の信頼度であり、全てのフレームでの再投影誤差の分散の逆数で与えられる。再投影誤差 Φ_{ijc} は、特徴点の画像上での検出座標 q_{ijc} と推定された三次元位置の画像上への投影座標 \hat{q}_{ijc} を用いて、以下のように表される。

$$\Phi_{ijc} = |q_{ijc} - \hat{q}_{ijc}| \quad (3)$$

このように外部指標として少数の基準点や GPS から得られる測位情報を用いることで、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とができる。

2.1.3 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

2.2 節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 2.1.2 節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，図 15 に示すように，世界座標系において，カメラの投影中心とランドマーク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．ただし，撮影姿勢に依存しない画像テンプレートを生成するためには，視差の影響が 1 画素以下である必要がある．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．

2.2 カメラ位置・姿勢の推定

本節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．オンライン処理では，まず初期フレームでのカメラ位置・姿勢を取得し，続いてランドマークの選択処理および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・姿勢推定処理を繰り返す．ここで，大江らの手法では，初期フレームのカメラ位置・

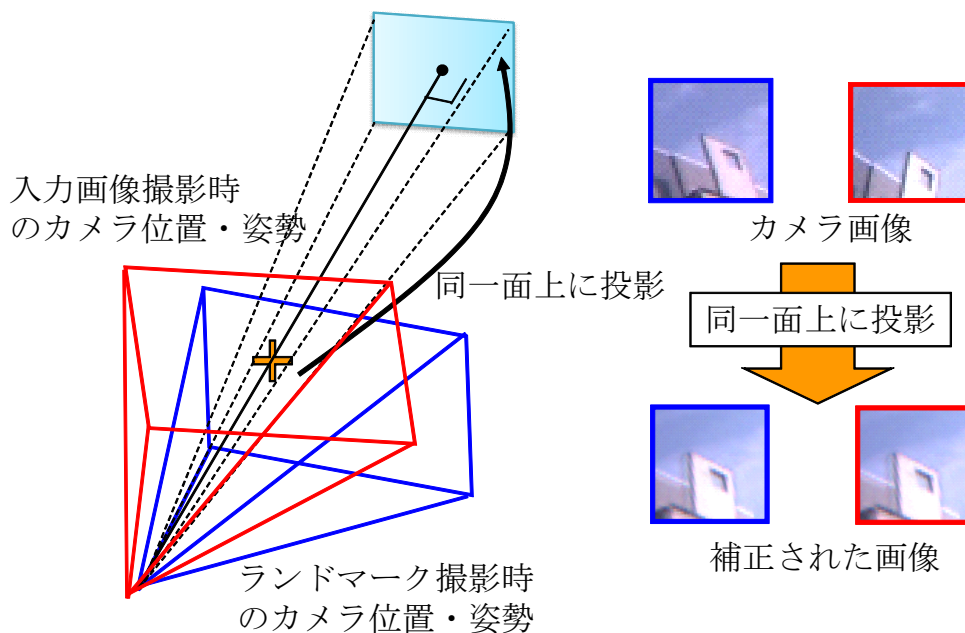


図 15 画像テンプレートの生成

姿勢は手動で与えることで推定処理を開始している．以下では，ランドマークの選択処理，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・姿勢推定処理について詳しく述べる．

2.2.1 ランドマークの選択

ここでは，前のフレームで推定されたカメラ位置・姿勢 M_{prev} 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 N 個選択する．

- 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，前フレームの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前フレームのカメラ位置との距離が閾値以下
- 入力画像上での他のランドマークとの距離が閾値以上

2.2.2 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

データベースより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、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、まず前のフレームで推定されたカメラ位置・姿勢 M_{prev} を用いて、データベースより選択されたランドマーク i を以下の式により入力画像上へ投影する。

$$\begin{bmatrix} a_i u_i \\ a_i v_i \\ a_i \end{bmatrix} = M_{prev} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (4)$$

ただし、 (x_i, y_i, z_i) はランドマーク i の世界座標系での三次元位置、 (u_i, v_i) はランドマーク i の画像上への投影位置、 a_i はカメラ座標系でのランドマーク i の奥行きを表す。次に、投影したランドマーク i の画像上の位置 (u_i, v_i) を中心とする一定ウィンドウ W 内の自然特徴点を検出する。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、自然特徴点周辺の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照合する。ここでは、オフラインでのデータベース構築時とオンラインでのカメラ位置・姿勢推定時における明るさの変化に対応するために正規化相互相関法により類似度の算出し、ウィンドウ W 内で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を用いて、PnP問題を解くことで現フレームのカメラの位置・姿勢を推定する [KSKE98]。現フレームのカメラの位置・姿勢は、LMedS 基準 [Rou84](付録 B 参照) 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。

従来手法においてランドマークと自然特徴点の対応付け処理にかかる計算コスト C_{conv} は、ウィンドウ W 内で検出される自然特徴点数を平均 F 個とすると

$$C_{conv} = NFA \quad (5)$$

となる。ここで A はランドマークと自然特徴点 1 個当たりの照合コストであり、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。

3. ランドマークへの優先度情報の付加と自然特徴点の追跡によるカメラ位置・姿勢推定の高速化

3.1 序言

本章では、従来の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 [大江 05] で問題となっていた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対応付け処理にかかる計算コストを削減することで実時間でカメラ位置・姿勢推定処理を実現する。対応付け処理にかかる計算コストは式 (5) に示したように、対応付けの候補となるランドマーク数 N および自然特徴点数 F 、ランドマークと自然特徴点 1 個当たりの照合コスト A に依存する。従来から、対応付け処理の計算コスト A を低減するために、様々な手法が提案されている [SL04, SLF05, WKR07, OCLF10]。ただし、これらの手法では、照合する画像は同一のカメラで撮影されていることを前提としており、異なるカメラで撮影された画像間での歪や画角の違いなどは考慮されない。しかし、事前知識を用いた拡張現実感システムでは、通常、事前知識獲得のために用いるカメラとユーザが利用するカメラは異なることが想定される。このように異なるカメラ間の照合を行う場合には、従来の計算コスト削減手法 [SL04, SLF05, WKR07, OCLF10] をそのまま用いることは難しい。ランドマークデータベースを用いたカメラ位置・姿勢推定手法も、データベース構築時に用いる全方位カメラとオンラインでのカメラ位置・姿勢推定時に用いる単眼カメラとの間のカメラ特性の違いを吸収する必要があるため、ランドマークと自然特徴点 1 個当たりの照合コスト A を低減することは困難である。そのため、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。

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 を削減するために、以下の 2 つの改良を加える。

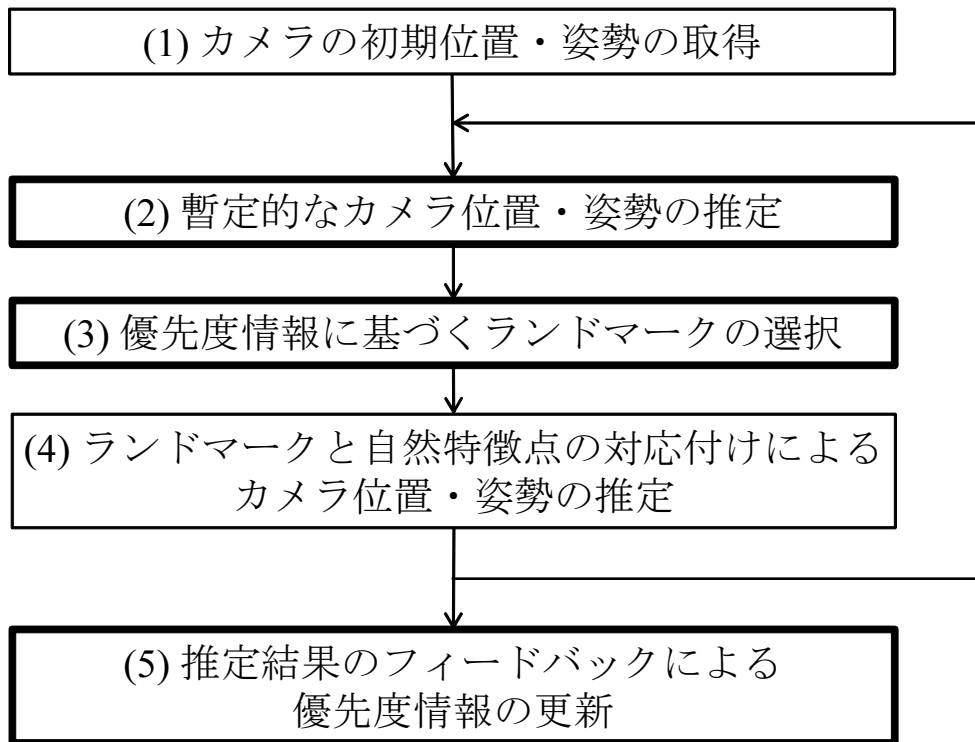


図 16 提案手法のオンライン処理の流れ

- (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

提案手法のオンラインでのカメラ位置・姿勢推定処理の流れを図 16 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。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・姿勢を取得し (1)、続いて逐次的なカメラ位置・姿勢推定処理 (2)~(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 (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 (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 (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 (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 (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 (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 (4) の処理コストを低減し，実時間でカメラ位置・姿勢推定を実現する．

3.2 ランドマークの優先度の定義

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，本手法では新たにランドマークへ優先度情報を付加する．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．

$$P_i = \frac{E_i}{D_i} \quad (6)$$

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，データベースの管理者が，あらかじめデータベース構築時にいくつかの学習データを与えることで，一般ユーザが利用する前に優先度を更新しておくことを想定する．この優先度の更新処理については 3.3.5 節で詳しく述べる．

3.3 優先度情報を用いた実時間でのカメラ位置・姿勢推定

本研究では，対応点候補となるランドマーク数と自然特徴点数を削減することで計算コストを低減する．図 17 に従来手法と提案手法の推定に用いる対応点候補の絞り込み方法の違いの概念図を示す．本研究では，推定精度を低下させず，計算コストを削減するために，推定に有効なランドマークおよび自然特徴点を残しながら対応点候補の削減を行う．具体的には，従来手法がランドマークの幾何学的な配置のみを考慮して選択処理を行っていたのに対して，提案手法では，優先度情報に基づいてデータベース中よりランドマークを選択することで，従来よりもランドマークの候補点数を削減しながら従来手法と同等または同等以上の数の推定に有効なランドマークを抽出する．また，対応点候補となる自然特徴点に対しても，従来手法が前フレームのカメラ位置・姿勢を用いて対応付けの候補となる自然特徴点を選択していたのに対して，提案手法では，より現フレームのカメラ位置・姿勢に近い暫定的なカメラ位置・姿勢を推定することによって，推定に有効な自然特徴点を残しながら対応点候補となる自然特徴点数を削減する．以下では，オンラインでのカメラ位置・姿勢推定の各処理について詳しく述べる．

3.3.1 初期カメラ位置・姿勢の推定

従来手法 [大江 05] では，初期フレームのカメラ位置・姿勢を手動で与える必要があったが，本研究では，初期フレームのカメラ位置・姿勢は，ランドマークデータベースを用いた静止画像からのカメラ位置・姿勢推定手法 [薄 08] によって自動で取得する (1)．そのため，データベースの構築処理において，Harris-Laplacian [MS04] によって算出される自然特徴点の固有スケール，SIFT-descriptor [Low04] による SIFT 特徴ベクトルを算出し，データベースへ登録しておく [薄 08]．本研究で用いるランドマークデータベースの構成要素を図 18 に示す．ここで，自然特徴点の固有スケールは，自然特徴点周辺のテクスチャの局所構造から決定されるものであり，これを用いることで，一定の空間領域に対応する画像の局所領域を切り出し，自然特徴点を正しく対応づけることができる．

初期フレームの推定処理では，まず，入力画像中の自然特徴点の固有スケール，

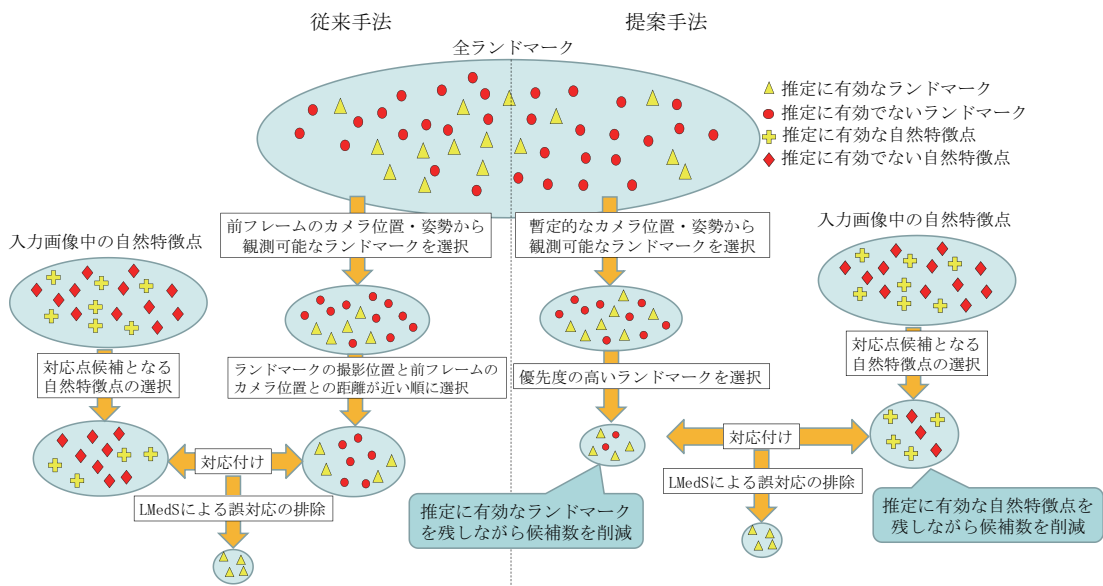


図 17 対応点候補の絞り込み方法の違い

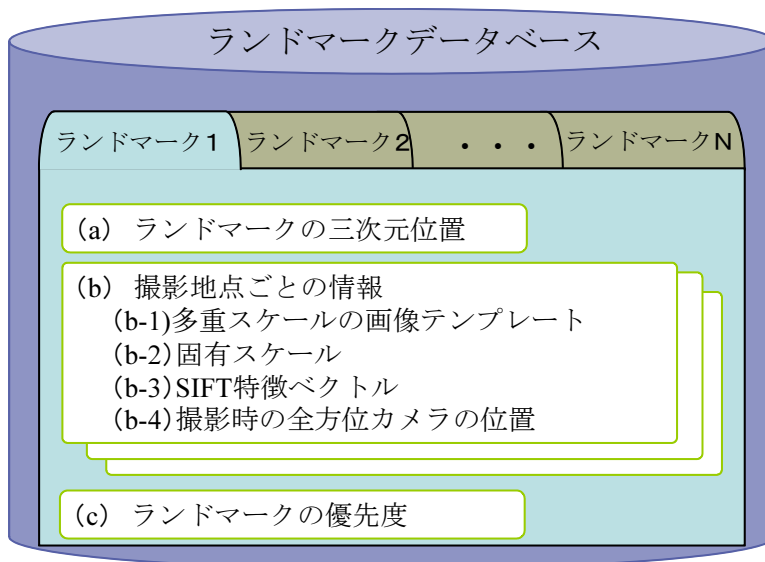


図 18 ランドマークデータベースの構成要素

SIFT 特徴ベクトルを算出した後，これらをデータベース中の全てのランドマークの SIFT 特徴ベクトルと照合することで自然特徴点とランドマークを対応付ける．次に，ランドマークの観測地点の整合性を利用して，対応付けられたランドマークから誤対応を排除する．最後に，得られた対応関係を用いて，PnP 問題 [KSKE98] を解くことで 6 自由度のカメラ位置・姿勢を推定する．

上記の初期位置・姿勢推定処理では，データベース中の全てのランドマークと入力画像上の自然特徴点を照合するため，推定処理の完了に数十秒を要する．従って，本研究では初期位置・姿勢の推定が完了するまでの間カメラを動かさないものとする．

3.3.2 自然特徴点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,

$$C_{track} = N_{track}FB + E_{LMedS} \quad (7)$$

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2.2.2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．

3.3.3 優先度情報に基づく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 (3) では, まず処理 (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．

3.3.4 ランドマークと自然特徴点の照合によるカメラ位置・姿勢の推定

ここでは，処理 (3) で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，まず処理 (2) で推定した暫定的なカメラパラメータ \hat{M} を用いて，処理 (3) で選択されたランドマーク i を入力画像上へ投影する．

$$\begin{bmatrix} a_i u_i \\ a_i v_i \\ a_i \end{bmatrix} = \hat{M} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (8)$$

次に，図 19 に示すように，投影したランドマーク i の画像上の位置 (u_i, v_i) 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照合する．ここでは，データベース構築時とカメラ位置・姿勢推定時の明るさの変化に対応するために，正規化相互相関法による照合を行い，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 (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位置・姿勢を推定する．現フレームのカメラ位置・姿勢は処理 (2) と同様に，LMedS 基準を用いて誤対応を排除した後に，再投影誤差を最小化することによって推定する．

3.3.5 推定結果のフィードバックによる優先度情報の更新

カメラの位置・姿勢推定の完了後，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．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

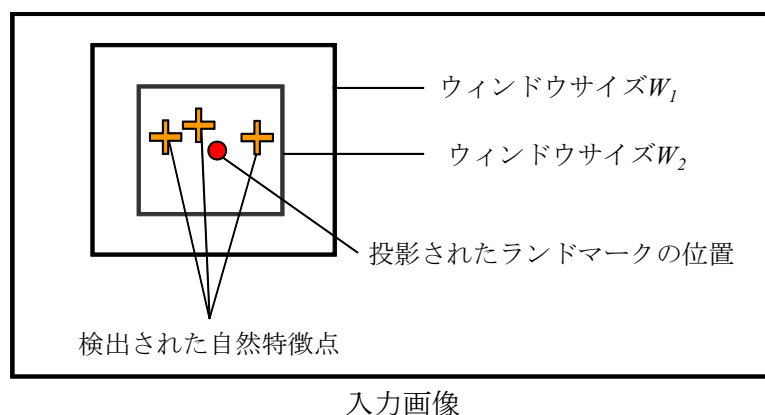


図 19 ランドマークの対応点の探索範囲

において，ランドマーク i が処理 (3) でデータベース中から選択された回数 D_{inew} と処理 (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．ランドマーク i の優先度 P_i は，ユーザより新たに取得した頻度情報 (D_{inew}, E_{inew}) と，過去に同じ環境内でカメラ位置・姿勢の推定を行ったユーザより得られている頻度情報 (D_{iold}, E_{iold}) を用いて，次式で更新する．

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

3.4 計算コストの比較

本節では，提案手法によるランドマークと自然特徴点の対応付けの理論的なコストの低減量について述べる．従来手法における対応付け処理の計算コストは $C_{prev} = NFA$ であるが，本研究では照合される自然特徴点数 F と対応付け処理に用いられるランドマーク数 N を削減することで，対応付けの計算コスト C_{prev} を低減する．提案手法では，ランドマークの探索範囲を限定するために，新たに暫定的なカメラパラメータを推定することが必要となるため，ランドマーク対応

付けの総コスト C_{new} は,

$$C_{new} = C_{track} + C_{proj} \quad (10)$$

となる．ただし, C_{track} , C_{proj} はそれぞれ, 暫定的なカメラパラメータの推定 (2), 処理 (4) における暫定的なカメラパラメータを用いたランドマークの対応付け, に必要な計算コストである．ここで, 処理 (2) の計算コスト C_{track} は, 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く, また, 追跡に用いるランドマーク数 N_{track} は N に対して小さいため, 提案手法で新たに導入した処理 (2) のオーバーヘッドは小さい．

処理 (4) では, 処理 (2) で推定された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, 対応点候補として用いられる自然特徴点の平均数を S_2/S_1 (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, また, 優先度に基づくランドマークの選択処理と, 処理 (2) で追跡されたランドマークを対応付けるランドマークから除外することにより, ランドマークの数 N を $(N_{prior} - N'_{track})/N$ に削減できる ($N_{prior} \leq N$)．これにより, 処理 (4) の計算コスト C_{proj} は,

$$C_{proj} = (N_{prior} - N'_{track}) \frac{S_2}{S_1} FA \quad (11)$$

となる．

ここで, 従来の対応付け処理に対する高速化の効果 (処理コストの比) は以下の式で算出できる．

$$\frac{C_{new}}{C_{prev}} = \frac{C_{track} + C_{proj}}{C_{prev}} \quad (12)$$

$$= \frac{C_{track}}{C_{prev}} + \frac{(N_{prior} - N'_{track}) S_2}{N S_1} \quad (13)$$

式 (13) において, 第 1 項は処理 (2) によるオーバーヘッドを, 第 2 項は処理 (4) における計算コストの削減効果を表している．実際には各処理の繰り返しにおいて一定のオーバーヘッドが生じるため, 第 2 項の計算コストの削減効果は 3.5 で述べる実験結果と完全には一致しない．

3.5 実験と考察

提案手法の有効性を示すために，屋外環境において実際にランドマークデータベースを構築し，実環境での実験を行った．ここでは，以下の4つの手法を比較し，カメラ位置・姿勢推定の処理速度向上効果について検証を行った．

(手法 A) 高速化処理無し (従来手法 [大江 05])

(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

(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 (提案手法)

ただし，現時点において構築したシステム (CPU: Core 2 Extreme 2.93GHz, メモリ: 2GB) では，初期位置・姿勢の推定に 30 秒程度を必要とするため，後述する実験では，初期位置・姿勢の推定中は環境内に固定された三脚にカメラを設置しておき，初期位置・姿勢推定の完了後にカメラを移動させている．

本実験ではビデオカメラ (SONY DSR-PD-150) を用いて撮影した 4 本の動画像 (解像度 720×480 画素，プログレッシブスキャン，15fps) に対してカメラの位置・姿勢推定を行った．そのうち，3 本の動画像の推定結果を優先度設定のために用い，残りの 1 本の動画像 (1000 フレーム) を評価用として用いた．カメラ位置・姿勢推定に用いた各パラメータを表 1 に示す．また，実験に用いたデータベースおよび真値の作成は以下のように行った．

データベースの構築：

本実験で用いるランドマークデータベースは，全方位型マルチカメラシステム (Point Grey Research 社 Ladybug) で屋外環境 (約 75m) を歩きながら撮影した全方位動画像 (1260 フレーム) を入力とし，佐藤らの手法 [佐藤 05] を用いることで構築した．図 20 に，データベースの構築に用いた全方位動画像の一部を示す．また，データベース構築時のカメラパスと復元されたランドマークの三次元位置を

表 1 実験で用いたパラメータ

	大江らの手法 [大江 05]	提案手法
追跡処理における 画像テンプレートのサイズ T (画素)	-	15
ウィンドウサイズ W_1 (画素)	-	120×60
ウィンドウサイズ W_2 (画素)	120×60	20×20
学習用シーケンス	-	3本

図 21 に示す．データベースに登録されたランドマーク数は約 12400 個であり，各ランドマークに対して，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．

真値の作成：

カメラ位置・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，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，真値作成の際に，三次元位置を与えた特徴点の画像上の再投影誤差が平均 1.5 画素以上となった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 6 点未満となったフレームについては真値が信頼できないため評価対象にしなかった．

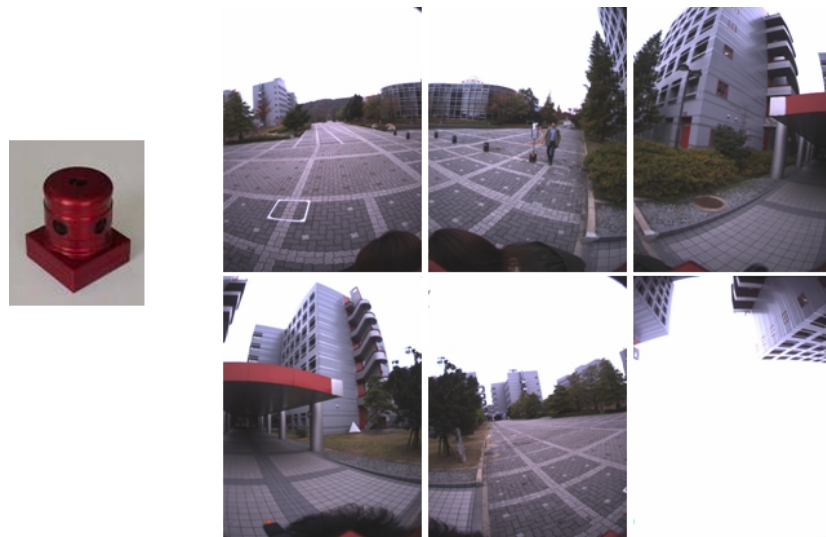


図 20 全方位型マルチカメラシステムとデータベースの構築に用いた全方位動画像の一部

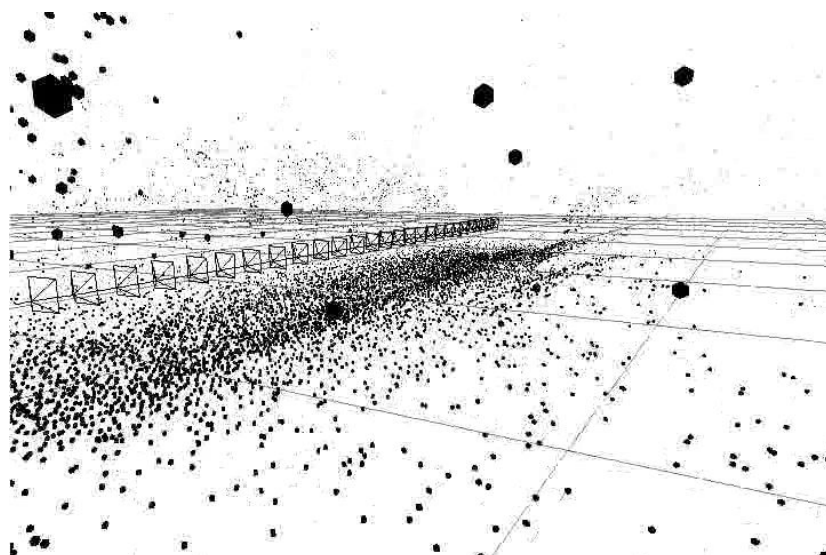


図 21 三次元復元の結果

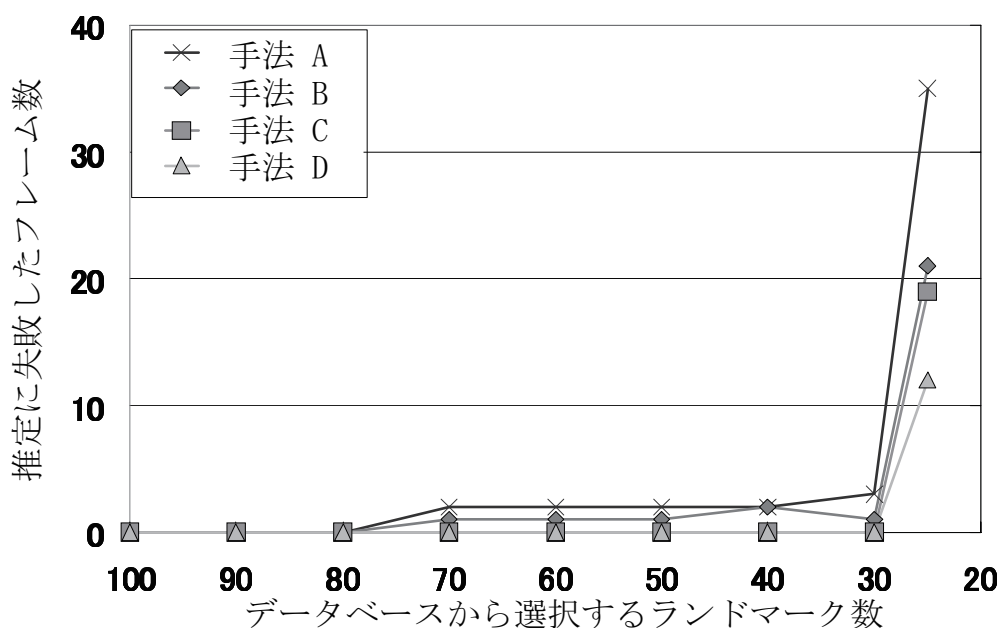


図 22 ランドマーク数と推定に失敗したフレーム数の関係

3.5.1 推定に用いるランドマーク数の決定

本実験では，データベースから選択するランドマークの最大数 N_{prior} を変化させ，各手法に対する推定失敗率を算出することで，処理 (3) において選択するランドマーク数 N_{prior} を決定した．ただし，優先度情報を用いない手法 A, B では，従来手法 [大江 05] と同様に幾何学的な制約のみを用いて撮影可能なランドマークを N 個選択した．

推定失敗率を算出するために，カメラ位置・姿勢の逐次推定処理において，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対応関係が 6 組以上得られなかった場合にカメラ位置・姿勢の推定に失敗したとみなした．また，推定に失敗した場合には，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失敗の回数を算出した．

図 22 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。優先度情報を用いない手法 A, B は、データベース中より選択するランドマーク数 N が 70 個以下の場合にカメラ位置・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, D では、ランドマーク数を 30 個まで減らした場合においても最終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。また、図 23 に従来手法においてカメラ位置・姿勢推定処理に失敗したフレームの例を示す。これらの画像のように、自然物が多く映り込み推定に利用できるランドマークが少ないシーンに対しても、提案手法では推定に失敗せずにカメラ位置・姿勢を推定することができた。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。この結果に基づき、3.5.2, 3.5.3 節で述べる処理速度および推定精度の比較実験では、データベースより選択するランドマーク数を手法 A, B では 80 個、手法 C, D では 30 個とした。



図 23 従来手法 [大江 05] で推定に失敗したフレーム

表 2 1 フレーム当たりの処理時間の比較 (ミリ秒)

手法	A	B	C	D
暫定的なカメラ位置・姿勢推定処理 (2)	-	26	-	21
ランドマークの選択処理 (3)	12	3	2	1
ランドマークと自然特徴点の対応付け処理 (4-1)	316	51	131	15
誤対応の排除とカメラ位置・姿勢の推定処理 (4-2)	61	16	16	17
オーバーヘッド	4	4	4	5
処理時間の合計	393	100	153	59

3.5.2 処理速度の比較

カメラ位置・姿勢の逐次推定処理 (2) ~ (4) の各処理時間を表 2 に示す。ただし、処理 (4) についてはランドマークと自然特徴点の対応付け処理 (4-1) と誤対応の排除と再投影誤差の最小化によるカメラ位置・姿勢の推定処理 (4-2) に分けて処理時間を計測した。

従来手法である手法 A と比較し、手法 B, C, D は、暫定的なカメラパラメータの推定による探索範囲の限定や優先度情報に基づくランドマークの選択を行うことによって、ランドマークと入力画像中の自然特徴点の対応付け処理 (4-1) の計算コストを削減できていることが確認できる。また、提案手法である手法 D では、全体の処理時間が平均 59ms となり、ビデオレートでのカメラ位置・姿勢推定を実現することができている。提案手法では、暫定的なカメラ位置・姿勢の推定 (2) に 21ms を要するが、これによってランドマークと入力画像中の自然特徴点の対応付け (4-1) の計算コストは従来手法 A の約 1/21 となった。なお、式 (13) を用いれば、提案手法によるランドマークと入力画像中の自然特徴点の対応付け処理 (4-1) の計算コストは、 $N = 80, N_{prior} - N_{track} \leq 30, S_2/S_1 = 1/18$ より、理論的には従来手法の 1/48 以下 (7ms 以下) となり実験結果 (15ms) と一致しない。これは、入力画像中に自然特徴点が一様に分布していないことや各処理の繰り返しにおいて一定のオーバーヘッドが生じるためであると考えられる。

3.5.3 推定精度の比較

各手法で推定されたカメラ位置・姿勢の精度を表3に示す。また、カメラ位置・姿勢推定処理においてアウトライヤとして排除されたランドマークの発生率とインライヤとして推定に用いられたランドマークの平均個数を示す。ここで、誤対応の発生率 $R_{outlier}$ は、ランドマークと自然特徴点の対応付け処理 (4-1) で対応づいたランドマーク数 N_{match} とカメラ位置・姿勢推定処理 (4-2) において誤対応として排除されたランドマーク数 $N_{outlier}$ を用いて以下の式で算出した。

$$R_{outlier} = \frac{N_{outlier}}{N_{match}} \quad (14)$$

有意水準を5%と設定したDunnett検定によって比較した結果、手法Aと手法B、C、Dの平均位置誤差の間には有意な差が見られ、従来手法Aに対して、提案手法B、C、Dではカメラ位置推定精度が向上していることを確認した。また、各手法の光軸方向の平均角度誤差の間には有意な差は見られず、提案手法B、C、Dは従来手法Aと同等の精度で推定できていることを確認した。提案手法B、C、Dにおけるカメラ位置推定精度の向上は、暫定的なカメラパラメータの推定による探索範囲の限定と優先度情報に基づくランドマークの選択を行うことによって、ランドマークと自然特徴点の対応付けにおける誤対応の発生率が低下したためであると考えられる。このことから、暫定的なカメラパラメータの推定による探索範囲の限定や優先度情報に基づくランドマークの選択は、ランドマークと入力画

表3 推定精度・誤対応の発生率の比較

手法	A	B	C	D
平均位置誤差 (mm)	360	257	231	256
位置誤差の標準偏差 (mm)	528	137	204	181
光軸方向の平均角度誤差 (度)	0.84	0.95	1.13	0.91
光軸方向の角度誤差の標準偏差 (度)	0.71	1.20	1.16	0.91
平均再投影誤差 (画素)	2.5	2.3	2.1	1.8

表 4 誤対応の発生率と推定に用いられたランドマーク数の比較

手法	A	B	C	D
誤対応の発生率 (%)	32	23	27	20
推定に用いられたランドマークの平均個数	26	27	21	23

像中の自然特徴点の対応付け処理 (4-1) の計算コストを削減できるだけでなく、カメラ位置推定精度の向上にも貢献していることが確認できる。

3.6 結言

本章では、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、安定性と処理速度の向上効果について検証した。実験の結果、従来手法と同等以上の精度で、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。

ただし、提案手法では、暫定的なカメラ位置・姿勢の推定に失敗した場合に、有効な対応点候補を選択することができなくなる。このような問題に対しては、文献 [RD06a, WKR07] などのようにランドマーク追跡処理の失敗を検出し、追跡処理に失敗する直前のカメラ位置・姿勢からの相対的な運動を推定することによって追跡処理に復帰することが考えられる。

4. 局所的な三次元形状を考慮したランドマークデータベースの構築によるカメラ位置・姿勢推定精度の向上

4.1 序言

本章では、ユーザに近い位置にCGを重畳する場合に、従来の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[大江05]では高精度に位置合わせすることが難しいという問題に対して、ランドマーク周辺の局所形状を考慮したランドマークの画像テンプレートを生成することで位置合わせ精度の向上を実現する。従来手法では、structure-from-motion法によって得られる疎な奥行き情報のみを用いてランドマークの見え方の情報を生成しているため、オンラインでのカメラ位置・姿勢推定時にユーザの近くに存在するランドマークの見え方を補正することが難しいという問題がある。そのため、図24に示す拡張現実感による文化遺産の復元のように、ユーザの近くに仮想物体を描画する場合には、表示されるCGにジッタが生じていた。

視点位置の変化に対してロバストな対応付けを実現する手法として、事前に特徴点周辺のパターンに対してアフィン変換や雑音を付与することによって、見え方の変化を統計的にモデル化する手法[OCLF10]、ランドマークを中心とする局所平面を仮定することで画像の補正を行った後にSIFT特徴量を算出する手法[WCL⁺08]、人工物などには直線が多いことから画像を行列とみなし行列のランクを最小化することで画像を補正する手法[ZLGM10]が提案されている。しかし、これらの手法を用いた場合にも、ランドマーク周辺の局所形状は平面であると仮定しているため、ユーザの近くに存在するランドマークのように、少しの視点移動で見え方が大きく変化するランドマークに対して、正しい対応付けを行うことは難しい。

そこで、本研究では、厳密な位置合わせが要求される場所におけるスポット的な位置合わせ精度の向上を目的として、全周レンジファインダにより取得した密な奥行き情報を用いることで、ランドマークの見え方を補正し、従来手法と比較してより多くのランドマークを画像上の自然特徴点と対応付ける。なお、取得した奥行き情報には、オクルージョンや計測範囲外にある物体などのようにレーザ



図 24 拡張現実感による文化遺産の復元

が照射されなかった未計測部分が存在する．そのため，本研究では，オンラインでのランドマークと入力画像中の自然特徴点の照合処理において，これらの未計測領域をマスクし，類似度を算出する際に除外する．

4.2 全周レンジファインダを用いたランドマークデータベースの構築

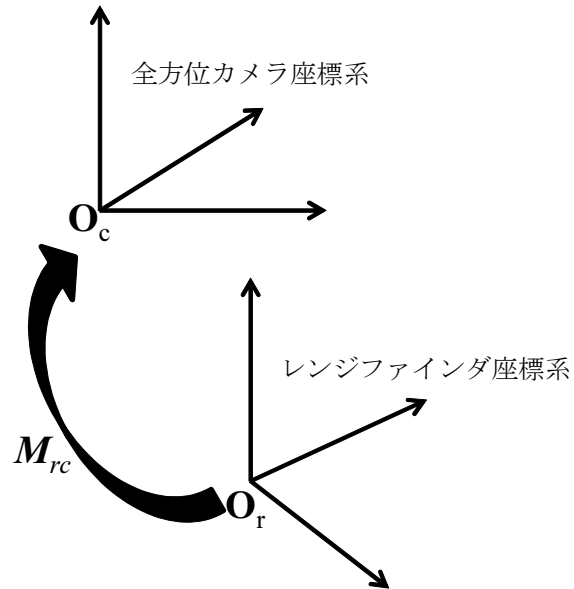
本節では，オフラインでのランドマークデータベースの構築処理について述べる．本研究では，ランドマークの見え方を補正するために，全周レンジファインダによって取得される密な奥行き情報を用いる．ランドマークデータベースの構築処理では，まず，全方位カメラと全周レンジファインダを用いて奥行き情報とテクスチャ情報の取得を行う．次に，これらの情報を用いてランドマーク情報を生成する．以下では，各処理について詳しく述べる．

全方位型マルチカメラシステム
(Point Grey Research社 Ladybug2)



全周レンジファインダ
(Reigl社 LMS-Z360)

(a) 計測システムの外観

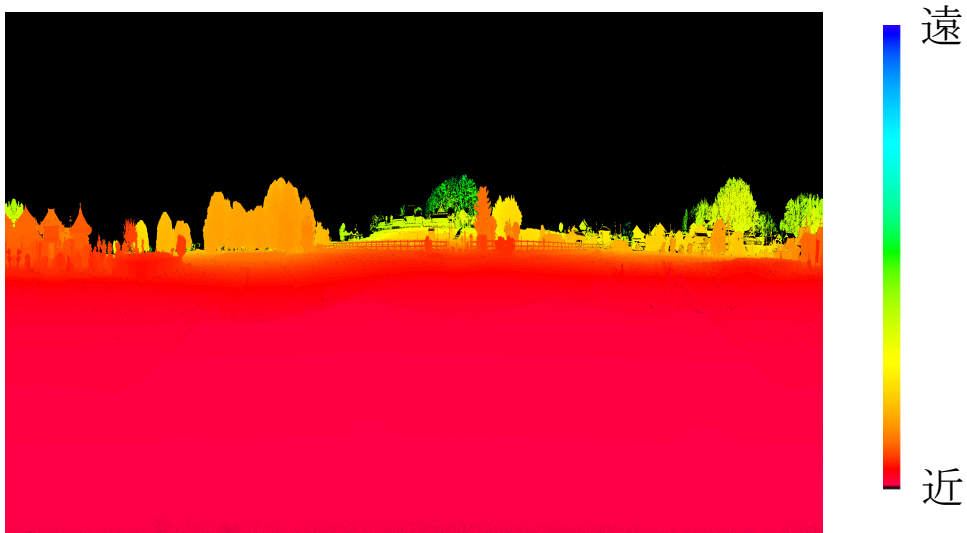


(b) 各センサの座標系の関係

図 25 計測システムと各センサの座標系の関係

4.2.1 奥行情報とテクスチャ情報の取得

本研究では、まず、全方位マルチカメラシステムと全周レンジファインダを用いて、利用環境のレンジデータおよびテクスチャ情報を取得する。データの取得を行う際、図 25 に示すように、全方位カメラと全周レンジファインダの幾何学的な関係は事前にキャリブレーションを行い固定しておく。全方位マルチカメラシステムと全周レンジファインダの相対関係 M_{rc} は、全周レンジファインダで計測した三次元点に対応する全方位画像上の点を手動で対応付け、全方位マルチカメラシステムに対する PnP 問題の解法 [佐藤 05] を適用することでレンジファインダ座標系におけるカメラ位置・姿勢を推定することによって算出する。図 26 に取得された全方位画像と全周レンジデータを示す。図 26(a) は全周レンジファインダで計測したレンジデータをセンサからの距離に応じた色付けを行って表示し



(a) 全周レンジデータ



(b) 全方位画像

図 26 全周レンジデータと全方位画像

ている．また，図 27 にセンサ間の相対関係 M_{rc} を用いて算出した図 26(b) の全方位画像に対応する奥行画像を示す．

ここで，取得した三次元形状情報には，空領域などの未計測部分が含まれており，これらの領域をパターンの照合時に単純に除外してしまうと，空と地上との

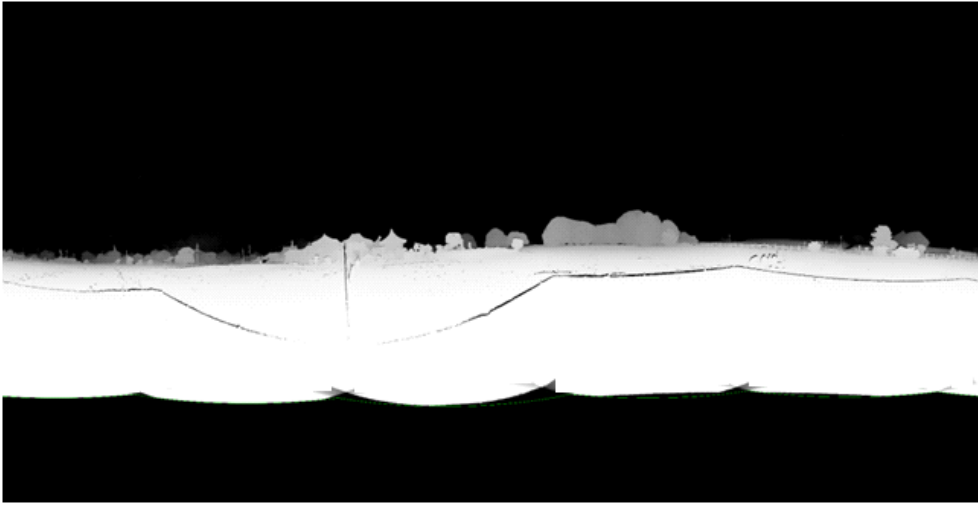


図 27 全方位奥行き画像



図 28 検出された空領域

境界部分に存在するランドマークにおいてアパーチャ問題が生じる。しかし、これらのランドマークは、カメラの姿勢を決定するために重要である。本研究では、利用環境の上空に遮蔽物が無い環境を想定し、アパーチャ問題を回避するために、

空領域には奥行き情報として無限遠を設定する．具体的には，取得した全方位奥行き画像上で，奥行き情報が設定されていない領域の内，最も大きな領域を空領域として検出する．図 28 に空領域の検出結果を示す．図中の赤色の領域が検出された空領域を表す．

4.2.2 局所形状を考慮したランドマーク情報の取得

本研究で用いるランドマークデータベースには従来手法と同様に，(a) 自然特徴点の三次元位置，(b) 撮影地点ごとの情報が保持される．撮影地点ごとの情報は，多重スケールの画像テンプレート，テンプレート作成時の視点位置から成る．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

本研究では，まず，取得した全方位画像より Harris オペレータ [HS88] を用いて自然特徴点を検出する．次に，全方位奥行き画像を用いて，自然特徴点の三次元位置を決定し，これらの自然特徴点をランドマークとしてデータベースへ登録する．

(b) 撮影地点ごとの情報の取得

撮影地点の違いによるランドマークの見え方の変化に対応するため，図 29 に示すように，地面を奥行き情報・テクスチャ情報の取得位置を中心とした格子状に分割し，各格子点上から見たランドマークの画像テンプレートを生成する．図 30 に，全方位画像から生成したランドマークの画像テンプレートと入力画像上のランドマーク周辺の画像パターンから生成した画像テンプレートを示す．ここで，画像テンプレートの生成位置は，入力画像上で自然特徴点の三次元位置を手動で対応付け，PnP 問題を解くこと推定したカメラ位置を用いた．図 30 左に，従来の structure-from-motion 法に基づく手法 [大江 05] によって生成されたランドマークの画像テンプレートを示す．従来手法では，structure-from-motion 法によって疎な奥行き情報しか得られないため，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を仮定し，画像テンプレート中の全ての画素に対して同一

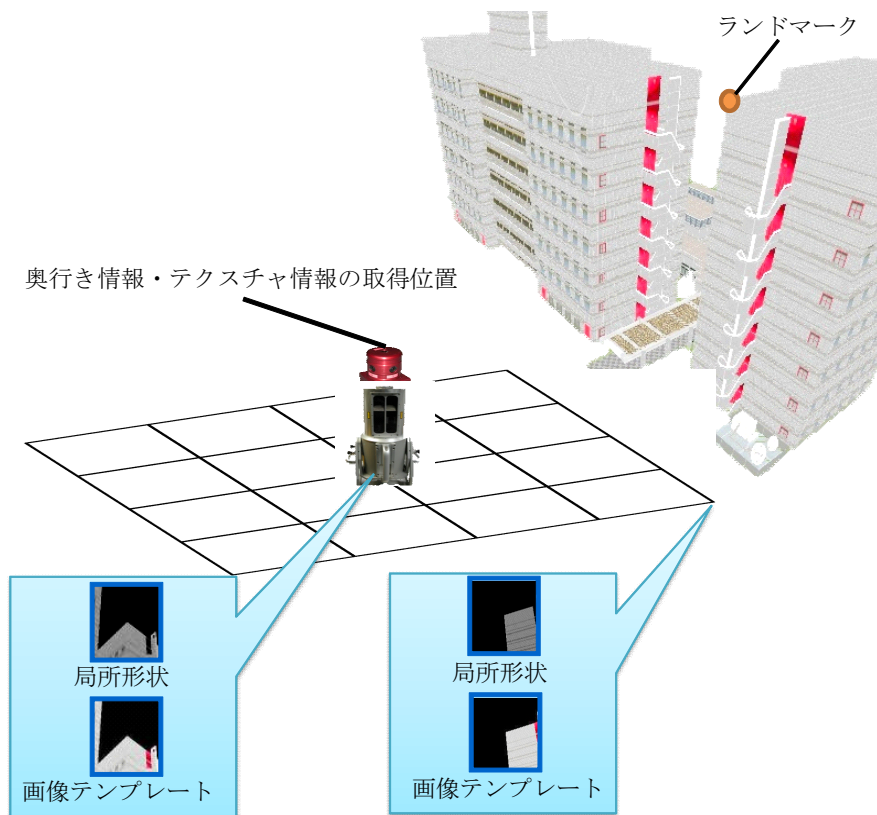


図 29 提案手法によるランドマーク情報の取得

の奥行き値を設定することで、テンプレートを作成していた。そのため、図 30 のように、ユーザの位置に近いランドマークの画像テンプレートは大きく歪んでしまうという問題がある。

本研究では、全周レンジファインダによって計測されたレンジデータを用いて、ランドマークの画像テンプレートを生成する。具体的には、まず、画像テンプレートの画素 i の奥行き値 d_i を、取得したレンジデータより算出する。次に、図 31 に示すように、算出した奥行き値 d_i を用いて、全方位画像を投影することで、画像テンプレートの各画素に対する輝度値を決定する。図 30 右に、提案手法によって生成された画像テンプレートを示す。提案手法では、密な奥行き情報を用いることで、適切な画像テンプレートを生成できている。

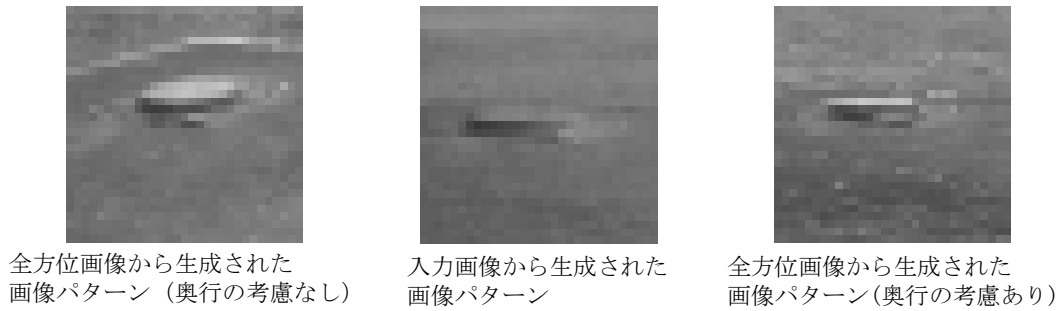


図 30 画像テンプレートの補正例

ただし，画像テンプレートの生成時点において，オクルージョンが生じている領域に関しては，オンラインでの照合処理の際に除外するためにマスク領域 Ω とする．具体的には，オンラインでのカメラ位置・姿勢推定処理において，入力画像中の自然特徴点の多重スケール画像パターン I とランドマークの多重スケール画像テンプレート T を照合する際に，以下に示す正規化相互相関により相関値 R_{NCC} を求める．

$$R_{NCC} = \frac{\sum_{i \notin \Omega} (T(p_i) - \bar{T})(I(p_i) - \bar{I})}{\sqrt{\sum_{i \notin \Omega} (T(p_i) - \bar{T})^2 \sum_{i \notin \Omega} (I(p_i) - \bar{I})^2}} \quad (15)$$

ただし， \bar{T} ， \bar{I} はそれぞれ T ， I の平均輝度値， $T(p_i)$ ， $I(p_i)$ はそれぞれテンプレート画像の画素 p_i に対する輝度値である．

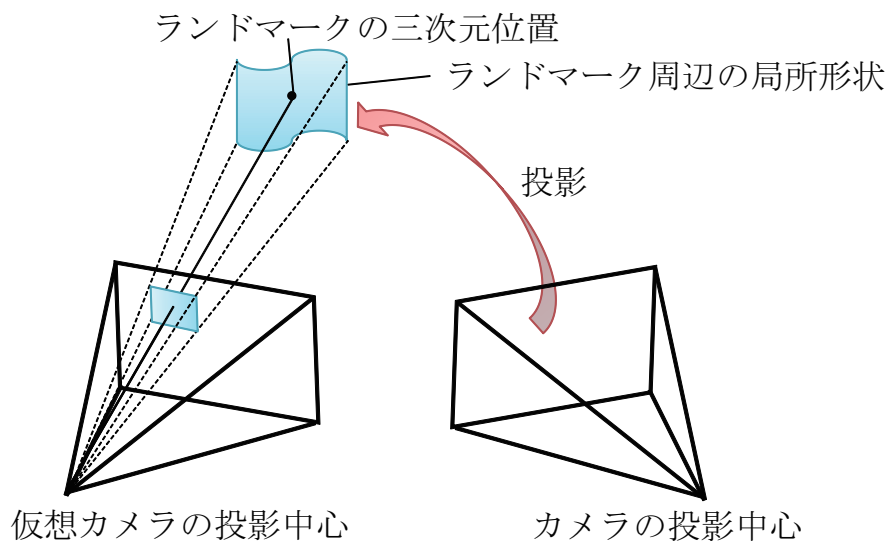


図 31 局所形状を考慮した画像テンプレートの生成

4.3 実験と考察

提案手法の有効性を示すため、屋外環境において実際にランドマークデータベースを構築し、実験を行った。ここでは、まず、ランドマーク周辺の局所形状を考慮することによる画像テンプレート補正効果について検証する。次に、ランドマーク周辺の局所形状を考慮して構築したデータベースとランドマーク周辺の局所形状を考慮せずに構築したデータベースを用いてカメラ位置・姿勢推定精度の比較を行った。本実験では、全方位マルチカメラシステム (Point Grey Research 社 Ladybug2) および全周レンジファインダ (Riegl 社 LMS-Z360) を用いて、屋外環境においてデータベースの構築を行った。本実験では、全周レンジファインダのレーザの有効距離を 150m、データの取得範囲を水平 360 度・垂直 90 度、解像度を 4500×2800 に設定した。データベース構築処理では、全方位データの取得位置を中心とする水平方向 $10\text{m} \times 10\text{m}$ の範囲を 1m 間隔の格子に分割し、各格子点上でランドマーク情報の生成を行った。本実験では、ビデオカメラ (SONY DSR-PD-150) を用いて撮影した動画 250 フレーム (解像度 720×480 画素、プログレッシブスキャン, 15fps) に対してカメラの位置・姿勢推定を行った。また、カ

表 5 相関値の比較

	局所形状の考慮あり場合	局所形状の考慮なし
平均	0.63	0.47
標準偏差	0.039	0.052

メラの内部パラメータはあらかじめ Tsai の手法 [Tsa86] によって校正し，初期フレームのカメラ位置・姿勢は画像上で自然特徴点の三次元位置を手動で対応付けることで与えた．

4.3.1 画像テンプレート補正効果の検証

まず，局所形状を考慮した場合の画像テンプレート補正効果について検証する．本実験では，図 30 に示すような全方位画像から生成したランドマークの画像テンプレートを入力画像上のランドマーク周辺の画像パターンと正規化相互相関法によって比較することで，提案手法によるパターンの補正効果を検証する．ここでは，画像テンプレートの補正に必要な入力画像の撮影位置を入力画像上で自然特徴点の三次元位置を手動で対応付け，PnP 問題を解くことで算出した．

ランドマーク 30 個の画像テンプレートに対して算出した正規化相互相関値の平均と分散を表 5 に示す．有意水準を 5% と設定した t 検定によって，奥行き情報を考慮して作成された画像テンプレートの相関値の平均値 (0.63) が奥行き情報を考慮しない従来の画像テンプレート補正方法による相関値の平均値 (0.47) には有意な差が見られ，提案手法は従来の画像テンプレート補正方法よりも入力画像上のランドマーク周辺の画像パターンに近い画像テンプレートを生成できていることが確認できる．これは，ランドマーク周辺の局所形状を考慮することで，適切に画像テンプレートの補正が行われたためだと考えられる．

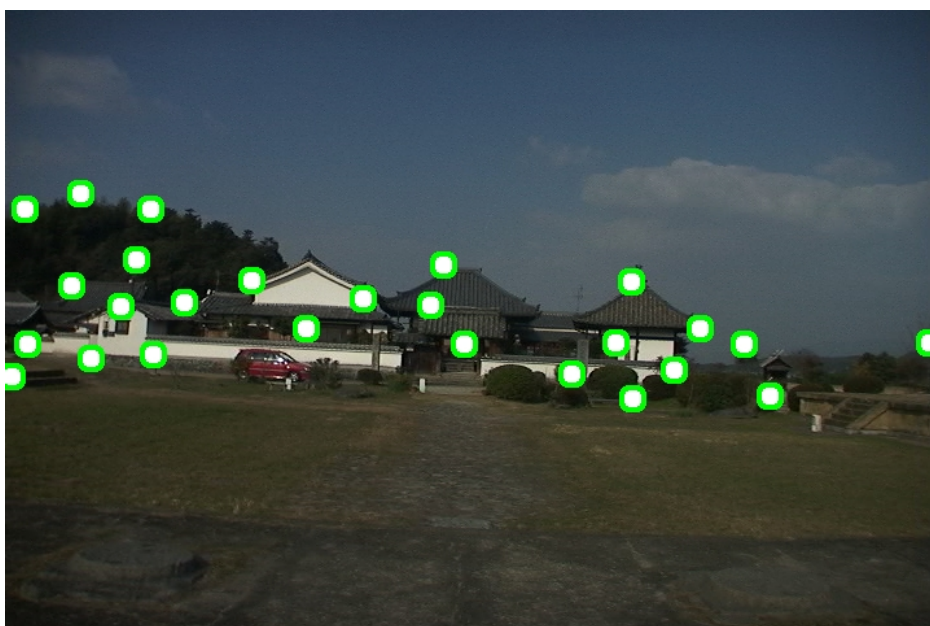
表 6 推定精度の比較

	局所形状の考慮あり	局所形状の考慮なし
平均位置誤差 (mm)	231	342
位置誤差の標準偏差 (mm)	107	164
光軸方向の平均角度誤差 (度)	1.11	1.41
光軸方向の角度誤差の 標準偏差 (度)	0.52	0.46

4.3.2 カメラ位置・姿勢推定精度の評価

次に、ランドマーク周辺の局所形状を考慮して構築したデータベースとランドマーク周辺の局所形状を考慮せずに構築したデータベースを用いて推定されるカメラ位置・姿勢推定精度を比較した。図 32, 図 33 に、カメラ位置・姿勢の推定に利用されたランドマークの画像上の位置を示す。ランドマーク周辺の局所形状を考慮しない場合では、地面に存在するランドマークが全く用いられていないが、ランドマーク周辺の局所形状を考慮する提案手法を用いた場合にはこれらのランドマークが推定に用いられている。これは、提案手法において、密な奥行き情報を用いて画像テンプレートを補正することで、ランドマークと自然特徴点が正しく対応づいた結果によると考えられる。

次に、推定されたカメラ位置・姿勢の定量的な評価を行った。カメラ位置・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、全周レンジファインダで計測された三次元位置を各フレームの画像上の特徴点と手動で対応付け、PnP 問題を解くことで算出した。ただし、真値作成の際に、三次元位置を与えた特徴点の画像上の再投影誤差が平均 1.5 画素以上となったフレームについては真値が信頼できないため評価対象にしなかった。推定されたカメラ位置・姿勢の精度を表 6 に示す。有意水準を 5% と設定した t 検定によって、提案手法と従来手法の平均位置誤差の間には有意な差が見られ、カメラ位置推定精度の向上が確認できる。一方で、光軸方向の平均角度誤差には有意な差は見られず、同等の精度



120 フレーム目

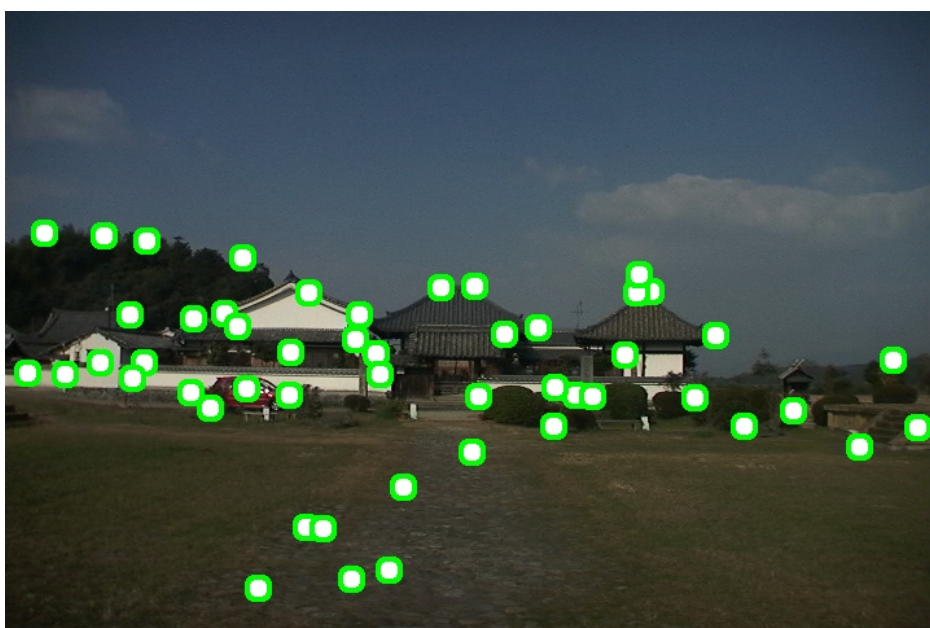


130 フレーム目

図 32 対応づいたランドマークの画像上の位置 (局所形状を考慮しない場合)



120 フレーム目



130 フレーム目

図 33 対応づいたランドマークの画像上の位置 (局所形状を考慮した場合)

で推定されていることを確認した。また、各フレームでの位置誤差、光軸方向の角度誤差の大きさを図 34、図 35 に示す。全てのフレームにおいて、提案手法では従来手法と比べカメラ位置の推定精度が同等以上となっていることが確認できる。これは、カメラの位置を決定するために重要となるユーザの近くに存在するランドマークが正しく対応付いたためであると考えられる。ここで、80 フレーム目から 90 フレーム目の間で、提案手法、従来手法ともに位置推定精度が低下している。これは、図 36 に示すように、この区間においてカメラが向いている方向には、利用できるランドマークがほとんど存在しないためであると考えられる。

4.4 結言

本章では、拡張現実感における幾何的位置合わせの精度を向上させることを目的として、全周レンジファインダによって取得される密な奥行き情報を用いてランドマークの画像テンプレートを補正する手法を提案した。ランドマーク周辺の局所形状を考慮して生成された画像テンプレートを用いることで、ユーザの近くに存在するランドマークを正しく対応付けることができ、これによって、より高精度なカメラ位置・姿勢推定を実現した。

実験では、ランドマーク周辺の局所形状を考慮した場合と考慮しない場合について画像テンプレートを作成し、入力画像上のランドマーク周辺の画像パターンと正規化相互相関法によって比較することで画像テンプレートの補正効果について検証した。また、カメラ位置・姿勢推定結果の定量評価によってユーザの近くに存在するランドマークを正しく対応付けることの重要性について明らかにした。

今後の課題として、広域な環境において効率的にデータベースを構築するために、structure-from-motion 法に基づくランドマークデータベース構築手法と全周レンジファインダを用いたランドマークデータベース構築手法を統合手法を開発することが挙げられる。

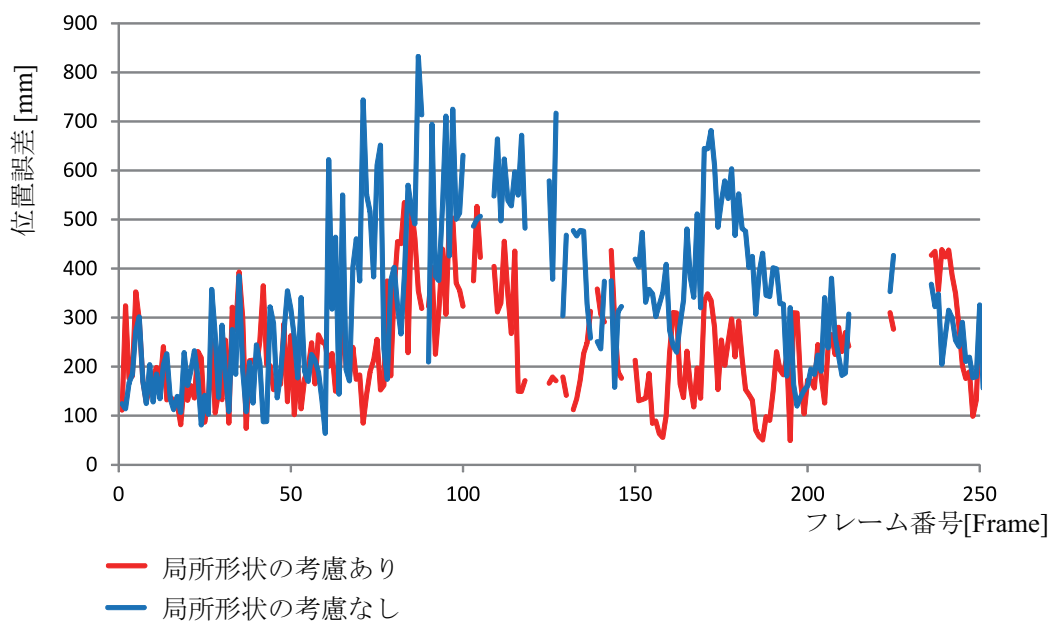


図 34 各フレーム毎の位置誤差

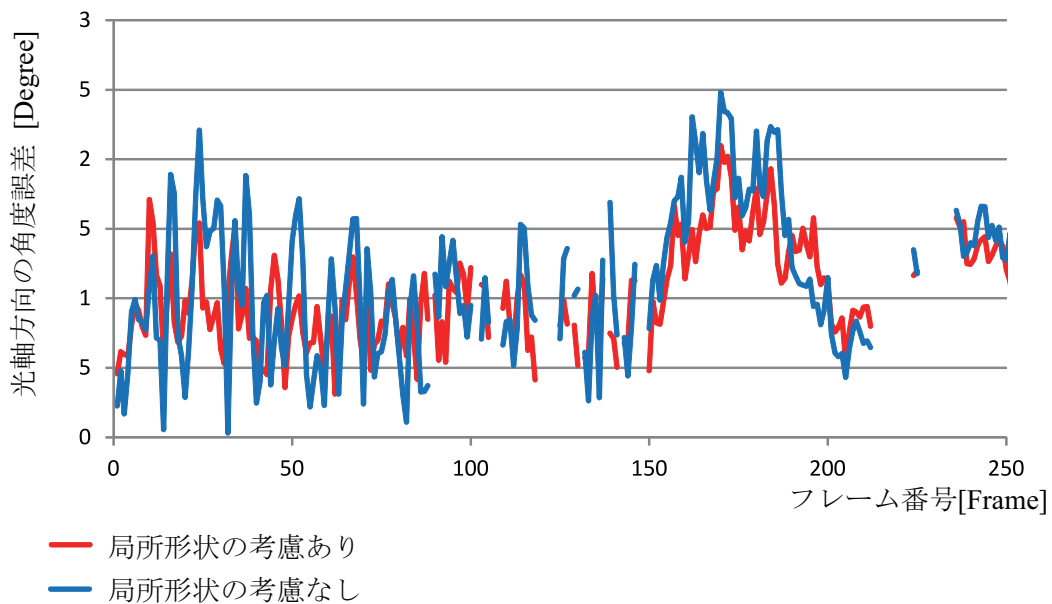
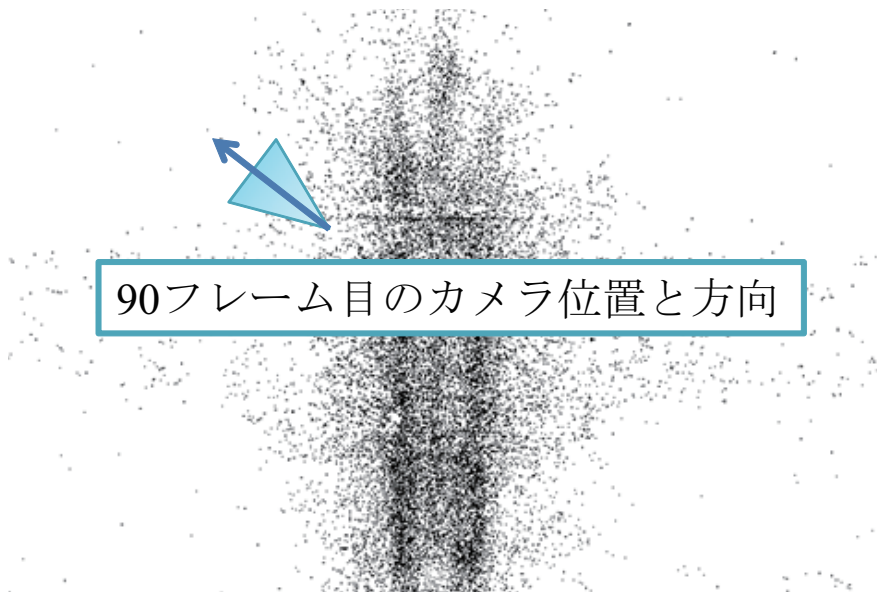


図 35 各フレーム毎の光軸方向の角度誤差



90フレーム目の画像



90フレーム目のカメラ位置と方向

ランドマークの位置

図 36 登録されているランドマークの位置と 90 フレーム目のカメラ位置の関係

5. 提案手法を用いた実アプリケーション

本章では、提案手法の有用性を示すために、以下に示すような位置に依存した情報の提示を必要とする拡張現実感のアプリケーションへの応用例を示す。

- ヒューマンナビゲーション
- カーナビゲーション
- MR-PreViz
- バーチャル歴史体験

ヒューマンナビゲーション、カーナビゲーションへの応用では、入力画像上へ案内情報を重畳表示することで従来の地図を用いたナビゲーションよりも直感的なナビゲーションを実現する。カーナビゲーションへの応用では、短時間での観測において車両の位置・姿勢は大きく変化しないことを利用し、カメラ位置・姿勢推定に車両の動き予測モデルを導入することによって、推定されるカメラ位置・姿勢を安定化する。また、MR-PreViz への応用では、事前に配置した CG キャラクタを重畳することによってカメラワークや俳優の演技のシミュレーションを実現する。バーチャル歴史体験への応用では、遺跡の CG モデルを入力画像中に重畳表示することによる遺跡の復元を実現する。

表 7 に各アプリケーションで用いた利用環境の三次元情報の取得方法を示す。ここで、バーチャル歴史体験への応用では、ユーザに近い位置に CG を配置する

表 7 各アプリケーションにおける三次元情報の取得方法

	三次元情報の取得方法
ヒューマンナビゲーション	structure-from-motion [佐藤 05]
カーナビゲーション	structure-from-motion [ISYY07]
MR-PreViz	structure-from-motion [佐藤 05]
バーチャル歴史体験	全周レンジファインダ

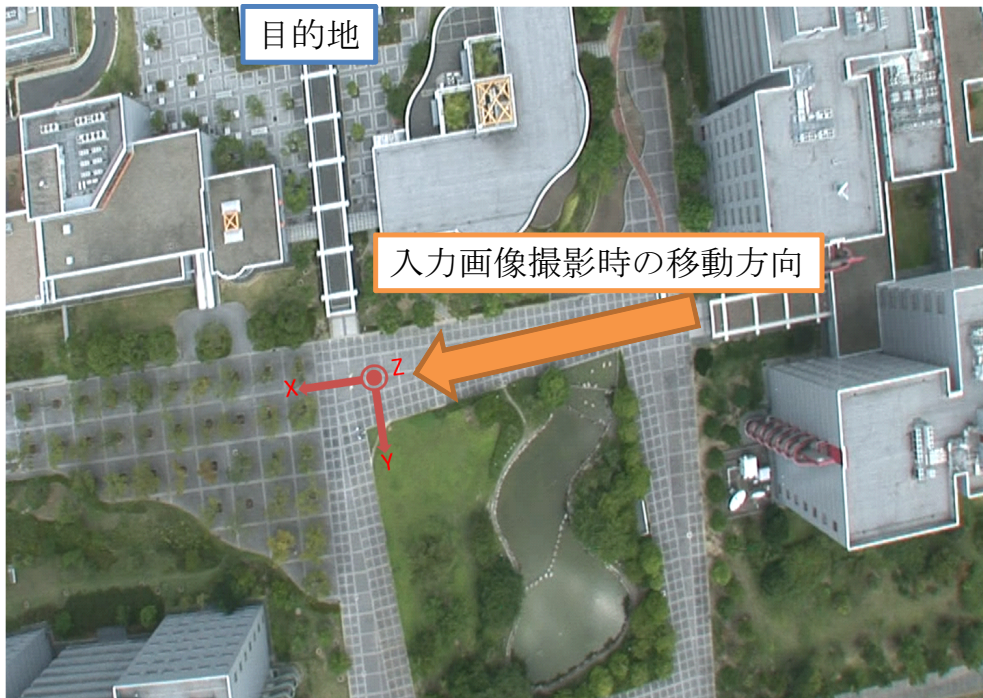


図 37 利用環境と撮影時のカメラパス

必要があるため，4章で述べた全周レンジファインダを用いたデータベースの構築を行った．以降では，それぞれのアプリケーションについて詳しく述べる．

5.1 ヒューマンナビゲーションへの応用

本節では，ランドマークデータベースを用いたカメラ位置・姿勢推定を用いたヒューマンナビゲーションへの応用について述べる．ここでは，3章の実験で構築したランドマークデータベースを用いて案内情報の提示を行った．また，実験では，3.5節と同一の機器構成を用いた．図37に利用環境の俯瞰画像と入力画像を撮影したカメラの移動方向を示す．重畳表示する案内情報は，事前に structure-from-motion で推定された利用環境の三次元情報を用いて手動で配置した．

図38に案内情報の重畳表示結果を示す．ここでは，図37中の目的地の方向を示す矢印と周辺の建物の名称を重畳表示することでナビゲーションを行った．拡

張現実感の技術を用いることにより，ユーザ視点の映像上に案内情報を提示することができるため，直感的なナビゲーションが可能となる．生成された拡張現実感画像から案内情報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．

ただし，本手法では，初期フレームのカメラ位置・姿勢推定が完了するまで30秒程度を必要とするため，初期位置・姿勢の推定中はカメラを固定しておく必要がある．そのため，より実用的なヒューマンナビゲーションを実現するためには，初期フレームのカメラ位置・姿勢推定処理高速化や初期位置・姿勢推定中のユーザの動きを推定する必要がある．

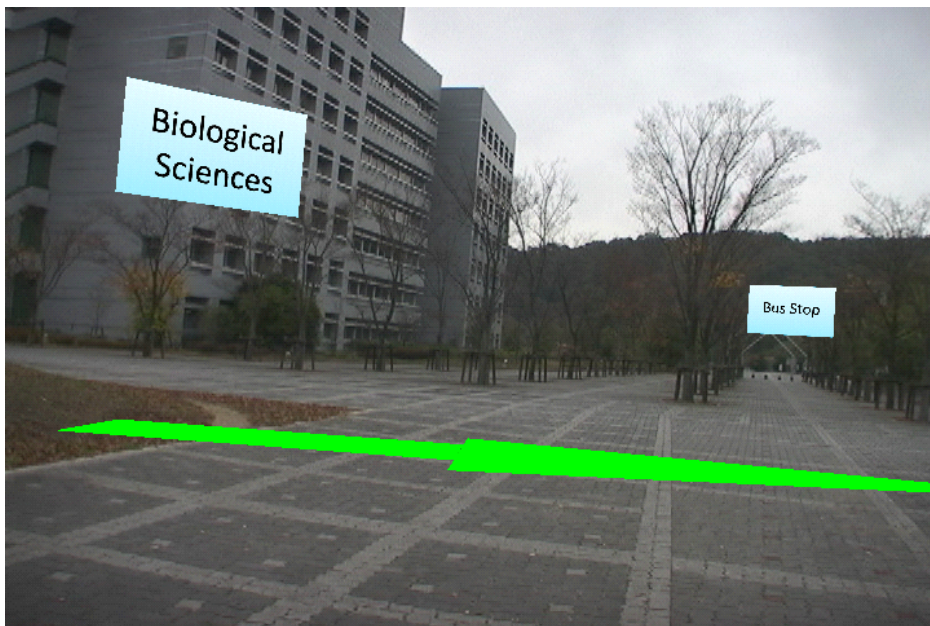
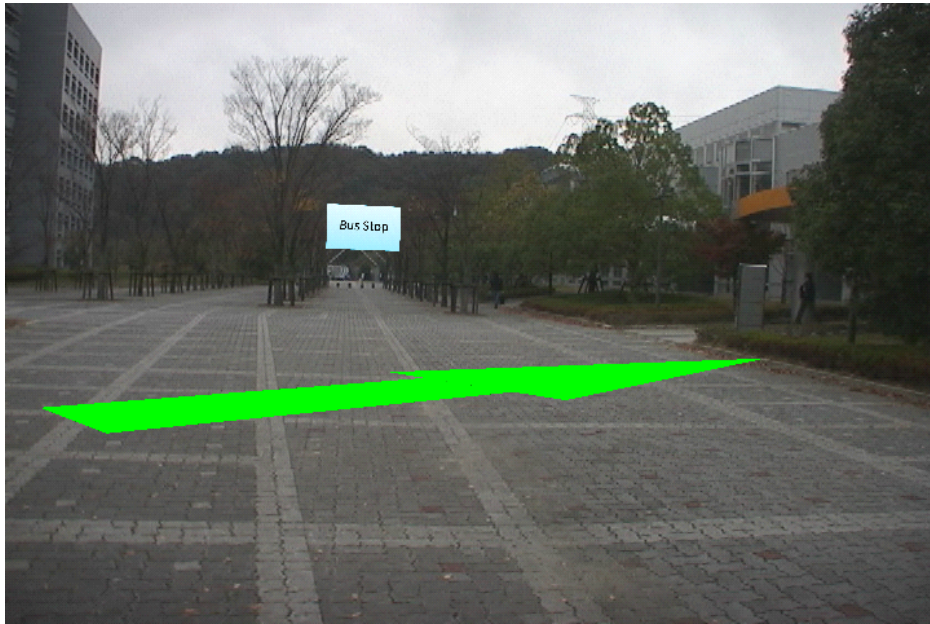


図 38 ヒューマンナビゲーションへの応用例

5.2 カーナビゲーションへの応用

現在、一般的に用いられているカーナビゲーションシステムでは、GPS および内界センサによって取得した位置と地図情報を利用することで自車位置を推定し、二次元の地図上でのナビゲーションを行っている。このような二次元の地図に基づくカーナビゲーションシステムでは、実環境と案内情報の対応関係を瞬時に把握することが難しく、ナビゲーション画面を注視することによる交通事故の発生が問題となっている。これに対して、実環境と案内情報の対応関係の把握を容易にするために、交差点などにおいて三次元の CG モデルを表示する機能を有する製品が販売されているが、CG モデルの作成コストが大きく都市部などの主要な交差点でのみ利用が可能な状態となっている。このような問題に対して、近年、車載カメラにより実環境を撮影した画像上に案内情報を合成提示する拡張現実感技術を用いたカーナビゲーションシステム（以下、AR カーナビ）に関する研究が行われている [赤穂 06, 澤野 06]。AR カーナビでは、実際の走行環境を撮影した画像上に案内情報の提示を行うため、実環境と案内情報の対応関係を直感的かつ瞬時に把握することができるという特長を持つ。本節では、ランドマークデータベースを用いたカメラ位置・姿勢推定によって走行車両の位置・姿勢を推定することで、AR カーナビを実現する。ここでは、走行時の車両の位置・姿勢は短時間では大きく変化しないことを利用し、カメラ位置・姿勢推定に車両の動き予測モデルを導入することによって、推定されるカメラ位置・姿勢を安定化する。

予測モデルの導入によるカメラ位置・姿勢推定の安定化

車両の位置・姿勢は連続フレーム間で急激に変化しないことから、図 39 に示すように車両の運動を局所的には等速運動で近似できるものとする。すなわち、現フレーム t のカメラ位置・姿勢 $M_{predict}$ を以下の式で予測する。

$$M_{predict} = M_{motion} M_{t-1} \quad (16)$$

ここで、 M_{t-1} は $t-1$ フレーム目で推定されたカメラ位置・姿勢を表し、 M_{motion} はフレーム間でのカメラの相対的な運動を表す。本研究ではフレーム間での細かな振動を抑制するため、 M_{motion} として、 $t-1$ フレームから $t-N$ フレームま

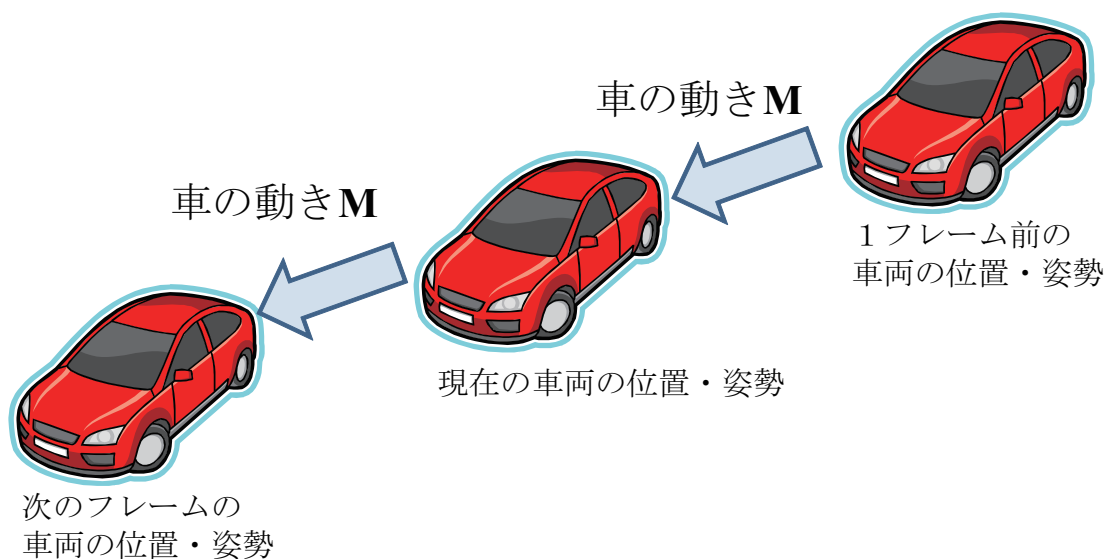


図 39 車両の動きのモデル化

での各フレームに対するカメラの相対運動を用いる。また、案内情報を入力画像上へ表示する際には、案内情報の画像上での提示位置を安定させるために、現フレーム t からフレーム $t - I$ までの案内情報の投影位置の重心に案内情報を提示する。

一般道における自車位置・姿勢推定実験

実際に、一般道においてランドマークデータベースを構築し、カメラ位置・姿勢の推定を行い走行車両の位置・姿勢推定の実験を行った。実験では、図 40 に示すように、自動車に全方位型マルチカメラシステム (Point Grey Research 社 Ladybug2) を固定し、屋外環境 (約 120m) を走行しながら撮影した全方位動画像 (500 フレーム) を用いて Ikeda らの手法 [ISYY07] によりランドマークデータベースの構築を行った。また、全方位動画像撮影の際に、助手席に単眼カメラ (SONY DSR-PD-150) を固定することで動画像 (解像度 720×480 画素, プログレッシブスキャン, 15fps, 400 フレーム) を撮影し、評価用として用いた。ここでは、カメラの動きの算出に用いるフレーム数を $N = 10$ とし、案内情報のレンダリングで



図 40 データ取得時のカメラ配置

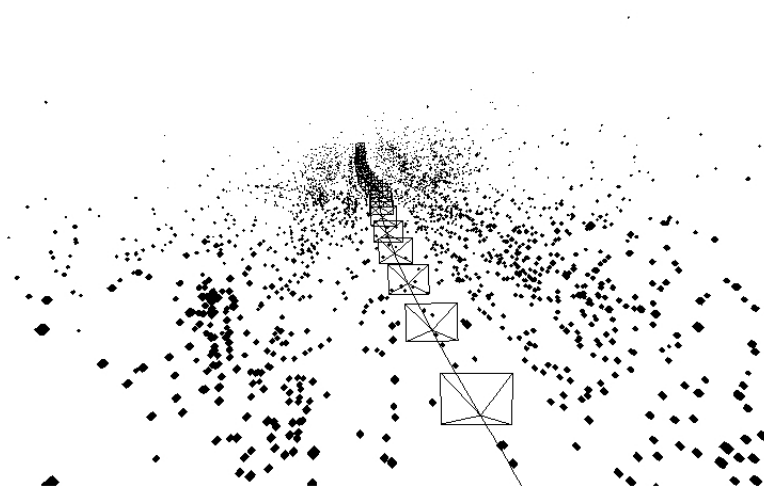
は $I = 5$ フレーム分の平均位置を案内情報の提示位置として用いた．図 41(a) に，データベースの構築に用いた全方位動画像の一部を示す．また，データベース構築時のカメラパスと復元されたランドマークの三次元位置を図 41(b) に示す．

予測モデルを導入しない場合に推定されたカメラパスを図 42 に，予測モデルを導入した場合に推定されたカメラパスを図 43 に示す．本実験では，データベースの構築用の全方位動画像と案内情報合成用の入力画像を同時に取得したため，本来は全方位カメラのカメラパスと単眼カメラのカメラパスがおおむね一致する．これに対して，予測モデルを用いずに推定されたカメラパスは全方位カメラのカメラパスから大きく外れており精度が低い．一方，予測モデルを用いて推定されたカメラパスは細かな振動はあるが，全方位カメラのカメラパスと同様の軌跡を描いていることが分かる．このことから，車両の動きに関する制約を入れることにより，安定したカメラ位置・姿勢推定処理を実現できていることが分かる．

図 44 に，予測モデルを用いて推定したカメラ位置・姿勢に基づいて案内情報を提示した結果を示す．生成された動画像から案内情報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．提案手法 (位置誤差：約 20cm，更新周期：1/15 秒) を用いることによって，車体の位置・姿勢を一般に用いられている GPS (位置誤差：約 10m，更新周期：1 秒) よりも高精度かつ高頻度に推定することができる．ただし，本手法は静的な環境を対象としているため，走行環境の



(a) 全方位画像



(b) 三次元復元の結果

図 41 走行環境のデータベースの構築

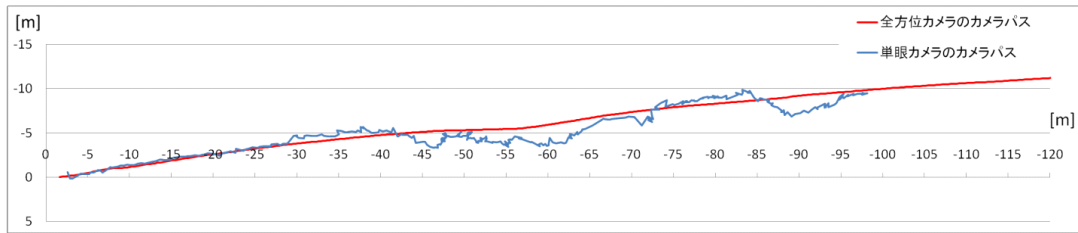


図 42 推定されたカメラパス (予測無しの場合)

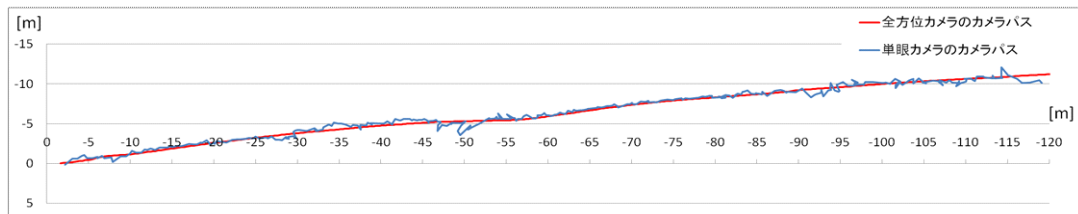


図 43 推定されたカメラパス (予測有りの場合)

ように動物体が多く存在するような環境ではロバストなカメラ位置・姿勢推定を実現することは難しい。そのため、より実用的な AR カーナビを実現するためには、車載センサ情報を統合することによって、動物体が多く存在する場合にもロバストにカメラ位置・姿勢を推定することが可能な手法を開発する必要がある。

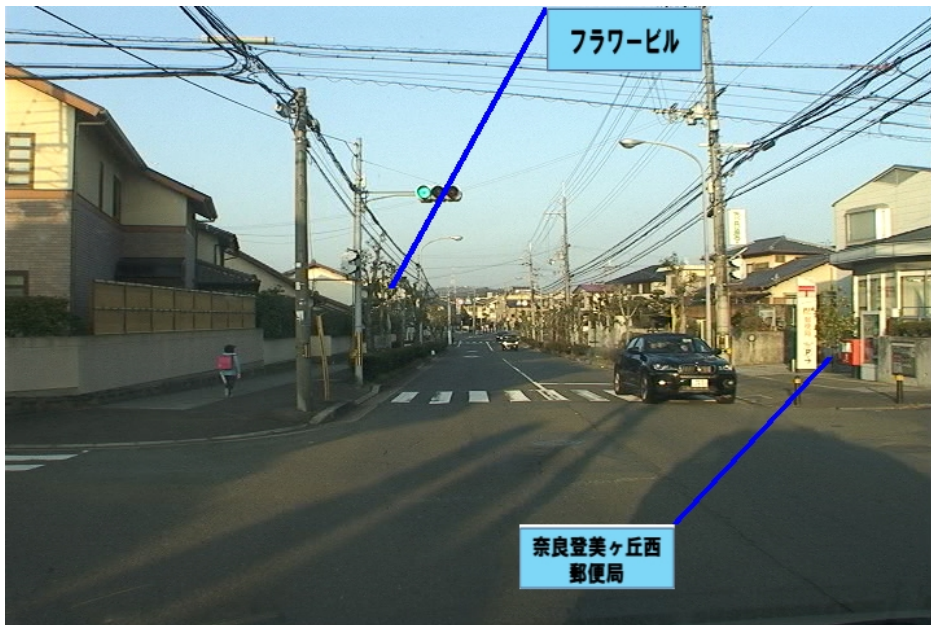
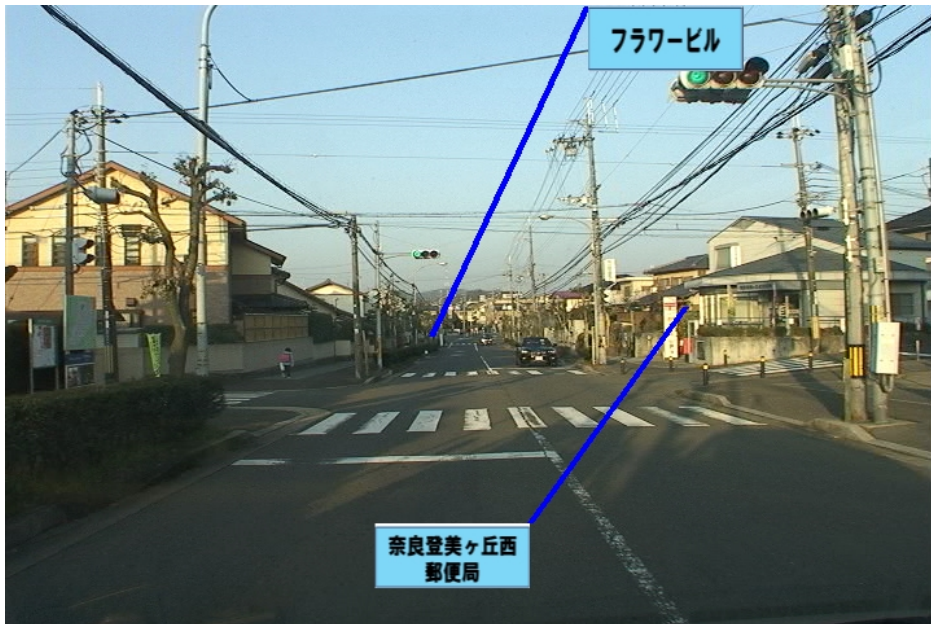


図 44 カーナビゲーション情報の提示結果

5.3 MR-PreViz への応用

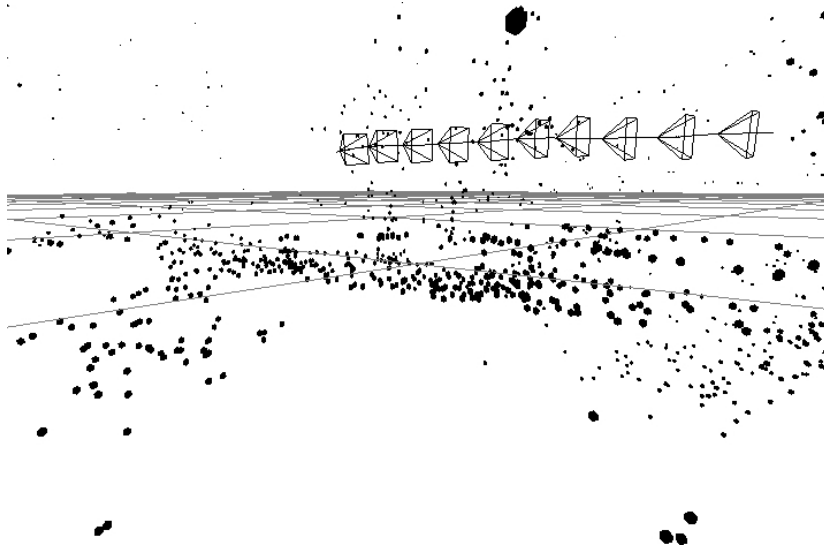
近年，映画制作の初期段階において，CG を用いてシーンの見え方をシミュレートするプレビジュアリゼーション (PreViz) の利用が進んでいる．通常 PreViz は CG のみによって実現されるが，実世界を撮影した画像中に CG などによって表現される仮想世界を位置合わせして描画することで情報を付加する複合現実感技術 (MR) を用いることで，実写の背景を用いた PreViz 映像を撮影現場で作成することが可能となる．複合現実型プレビジュアリゼーション (MR-PreViz) は通常の PreViz と異なり実際の背景を用いるため，監督や俳優は，CG のみで作られた PreViz 映像よりも直感的に本番のカメラワークや演技を把握することができる．しかし，従来提案されている MR-PreViz の手法 [SsKP05, Tho05] は，屋内での利用を想定しており，CG の位置合わせを実現するためにマーカを必要とするため，広域な屋外シーンを対象として，撮影現場における実時間での MR-PreViz 映像を作成することが困難であるという問題がある．これに対して，本節では，ランドマークデータベースに基づくカメラ位置・姿勢推定を実時間で行うことで，屋外環境において MR-PreViz 映像を実時間で生成する．ここでは，推定された現フレームのカメラ位置・姿勢を用いて，事前に三次元位置や動きが与えられた役者の CG モデルをカメラで撮影した画像上にレンダリングすることで MR-PreViz 画像を生成する．合成に用いる CG モデルの三次元位置・姿勢・スケール等は，ランドマークと同一座標系であらかじめ定義しておく．

MR-PreViz 映像を生成するために，全方位型マルチカメラシステム (Point Grey Research 社 Ladybug2) で屋外環境 (約 10m) を歩きながら撮影した全方位動画像 (200 フレーム) を用いて佐藤らの手法 [佐藤 05] によってランドマークデータベースの構築を行った．図 45(a) に，データベースの構築に用いた全方位動画像の一部を示す．また，データベース構築時のカメラパスと復元されたランドマークの三次元位置を図 45(b) に示す．

MR-PreViz 映像の生成実験では，図 46 に示す試作システムを用いて，CG キャラクタの合成実験を行った．同図に示すように，試作システムはカメラ位置・姿勢推定と CG モデルのレンダリングの負荷を分散させるために 2 台の PC から構成されている．試作システムでは，まず，カメラで撮影された画像をカメラ位置・姿



(a) 全方位画像



(b) 三次元復元の結果

図 45 MR-PreViz 環境のデータベースの構築

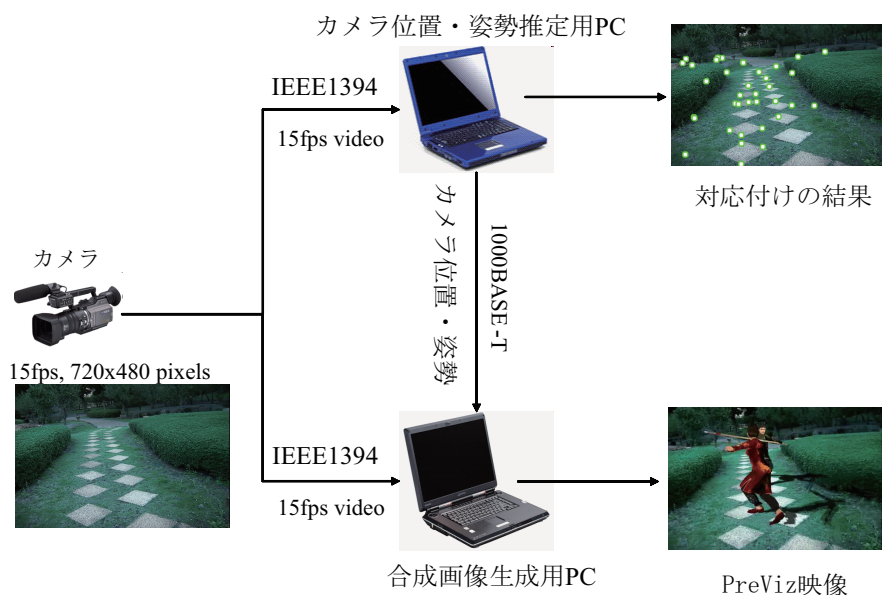


図 46 試作システムの機器構成

勢推定用 PC(Core 2 Extreme 2.93GHz, メモリ 2GB) と合成画像生成用 PC(Core 2 Duo 2GHz, メモリ 2GB) に分配する．次に，カメラ位置・姿勢推定用 PC でカメラ位置・姿勢の推定を行い，推定されたカメラパラメータを 1000BASE-T の有線 LAN 経由で合成画像生成用 PC に送信する．最後に，合成画像生成用 PC で，推定されたカメラパラメータを用いて CG キャラクタを実写画像に合成し提示することで実時間での PreViz を実現する．

試作システムを用いて，屋外環境で実写背景に CG キャラクタを合成した結果を図 47 に示す．図 48 に推定に用いられたランドマークの画像上での位置を示す．提案手法では，図 48 に示すような，自然物が多く存在する複雑な環境でも，多くのランドマークが自然特徴点と対応付けられ，1 秒間に 15 フレームのカメラの位置・姿勢を推定し，PreViz 映像を提示できることを確認した．なお，本手法では撮影から MR-PreViz 映像の提示までの間に，理論上，最低でも 1/15 秒以上の遅延が発生するが，主観的には MR-PreViz での利用において問題になるレベルの遅延は感じられなかった．

さらに、本手法は立命館大学 田村研究室へ技術移転を行い、MR-PreViz へ特化した手法 [TMT⁺09] へと変更を加えた。具体的には、MR-PreViz 映像を作成する際のカメラパスは事前にある程度決まっていることから、データベース構築時に MR-PreViz 映像作成で利用するカメラパスと同様のパスで撮影しておくことでオンラインでのカメラ位置・姿勢推定処理においてカメラの存在する領域を制限し、高速な初期カメラ位置・姿勢推定および復帰処理を実現した。また、データベースの構築とオンラインでのカメラ位置・姿勢推定に同一のカメラを用いることで、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高速な対応付け処理を実現した。

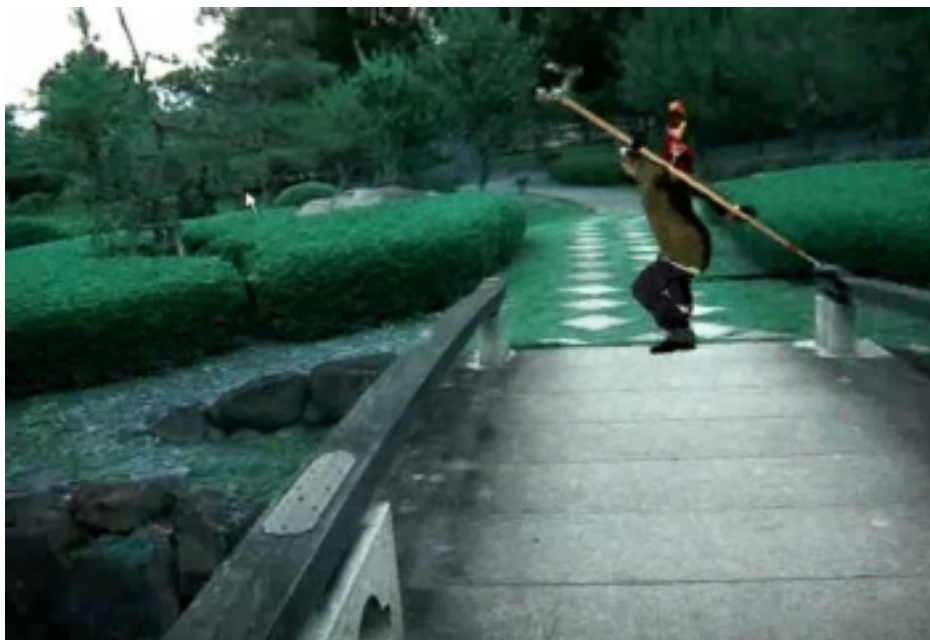


図 47 CG キャラクタの合成結果 (CG モデルは立命館大学 田村研究室より提供)

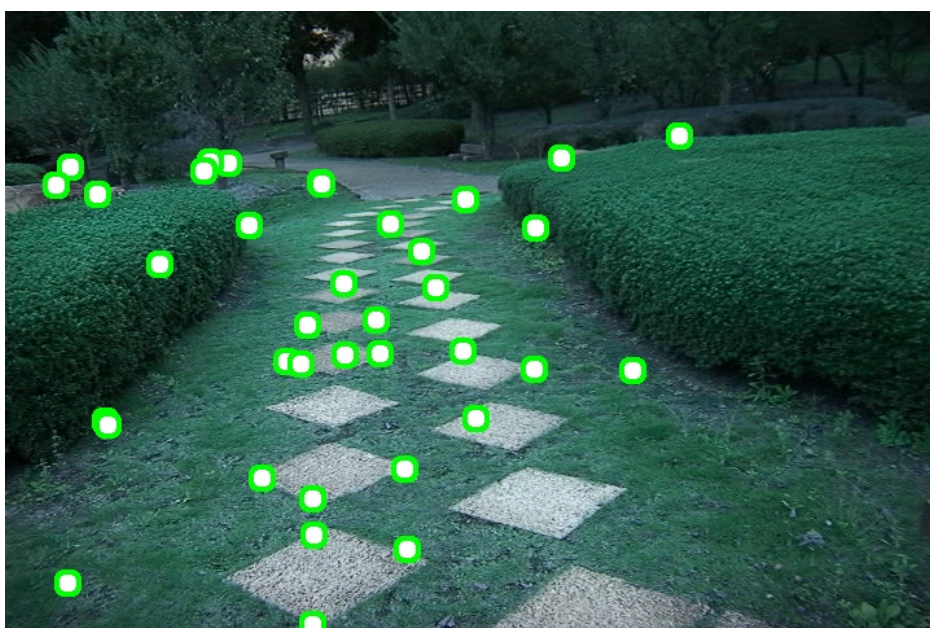
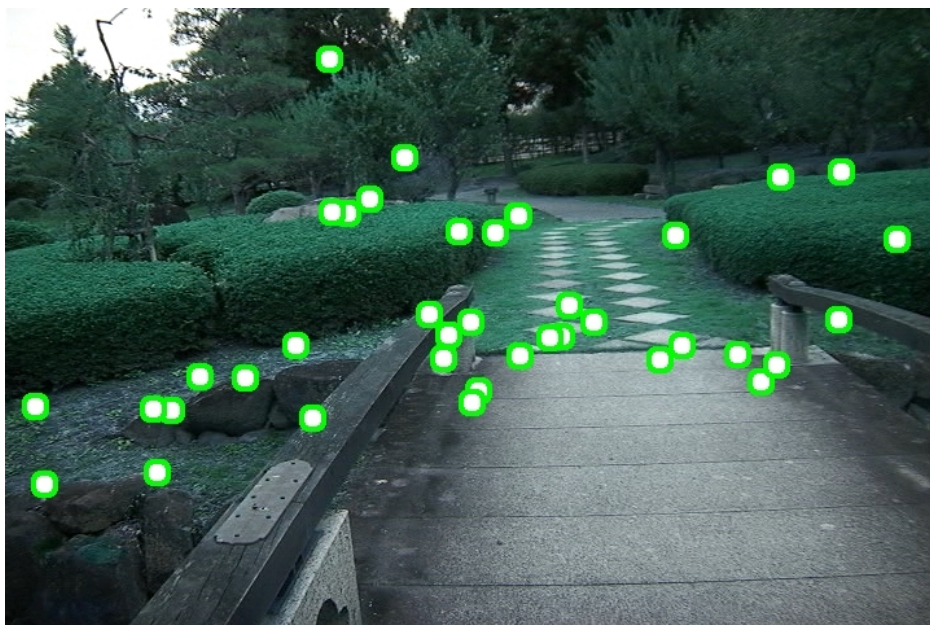


図 48 ランドマークと対応づいた自然特徴点の位置

5.4 バーチャル歴史体験への応用

本節では、拡張現実感技術を用いたバーチャル歴史体験への応用について述べる。日本各地には、多くの文化的な遺産が存在している。しかし、これらの文化遺産のうち、当時の形をそのまま残しているものは少ない。このような場所では、観光客などは当時の様子を簡単に想像することが難しいという問題がある。このようなことから、これまで、失われた文化遺産に対してレプリカやCGによる再現が一般的に行われている。遺構上に実寸大のレプリカを復元することによって観光客などは当時の雰囲気や空間を簡単に想像することが可能となるが、実寸大のレプリカを建造するためには多くの人的コストや金銭的成本が必要となる。一方で、CGによる再現では、レプリカを建造する場合に比べ低い人的コストや金銭的成本で文化遺産を再現することが可能であるが、図 49 に示すようなCGのみでは高い臨場感を与えることが難しく、観光客などは当時の様子を想像することが難しい。

このような問題に対して、近年、複合現実感技術を用いた文化遺産の復元が行われている [DK02, 角田 08]。複合現実感技術を用いて遺構上にかつて存在した建物のCGなどを重畳表示することによって、観光客などは当時の様子を簡単に想像することが可能となる。しかし、従来の複合現実感技術を用いた文化遺産の復元手法 [DK02, 角田 08] では、センサを用いた現実環境と仮想環境の位置合わせ手法を採用しているため、高精度な位置合わせが難しいという問題があった。そこで、本研究では、提案手法を用いることによって、画像に基づく高精度な位置合わせを実現する。

奈良県明日香村の川原寺跡地において複合現実感技術を用いた文化遺産の復元を行った。川原寺は、飛鳥時代に飛鳥寺、大官大寺、薬師寺とともに四大寺に数えられた寺院である。しかし、現在は焼失によって南大門、中門、廻廊などの位置のみがわかるように整備されているため、当時の様子を想像することが難しい。そこで、本研究では、図 50 に示す川原寺跡地において拡張現実感技術を用いた川原寺の復元を行う。合成に用いる川原寺のCGモデルの三次元位置・姿勢・スケール等は、データベース構築の際に計測した全周レンジデータを用いてあらかじめ手動でランドマークと同一の座標系で定義しておく。ここでは、4章の実験

で構築したランドマークデータベースを用いてカメラ位置・姿勢の推定を行い川原寺のCGを重畳表示した。図51に本実験で用いた機器の構成を示す。また、図52にユーザの外観を示す。川原寺周辺では、ランドマークとして利用できる特徴点が少ないため、磁気方式による位置・姿勢センサを併用した。具体的には、ランドマークを利用することが難しい場合には磁気センサによる位置・姿勢情報を用いてCGを描画し、ランドマークが利用できる場合にはランドマークを用いた位置・姿勢推定を行う。また、本節では5.3節と同様にカメラ位置・姿勢推定とCGモデルのレンダリングの負荷を分散させるために2台のPCを用いた。CGモデルの描画では、東京大学池内研究室の光源環境推定システムを用いることで光源環境の推定を行い、これを用いたレンダリングを行った[角田08]。図53、図54に入力画像と川原寺のCGモデルを合成した画像を示す。生成された合成画像からCGモデル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。4章で提案したランドマーク周辺の局所形状を考慮したランドマークデータベースを用いることで、図53のようにユーザの近くにCGを描画する場合にも正しい位置に描画されていることが確認できる。このような合成画像の提示により、CGのみを用いた場合に比べより臨場感の高いバーチャル歴史体験を実現することができる。

ただし、本実験では、磁気センサを補助的に利用しているため、合成画像の提示は限定された場所でのみ可能となっている。そのため、より実用的な観光アプリケーションを実現するためには、初期化処理の高速化や復帰処理を組み込むことにより広域な環境での利用を可能にする必要がある。

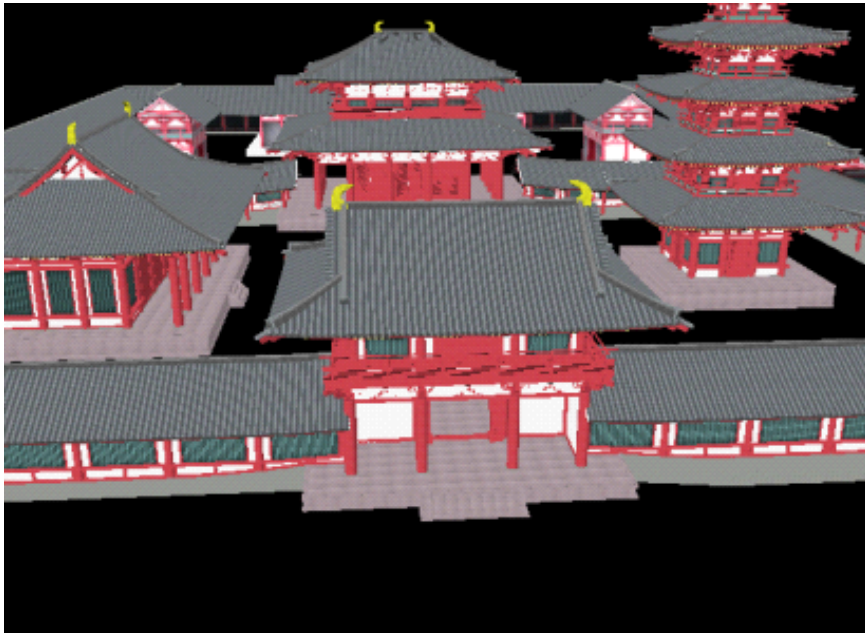


図 49 復元された川原寺のCGモデル（東京大学 池内研究室より提供）



図 50 川原寺跡地

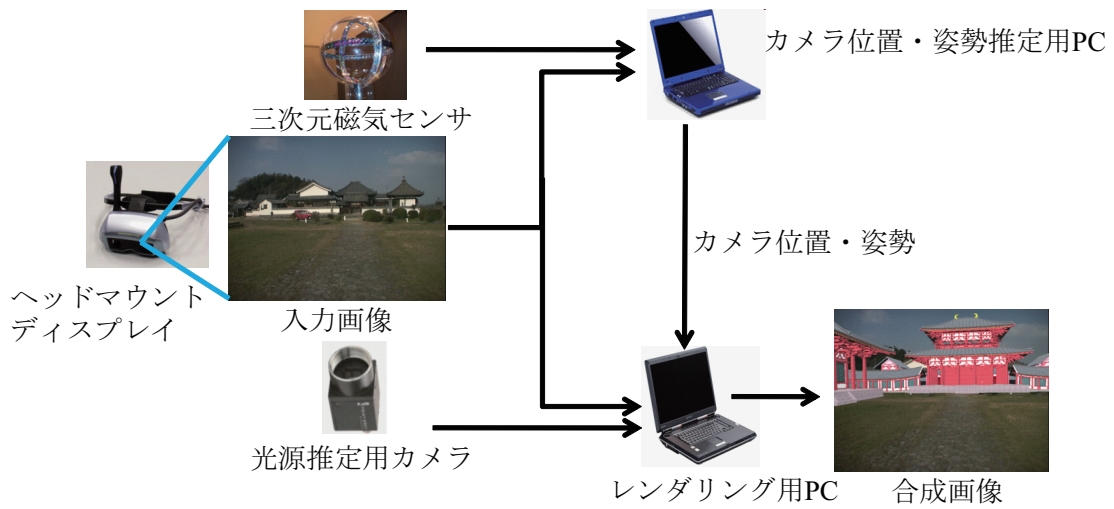


図 51 バーチャル歴史体験システムの機器構成



図 52 ユーザの外観



图 53 入力画像 (川原寺)



図 54 川原寺の CG モデルの合成結果

6. 結論

6.1 まとめと考察

本論文では、拡張現実感技術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、従来提案されてい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 [大江 05] を基礎とし、従来手法で問題となっていた計算コストやデータベース構築時のカメラパスから離れた地点でのランドマークと入力画像中の自然特徴点の対応付け処理を解決することで高速・高精度な位置合わせを実現した。

提案手法では、まず、計算コストの問題を解決するために、対応付けの候補としてデータベースから選択されるランドマーク数と入力画像中から検出される自然特徴点数を削減した。具体的には、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、対応付け処理にかかる計算コストを従来手法の約 $1/21$ に削減することができ、1秒間にビデオレートでの処理を行うことが可能となった。ただし、提案手法では、暫定的なカメラ位置・姿勢の推定に失敗した場合に、有効な対応点候補を正しく選択することができなくなる。このような問題に対しては、ランドマーク追跡を評価し、追跡処理の失敗を検出することによって、失敗する直前のカメラ位置・姿勢からの相対運動を推定することによって追跡処理に復帰することが考えられる。

次に、従来手法でデータベース構築時のカメラパスから離れた地点でのランドマークと入力画像中の自然特徴点の対応付けが難しく、高精度なカメラ位置・姿勢の推定が難しいという問題に対して、全周レンジファインダによって取得される密な奥行き情報を用いてランドマークの画像テンプレートを補正する手法を提案した。提案手法では、高精度な位置合わせが要求される場所において、スポット的な位置合わせ精度の向上を目的とし、取得された密な奥行き情報を用いてラ

ランドマーク周辺の局所形状を考慮したランドマークの画像テンプレートを生成することでユーザに近い位置に存在するランドマークを効率的に対応付け、従来手法より高精度なカメラ位置・姿勢推定を実現した。また、実験により、ユーザの近くに存在するランドマークを正しく対応付けることの重要性について明らかにした。

最後に、提案手法を拡張現実感技術を用いたヒューマンナビゲーション、カーナビゲーション、MR-PreViz、バーチャル歴史体験への応用を示し、提案手法がこれら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おいて有用であることを示した。

6.2 今後の展望

本論文では、ランドマークデータベースを用いた高速・高精度なカメラ位置・姿勢推定手法を提案したが、本手法をより実用的なアプリケーションに適用するためには、以下の検討が必要である。

初期カメラ位置・姿勢の推定および復帰処理

本研究では、初期フレームのカメラ位置・姿勢の推定にランドマークデータベースを用いた静止画像からのカメラ位置・姿勢推定手法 [薄 08] を適用している。しかし、現在、初期フレームのカメラ位置・姿勢推定処理に約 30 秒程度を要する。そのため、より実用的な拡張現実感のアプリケーションを実現するためには、初期フレームのカメラ位置・姿勢推定中のユーザの移動などを考慮する必要がある。推定中のユーザの移動を計測するためには、Visual-SLAM やジャイロや加速度センサなどのカメラ以外のセンサを用いることでカメラの相対運動を推定することが考えられる。また、カメラ以外のセンサ情報を用いることにより、よりロバストな逐次的なカメラ位置・姿勢の推定処理を実現できると考えられる。

自立計測手法との統合

ランドマークデータベースを用いた手法は比較的容易に広域で複雑な環境に適用することが可能であるが、より広域な環境で利用するためには自律計測による

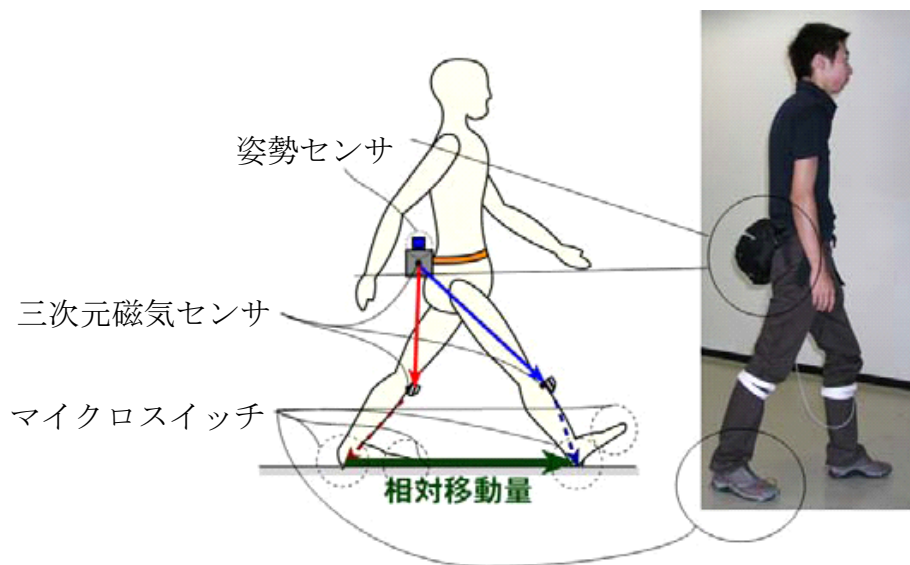


図 55 センサを用いた移動量の計測 [神原 08]

位置・姿勢推定手法などと統合する必要がある．ランドマークデータベースを用いたカメラ位置・姿勢の推定が出来ない区間でもカメラ位置・姿勢の推定を可能とするために，拡張現実感技術を用いたヒューマンナビゲーションでは，図 55 に示すようなセンサを用いることにより，ユーザの歩数を計測することにより移動量を算出する手法 [Fox05, 神原 08] などを応用することが考えられる．また，屋外環境では，GPS と動画像を併用したカメラ位置・姿勢推定手法 [KTSY10] を併用することが考えられる．これらの手法と提案手法を組み合わせることにより，より広域な環境において拡張現実感環境を効率的に構築することが可能となる．

データベースの更新処理

屋内外の広域環境での利用を想定した場合，図 56 に示すような撮影日時の違いによる環境の変化に対応しなければならない．しかし，ランドマークデータベースを用いたカメラ位置・姿勢推定において，ランドマークの優先度以外の情報は，データベース構築時以降更新されない．そのため，季節の変化や撮影時間の違いなどにより利用環境に変化が生じた場合には，ランドマークと自然特徴点の対応



図 56 撮影日時の違いによる環境(照明条件)の変化の例

付けが正しく行えない。現状の手法において、これを解決するためには再度、データベースを構築する必要がある。

あらかじめ変動する要素を決定しておき、それぞれの要素について学習しておくことによって、変動にロバストな認識手法 [VT05] を応用することや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より、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により生成することで環境の変化に対応することが考えられる。

謝辞

本研究を行う機会を与えて頂き，研究の全過程を温かく見守って下さり，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 教授に心より感謝の意を表します．特に，博士前期課程・後期課程を通して，国内外で数多くの発表の機会を与えて頂き，その度に論文の添削などきめ細やかな指導をして頂きました．

また，副指導教官として御助言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，並びに視覚情報メディア講座 山澤一誠 准教授に深く感謝いたします．

本研究の全過程において，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，本研究を行うに当たり，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 助教に心より感謝いたします．発表論文の添削，実験，発表練習に至るまで細やかな御指導を頂きました．

実験で用いたCGモデルを提供して頂いた立命館大学 リアリティメディア研究室（田村・木村研究室）および東京大学 生産技術研究所 池内研究室に心より御礼申し上げます．

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，中村美奈 女史に心より感謝申し上げます．研究活動において常に温かい御助言を頂いた視覚情報メディア講座の諸氏に感謝いたします．

最後に，長い間温かく見守って下さった両親，そして，日々の学生生活を支えてくれた妻 由希子，娘 麗舞に感謝いたします．

参考文献

- [AHNS99] R. Azuma, B. Hoff, H. Neely, and R. Sarfaty. A motion-stabilized outdoor augmented reality system. *Proc. IEEE Virtual Reality*, pp. 252–259, 1999.
- [ARFPB90] J. M. Airey, J. H. Rohlf, and Jr F. P. Brooks. Towards image realism with interactive update rates in complex virtual building environments. *Proc. Symp. on Interactive 3D graphics*, pp. 41–50, 1990.
- [AWK⁺09] C. Arth, D. Wagner, M. Klopschitz, A. Irschara, and D. Schmalstieg. Wide area localization on mobile phones. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 73–82, 2009.
- [Azu97] R. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, pp. 355–385, 1997.
- [Beh99] R. Behringer. Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors. *Proc. IEEE Virtual Reality*, pp. 244–251, 1999.
- [BETG08] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346–359, 2008.
- [BWS06] 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.
- [CGCMC07] D. Chekhlov, A. P. Gee, A. Calway, and W. Mayol-Cuevas. Ninja on a plane: Automatic discovery of physical planes for augmented reality using visual slam. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 153–156, 2007.

- [CMPC06] A. Comport, E. Marchand, M. Pressigout, and F. Chaumette. Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 12, No. 4, pp. 615–628, 2006.
- [CRT04] R. Cipolla, D. Robertson, and B. Tordoff. Image-based localization. *Proc. Int. Conf. Virtual Systems and Multimedia*, pp. 22–29, 2004.
- [DC02] T. Drummond and R. Cipolla. Real-time visual tracking of complex structure. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 7, pp. 932–946, 2002.
- [DK02] P. Dähne and J. Karigiannis. Archeoguide: System architecture of a mobile outdoor augmented reality system. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 263–264, 2002.
- [DMM03] A. Davison, W. Mayol, and D. Murray. Real-time localization and mapping with wearable active vision. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 18–27, 2003.
- [ED06] E. Eade and T. Drummond. Scalable monocular SLAM. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 469–476, 2006.
- [Fox05] E. Foxlin. Pedestrian tracking with dhoe-mounted inertial sensors. *IEEE Computer Graphics and Applications*, Vol. 25, No. 6, pp. 38–46, 2005.
- [GRS⁺02] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar, and N. Navab. Marker-less tracking for AR: a learning-based approach. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 295–304, 2002.
- [HFP99] T. Hóllerer, S. Feiner, and J. Pavlik. Situated documentaries: embedding multimedia presentations in the real world. *Proc. Int.*

Symp. on Wearable Computers, pp. 79–86, 1999.

- [HHS⁺02] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The anatomy of a context-aware application. *J. of Wireless Networks*, Vol. 8, pp. 187–197, 2002.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. *Proc. Alvey Vision Conf.*, pp. 147–151, 1988.
- [HSHS05] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith. Mixed reality in education, entertainment, and training. *IEEE Computer Graphics and Applications*, Vol. 25, No. 6, pp. 24–30, 2005.
- [ISYY07] 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.
- [IZFH09] A. Irschara, C. Zach, J. Frahm, and B. Horst. From structure-from-motion point clouds to fast location recognition. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2599–2606, 2009.
- [KB99] 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.
- [KBG03] H. Kato, K. Billinghurst, and M. Grafe. A registration method based on texture tracking using ARToolKit. *Proc. Augmented Reality Toolkit Workshop*, pp. 77–85, 2003.
- [KD07] H. Kaufmann and A. Dunser. Summary of usability evaluations of an educational augmented reality application. *Proc. Int. Conf. on Virtual Reality*, No. 10, pp. 660–669, 2007.

- [CLKK⁺02] 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.
- [KM07] 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.
- [KM08] G. Klein and D. Murray. Compositing for small cameras. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 57–60, 2008.
- [KM09] G. Klein and D. Murray. Parallel tracking and mapping on a camera phone. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 83–86, 2009.
- [KS04] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 257–263, 2004.
- [KSke98] R. Klette, K. Schluns, and A. Koschan Eds. *Computer Vision: Three-dimensional Data from Image*. Springer, 1998.
- [KTSY10] H. Kume, T. Taketomi, T. Sato, and N. Yokoya. Extrinsic camera parameter estimation using video images and GPS considering GPS positioning accuracy. *Proc. Int. Conf. on Pattern Recognition*, pp. 3923–3926, 2010.
- [LH09] T. Lee and T. Höllerer. Multithreaded hybrid feature tracking for markerless augmented reality. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 15, No. 3, pp. 355–368., 2009.
- [Lin98] T. Lindeberg. Feature detection with automatic scale selection. *Int. J. of Computer Vision*, Vol. 30, No. 2, pp. 79–116, 1998.

- [LMW⁺04] F. Liarokapis, N. Mourkoussis, M. White, J. Darcy, M. Sifniotis, P. Petridis, A. Basu, and P. F. Lister. Web3D and augmented reality to support engineering education. *World Trans. on Engineering and Technology Education*, Vol. 3, No. 1, pp. 11–14, 2004.
- [Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, Vol. 60, No. 2, pp. 91–100, 2004.
- [LVTF04] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua. Stable real-time 3d tracking using online and offline information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1391–1402, 2004.
- [MMT⁺08] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht. An augmented reality museum guide. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 103–106, 2008.
- [MS01] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. *Proc. Int. Conf. on Computer Vision*, pp. 525–531, 2001.
- [MS04] K. Mikolajczyk and C. Schmid. Scale & affine harris laplace invariant interest point detectors. *Int. J. of Computer Vision*, Vol. 60, No. 1, pp. 63–86, 2004.
- [NF02] 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.
- [NPD07] J. Neubert, J. Pretlove, and T. Drummond. Semi-autonomous generation of appearance-based edge models from image sequences. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 79–89, 2007.

- [NS06] D. Nistér and H. Stewenius. Scalable recognition with a vocabulary tree. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2161–2168, 2006.
- [OCLF10] M. Özuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, pp. 448–461, 2010.
- [OKY06] B. Okumura, M. Kanbara, and N. Yokoya. Augmented reality based on estimation of defocusing and motion blurring from captured images. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 219–225, 2006.
- [PCB00] 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.
- [PL07] J. Platonov and M. Langer. Automatic contour model creation out of polygonal cad models for markerless augmented reality. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 75–78, 2007.
- [RC02] D. P. Robertson and R. Cipolla. Building architectural models from many views using map constraints. *Proc. European Conf. on Computer Vision*, pp. 155–169, 2002.
- [RD05] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. *Proc. Int. Conf. on Computer Vision*, pp. 1508–1511, 2005.
- [RD06a] 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.

- [RD06b] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. *Proc. European Conf. on Computer Vision*, Vol. 1, pp. 430–443, 2006.
- [RDLW95] S. Ravela, B. Draper, J. Lim, and R. Weiss. Adaptive tracking and model registration across distinct aspects. *Proc. Int. Conf. on Intelligent Robots and Systems*, pp. 174–180, 1995.
- [RED07] G. Reitmayr, E. Eade, and T. Drummond. Semi-automatic annotations in unknown environment. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 67–70, 2007.
- [Rou84] P. J. Rousseeuw. Least median of squares regression. *J. of the American Statistical Association*, Vol. 79, pp. 871–880, 1984.
- [RPD10] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, No. 1, pp. 105–119, 2010.
- [SB02] G. Simon and M.-O. Berger. Pose estimation from planar structures. *IEEE Computer Graphics and Applications*, Vol. 22, No. 6, pp. 46–53, 2002.
- [SFPG06] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc. Gpu-based video feature tracking and matching. *Proc. Workshop on Edge Computing Using New Commodity Architectures*, 2006.
- [SFZ00] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 120–128, 2000.

- [SHTH07] 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.
- [SL04] 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.
- [SLF05] I. Skrypnyk, P. Lager, and P. Fua. Randomized trees for real-time keypoint recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 775–781, 2005.
- [SMB98] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. *Proc. IEEE Int. Conf. on Computer Vision*, pp. 230–235, 1998.
- [SSH06] P. Supan, I. Stuppacher, and M. Haller. Image based shadowing in real-time augmented reality. *Int. J. of Virtual Reality*, Vol. 5, No. 3, pp. 1–7, 2006.
- [SsKP05] M. Shin, B. s. Kim, and J. Park. AR storyboard: An augmented reality based Interactive storyboard authoring tool. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 198 – 199, 2005.
- [ST94] J. Shi and C. Tomasi. Good features to track. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.
- [TCD⁺00] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De 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.
- [Tho05] G. Thomas. Mixed reality techniques for TV and their application for on-set and pre-visualization in film production. *Proc. Int.*

Workshop on Mixed Reality Technology for Filmmaking, pp. 31–36, 2005.

- [TKY04] 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.
- [TMT⁺09] Wataru Toishita, Yutaka Momoda, Ryuhei Tenmoku Fumihisa Shibata Hideyuki Tamura, Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya. A novel approach to on-site camera calibration and tracking for mr pre-visualization procedure. *Proc. Human-Computer Interaction International*, pp. 492–502, 2009.
- [TNA⁺04] R. Tenmoku, Y. Nakazato, A. Anabuki, M. Kanbara, and N. Yokoya. Nara palace site navigator: Deviceindependent human navigation using a networked shared database. *Int. Conf. on Virtual Systems and Multimedia*, pp. 1234–1242, 2004.
- [Tsa86] 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.
- [TYK01] H. Tamura, H. Yamamoto, and A. Katayama. Mixed reality: Future dreams seen at the border between real and virtual worlds. *IEEE Computer Graphics and Applications*, Vol. 21, pp. 64–70, 2001.
- [UK91] M. Uenohara and T. Kanade. Vision-based object registration for real-time image overlay. *J. of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.
- [USSM09] H. Uchiyama, H. Saito, M. Servieres, and G. Moreau. Augmented

- text document by on-line learning of local arrangement of keypoints. *Proc. Int. Symp. Mixed and Augmented Reality*, pp. 95–98, 2009.
- [VLF04] 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.
- [VT05] M. A. O. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 547–553, 2005.
- [WCL⁺08] C. Wu, B. Clipp, X. Li, J. Frahm, and M. Pollefeys. 3D model matching with viewpoint-invariant patches (VIP). *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [WKR07] B. Williams, G. Klein, and I. Reid. Real-time SLAM relocalisation. *Proc. Int. Conf. on Computer Vision*, 2007.
- [WRM⁺08] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 125–134, 2008.
- [WS03] D. Wagner and D. Schmalstieg. First steps towards handheld augmented reality. *Proc. Int. Symp. on Wearable Computers*, pp. 21–23, 2003.
- [ZLGM10] Z. Zhang, X. Liang, A. Ganesh, and Y. Ma. Tilt: Transform invariant low-rank textures. *Proc. Asian Conf. on Computer Vision*, pp. 1617–1630, 2010.
- [赤穂 06] 赤穂賢吾, 加藤博一, 西田正吾. 実写ベースカーナビにおける交差点案内の検討と評価. 電子情報通信学会 技術研究報告 MVE, pp. 1–6, 2006.

- [岩佐 01] 岩佐英彦, 栗飯原述宏, 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.
- [大江 05] 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.
- [角田 08] 角田哲也, 大石岳史, 池内克史. 高速陰影表現手法を用いた飛鳥京MR コンテンツの開発とその評価. 映像情報メディア学会誌, Vol. 62, No. 9, pp. 1466–1473, 2008.
- [神原 04] 神原誠之, 横矢直和. 現実環境の照明条件と奥行きの実時間推定による仮想物体の陰影表現が可能な拡張現実感. 画像の認識・理解シンポジウム講演論文集, Vol. 2, pp. 247–252, 2004.
- [神原 05] 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・理解シンポジウム講演論文集, pp. 933–938, 2005.
- [神原 08] 神原誠之, 濱口明宏, 山中一樹, 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位置の自律計測. 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 4, pp. 439–449, 2008.
- [興梠 01] 興梠正克, 蔵田武志, 坂上勝彦, 村岡洋一. パノラマ画像群を位置合わせに用いたライブ映像上への注釈提示とその実時間システム. 電子情報通信学会論文誌 (D-II), Vol. J84-D-II, No. 10, pp. 2293–2301, 2001.
- [佐藤 03] 佐藤清秀, 内山晋二, 田村秀行. 複合現実感における位置合わせ手法. 日本バーチャルリアリティ学会論文誌, Vol. 8, No. 2, pp. 171–180, 2003.

- [佐藤 05] 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.
- [澤野 06] 澤野弘明, 岡田稔. 車載カメラによる実時間画像処理とその AR 技術に基づく表示方式によるカーナビへの応用. 芸術科学会論文誌, Vol. 5, No. 2, pp. 57–68, 2006.
- [薄 08] 薄充孝, 中川知香, 佐藤智和, 横矢直和. ランドマークデータベースに基づく静止画像からのカメラ位置・姿勢推定. 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 161–170, 2008.
- [天目 09] 天目隆平, 西上彰人, 柴田史久, 木村朝子, 田村秀行. ポスタを利用した複合現実感用幾何位置合わせ. 日本バーチャルリアリティ学会論文誌, Vol. 14, No. 3, pp. 351–360, 2009.
- [樋下 09] 樋下航, 一刈良介, 天目隆平, 柴田史久, 田村秀行. ランドマークデータベースに基づくカメラトラッキング法の高速化と安定化. 電子情報通信学会 技術研究報告 PRMU2009-192, 2009.
- [中里 05] 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.
- [羽原 02] 羽原寿和, 町田貴史, 小川剛史, 竹村治雄. 画像マーカを用いた屋内位置検出機構とその評価. 電子情報通信学会 技術研究報告 MVE, pp. 65–70, 2002.
- [安室 03] 安室喜弘, 石川悠, 井村誠孝, 南広一, 眞鍋佳嗣, 千原國宏. 立体マーカを用いた実空間における仮想物体の調和的表現 ~インタラクティブ MR インテリアデザイン~. 映像情報メディア学会誌, Vol. 57, No. 10, pp. 1307–1313, 2003.

研究業績

学術論文誌

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, Vol. J92-D, No. 8, pp. 1440-1451, 2009. (第3章に関連)

査読付き国際会議

1. S. Ikeda, T. Taketomi, B. Okumura, T. Sato, M. Kanbara, N. Yokoya, and K. Chihara, “Real-time outdoor pre-visualization method for videographers -real-time geometric registration using point-based model-”, *Proc. IEEE Int. Conf. on Multimedia and Expo*, pp. 949-952, 2008.
2. T. Taketomi, T. Sato, and N. Yokoya, “Real-time camera position and posture estimation using a feature landmark database with priorities”, *Proc. Int. Conf. on Pattern Recognition*, 2008. (第3,5章に関連)
3. T. Taketomi, T. Sato, and N. Yokoya, “Real-time geometric registration using feature landmark database for augmented reality applications”, *Proc. SPIE Electronic Imaging*, Vol. 7238, 2009. (第3,5章に関連)
4. H. Kume, T. Taketomi, T. Sato and N. Yokoya, “Extrinsic camera parameter estimation using video images and GPS considering GPS positioning accuracy”, *Proc. Int. Conf. on Pattern Recognition*, pp. 3923-3926, 2010.
5. W. Toishita, Y. Momoda, R. Tenmoku, F. Shibata, H. Tamura, T. Taketomi, T. Sato, and N. Yokoya, “A novel approach to on-site camera calibration and tracking for MR pre-visualization procedure”, *Proc. Human-Computer Interaction International*, pp. 492-502, 2009.

6. T. Taketomi, K. Makita, T. Sato, and N. Yokoya, “Cultural heritage tourism with augmented reality”, *Proc. 3rd Korea-Japan Workshop on Mixed Reality*, pp. 13-23, 2010. (第 4,5 章に関連)
7. T. Taketomi, T. Sato, and N. Yokoya, “AR cultural heritage reconstruction based on feature landmark database constructed by using omnidirectional range sensor”, *Proc. ACCV 2010 Workshop on e-Heritage (Electronic Cultural Heritage)*, 2010. (第 4,5 章に関連)

査読無し国際会議

1. T. Taketomi, T. Sato, S. Ikeda, and N. Yokoya, “Real-time geometric registration of real and virtual worlds using a feature landmark database with priorities”, *Int. Workshop on Ubiquitous Virtual Reality*, 2008. (第 3,5 章に関連)

研究会・シンポジウム

1. 武富貴史, 佐藤智和, 横矢直和, “優先度情報を付加したランドマークデータベースによる動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”, 電子情報通信学会 技術研究報告 PRMU2007-202, 2008. (第 3 章に関連)
2. 武富貴史, 佐藤智和, 横矢直和, “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に基づく実時間でのカメラ位置・姿勢推定”, 画像の認識・理解シンポジウム講演論文集, pp. 860-865, 2008. (第 3,5 章に関連)
3. 糸秀行, 武富貴史, 佐藤智和, 横矢直和, “動画像と GPS の併用によるカメラ位置・姿勢推定における GPS 測位精度の考慮”, 第 14 回パターン計測シンポジウム講演論文集, pp. 33-38, 2009.

4. 武富貴史, 佐藤智和, 横矢直和, “AR 観光システムのための全周レンジファインダを用いたランドマークデータベース構築に基づく幾何的位置合わせ”, 画像の認識・理解シンポジウム講演論文集, pp. 1947-1953, 2010. (第 4,5 章に関連)
5. 糸秀行, 武富貴史, 佐藤智和, 横矢直和, “GPS 測位精度を考慮した動画像と GPS の併用によるカメラ位置・姿勢推定”, 画像の認識・理解シンポジウム講演論文集, pp. 1198-1205, 2010.
6. 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 技術研究報告 CVIM173-36, 2010.

全国大会

1. 武富貴史, 佐藤智和, 横矢直和, “位置に依存した固有空間上でのランドマークの照合によるカメラの位置・姿勢推定の高速化”, 情報科学技術フォーラム 一般講演論文集, Vol. 3, pp. 95-96, 2007.
2. 武富貴史, 佐藤智和, 池田聖, 横矢直和, “複合現実感による映画制作支援のための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”, 日本バーチャルリアリティ学会第 13 回大会論文集, pp. 367-370, 2008. (第 3,5 章に関連)
3. 武富貴史, 佐藤智和, 池田聖, 横矢直和, “ランドマークデータベースを用いたカメラ位置・姿勢推定による実時間での複合現実型プレビジュアルイゼーションの実現”, 情報処理学会関西支部 支部大会 講演論文集, pp. 207-210, 2008. (第 3,5 章に関連)
4. 樋下航, 天目隆平, 柴田史久, 田村秀行, 武富貴史, 佐藤智和, 横矢直和, “局所不変特徴量を用いた屋外 MR トラッキング法の高速化と安定化”, 情報処理学会第 71 回全国大会講演論文集, No. 4Y-9, 2009.

5. 武富貴史, 佐藤智和, 山口晃一郎, 横矢直和, “AR カーナビゲーションのためのランドマークデータベースを用いた高精度な自転車位置・姿勢の推定”, 電気学会 電子・情報・システム部門大会 講演論文集, pp. 917-920, 2009. (第5章に関連)
6. 糸秀行, 武富貴史, 佐藤智和, 横矢直和, “GPS の併用による測位精度を考慮した動画像からのカメラ位置・姿勢推定”, 日本写真測量学会 平成 21 年度秋季学術講演会 発表論文集, pp. 127-128, 2009.

展示発表

1. T. Taketomi, B. Okumura, S. Ikeda, T. Sato, M. Kanbara, and N. Yokoya, “Geometric registration using feature landmark database”, *Int. Symp. on Mixed Augmented Reality, Demonstration*, 2007. (第3,5章に関連)

付録

A. 再投影誤差最小化によるカメラ位置・姿勢推定

特徴点の世界座標系における三次元位置 $F_i = (X_i, Y_i, Z_i)$ と画像上の検出座標 $\mathbf{x}_i = (x, y)$ の n 組の対応から，以下の式で定義される再投影誤差 E を最小化することでカメラ位置・姿勢 M を推定する．

$$E = \sum_{i=1}^n |\hat{\mathbf{x}}_i - \mathbf{x}_i|^2 \quad (\text{A1})$$

ここで， $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i)$ は特徴点 F_i をカメラパラメータ M を用いて画像上へ再投影した座標を表す．また，カメラ位置・姿勢 M は，カメラ座標系の 3 軸に対する回転角 (r_1, r_2, r_3) とカメラ座標系における平行移動成分 $t = (t_1, t_2, t_3)^t$ を用いて以下のように表すことができる．

$$M = \begin{pmatrix} c_2 c_3 + s_1 s_2 s_3 & c_1 s_3 & -s_2 c_3 + s_1 c_2 s_3 & t_1 \\ -c_2 s_3 + s_1 s_2 c_3 & c_1 c_3 & s_2 s_3 + s_1 c_2 c_3 & t_2 \\ c_1 s_2 & -s_1 & c_1 c_2 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A2})$$

ただし，

$$\begin{aligned} s_1 &= \sin r_1, s_2 = \sin r_2, s_3 = \sin r_3 \\ c_1 &= \cos r_1, c_2 = \cos r_2, c_3 = \cos r_3 \end{aligned} \quad (\text{A3})$$

である．また，カメラ座標系における特徴点 F_i の三次元位置 (X'_i, Y'_i, Z'_i) は

$$\begin{pmatrix} X'_i \\ Y'_i \\ Z'_i \\ 1 \end{pmatrix} = M \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \quad (\text{A4})$$

となる．ここで，カメラの内部パラメータである焦点距離 f ，CCD のサイズ ccd_x, ccd_y ，画像サイズ scr_x, scr_y ，投影中心座標 c_x, c_y ，画像スケール係数 s_x ，およびレンズ歪

み率 κ_1, κ_2 とすると，レンズ歪みの無い画像上への投影座標 (u_i, v_i) は，

$$(u_i, v_i) = \left(-\frac{scr_x}{ccd_x} \frac{X'_i}{Z'_i} f s_x + c_x, -\frac{scr_y}{ccd_y} \frac{Y'_i}{Z'_i} f + c_y \right) \quad (A5)$$

となり，レンズ歪みを考慮した画像上への投影座標 \hat{x}_i は，

$$(\hat{x}_i, \hat{y}_i) = ((1 + \kappa_1 \rho^2 + \kappa_2 \rho^4)(u - c_x) + c_x, (1 + \kappa_1 \rho^2 + \kappa_2 \rho^4)(v - c_y) + c_y) \quad (A6)$$

$$\rho = \sqrt{\left(\frac{ccd_x}{scr_x s_x} (u - c_x) \right)^2 + \left(\frac{ccd_y}{scr_y} (v - c_y) \right)^2} \quad (A7)$$

となる．カメラ位置・姿勢の推定では，カメラ座標系の3軸に対する回転角 (r_1, r_2, r_3) とカメラ座標系における平行移動成分 t の6パラメータを変数とし，再投影誤差の総和である E (式(A1)) を最小化することでカメラの外部パラメータ M を推定する．

B. 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．

$$R_i = |x_i - \hat{x}_i|^2 \quad (\text{B1})$$

$$RM = \text{median}(R_1, R_2, \dots, R_n) \quad (\text{B2})$$

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し，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．