

修士論文

自然特徴点ランドマークデータベースと
姿勢センサを用いた
高速な回転にロバストなカメラ位置・姿勢推定

松田 幸大

2007年3月9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。

松田 幸大

審査委員：

横矢 直和 教授 (主指導教員)

千原 國宏 教授 (副指導教員)

自然特徴点ランドマークデータベースと 姿勢センサを用いた 高速な回転にロバストなカメラ位置・姿勢推定*

松田 幸大

内容梗概

カメラの位置及び姿勢を実時間で推定する技術は、拡張現実感 (Augmented Reality; AR) やロボットナビゲーションなど、様々な分野への応用が可能である。これらの分野では、マーカやセンサなど物理的なインフラの整備を必要とせず、誤差が蓄積しない絶対的な位置・姿勢の推定手法が望まれている。従来、事前に環境内の撮影及び計測を行うことで景観内の自然特徴点の三次元情報や画像テンプレートをランドマークとしてデータベース化しておき、入力画像上の特徴点とデータベース内のランドマークを複数組対応付けることでカメラの位置と姿勢を推定する手法が提案されている。しかし、従来手法では、カメラが高速に回転する場合、特徴点とランドマークを正しく対応付けられず、カメラの位置・姿勢の推定誤差が増大する問題がある。本論文では、自然特徴点ランドマークデータベースと姿勢センサにより得られる角速度情報を併用することで、カメラが高速に回転する場合においても位置と姿勢をロバストに推定する手法を提案する。提案手法では、まず角速度情報を用いることで、データベース中の多数のランドマークから推定に用いるランドマークを効果的に選択し、データベース内の画像テンプレートに対して入力画像と同様のモーションブラーを再現する。次に、テンプレートマッチングによって入力画像内の特徴点とランドマークを対応付け、

* 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-MT0551113, 2007年 3月 9日.

カメラの位置・姿勢を推定する．実験では，カメラに高速な回転を与えて取得した動画像からカメラの位置・姿勢を推定し，特徴点とランドマークの対応や推定されたカメラの位置・姿勢の推定精度を検証する．

キーワード

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

Robust Estimation of Position and Posture of Camera with High-speed Rotation Using Feature Landmark Database and Inertial Sensor*

Kodai Matsuda

Abstract

This paper describes a novel camera position and posture estimation method using an inertial sensor and feature landmark database constructed of three-dimensional positions of natural feature points (landmarks) and their image templates. This method is widely applicable to a number of fields such as augmented reality and robot navigation. In such fields, it is frequently required to estimate absolute position and posture of cameras without using any physical infrastructures such as artificial markers. A conventional method using feature landmark database has already been satisfying these requirements. In this method, camera position and posture are estimated by finding correspondences between feature points in the current input image and landmarks in the database by template matching. However, the estimation error increases if incorrect correspondences occur due to high-speed rotation of camera. To solve this problem, in the proposed method, the landmarks are first selected from the database by using angular velocity acquired with an inertial sensor. The motion blur is then reproduced on image templates of landmarks, and blurred templates are used for matching with

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0551113, March 9, 2007.

those in the input images. In experiments, camera position and posture have been successfully estimated in real environments.

Keywords:

camera position and posture estimation, natural features, landmark database, inertial sensor, augmented reality

目次

| | |
|--|----|
| 1. はじめに | 1 |
| 2. カメラ位置・姿勢推定に関する従来研究と本研究の位置付け | 3 |
| 2.1 センサを用いたカメラ位置・姿勢計測 | 3 |
| 2.2 動画像からのカメラ位置・姿勢推定 | 5 |
| 2.2.1 人工的なマーカを用いた手法 | 6 |
| 2.2.2 人工的な三次元モデルを用いた手法 | 6 |
| 2.2.3 自然特徴点ランドマークデータベースを用いた手法 | 8 |
| 2.3 センサ情報と動画像からのカメラ位置・姿勢推定 | 10 |
| 2.4 本研究の位置付けと方針 | 10 |
| 3. 自然特徴点ランドマークデータベースと姿勢センサを用いたカメラ位置・姿勢推定 | 14 |
| 3.1 提案手法の概要 | 14 |
| 3.2 自然特徴点ランドマークデータベースの構築 | 14 |
| 3.3 カメラ位置及び姿勢の推定手法 | 15 |
| 3.3.1 姿勢センサによるカメラ姿勢の算出・ステップ (A) | 17 |
| 3.3.2 カメラ位置・姿勢推定に用いるランドマークの選択・ステップ (B) | 19 |
| 3.3.3 画像テンプレートに対するモーションブラーの再現・ステップ (C) | 22 |
| 3.3.4 入力画像とランドマークの対応付け・ステップ (D) | 25 |
| 3.3.5 誤対応の排除とカメラ位置・姿勢推定・ステップ (E) | 26 |
| 4. カメラ位置・姿勢推定の精度評価実験 | 28 |
| 4.1 実験環境 | 28 |
| 4.2 定量的な評価実験 | 30 |
| 4.2.1 回転のみを伴うカメラで撮影された動画像 (入力 1) | 34 |
| 4.2.2 回転と並進を伴うカメラで撮影された動画像 (入力 2) | 34 |

| | |
|-------------------------------------|----|
| 5. まとめと今後の課題 | 42 |
| 謝辞 | 43 |
| 付録 | 44 |
| A. カメラと姿勢センサの同期とキャリブレーション | 44 |
| A.1 同期手法 | 44 |
| A.2 カメラと姿勢センサのキャリブレーション手法 | 45 |
| B. カメラの投影モデル | 48 |
| 参考文献 | 50 |

目 次

| | | |
|----|--|----|
| 1 | 無線 LAN の電波強度による位置測定 [8] | 4 |
| 2 | ARToolkit とモバイルコンピュータを用いた拡張現実感システム [12] | 7 |
| 3 | 物体の三次元モデルとモデルの周囲を中心とした環境マップ群 [26] | 8 |
| 4 | Gordon らの手法による物体の三次元復元 [28] | 9 |
| 5 | PhotoBuilder で構築した環境のポリゴンモデル [30] | 11 |
| 6 | 自然特徴点ランドマークデータベースの構成要素 | 15 |
| 7 | ランドマークとその多重スケールの画像テンプレート | 16 |
| 8 | カメラ位置及び姿勢推定の処理の流れ | 17 |
| 9 | カメラと姿勢センサの座標系の定義 | 19 |
| 10 | カメラと選択されるランドマークの位置関係 | 21 |
| 11 | モーションブラー再現による入力画像との正規化相互相関値の変化 | 23 |
| 12 | モーションブラーのベクトルの算出と画像テンプレートに対する モーションブラーの再現 | 24 |
| 13 | カメラと姿勢センサの位置関係 | 29 |
| 14 | 全方位型マルチカメラシステムと全方位画像 | 30 |
| 15 | 入力画像 (入力 1) | 31 |
| 16 | 入力画像 (入力 2) | 33 |
| 17 | 入力 1 におけるカメラ位置・姿勢推定結果を用いた match move | 35 |
| 18 | 入力 1 における match move の拡大図と被写体が動く方向 | 35 |
| 19 | 角速度の大きさと対応付けられたランドマーク数 (入力 1) | 36 |
| 20 | 角速度の大きさと推定位置誤差 (入力 1) | 37 |
| 21 | 入力 2 におけるカメラ位置・姿勢推定結果を用いた match move | 39 |
| 22 | 入力 2 における match move の拡大図と被写体が動く方向 | 39 |
| 23 | 角速度の大きさと対応付けられたランドマーク数 (入力 2) | 40 |
| 24 | 角速度情報 ω^g , ω^c のデータフィッティング | 46 |
| 25 | キャリブレーションデータの取得 | 47 |
| 26 | マーカの投影誤差のヒストグラム | 47 |
| 27 | 世界座標系とスクリーン座標系の関係 | 49 |

表 目 次

| | | |
|---|---|----|
| 1 | 従来手法の比較 | 12 |
| 2 | カメラ位置・姿勢推定に用いた機器 | 28 |
| 3 | 姿勢センサの仕様 | 29 |
| 4 | 推定に用いた入力画像と算出する推定誤差 | 32 |
| 5 | 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 推定の各処理における条件 | 32 |
| 6 | 平均推定誤差と標準偏差 (入力 1) | 38 |
| 7 | 平均推定誤差 (入力 2) | 38 |

1. はじめに

カメラの位置及び姿勢を実時間で推定する技術は、拡張現実感 [1,2] やロボットナビゲーション [3,4] など、様々な分野への応用が可能である。特に、屋外などの広範囲な環境を対象としたシステムでは、カメラの絶対的な位置及び姿勢の取得が必要である。従来から、赤外線マーカなどのセンサ [5-10]、画像マーカ [11-21] などの物理的なインフラを用いた手法、予め作成された環境のモデルを用いた動画像処理による手法 [22-29] など、多くの手法が提案されてきた。物理的なインフラを用いた手法 [5-21] では、センサからの信号やマーカのパターンから個々を識別することが容易であるため、時系列情報なしでカメラの位置及び姿勢の推定が可能であり、高速な動きに対してロバストな推定が実現できる。しかし、これらの手法では、環境中に物理的なインフラを整備する必要があり、自然環境などインフラの設置が困難な環境においては利用できない。一方、環境のモデルを用いた手法では、センサやマーカなどのインフラを用いずに、カメラの位置及び姿勢の推定が可能である。環境のモデルとして、ワイヤースケルトンモデル [22-25] や自然特徴点群 [28, 29]、テクスチャ付きのポリゴンモデルを用いる手法 [26, 27] が提案されているが、これらの手法では、入力画像中のエッジとモデルとを対応付けることによりカメラの絶対的な位置及び姿勢が推定される。一般的に、こういった環境モデルの生成には多大な労力を要するが、自然特徴点を用いた手法 [28, 29] では、動画像から structure-from-motion の技術を用いてモデルが自動生成されるため、広範囲な景観内の自然特徴点のモデルを容易に構築できるという特長を持つ。しかし、これらの環境のモデルを用いた手法 [24, 27-29] では、モデルと画像中の個々の特徴を一意に識別することが難しく、フレーム間におけるカメラの回転及び並進が大きい場合、モーションブラーや特徴の探索範囲の拡大によりモデルと入力画像において誤対応が生じ、カメラの位置及び姿勢の推定誤差が増大する問題がある。この問題を解決するために、環境のモデルと姿勢センサから得られる情報を併用してカメラの位置及び姿勢を推定するハイブリッドな手法 [30-32] が研究されている。しかし、従来のポリゴンモデル [31] やワイヤースケルトンモデル [30, 32] とセンサによる情報を併用した手法では、対象とする環境の規模が大きく複雑な場合、モデルの作成自体が難しいという問題がある。

本論文では，自然特徴点群の三次元位置と画像テンプレートをデータベース化した自然特徴点ランドマークデータベース [29] 及び姿勢センサにより得られる角速度情報を併用することで，カメラの高速な回転が生じた場合においても位置及び姿勢をロバストに推定する手法を提案する．提案手法では，まず，現フレームにおけるカメラの暫定的な位置及び姿勢を求め，フレーム毎にデータベース中の多数のランドマークから推定に用いるランドマークを適切に限定する．次に，限定されたランドマークの画像テンプレートに対して入力画像と同様のモーションブラーを再現し，モーションブラーによる入力画像とランドマーク間の類似度の低下を防ぐ．さらに，カメラの並進を考慮した大きさのウィンドウによって入力画像内の特徴点を検出し，検出された特徴点とランドマークをテンプレートマッチングによって対応付けることで，カメラの位置及び姿勢を推定する．実験では，並進移動するカメラに高速な回転を与えて取得した動画像からカメラの位置及び姿勢を推定し，特徴点とランドマークの対応や推定されたカメラの位置及び姿勢の推定精度を検証する．

以下，2章では，カメラの位置及び姿勢の推定に関する従来研究と本研究の位置付けと方針について述べる．3章では，自然特徴点ランドマークデータベースと姿勢センサにより得た角速度情報を併用した逐次的なカメラの位置及び姿勢の推定手法について述べる．4章では，実環境を撮影した画像を用いたカメラの位置及び姿勢の推定実験について報告する．5章では，まとめと今後の課題について述べる．

2. カメラ位置・姿勢推定に関する従来研究と本研究の位置付け

本章では、まず、絶対的なカメラの位置及び姿勢の推定に関する従来研究を、赤外線センサや超音波センサなどのセンサのみを用いて計測する手法、ビデオカメラからの入力画像のみを用いて推定する手法、姿勢センサなどのセンサから得られる情報と入力画像を併用することで推定する手法に大別し、それぞれの手法の特徴と問題点について述べる。次に、本研究の位置付けと方針について述べる。

2.1 センサを用いたカメラ位置・姿勢計測

センサによる情報のみからカメラの位置及び姿勢を計測する手法は、環境中の基準点に離散的に設置した赤外線マーカなどのインフラを用いた手法 [5-10]、インフラを用いずに絶対的な位置及び姿勢を取得可能なセンサと自律的に位置及び姿勢の変位を計測可能なセンサを併用した手法 [1, 2, 33] に分類できる。インフラを用いた手法では、カメラに取り付けられた超音波センサ [5-7] や無線 LAN の端末 [8]、赤外線カメラ [9, 10] などのセンサとインフラ間において、超音波や電波などによる信号を送受信することで基準点を認識する。認識された基準点からカメラの絶対的な位置を取得し、電子コンパスや姿勢センサによって姿勢を計測する手段が一般的である。これらの手法の中でも三辺測量による位置計測手法として、超音波センサを用いる ActiveBat [6] では、ユーザが装着する超音波ビーコンから発する超音波を環境中の 3 台以上の超音波センサで受信し、超音波の到達時間から三辺測量を行うことで対象物の 3 次元位置が計測される。また、無線 LAN を用いる RADAR [8] では、図 1 のように、無線 LAN 端末から発する信号の強度を複数の基地局で測定し、受信電波強度分布より算出した各基地局からユーザまでの距離を用い、三辺測量によってユーザの位置が計測される。三辺測量以外の計測手法としては、環境に設置した多数の赤外線 LED などの赤外線信号を発するインフラを用い、赤外線カメラによりインフラの検出を行うことでカメラの位置を計測する、赤外線マーカを用いた手法 [9] が挙げられる。

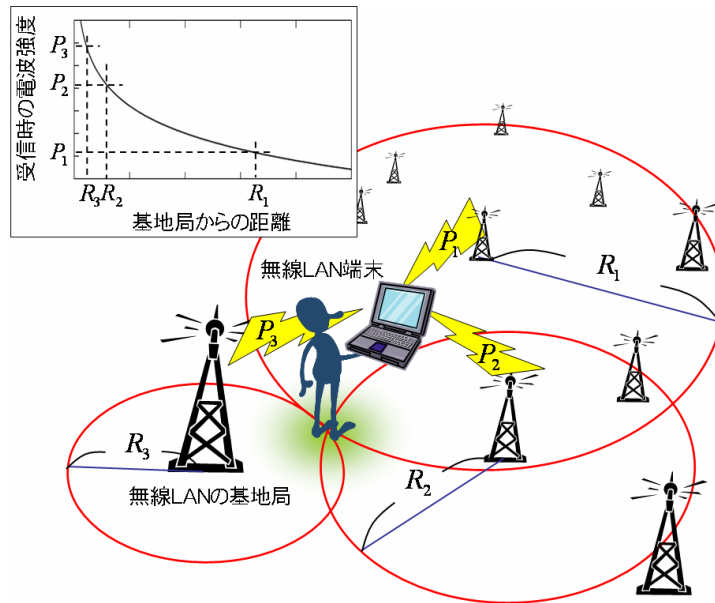


図 1 無線 LAN の電波強度による位置測定 [8]

以上のインフラを用いた手法では、超音波や電波にインフラ側のセンサを識別する情報を付加することが出来るため、個々のセンサを識別することが容易である。このため、時系列情報なしでカメラの位置及び姿勢が推定でき、高速な動きに対してロバストである。しかし、大量のセンサを環境中に設置しなければならないため、設置や保守に人的コストがかかる点や自然環境などのインフラの設置が困難な環境に対しては適用できない点が問題である。

既存のインフラを利用する手法では、絶対的な位置を計測可能な GPS(Global Positioning System) と姿勢センサ [2,34] や慣性センサ [1,33] などの自律計測型のセンサを組み合わせる手法が主流となっている。MARS [1] では、ディファレンシャル GPS を用いる位置計測と磁気センサ・慣性ヘッドトラッカーによる姿勢計測を統合することでカメラの位置及び姿勢を計測し、広範囲な屋外環境を移動するユーザが見ている現実環境の映像に対して、正しい位置に注釈情報を合成して提示する拡張現実感システムを実現している。また、Daehne ら [2] は地磁気センサで計測した姿勢とディファレンシャル GPS で計測した位置を直接使い、屋

外に存在する遺跡に復元された当時の建物の CG を重畳表示する拡張現実感システムを実現している。以上のように，GPS と姿勢センサを用いた手法では，ユーザが装着した機器のみにより位置及び姿勢が計測されるため，事前の作業が必要なく，広範囲な屋外においてカメラの位置及び姿勢が計測できる。しかし，絶対的な位置の測定に利用される GPS は屋内での利用が困難であり，絶対的な姿勢を計測できる磁気センサは建築構造物などによって容易に計測誤差が増大するなど，センサによって使用出来る環境に制限が多い。このため，環境に依存せずに位置及び姿勢を計測するには複数のセンサを組み合わせることが必要となり，システムが複雑になる問題がある。また，インフラを用いない手法として，自律計測型のセンサのみの組み合わせによる計測手法 [35, 36] も提案されている。これらの手法は，インフラを用いることなく環境によらずにカメラの位置及び姿勢を計測出来るが，相対的な位置や姿勢の計測なので，長時間使用する場合，広範囲な環境における計測では累積誤差が無視できない。

2.2 動画像からのカメラ位置・姿勢推定

カメラからの入力画像を用いる手法としては，人工的なマーカを用いる手法 [12–16]，事前に作成した環境内の物体の人工的な三次元モデルを用いる手法 [22–27]，自然特徴点における三次元位置などの情報をランドマークとして格納した自然特徴点ランドマークデータベースを用いる手法 [28, 29] などが挙げられ，三次元位置が既知である環境中の基準点をカメラ位置及び姿勢推定に用いることが共通した特徴として挙げられる。これらの手法では，三次元位置が既知である点と画像中の点を複数対応付け，三次元点の画像面への投影位置と画像中で検出した点との投影誤差が小さくなるカメラの位置・姿勢パラメータを計算することで，各フレームにおいて絶対的なカメラの位置及び姿勢が推定される。以下では，それぞれの手法における，三次元位置が既知である基準点の特徴やカメラ位置・姿勢推定時における問題点について述べる。

2.2.1 人工的なマーカを用いた手法

人工的なマーカを用いたカメラの位置及び姿勢の推定手法として, ARToolkit [11] などの形状やパターンなどが既知の画像マーカを用いた手法 [12-16] が研究されている。これらの手法は, 図 2 に示すように, 環境中に事前に設置した三次元位置が既知の画像マーカを撮影した画像から, マーカ上の点の三次元座標と入力画像上の二次元座標を対応付けることで位置及び姿勢を推定する。また, パターンから個々のマーカを識別することが容易であり, 2.1 節で述べたセンサを用いた手法と同様に, 時系列情報を用いずに現フレームの画像のみからカメラの位置及び姿勢が推定できるため, カメラの高速な動きに対してロバストである特長を持つ。しかし, これらの画像マーカを用いた手法では, 設置や整備にかかる人的コストや景観を損ねる問題がある。これらの問題を解決する手法として, Naimark ら [13] は, 設置における人的コストの問題を解決するために, 基準となる少数のマーカの三次元位置と相対的な位置関係から周囲のマーカの三次元位置を求めることで, 全てのマーカに対して三次元位置を測定する手間をなくした。また, 中里ら [14] は, 不可視マーカを用いることで, 環境中に設置したマーカが景観を損ねる問題を解決している。

だが, これらのマーカを用いた手法では, 2.1 節において述べたインフラを用いた手法と同様に, 大量のマーカを環境中に設置しなくてはならないため, 屋外の自然環境などのマーカの設置が困難な環境に利用できないといった問題がある。

2.2.2 人工的な三次元モデルを用いた手法

環境の人工的な三次元モデルを用いた手法では, 事前に幾何学的・光学的情報を取得した後は, センサやマーカなどの物理的なインフラを用いずに, カメラの絶対的な位置及び姿勢の推定が可能である。環境のモデルとして, ワイヤーフレームモデル [22-25] やテクスチャつきのポリゴンモデルを用いる手法 [26, 27] が提案されており, これらの手法では, エッジや自然特徴点などの入力画像中の特徴とモデルを対応付けることによりカメラの位置及び姿勢が推定される。だが, 環境のモデルを用いた手法の特徴としては, 画像中の個々の特徴点を一意に識別することが難しい点が挙げられ, 画像中の動きやオクルージョンによってモデルと入



図 2 ARToolkit とモバイルコンピュータを用いた拡張現実感システム [12]

力画像間において容易に誤対応が生じる．この問題に対し，Pressigout ら [22] は投影誤差を最小化する過程で，全データに対して均等な重みを用いるのではなく例外値には小さな重みを与えて 2 乗誤差を求め，投影誤差が最小となるカメラの位置及び姿勢を推定する M-Estimator によって，入力画像中のエッジとワイヤーフレームモデルの誤対応に対してロバスト性を高めている．また，Rosten らは，ワイヤーフレームモデルと入力画像中のエッジの対応付けだけでなく，入力画像上における特徴の動きに比較的ロバストな特徴点追跡を行うことで，推定のロバスト性を高める手法を提案している [24]．図 3 に示すような，CAD で作成した物体の三次元モデルと物体を複数の方向から撮影したキーフレーム画像やモデルの周囲を中心とした環境マップ群を用いる手法 [26, 27] では，オクルージョンに対してロバストなカメラ位置及び姿勢推定が可能である．

しかし，フレーム間におけるカメラの並進及び回転が大きい場合，モーションブラーや特徴の探索範囲の拡大によりモデルと入力画像において誤対応が生じるためにカメラの位置及び姿勢の推定誤差が増大する問題や，人工的な環境モデルの生成には多大な労力を要するといった問題がある．

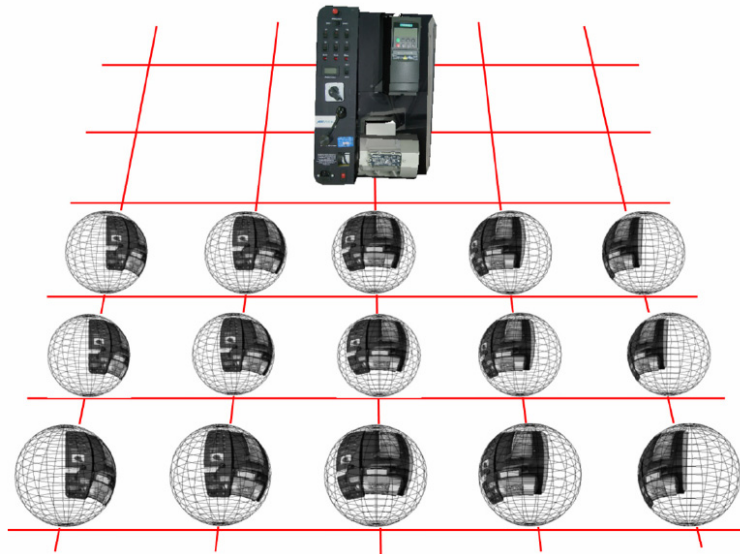


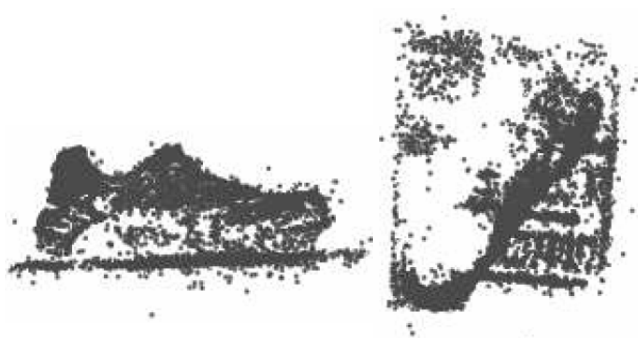
図 3 物体の三次元モデルとモデルの周囲を中心とした環境マップ群 [26]

2.2.3 自然特徴点ランドマークデータベースを用いた手法

環境のモデルを用いたカメラの位置及び姿勢の推定手法として、三次元位置が既知である環境中の自然特徴点と入力画像中の自然特徴点を対応付けることで、物理的なインフラを用いずにカメラの絶対的な位置及び姿勢を推定する手法 [28, 29] が提案されている。この自然特徴点群からなる環境のモデルを、以後、自然特徴点ランドマークデータベースと呼び、登録された自然特徴点をランドマークとする。自然特徴点ランドマークデータベースを用いた手法では、動画像から structure-from-motion の技術を用いてモデルが自動生成されるため、2.2.2 節で述べた CAD などを用いて人工的に構築した環境のモデルを用いた手法に比べ、広範囲な景観内の自然特徴点の三次元モデルを容易に構築できるという特長を持つ。自然特徴点ランドマークデータベースには、自然特徴点の三次元情報と共にそれらを識別するための特徴ベクトル [28] や画像テンプレート [29] などの情報が登録されており、ランドマークと入力画像中の特徴点の対応付けに用いられる。自然特徴点ランドマークデータベース構築時において、図 4 に示すようなモ



三次元復元における入力の一部



特徴点の三次元復元の結果を
横から見た図(左)と上から見た図(右)

図 4 Gordon らの手法による物体の三次元復元 [28]

デル化する対象物体を撮影した複数枚の画像のみから特徴点の三次元復元を行う手法 [28] を Gordon らは提案している．この手法における自然特徴点ランドマークデータベースには，スケールや回転に不変な SIFT 特徴量が登録されており，これにより入力画像とランドマークを対応付けることで，カメラの絶対的な位置及び姿勢が推定される．しかし，複数枚の画像のみからモデルを構築した場合，広範囲な環境を対象とした場合にはモデル中の特徴点の三次元位置に推定誤差が累積する問題がある．そこで，大江ら [29] は，広範囲な環境において三次元位置の推定精度が高い多数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手法として，Sato らの手法 [37] を用いて，事前に計測した環境中の複数の自然特徴点と全方位カメラによって撮影した入力画像から環境中の自然特徴点を三次元復元する手法を提案している．

しかし，自然特徴点ランドマークデータベースを用いた手法は，画像中の個々の特徴を一意に識別することが難しく，フレーム間におけるカメラの並進及び回転が大きい場合，モーションブラーや探索範囲の拡大によりモデルと入力画像において誤対応を生じ，位置及び姿勢の推定誤差が増大する問題がある．

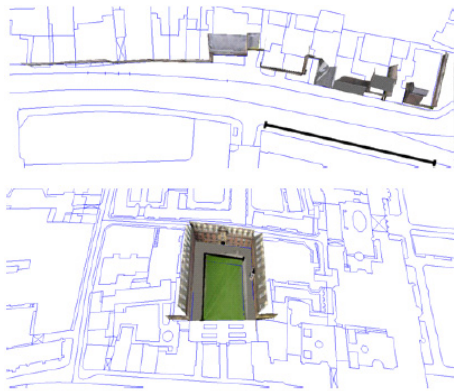
2.3 センサ情報と動画像からのカメラ位置・姿勢推定

カメラの高速な回転や並進などの運動によって2.2.2節で述べた人工的な三次元モデルを用いた手法で生じる誤対応の問題を解決するため、環境の人工的なモデルとセンサから得られる情報を併用して、カメラの位置及び姿勢を推定する手法 [30–32] が提案されている。これらの手法では、センサによる相対的な位置及び姿勢の変化の計測と動画像による絶対的な位置及び姿勢の推定を統合した枠組みが一般的である。特徴の探索範囲の拡大による誤対応を防ぐ手法として、Jiangら [30] や Reitmayrら [31] は、姿勢センサによる姿勢の計測結果と、ワイヤースケルトンやポリゴンモデルを用いた動画像処理による推定を拡張カルマンフィルタによって統合し、カメラの回転やオクルージョンなどの環境変化による誤対応にロバストな手法を提案している。カメラの位置・姿勢推定に用いるモデルの例として、Reitmayrら [31] の手法で用いる環境モデルを図5に示す。このポリゴンモデルは、PhotoBuilder [38] を用いて画像と地図から取得した奥行き情報により構築することができ、人工的なモデルの中では比較的容易に構築することが可能であるが、モデル内の細かな特徴が平面に近似されて登録されているために厳密なカメラ位置及び姿勢推定には向いていない。また、カメラが高速な回転を伴う際のモーションブラーによってワイヤースケルトンモデルと入力画像中の物体のエッジの対応付けが失敗する問題を解決するために、Kleinら [32] は姿勢センサにより得られる角速度情報からモーションブラー中のエッジを検出する探索ウィンドウを作成することで、モーションブラーによって検出が困難であったエッジを入力画像中から検出し、ワイヤースケルトンと対応する手法を提案している。

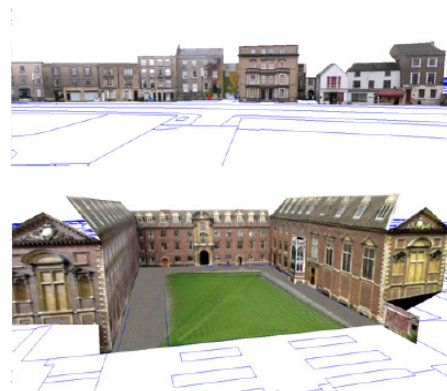
しかし、従来のワイヤースケルトンモデル [30, 32] やポリゴンモデル [31] とセンサによる情報を併用した手法では、対象とする物体や環境の規模が大きく、自然物のように複雑である場合には、モデルの作成自体が難しいという問題がある。

2.4 本研究の位置付けと方針

ウェアラブルコンピュータを用いた拡張現実感やロボットナビゲーションなどの対象となる、広範囲な環境において一般的な運動を伴うカメラの位置及び姿勢



環境のモデルを上から見た図



環境のモデルを横から見た図

図 5 PhotoBuilder で構築した環境のポリゴンモデル [30]

の推定には、以下の 5 点を満たすことが重要である。

- a. マーカやセンサなど物理的なインフラが不要
- b. 適用可能な環境が限定されない
- c. 絶対的なカメラの位置及び姿勢が推定される
- d. 事前知識として環境のモデルを用いる場合、モデルの構築が容易
- e. 並進や回転などのカメラの動きに対してロバスト

しかし、表 1 に示すように、従来のカメラの位置・姿勢推定法では、屋外環境のみで利用可能なものやフレーム間におけるカメラの動きが小さいことが前提であるものなど、それぞれ適用可能な条件や範囲が異なり、5 つの要求事項を同時に満たすことは困難であった。要求事項の中でも、インフラの設置が必要な手法や環境モデリングが必要な手法は前提条件として人手による事前作業が必要なため、広範囲な環境を対象とした場合に人的コストの増大は避けられない。また、適用環境が限定される手法においても、屋内、屋外を問わず様々な環境において位置及び姿勢を推定するためには他の手法と組み合わせなければならず、システムが

表 1 従来手法の比較

| カメラの位置及び姿勢の推定手法 | | 環境インフラ | 適用環境の限定 | 累積誤差 | 環境モデリング | 動きに対するロバスト性 |
|-----------------|-----------------------|---------|---------|-------|-------------------|-------------|
| 入力 | | | | | | |
| センサ情報 | 環境インフラ | ×: 必要あり | ×: 限定あり | ○: なし | ○: 必要なし | ○: 高い |
| | GPS+姿勢計測 | ○: 必要なし | ×: 限定あり | ○: なし | ○: 必要なし | ○: 高い |
| | 自律計測 | ○: 必要なし | ○: 限定なし | ×: あり | ○: 必要なし | ○: 高い |
| 画像+モデル | 画像マーカ | ×: 必要あり | ×: 限定あり | ○: なし | ○: 必要なし | ○: 高い |
| | 人工的なモデル | ○: 必要なし | ×: 限定あり | ○: なし | ×: 必要あり | ×: 低い |
| | 自然特徴点 ランドマークデータベース | ○: 必要なし | ○: 限定なし | ○: なし | △: 必要あり (自動生成) | ×: 低い |
| 画像+モデル+センサ情報 | 人工的なモデル+センサ | ○: 必要なし | ×: 限定あり | ○: なし | ×: 必要あり | ○: 高い |

複雑になる問題がある．累積誤差が生じる手法も同様に，絶対的なカメラの位置及び姿勢を推定するには他の手法と組み合わせる必要がある．しかし，動きに対するロバスト性については，センサ情報を併用することで環境の人工的なモデルを用いた手法のロバスト性を高める手法が存在するように，センサ情報を用いることで解決可能であることが分かっている．

そこで，本研究では要求事項 (a) ~ (e) を同時に満たすカメラの位置及び姿勢の推定手法の開発を目的とし，自然特徴点ランドマークデータベースと姿勢センサから得られる角速度情報を併用した手法を提案する．提案手法では，構築が容易な環境のモデルである自然特徴点ランドマークデータベースを用いることで，インフラを用いずに絶対的なカメラの位置及び姿勢が推定される．本手法では，全方位動画と事前に計測した基準マーカを用いた Sato ら [37] の手法により環境中の自然特徴点を三次元復元することで，広範囲な環境において正確な三次元位置のランドマークが登録された自然特徴点ランドマークデータベースを用いる．さらに，一般的に生じるカメラの高速な回転に対するロバスト性向上のため，姿勢センサにより得られる角速度情報を用いる．本研究では，まず，角速度情報を用いてデータベース中から推定に用いるランドマークを効果的に選択し，入力画像中における選択されたランドマークの対応点の探索範囲を限定する．また，モーションブラーを考慮したテンプレートマッチングにより，ランドマークと自然特

徴点を対応付ける．これらの処理によって，ランドマークと自然特徴点の誤対応を防ぎ，高速な回転に対してロバストに絶対的なカメラ位置及び姿勢を推定する．

3. 自然特徴点ランドマークデータベースと姿勢センサを用いたカメラ位置・姿勢推定

3.1 提案手法の概要

提案手法では，オフライン処理により予め構築した自然特徴点ランドマークデータベースを用いてカメラ位置及び姿勢を推定する．以下では，まず，自然特徴点ランドマークデータベースの構築について述べる．次に，オンライン逐次処理の枠組みである，自然特徴点ランドマークデータベースと姿勢センサによる角速度情報を用いたカメラの位置及び姿勢の推定手法について述べる．なお，本研究では全方位カメラ及びビデオカメラの内部パラメータは既知であるとする．

3.2 自然特徴点ランドマークデータベースの構築

自然特徴点ランドマークデータベースは，図6に示すように多数のランドマークから成り，各ランドマークの情報として，ランドマークの3次元位置と外観情報を保持している．様々な位置からのランドマークの見え方の変化に対応するために，外観情報は各ランドマークの撮影地点ごとに以下の情報が保持されている [29] ．

a. 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

入力画像中からランドマークの対応点を探索する際に用いられる．本研究では，図7に示すように，各撮影地点において基本スケールとその2倍，4倍のスケールで作成された多重解像度の画像テンプレートを保持する．ただし，基本スケールとは，入力画像と画像テンプレートの解像度がほぼ等しくなるような画像テンプレートの実空間中での大きさを示す．

b. ランドマーク撮影時の全方位カメラの位置及び姿勢

入力画像と対応付けるランドマークの選択に用いられ，全方位カメラの位置を世界座標系で保持する．

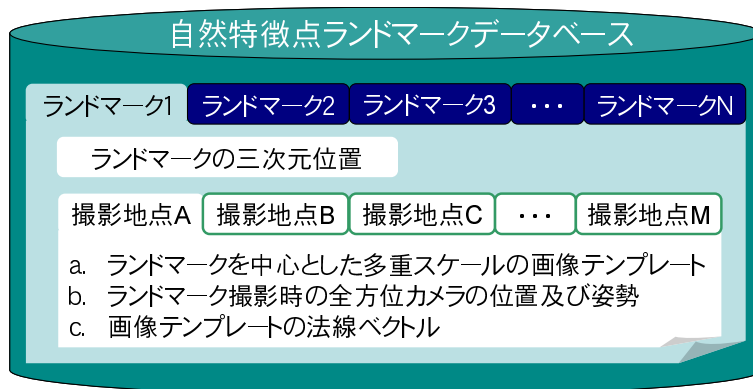


図 6 自然特徴点ランドマークデータベースの構成要素

c. 画像テンプレートの法線ベクトル

図 7 に，実空間中における画像テンプレートの法線方向を示す．各ランドマークに対して複数保持される撮影地点毎の情報から，対応点の探索時に用いる画像テンプレートとその法線ベクトルなどを選択する基準として用いられる．また，3.3.4 節で述べる入力画像とランドマークの対応付け処理において，入力画像とのテンプレートマッチングに用いる照合用の画像テンプレートをランドマークの画像テンプレートから作成する際に用いられる．

以上の情報は，事前に環境の撮影及び基準となる自然特徴点の三次元位置の計測を行い，環境内の自然特徴点を三次元復元することで取得する [37] ．

3.3 カメラ位置及び姿勢の推定手法

上記の自然特徴点ランドマークデータベース，姿勢センサからの角速度情報及び入力画像から，カメラの絶対的な位置及び姿勢を推定する手法について述べる．本手法の処理の流れを図 8 に示す．各フレームにおける処理は以下の通りである．まず，姿勢センサで得られた角速度情報から暫定的なカメラの位置及び姿勢を算出し (A)，データベースから推定に利用するランドマークを限定する (B) ．限定さ

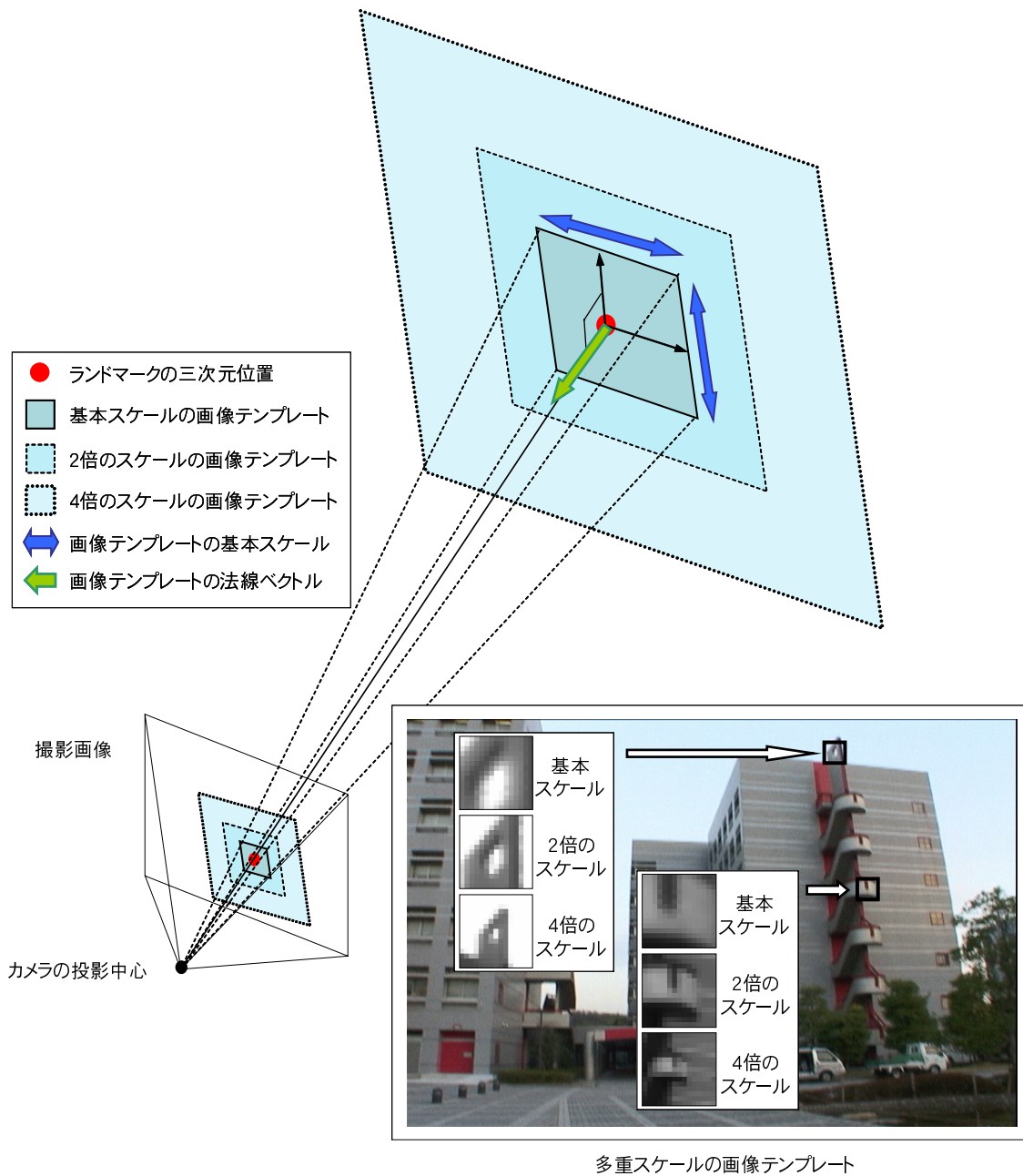


図 7 ランドマークとその多重スケールの画像テンプレート

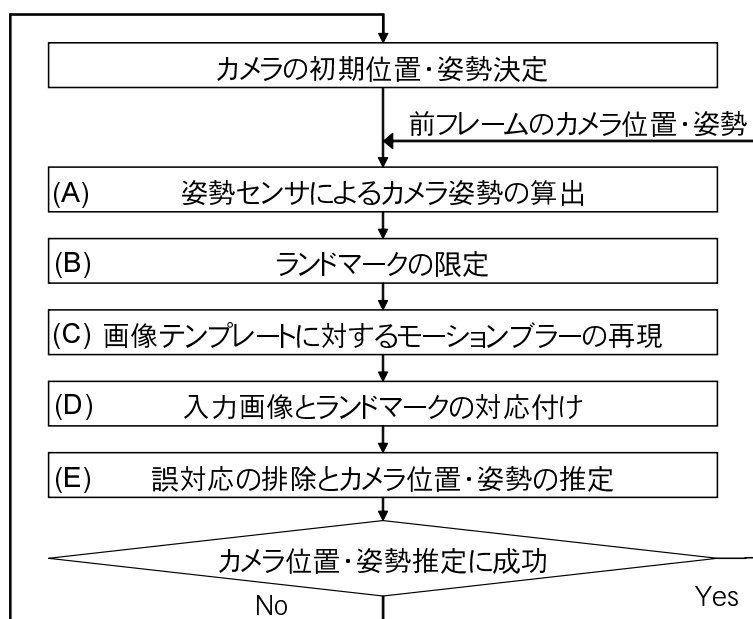


図 8 カメラ位置及び姿勢推定の処理の流れ

れたランドマークの画像テンプレートに対して入力画像と同じモーションブラーを再現する (C) . 生成したテンプレートと入力画像のテンプレートマッチングにより, 入力画像中の自然特徴点とランドマークとを対応付け (D) , 最後に, それらの対応関係を用いて PnP 問題 [39] を解くことによりカメラの位置及び姿勢を推定する (E) . ただし, 本稿では, 初期フレームのカメラの位置及び姿勢は何らかの手法によって取得出来ることを前提としており, カメラ位置・姿勢推定が失敗したと判断された場合には, 再度, カメラの初期位置・姿勢の推定から各ステップの処理を繰り返す. このカメラ位置・姿勢推定の成功・失敗判定については, ステップ (E) の説明で述べる. 以下, ステップ (A) ~ (E) について詳述する.

3.3.1 姿勢センサによるカメラ姿勢の算出・ステップ (A)

図 9 に示す座標系において, 第 i フレームにおける姿勢センサから得られる角速度 ω_i と世界座標系から第 $(i-1)$ フレームのカメラ座標系への 3×3 の変換行

列（回転行列） \mathbf{R}_{i-1}^c を用いて，第 i フレームにおけるカメラの暫定的な姿勢 $\tilde{\mathbf{R}}_i^c$ を算出する処理について述べる．ただし，ここで，センサから得られる角速度情報のみを用いて姿勢を推定すると，ドリフトによる累積的な誤差の影響を受けるため，本手法では，第 $(i-1)$ フレームで推定された絶対的なカメラの姿勢 \mathbf{R}_{i-1}^c と角速度 ϖ_i ，および予めキャリブレーションにより得たカメラ座標系からセンサ座標系への変換行列 \mathbf{R}^{cg} のみからカメラの姿勢 $\tilde{\mathbf{R}}_i^c$ を算出することでドリフト誤差の影響を軽減する．キャリブレーション手法については付録Aを参照．以下では，カメラの姿勢 $\tilde{\mathbf{R}}_i^c$ の算出方法について述べる．

図9に示すように，第 i フレームにおける姿勢センサの姿勢 \mathbf{R}_i^g は，カメラの姿勢 \mathbf{R}_i^c とカメラ座標系から姿勢センサ座標系への回転行列 \mathbf{R}^{cg} から $\mathbf{R}_i^g = \mathbf{R}^{cg}\mathbf{R}_i^c$ で表される．また，第 $(i-1)$ フレームから第 i フレームへの姿勢センサの回転を $\mathbf{R}_{i-1,i}^g$ としてあらわすと， $\mathbf{R}_i^g = \mathbf{R}_{i-1,i}^g\mathbf{R}_{i-1}^g$ が成立する．以上より，第 i フレームにおけるカメラの暫定的な姿勢 $\tilde{\mathbf{R}}_i^c$ は以下の式で算出される．

$$\tilde{\mathbf{R}}_i^c = (\mathbf{R}^{cg})^{-1}\mathbf{R}_{i-1,i}^g\mathbf{R}^{cg}\mathbf{R}_{i-1}^c \quad (1)$$

第 $(i-1)$ フレームから第 i フレームにおける微小な回転を $\Omega_i = (\mathbf{R}_{i-1}^g)^{-1}\mathbf{R}_{i-1,i}^g$ で表せば，式(1)は次式のように変形できる．

$$\begin{aligned} \tilde{\mathbf{R}}_i^c &= (\mathbf{R}^{cg})^{-1}\mathbf{R}_{i-1,i}^g\mathbf{R}^{cg}\mathbf{R}_{i-1}^c \\ &= (\mathbf{R}^{cg})^{-1}\mathbf{R}_{i-1}^g\{(\mathbf{R}_{i-1}^g)^{-1}\mathbf{R}_{i-1,i}^g\}\mathbf{R}^{cg}\mathbf{R}_{i-1}^c \\ &= (\mathbf{R}^{cg})^{-1}\mathbf{R}_{i-1}^g\Omega_i\mathbf{R}^{cg}\mathbf{R}_{i-1}^c \\ &= \mathbf{R}_{i-1}^c\Omega_i\mathbf{R}^{cg}\mathbf{R}_{i-1}^c \end{aligned} \quad (2)$$

微小な回転は各軸周りの角速度ベクトルとの外積で表すことができるため， Ω_i は第 i フレームにおける角速度 $\varpi_i = (\varpi_{ix}, \varpi_{iy}, \varpi_{iz})$ の要素のみからなる行列であり，以下の様に表現できる．

$$\Omega_i = (\mathbf{R}_{i-1}^g)^{-1}\mathbf{R}_{i-1,i}^g = \begin{pmatrix} 0 & -\varpi_{iz} & \varpi_{iy} \\ \varpi_{iz} & 0 & -\varpi_{ix} \\ -\varpi_{iy} & \varpi_{ix} & 0 \end{pmatrix} \quad (3)$$

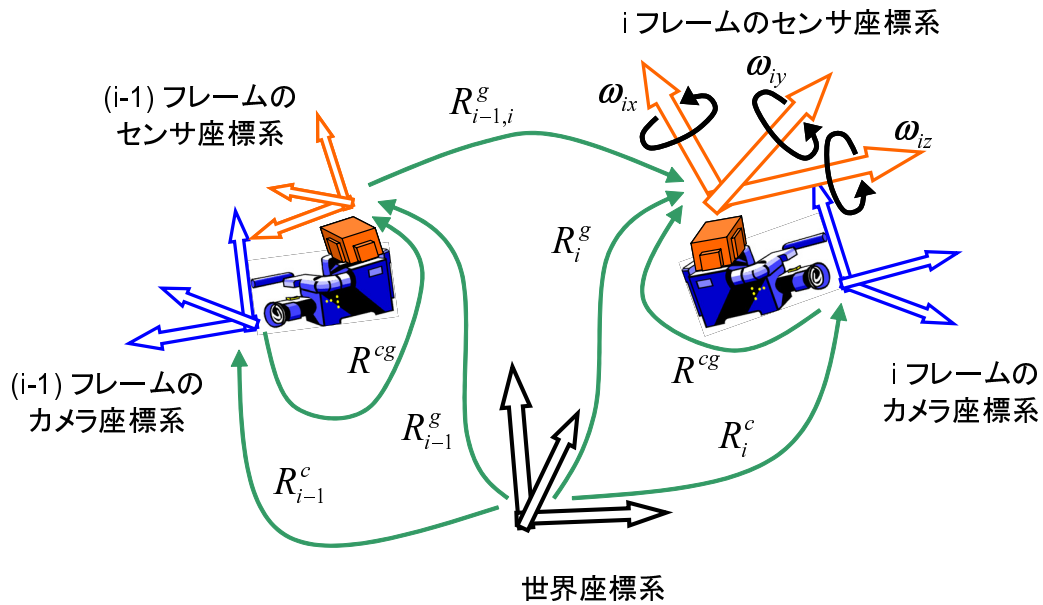


図 9 カメラと姿勢センサの座標系の定義

3.1 節で述べたように、姿勢センサとカメラ間の変換行列 R^{cg} はキャリブレーション済みで既知であり、カメラの姿勢 R_{i-1}^c 及び行列 Ω は累積誤差のない回転をあらわすので、式 (2) を用いることで前フレームのカメラ姿勢と現フレームにおける角速度情報から絶対的なカメラの姿勢を表す \tilde{R}^c を算出できる。

3.3.2 カメラ位置・姿勢推定に用いるランドマークの選択・ステップ (B)

ステップ (A) で算出した暫定的なカメラの姿勢 \tilde{R}_i^c を用いて、自然特徴点ランドマークデータベースに登録されている多数のランドマークから、入力画像との対応付けに用いるランドマークを限定する処理について述べる。ここでは、第 i フレームに対して予測されるカメラの位置 \tilde{P}_i^c 及び姿勢 \tilde{R}_i^c から、画角内に入るランドマークを選択し、入力画像と対応付ける候補とする。具体的には、第 $(i-1)$ フレームでのカメラの位置を P_{i-1}^c 、ランドマーク j の三次元位置を $l_{(j)}$ とすると、第 i フレームにおける画像上へのランドマーク j の予測投影位置 $\tilde{c}_{(j)i} = [x_{(j)i}, y_{(j)i}]^t$

は以下の式で表される．

$$\begin{pmatrix} a_{(j)i}x_{(j)i} \\ a_{(j)i}y_{(j)i} \\ a_{(j)i} \end{pmatrix} = \tilde{\mathbf{R}}_i^c \mathbf{l}_{(j)} + \tilde{\mathbf{P}}_i^c \quad (4)$$

ただし，第 i フレームにおけるカメラの予測位置 $\tilde{\mathbf{P}}_i^c$ は，カメラが等速直線運動をしていると仮定し，次式により算出する．

$$\tilde{\mathbf{P}}_i^c = 2\mathbf{P}_{i-1}^c - \mathbf{P}_{i-2}^c \quad (5)$$

また， $a_{(j)i}$ は第 i フレームにおけるカメラ座標系でのランドマーク j の奥行きを表す．本処理では，図 10 に示すように，上式で得られる予測投影位置 $\tilde{c}_{(j)i}$ がカメラの画角内に存在するランドマークのみを選択し，推定に利用するランドマークとする．ただし，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る場合，このランドマークを用いない．具体的には，格子状に分割された入力画像上において， \tilde{c}_i から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るか判定し，以降の処理で用いるかを決定する．

以上の処理で選択されたランドマークを入力画像と対応付ける処理 (3.3.4 節で後述) において，テンプレートマッチングによる誤対応を防ぐため，入力画像と最も見え方の近い画像テンプレートを選択する．具体的には，複数の撮影地点毎の情報に対して画像テンプレートの法線ベクトルとランドマークの三次元位置 \mathbf{l} から入力画像の撮影カメラ位置 $\tilde{\mathbf{P}}^c$ へのベクトルのなす角 θ を算出し，角度 θ が最小であった法線ベクトルを持つ画像テンプレートを以降の処理で用いる．ただし，角度 θ が閾値以上であった場合，そのランドマークは推定に用いない．なお，誤対応の排除とカメラ位置・姿勢推定処理 (3.3.5 節で後述) で用いる PnP 問題の解法において，入力画像との誤対応を含まない 6 点以上のランドマークが用いられればカメラ位置・姿勢推定が可能であるため，選択されるランドマークの個数に上限を設ける．選択されたランドマークの数が設定した上限を上回る場合，入力画像上での見え方の変化が小さいと考えられるランドマークを優先的に選択するために，ランドマーク撮影時の全方位カメラの位置と暫定的なカメラ位置 $\tilde{\mathbf{P}}^c$ の距離が小さい順にランドマークを選択する．

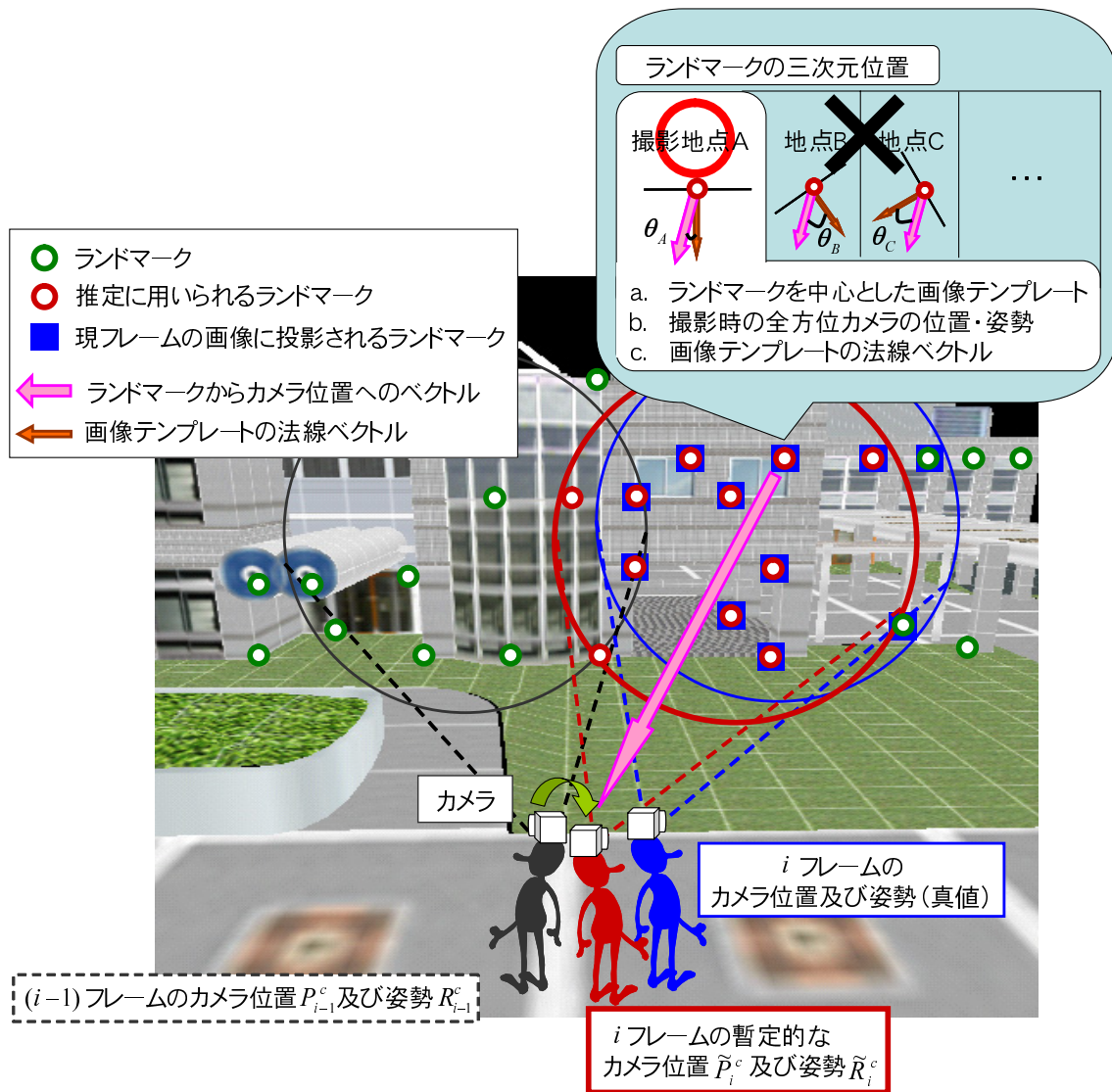


図 10 カメラと選択されるランドマークの位置関係

3.3.3 画像テンプレートに対するモーションブラーの再現・ステップ (C)

ステップ (B) により選択されたランドマークの画像テンプレートから，入力画像のモーションブラーを再現した画像テンプレートを生成する処理について述べる．なお，図 12 に示すように，本手法では第 i フレームを取得した時点に最も近い瞬間のカメラの位置及び姿勢を推定するため，第 i フレームにおいて露光が終了する瞬間のカメラの位置及び姿勢を推定する．本研究では，等速直線運動を仮定した PSF モデルを用いて，各ランドマークに対してデータベースの画像テンプレートにモーションブラーを再現する．この処理により，図 11 に示すように，入力画像とランドマーク間の画像テンプレートの類似度を高めることが出来る．同図から，特に，高周波成分を多く含む画像がブラーによって変化しやすいため，複雑な環境の画像テンプレートに対しては，より効果的であることが分かる．モーションブラーの大きさ，方向は，以下のベクトル $\mathbf{b}_{(j)}$ で表される．

$$\mathbf{b}_{(j)} = (b_{(j)x}, b_{(j)y}) = (\mathbf{c}_{(j)i-1} - \tilde{\mathbf{c}}_{(j)i})s\Delta t \quad (6)$$

ただし， $\mathbf{c}_{(j)i-1}$ は第 $(i-1)$ フレームにおけるカメラ位置 \mathbf{P}_{i-1}^c 及び姿勢 \mathbf{R}_{i-1}^c と三次元位置 $\mathbf{l}_{(j)}$ から算出したランドマークの第 $(i-1)$ フレームにおける画像への投影座標， Δt はフレーム間の時間， $b_{(j)x}$ はベクトル $\mathbf{b}_{(j)}$ の入力画像上における x 成分， $b_{(j)y}$ はベクトル $\mathbf{b}_{(j)}$ の入力画像上における y 成分を表し，シャッタースピード s は既知とする．

算出されたベクトル $\mathbf{b}_{(j)}$ を用いて画像テンプレートに対しモーションブラーを再現する処理について詳述する．ランドマーク j の画像テンプレートの領域を $\Psi_{(j)}$ とすると， $\mathbf{b}_{(j)}$ から算出されたモーションブラーを再現するウィンドウ B を畳み込むことにより， $\Psi_{(j)}(\mu, \nu)$ の輝度値 $I_{(j)}(\mu, \nu)$ は， $\Psi_{(j)}(\mu, \nu)$ 周囲の輝度値とベクトル $\mathbf{b}_{(j)}$ から以下の式により算出される．

$$I_{(j)}(\mu, \nu) = \sum_{k=1}^{|\mathbf{b}_{(j)}|} I_{(j)}\left(\left(\mu + \frac{kb_{(j)x}}{|\mathbf{b}_{(j)}|}\right), \left(\nu + \frac{kb_{(j)y}}{|\mathbf{b}_{(j)}|}\right)\right) \frac{1}{|\mathbf{b}_{(j)}|} \quad (7)$$

ただし， μ, ν はランドマークの画像テンプレートにおける座標を表す． $\Psi_{(j)}$ に対してウィンドウ B を走査することにより，ランドマーク j が中心であるモーションブラーを再現済みの画像テンプレートが生成される．


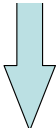



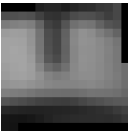
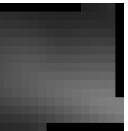

| | ブラーの再現なし | ブラーの再現あり | 入力画像 | ブラーの再現なし | ブラーの再現あり | 入力画像 |
|---|---|---|---|--|---|---|
| スケール小  多重スケールの画像テンプレート  スケール大 |  |  |  |  |  |  |
| 正規化相互相関値 | 0.40 | 0.72 | | 0.76 | 0.95 | |
| | 高周波成分を多く含む画像テンプレート | | | 高周波成分の少ない画像テンプレート | | |

図 11 モーションブラー再現による入力画像との正規化相互相関値の変化

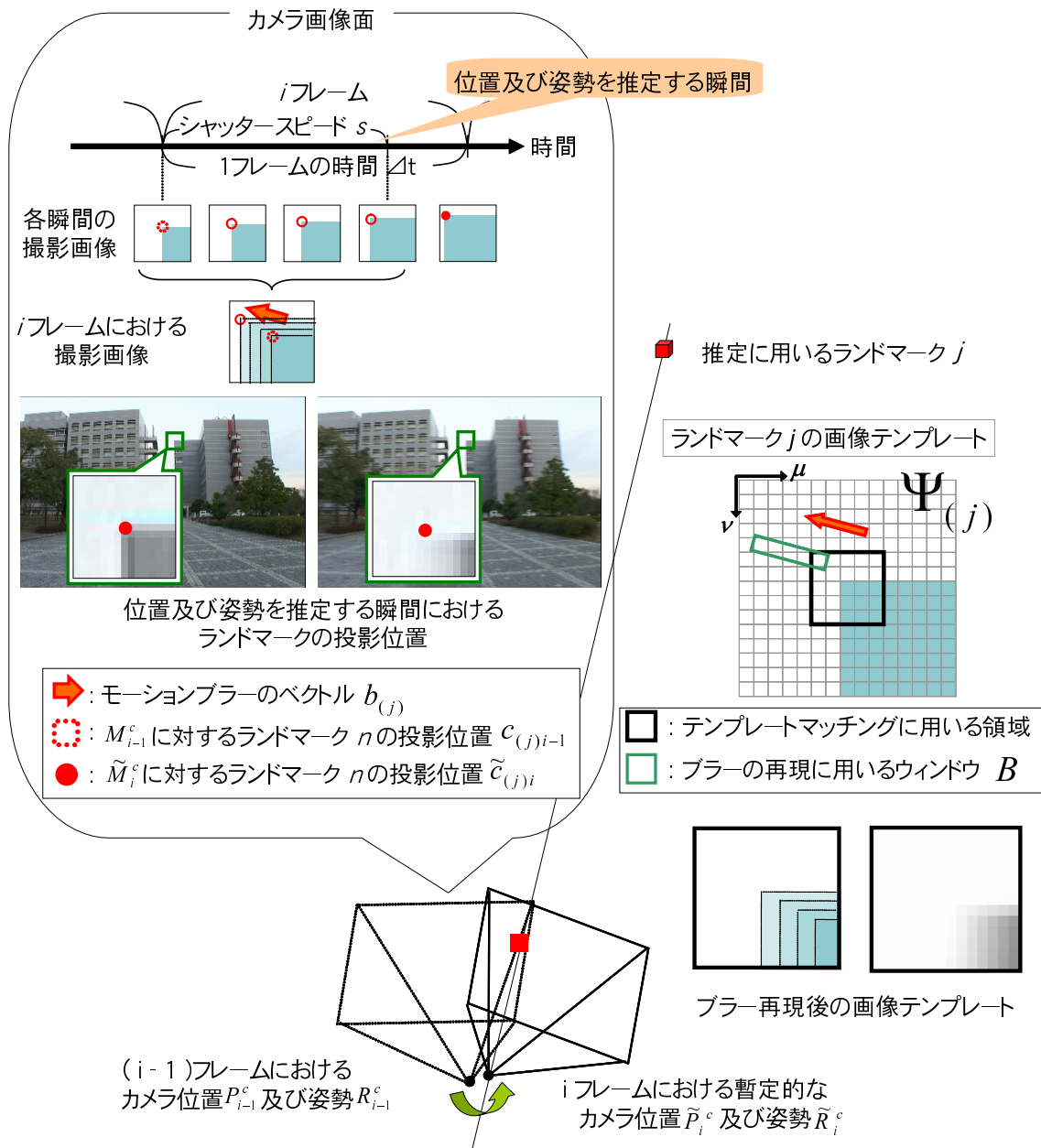


図 12 モーションブラーのベクトルの算出と画像テンプレートに対するモーションブラーの再現

3.3.4 入力画像とランドマークの対応付け・ステップ (D)

ステップ (B) で限定された各ランドマークに対して入力画像中の自然特徴点を対応付ける処理について述べる．具体的には，ステップ (B) で算出されたランドマーク j の予測投影位置 $\tilde{c}_{(j)i}$ を中心とする大きさ δ のウィンドウ内において，ステップ (C) で生成されたテンプレートを用いて，正規化相互相関を指標としたテンプレートマッチングを行う．

1. 入力画像中の対応点の探索

ウィンドウの大きさ δ は，カメラの並進移動と姿勢センサの計測精度を考慮して設定する．具体的には，ランドマーク j のカメラからの奥行き $a_{(j)i}$ ，アプリケーションに合わせて設定されたカメラの最大並進速度 V_{max} ，姿勢センサの精度から設定される定数 δ' を用いて検出ウィンドウの大きさ $\delta_{(j)i}$ は以下の式で算出される．

$$\delta_{(j)i} = \frac{V_{max}\Delta t}{a_{(j)i}} + \delta' \quad (8)$$

2. 画像テンプレートのマッチングによる対応付け

ランドマークと入力画像中の自然特徴点を対応付けるために，入力画像とランドマークの画像テンプレートのテンプレートマッチングを以下の手順で行い，前項で設定したウィンドウ内の各画素とランドマークを識別する．

- (a) 対応候補点である探索ウィンドウ内の各画素について，入力画像からその画素を中心とした画像テンプレートを作成する．
- (b) 作成した画像テンプレートとステップ (C) においてモーションブラーを再現した画像テンプレートを正規化相互相関によって比較し，最も相関値が高い点をランドマークと対応付ける．

以上の処理の詳細を以下で述べる．

正規化相互相関に用いるランドマークの画像テンプレートは，ステップ (C) において作成した画像テンプレートを入力画像面に投影したものであり，カ

メラ位置 \tilde{P}^c 及び姿勢 \tilde{R}^c と多重スケールの各画像テンプレート, ランドマークの三次元位置 l , 画像テンプレートの法線ベクトルから作成する. 一方, 対応候補点の画像テンプレートは, 対応付けるランドマークの画像テンプレートと同一の大きさの領域のものと, その2倍, 4倍の多重スケールで構成される. これらの多重スケール画像テンプレート間の正規化相互相関は, 各テンプレート内の全ての画素の輝度値を並べたベクトルを p, q とし, その平均値を並べたベクトルを \bar{p}, \bar{q} とした時, 以下の式により算出される.

$$C(p, q) = \frac{(p - \bar{p}) \cdot (q - \bar{q})}{|p - \bar{p}| |q - \bar{q}|} \quad (9)$$

この相関値が最大かつ閾値以上である対応候補点をランドマークと対応付ける. ここで得られた対応関係から, データベース内の自然特徴点の三次元座標と入力画像上の二次元座標を対応付けることができる.

3.3.5 誤対応の排除とカメラ位置・姿勢推定・ステップ (E)

本ステップでは, ステップ (D) で対応付けられたランドマークと入力画像中の特徴点の対応関係を用いて絶対的なカメラの位置及び姿勢を推定する. 三次元位置と二次元座標の対応組を6組以上用い, PnP問題 [39] を解くことで, カメラ位置及び姿勢は推定される. しかし, これまでの処理で対応付けられたランドマークと特徴点の対応組に誤対応が含まれている場合, 位置及び姿勢の推定精度が低下する問題がある. そのため, RANSAC [40] により誤対応を排除してカメラ位置及び姿勢を推定する処理を繰り返し, 再投影誤差が最小となるカメラの位置及び姿勢を, i フレームにおける最終的なカメラ位置 P_i^c 及び姿勢 R_i^c の推定結果とする. カメラの投影モデルについては付録 B を参照. RANSAC とは, 正しいデータと誤ったデータが混合したデータにおいて, ランダムなサンプルデータの選択, サンプルデータから算出したパラメータの評価の2処理を繰り返し, 最も評価の高いサンプルデータを正しいデータとする統計的手法である. 理論的には, 過半数の正しいデータが含まれれば誤ったデータを除去することが出来るため, 正しいデータを取り出す代表的な手法とされる. 以下に, RANSAC による具体的な処理について詳述する.

- (1) ステップ(D)で対応付けられた三次元位置と二次元座標の対応組からランダムに6組以上を選択し、暫定的なカメラ位置及び姿勢を推定する。
- (2) (1)で得られた暫定的なカメラ位置及び姿勢に対し、全ての対応組において再投影誤差 N を算出し、その中央値 \bar{N} を求める。ランドマーク j と対応付けた二次元座標を (j_x, j_y) 、ランドマーク j の三次元座標 $l_{(j)}$ を暫定的なカメラ位置及び姿勢で画像上に投影した座標を $(\tilde{j}_x, \tilde{j}_y)$ とした時、再投影誤差 N は、 $N = (j_x - \tilde{j}_x)^2 + (j_y - \tilde{j}_y)^2$ として定義する。
- (3) (1), (2)の処理を複数回繰り返して、再投影誤差の中央値 \bar{N} が最小となる暫定的なカメラ位置及び姿勢を選択する。
- (4) 選択されたカメラパラメータによって全対応組の投影誤差を評価し、閾値を超える組み合わせを誤対応として削除する。

以上の処理によって得られた誤対応が排除された三次元位置と二次元座標の対応関係を用いて再投影誤差を算出し、これを最小化することで第 i フレームにおけるカメラの位置 P_i^c 及び姿勢 R_i^c の推定を行う。再投影誤差の最小化は非線形最小化問題であり、局所解の問題や計算量の問題を伴うため、まず線形解法を用いてまずカメラ位置・姿勢の初期値を求め [39]、次に、非線形最小化によって再投影誤差 N を最小化する。ここで、最終的に利用可能な対応点数が6点を下回るか再投影誤差が十分に最小化されない場合、推定が失敗したとみなす。

4. カメラ位置・姿勢推定の精度評価実験

提案手法の有効性を検証するために，実動画像を用いた精度評価実験を行った．以下，まず実験環境について述べ，次に，提案手法の定量的な評価実験について述べる．

4.1 実験環境

本実験で入力として用いた動画像，角速度情報，ランドマークデータベースのデータの取得方法について述べる．以下に，これらのデータの取得方法について述べる．使用する機器構成を表 2 に示す．図 13 に示すように，動画像は手持ちのビデオカメラ，姿勢情報はカメラに固定された姿勢センサで得た．姿勢センサの主な仕様を表 3 に示す．自然特徴点ランドマークデータベースの生成には，図 14 左に示すような放射状に固定された複数のカメラにより全天球の動画像を取得できる全方位型マルチカメラシステム (Point Grey Research 社 Ladybug，解像度 $768 \times 1024 \times 6$ 台) を用いた．このカメラシステムにより屋外環境を図 14 右に示すような画像系列として移動撮影し，佐藤らの手法 [37] によりカメラパスと自然特徴点の三次元位置を推定することにより自然特徴点ランドマークデータベースを構築した．自然特徴点ランドマークデータベース中のランドマーク数は約 7500 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 8ヶ所であった．なお，定量的な評価に用いるカメラ位置及び姿勢の真値は，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．

表 2 カメラ位置・姿勢推定に用いた機器

| | |
|---------|------------------------------------|
| 入力撮影カメラ | SONY DSR-DP-150 |
| 広角レンズ | SONY VCL-HG0758 |
| 姿勢センサ | TOKIMEC TISS-5-40 |
| マシンスペック | CPU:PentiumM 2.13GHz, Memory:2.0GB |



図 13 カメラと姿勢センサの位置関係

表 3 姿勢センサの仕様

| | |
|--------|---|
| 方位角誤差 | $\pm 2^\circ$ ドリフト: $3^\circ/\text{h}$ スケールファクタ: $\pm 0.11\%$ |
| ロール角誤差 | $\pm 0.5^\circ$ |
| ピッチ角誤差 | $\pm 0.5^\circ$ |
| 最大角速度 | $458^\circ/\text{s}$ |
| 計測レート | 50Hz |



図 14 全方位型マルチカメラシステムと全方位画像

4.2 定量的な評価実験

上記の実験環境において、以下に述べる 2 つの比較手法と提案手法で推定されたカメラの位置及び姿勢を真値と比較することで推定精度を調べた。ただし、モーションブレンダーが生じた画像から直接カメラの位置・姿勢の真値を求め、推定誤差を計算することは困難なので、推定に用いた入力動画は、以下の 2 種類の方法で撮影したものをを用いた。

入力 1 回転のみを伴うカメラによる撮影

入力 2 回転と並進を伴うカメラによる撮影

回転のみを伴うカメラによって得られた動画(入力 1)は、モーションブレンダーが生じていない初期フレームから算出したカメラ位置を真値とすることで、動画中の全フレームに対して推定位置誤差を算出した。一方、回転と並進を伴うカメラによって得られた動画(入力 2)は、推定誤差を算出する各フレームにおいてカメラの位置・姿勢の真値を算出し、推定位置誤差・推定姿勢誤差を算出した。

ただし，真値計算時の再投影誤差が1.5pixel以上となったフレームは，信頼性が低いため，評価対象から除外した．入力1,2に対応する動画像の一部のフレームをそれぞれ図15, 図16に示す．これらは，入力画像撮影時において自動設定されたシャッタースピード1/15秒で撮影した．また，本実験において，カメラの最大並進速度 V_{max} は，一般的な人間のジョギングの速度である $8km/h$ とし，姿勢センサの精度から算出した検出ウィンドウの大きさ δ' を16画素に設定した．本実験で使用したカメラ位置・姿勢推定処理の各ステップにおける推定条件を表5に示す．

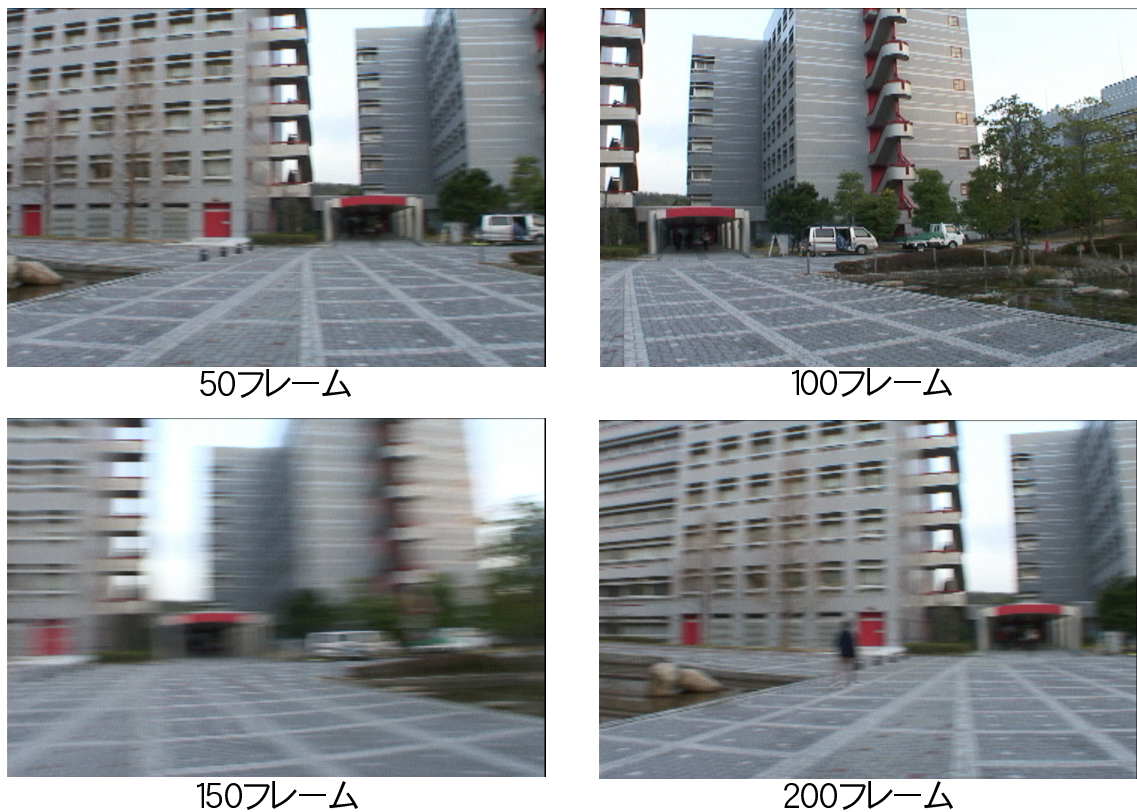


図15 入力画像(入力1)

本実験では，上記の入力画像を用い，提案手法と2つの比較手法とを比較する．2つの比較手法は，本研究で新たに導入したステップ(B)，(C)の有効性を確認す

表 4 推定に用いた入力画像と算出する推定誤差

| | 評価するパラメータ | 対象フレーム |
|------|-----------|-------------------------|
| 入力 1 | 位置 | 全フレーム |
| 入力 2 | 位置・姿勢 | 再投影誤差が 1.5pixel 以下のフレーム |

表 5 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・姿勢推定の各処理における条件

| | | |
|----------|------------------------------------|-----------|
| | 入力画像の画素数 (画素) | 720 × 480 |
| | 入力画像のフレームレート (フレーム/秒) | 15 |
| ステップ (B) | 画像テンプレートとカメラの光軸のなす最大角 θ (度) | 30 |
| ステップ (B) | データベースから選択するランドマーク数 (個/フレーム) | 100 |
| ステップ (B) | 入力画像上に設定する格子の間隔 (画素) | 40 × 40 |
| ステップ (D) | カメラの最大並進速度 V_{max} (km/h) | 8 |
| ステップ (D) | センサの誤差を考慮した探索範囲の大きさ δ' (画素) | 16 |
| ステップ (E) | RANSAC の繰り返し回数 (回) | 500 |



50フレーム



100フレーム



150フレーム



200フレーム

図 16 入力画像 (入力 2)

るために、提案手法中のステップ (B)、(C) のうち、1 つのステップを以下に示す処理に置き換えたものとする。

比較手法 (1) ステップ (B) において、カメラの予測位置 \tilde{P}_i^c 及び姿勢 \tilde{R}_i^c の代わりに、前フレームで推定したカメラの位置 P_{i-1}^c 及び姿勢 R_{i-1}^c を用いて推定に用いるランドマークを限定する。

比較手法 (2) ステップ (C) において、ランドマークの画像テンプレートに対してモーションブラーを再現しない。

4.2.1 回転のみを伴うカメラで撮影された動画像(入力1)

推定されたカメラ位置・姿勢を用いて仮想物体を一定位置に重畳させた match move の結果を図 17 に，モーションブラーが大きい150 フレームにおける被写体が動く方向と match move の結果を一部拡大したものを図 18 に示す．ステップ (C) で述べたように，提案手法では露光が終了する瞬間のカメラ位置・姿勢を推定することから，図 17, 18 より，CG で重畳したワイヤーフレームが入力画像と幾何学的に正しく合成されており，カメラ位置・姿勢推定に大きな誤差が含まれていないことが分かる．各フレームにおける角速度と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図 19 に，推定位置誤差を図 20 に，各手法の平均推定誤差と標準偏差を表 6 に示す．比較手法 (1) では，277 フレームにおいてカメラ位置姿勢推定が途中で失敗した．図 19 から，比較手法 (1) の推定処理の失敗は，大きな角速度が発生した直後に発生していることが分かる．比較手法 (2) では，大きな角速度が発生した70 フレームや270 フレーム付近における対応付けられたランドマーク数の減少が見られる．一方，提案手法にも対応付けられたランドマーク数の変化が見られるが比較的安定しており，全てのフレームで推定が成功した．また，図 20 から，比較手法は角速度が大きいフレームにおいて推定誤差が増大しているが，提案手法は対応付けられたランドマーク数と同様に，比較手法と比べて安定していることが分かる．このことから，姿勢センサを用いたステップ (B), (C) の導入により，高速な回転に対するロバスト性が向上していることが分かる．また，各比較手法に比べ，提案手法は各フレームにおける誤差のばらつきが少なく，大きな誤差を生じることなくカメラの位置を推定できていることが分かる．統計的に推定誤差の検証を行うため，有意水準を 5%として提案手法に対して t 検定を行った結果，比較手法 (1) との t 値が-7.64，比較手法 (2) との t 値が-7.93 であり，提案手法と比較手法の推定誤差に有意な差があることを確認した．

4.2.2 回転と並進を伴うカメラで撮影された動画像(入力2)

推定されたカメラ位置・姿勢を用いて仮想物体を一定位置に重畳させた match move の結果を図 21 に，モーションブラーが大きい200 フレームにおける被写体

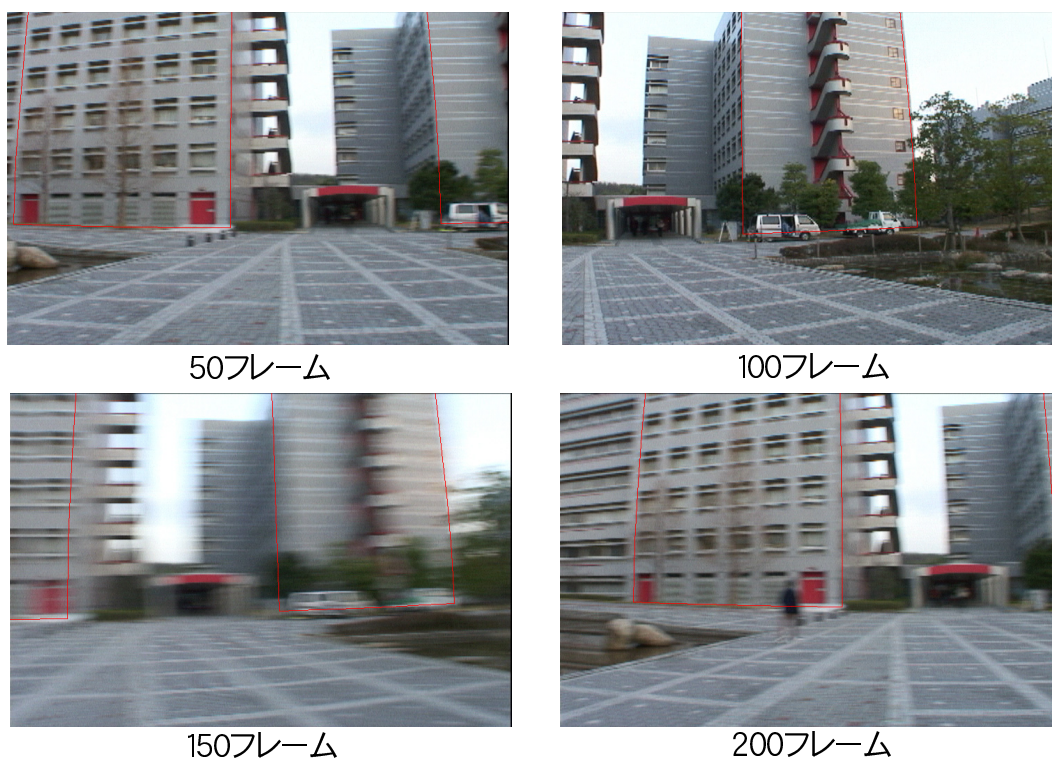


図 17 入力 1 におけるカメラ位置・姿勢推定結果を用いた match move

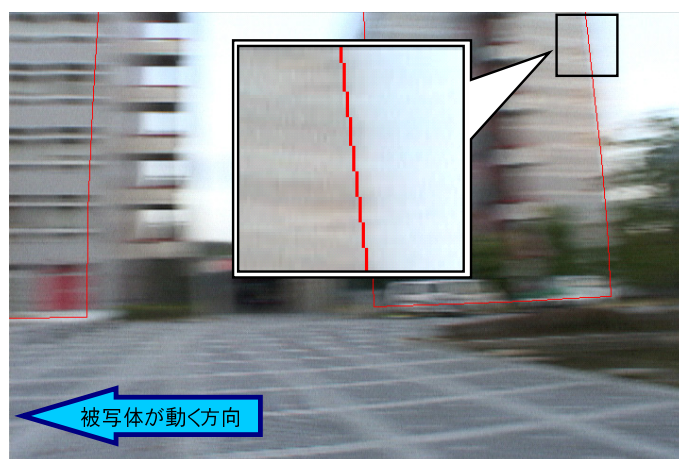


図 18 入力 1 における match move の拡大図と被写体が動く方向

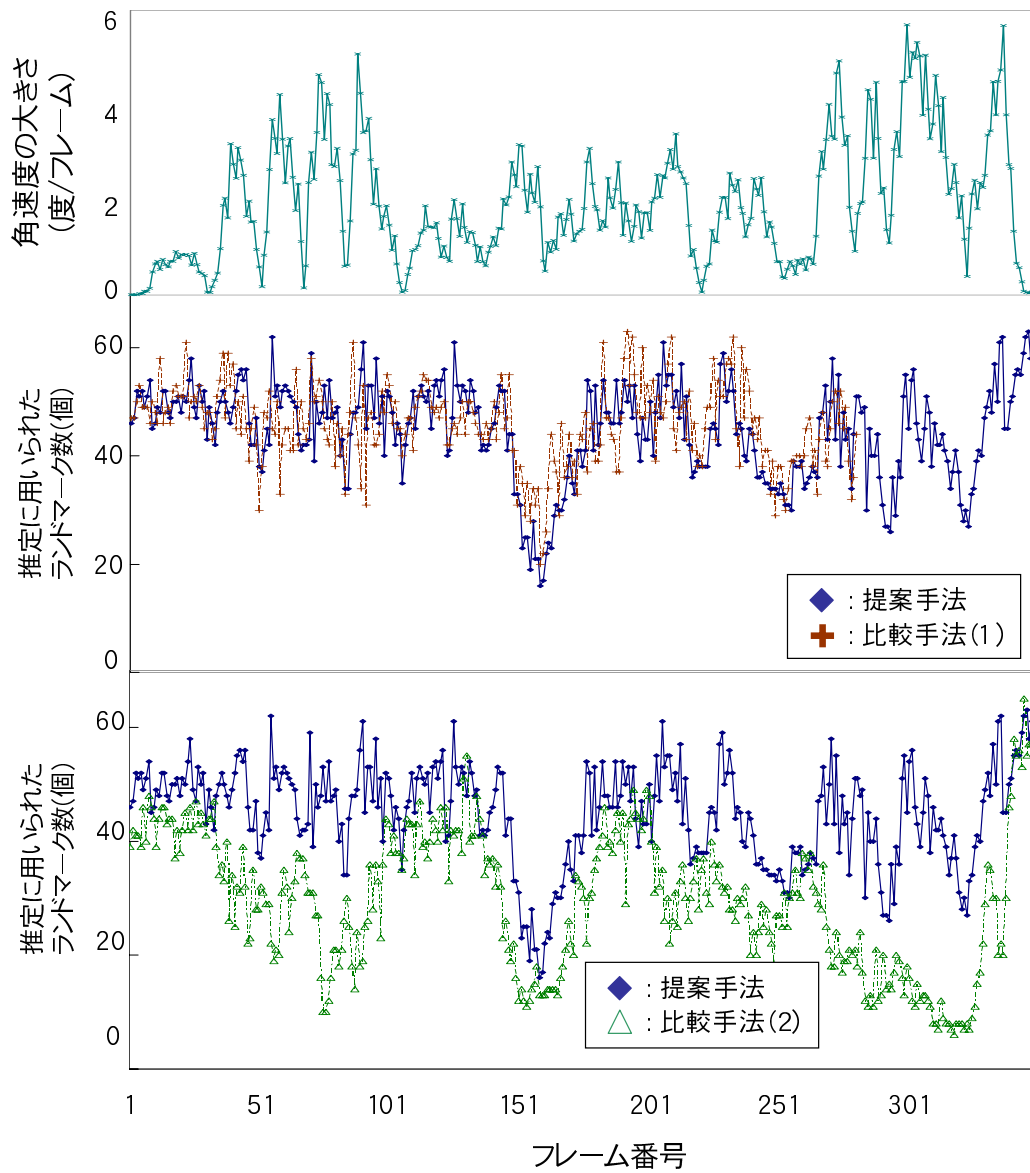


図 19 角速度の大きさと対応付けられたランドマーク数 (入力 1)

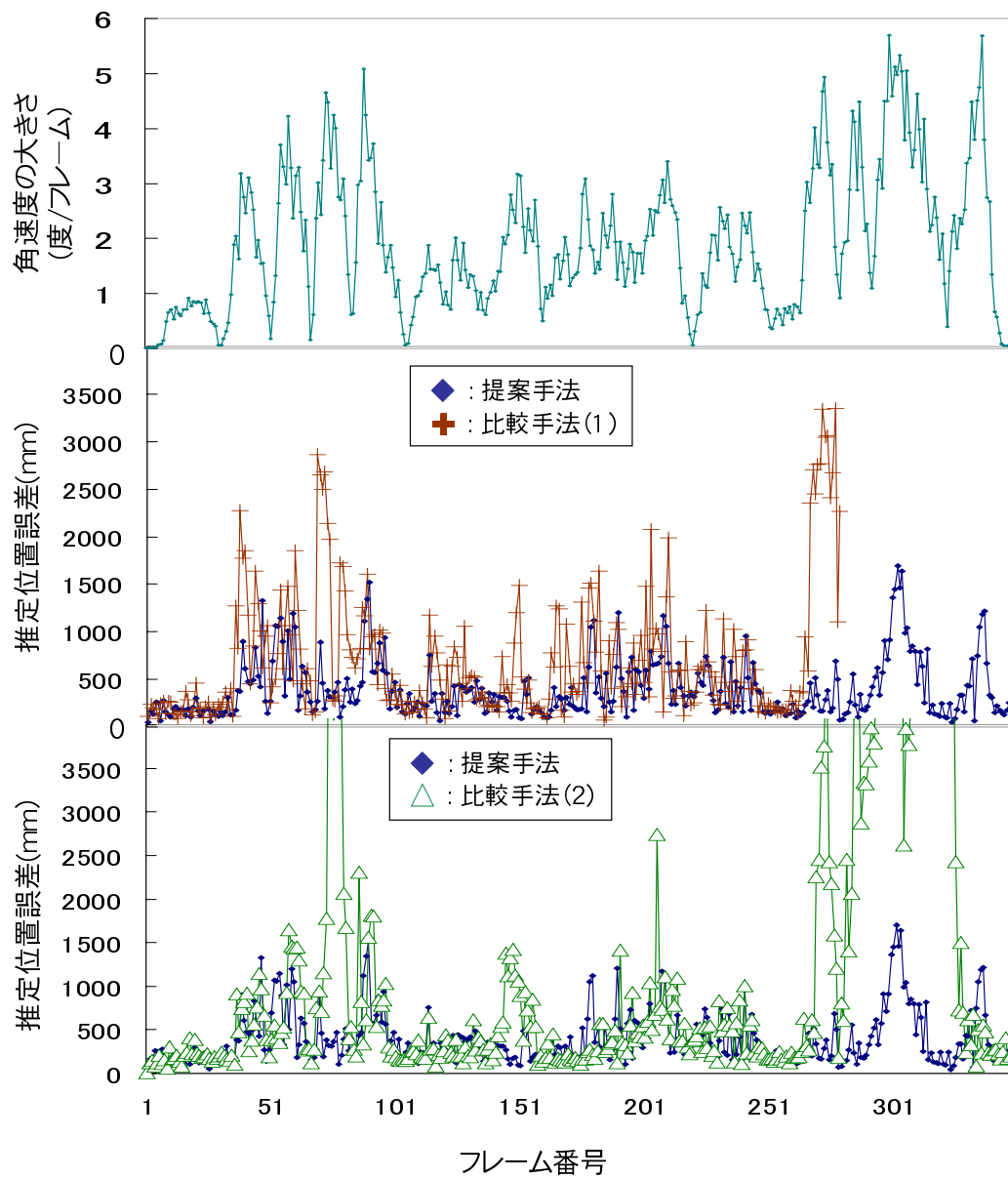


図 20 角速度の大きさと推定位置誤差 (入力 1)

表 6 平均推定誤差と標準偏差 (入力 1)

| | 平均位置誤差 (mm) | 標準偏差 (mm) |
|----------|-------------|-----------|
| 提案手法 | 394 | 309 |
| 比較手法 (1) | 706 | 607 |
| 比較手法 (2) | 1239 | 1960 |

表 7 平均推定誤差 (入力 2)

| | 平均位置誤差 (mm) | 平均姿勢誤差 (度) |
|----------|-------------|------------|
| 提案手法 | 459 | 0.36 |
| 比較手法 (1) | 1406 | 1.42 |
| 比較手法 (2) | 571 | 0.39 |

が動く方向と match move の結果を一部拡大したものを図 22 に示す。入力 1 に対する実験と同様に、CG で重畳したワイヤーフレームが入力画像と幾何学的に正しく合成されており、カメラ位置・姿勢推定に大きな誤差が含まれていないことが分かる。各フレームにおける角速度と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を図 23 に、各手法の平均推定誤差を表 7 に示す。比較手法 (1) 及び (2) では、それぞれ 44 フレーム、84 フレームにおいて、カメラの位置及び姿勢の推定が途中で失敗した。これに対して、提案手法では、最終フレームまでカメラ位置姿勢を推定することが出来た。図 23 から、比較手法 (1) では、前述で述べた実験と同様に、大きな角速度が発生した直後に推定処理の失敗が発生し、表 7 から位置及び姿勢の推定誤差も増大していることが分かる。比較手法 (2) では、表 7 から提案手法と同程度の精度が得られているが、比較手法 (1) と同様に、大きな角速度が発生した直後に推定処理の失敗が見られる。今回の実験では、モーションブラーが生じたフレームでは推定誤差が検証できないため、表 7 はブラーの発生していないフレームにおける平均推定誤差を表しているが、比較手法 (1)、(2) における推定の失敗は、いずれもモーションブラーが生じたフレームにおいて発生しており、推定に用いられるランドマーク数が減少しているため、これらのフレーム付近において比較手法の推定誤差が増大している可能性がある。

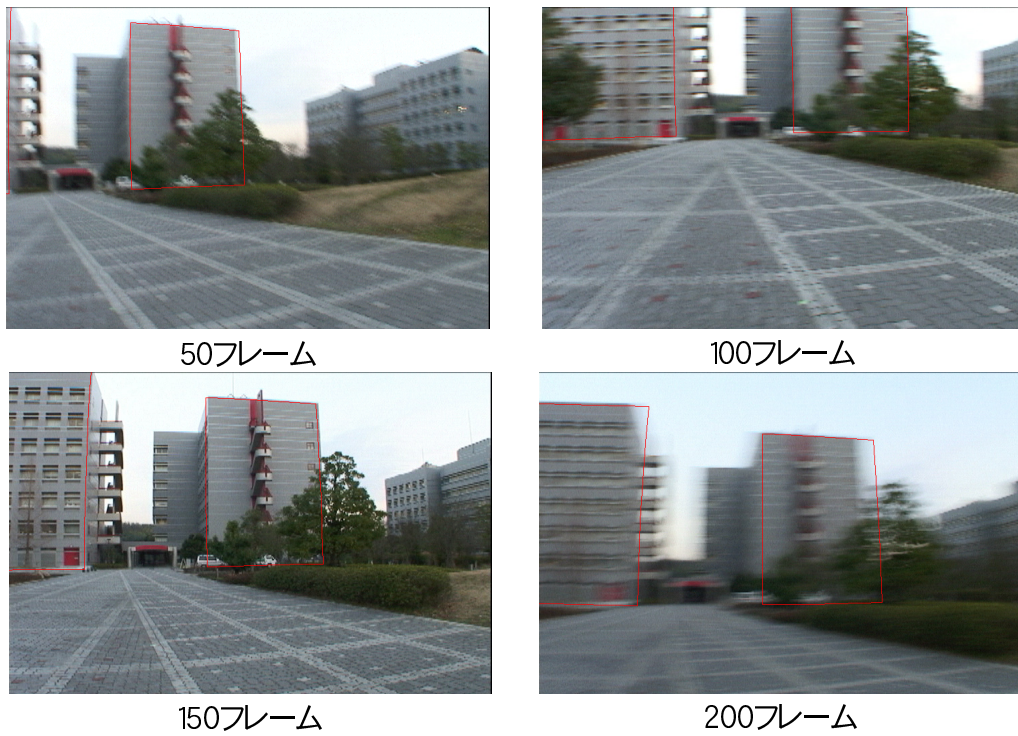


図 21 入力 2 におけるカメラ位置・姿勢推定結果を用いた match move

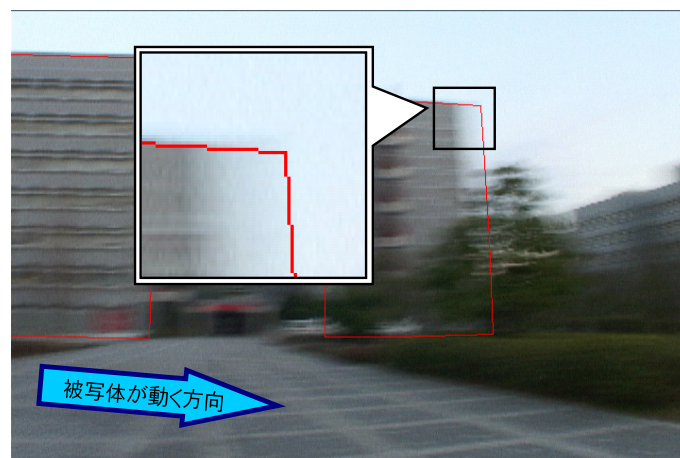


図 22 入力 2 における match move の拡大図と被写体が動く方向

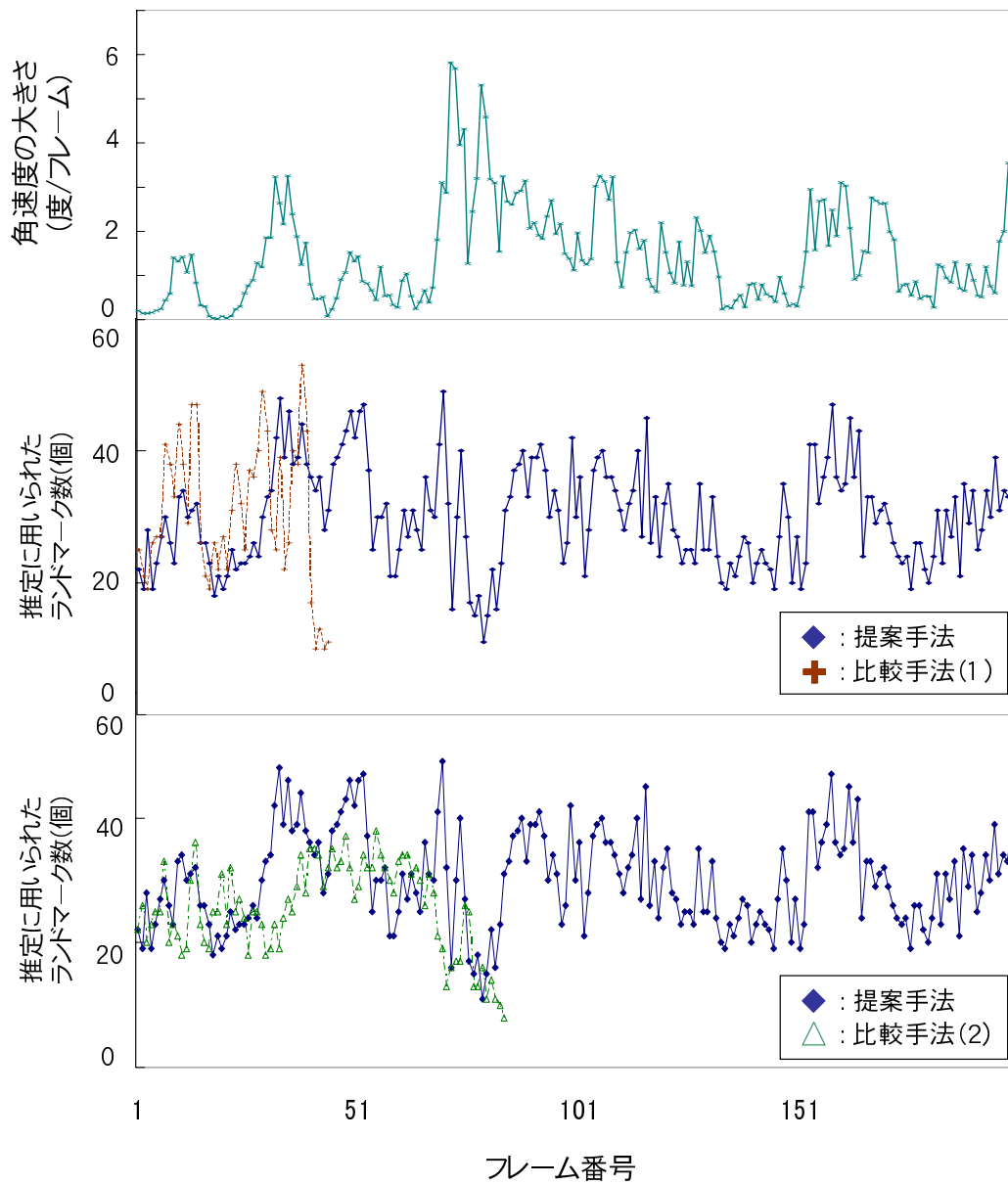


図 23 角速度の大きさと対応付けられたランドマーク数 (入力 2)

以上のことから，位置及び姿勢の推定精度，推定のロバスト性において，提案手法の有効性を確認した．今回の実験において，提案手法における1枚の入力に対する処理時間は平均16.8秒であり，主にステップ(c)におけるモーションブラーの再現に時間がかかった．自然特徴点ランドマークデータベース登録時におけるモーションブラーの影響を受けやすいランドマークの識別や，モーションブラーの再現におけるGPUの使用といった処理の高速化により，カメラ位置・姿勢推定のリアルタイム化が実現出来れば，本手法は，屋内などの環境中に光が少なくモーションブラーが生じやすい状況や，ウェアラブルコンピュータを用いた拡張現実感やロボットナビゲーションなどの高速な回転を伴う状況においてカメラの位置及び姿勢をロバストに推定することが出来るため，それらのシステムにおいて厳密な位置あわせを必要としないナビゲーションなどのアプリケーションに用いることが可能であると考えられる．

5. まとめと今後の課題

本論文では，自然特徴点ランドマークデータベースと姿勢センサによる角速度情報，動画像を入力とし，データベースと入力画像中の自然特徴点を対応付けることにより，カメラの位置及び姿勢を推定する手法を提案した．カメラ位置及び姿勢推定に用いた自然特徴点ランドマークデータベースは，全方位動画像を用いて環境中の多数の自然特徴点をランドマークとして半自動的にデータベース内に登録することで構築した．また，入力画像とランドマークの対応付けのために，撮影時のカメラの位置及び姿勢を考慮した多重スケール画像テンプレートを用いた．提案手法では，推定に用いるランドマークの効果的な選択やデータベース中の画像テンプレートに対するモーションブラーの再現により，物理的なインフラを用いることなく，高速な回転を伴うカメラの絶対的な位置及び姿勢をロバストに推定することができる．

実験では，並進移動するカメラに高速な回転を与えて取得した動画像からカメラの位置及び姿勢を推定し，真値と比較することによって，カメラの位置及び姿勢の推定精度を検証した．評価実験の結果，精度は従来手法と同程度であったが，高速な回転に対するロバスト性が向上したことを確認した．

今後の課題として，処理速度の向上による逐次推定のリアルタイム化が挙げられる．この課題を解決することができれば，広範囲な環境において高速な回転を伴う拡張現実感やロボットナビゲーションなどに利用できると考えられる．

謝辞

本研究を進めるにあたり，終始暖かく御指導を頂いた視覚情報メディア講座横矢直和教授に厚く御礼申し上げます．本学での研究におけるあらゆる面で，多大な御助言，御協力を賜りましたことを深く感謝いたします．

副指導教官として御助言を頂いた像情報処理学講座千原國宏教授，ならびに視覚情報メディア講座山澤一誠助教授に深く感謝致します．

公私にわたり大変御世話になった視覚情報メディア講座神原誠之助手に心より感謝申し上げます．

本研究の遂行にあたり，様々な相談に応じて頂き，終始有意義な御助言と御鞭撻を頂いた視覚情報メディア講座佐藤智和助手に深く感謝いたします．

本研究にあたってご指導を頂きました池田聖特任助手に厚く御礼申し上げます．研究方法の初歩から，論文の執筆に至るまでご指導いただきましたことを感謝いたします．

さらに，日々の研究活動を支えて下さった，視覚情報メディア講座守屋知代女史に感謝いたします．

山中一樹君を始め視覚情報メディア講座の学生諸兄，並びに佐賀野正行君を始め像情報処理学講座の学生諸兄には常に刺激的な議論を頂き，精神的にも支えられました．本当に有り難うございました．学生生活で苦楽をともにした同窓生の皆さんの精神的な支えが無ければ，この論文を完成させる事は出来ませんでした．感謝の念にたえません．

最後になりましたが，終始御援助を賜り，温かく見守ってくださった両親に厚く感謝の意を表します．

付録

A. カメラと姿勢センサの同期とキャリブレーション

本項では、カメラによりデータと姿勢センサから得られるデータの同期手法、また、両センサのキャリブレーション手法について述べる。本研究では高速な回転に対するロバスト性を高めるための最初の処理として、姿勢センサにより暫定的なカメラ姿勢 \tilde{R}_c^g の算出を行う。しかし、ビデオカメラによる画像の取得レートと姿勢センサによる角速度情報の取得レートが異なるため、まず両センサを同期させる。次に、姿勢センサから得られるセンサ情報は姿勢センサの座標系に対する角度の変化量であるので、両センサの座標系における軸の向きを一致させるため、カメラ座標系から姿勢センサの座標系への回転行列 R^{cg} を求める。以下に、同期手法とキャリブレーション手法の詳細について述べる。

A.1 同期手法

本手法では、姿勢センサから得た角速度情報 ω^g と画像マーカを回転のみを伴うカメラで撮影した動画像から算出された角速度情報 ω^c を用い、以下の手順により両センサの角速度をフィッティングしてデータを対応付け、同期させる。各手順におけるセンサ情報の変位を図 24 として示す。

1. 画像マーカ撮影時における姿勢センサの開始データ、終了データの決定

姿勢センサから得られた角速度情報において、カメラが画像マーカを撮影し始める際のデータと撮影を終了する際のデータを暫定的に決定する。以降、各データを開始データ、終了データとする。

2. 1 フレームあたりの角速度情報の算出

開始データと終了データ、カメラが画像マーカを撮影したフレーム数から、カメラと姿勢センサの計測レートの比率を算出する。得られた計測レートの比率と角速度情報 ω^g から、1 フレームあたりの角速度情報 $\dot{\omega}^g$ を算出する。

3. 両センサにおける角速度の差分の算出

画像マーカを撮影した全フレームにおいて，両センサの角速度の差分 $D = \sum_i^N \sqrt{(\dot{\omega}_i^g - \dot{\omega}_i^c)^2}$ を算出する．ただし， i はフレーム， N は画像マーカを撮影したフレーム数を示す．

以上の処理を，開始データ，終了データを変化しつつ繰り返し， D が最小となる開始データ，終了データから画像フレームと姿勢センサのデータを同期させる．

A.2 カメラと姿勢センサのキャリブレーション手法

姿勢センサとカメラのキャリブレーション手法について説明する．本手法では，両センサの回転軸を一致させるため，以下の2項目の投影誤差 N が最小となるカメラ座標系から姿勢センサの座標系への変換行列 R^{cg} を算出する．

- ・ 基準点 j のカメラ座標系における位置と i フレームにおける暫定的なカメラの位置 \tilde{P}_{i-1}^c 及び姿勢 \tilde{R}_{i-1}^c ，回転行列 R^{cg} から算出した，基準点 j の投影位置 $\tilde{c}_{(j)i}$
- ・ 基準点 j の画像上での検出位置 $d_{(j)i}$

これらの投影誤差を $N_{(j)} = (\tilde{c}_{(j)i,x} - d_{(j)i,x})^2 + (\tilde{c}_{(j)i,y} - d_{(j)i,y})^2$ として定義し， $\sum_{j=0}^D N_{(j)}$ が最小となる R^{cg} を算出する．ただし， x, y は入力画像の座標， D は基準点の数を表す．

なお，図 26 に示すようなキャリブレーション時における投影誤差のヒストグラムから，入力画像とランドマークの対応付け処理(3.3.4)において用いた，姿勢センサの精度から設定される探索範囲の大きさ δ' を算出する．具体的には，設定した割合を超える度数の階級を δ' に設定する．

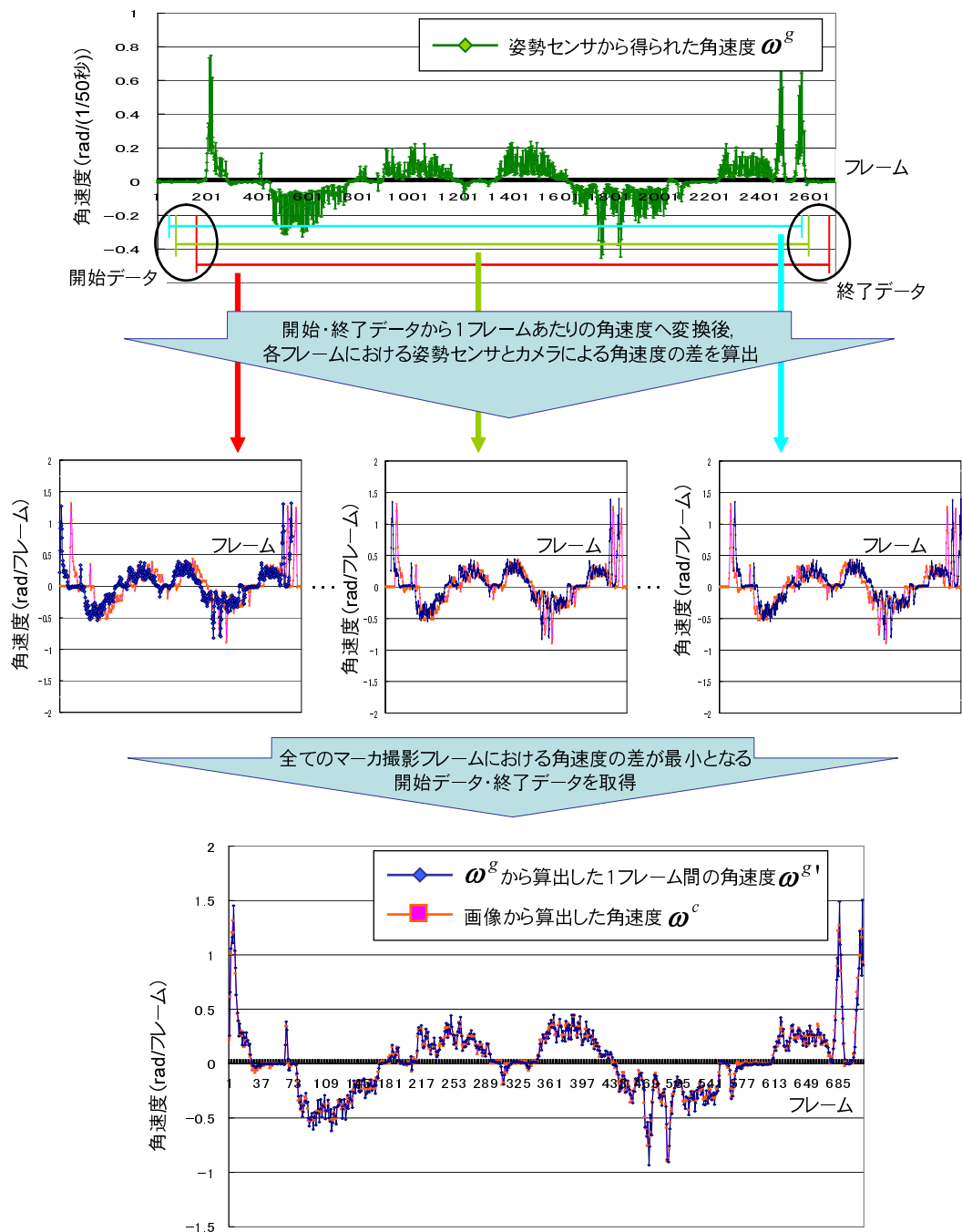


図 24 角速度情報 ω^g , ω^c のデータフィッティング

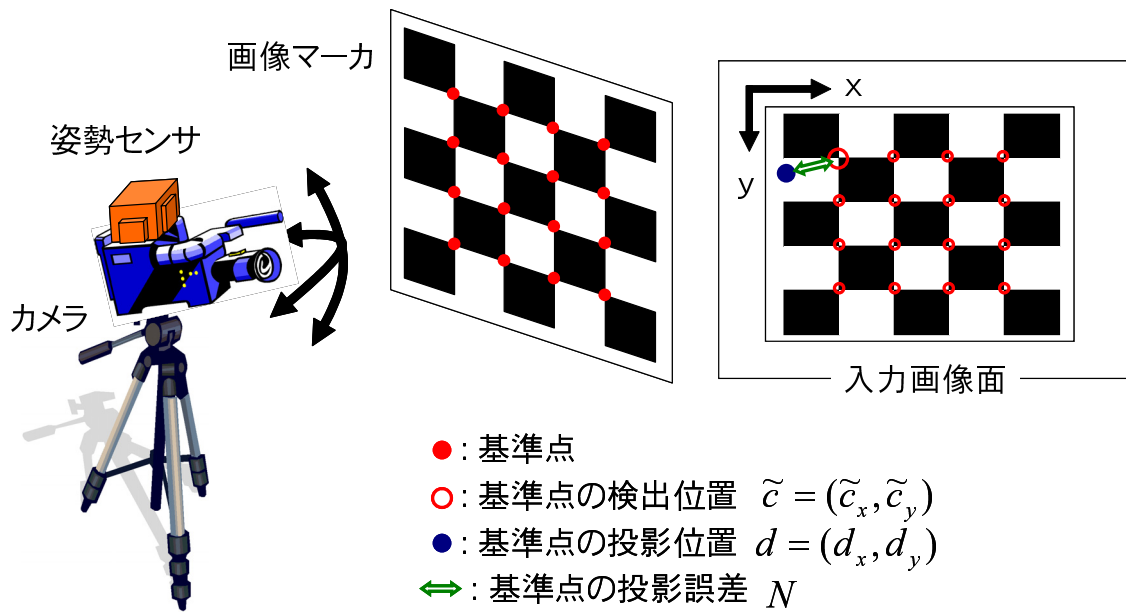


図 25 キャリブレーションデータの取得

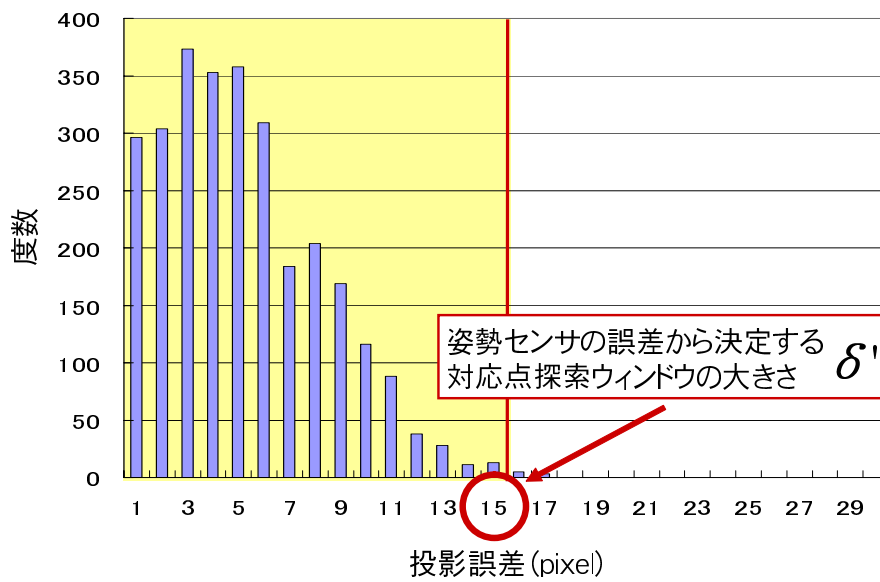


図 26 マーカの投影誤差のヒストグラム

B. カメラの投影モデル

ピンホールカメラモデルとして仮定した，本手法で用いたカメラの投影モデルについて述べる．点 P における世界座標 (x^W, y^W, z^W) と画像座標 (x^S, y^S) 間の関係を示すため，世界座標系とカメラ座標系，スクリーン座標系間における座標の変換を以下のように定義する．ただし，今回はレンズ歪みによる投影位置の変化は考えないものとする．

環境中における座標系と点を以下のように定義する．図 27 に示すように基準となる世界座標系 $O^W - X^W - Y^W - Z^W$ における点 P の座標を (x^W, y^W, z^W) と表し，カメラ座標系 $O^C - X^C - Y^C - Z^C$ における点 P の座標を (x^C, y^C, z^C) と表す．ただし， O^C はレンズ中心（ピンホール）を表し， Z^C 軸はレンズの光軸に一致するように設定する．

各座標系における点の座標を以下のように定義する． $X^C - Y^C$ 平面に平行で Z^C 座標が f の位置に画像中心が O^S ，座標軸が $X^S - Y^S$ の画像平面を考える．理想的なピンホールカメラの場合には，点 P の座標 (x^C, y^C, z^C) は画像平面状で (x^S, y^S) と表される．

1. (x^W, y^W, z^W) から (x^C, y^C, z^C) への変換

回転行列 \mathbf{R}^{WC} と平行移動ベクトル \mathbf{P}^{WC} を用いて表す．

$$\begin{pmatrix} x^C \\ y^C \\ z^C \end{pmatrix} = \mathbf{R}^{WC} \begin{pmatrix} x^W \\ y^W \\ z^W \end{pmatrix} + \mathbf{P}^{WC} \quad (10)$$

ただし，

$$\mathbf{R} = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix}, \mathbf{P} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} \quad (11)$$

2. 透視変換による (x^C, y^C, z^C) から (x^S, y^S) への変換

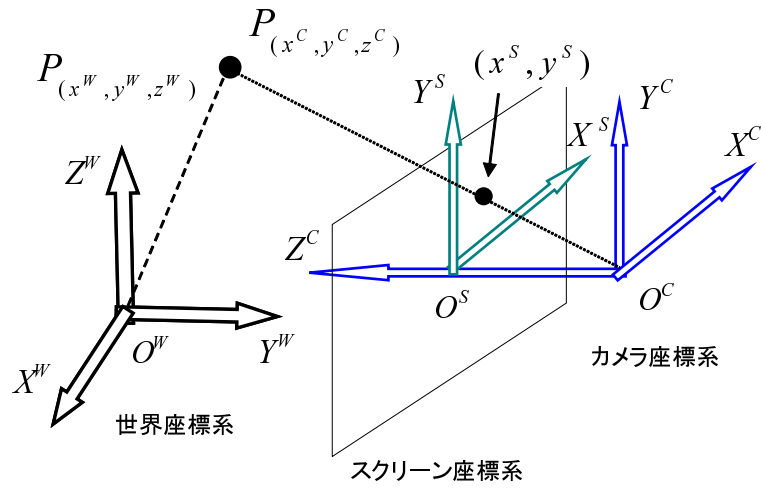


図 27 世界座標系とスクリーン座標系の関係

焦点距離 f を用いて表す .

$$x^s = f \frac{x^c}{z^c} \quad y^s = f \frac{y^c}{z^c} \quad (12)$$

参考文献

- [1] S. Feiner, B. MacIntyre, T. Höllerer and A. Webster: “A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment,” Proc. 1st IEEE Int. Symp. on Wearable Computers, pp. 74–81, 1997.
- [2] P. Daehne and J. Karigiannis: “Archeoguide: System architecture of a mobile outdoor augmented reality system,” Proc. 1st IEEE Int. Symp. on Mixed and Augmented Reality, pp. 263–264, 2002.
- [3] 松本, 稲葉, 井上: “ビューベースアプローチに基づく移動ロボットナビゲーション”, 日本ロボット学会誌, Vol.20, No.5, pp. 506–514, 2002.
- [4] 石岡, 矢晃, 油田: “屋内外に渡って人間を誘導するキャンパスガイドロボットの開発”, 日本機械学会ロボティクス・メカトロニクス講演会, pp. 381–395, 2005.
- [5] A. Ward, A. Jones and A. Hopper: “A new location technique for the active office,” IEEE Personal Communications, Vol.4, No.5, pp. 42–47, 1997.
- [6] A. Harter, A. Hopper, P. Steggles, A. Ward and P. Webster: “The anatomy of a context-aware application,” Proc. 5th IEEE Int. Conf. on Mobile Computing and Networking, pp. 59–68, 1999.
- [7] InterSense Inc.: “Is-900,” <http://www.isense.com/products/prec/is900/>.
- [8] P. Bahl and V. N. Padmanabhan: “Radar: An in-building rf-based user location and tracking system.,” Proc. 19th IEEE Conf. on Communications Society, pp. 775–784, 2000.
- [9] G. Welch, G. Bishop, L. Vicci, S. Brumback, K. Keller and D. Colucci: “High-performance wide-area optical tracking -the hiball tracking system,”

- Presence: Teleoperators and Virtual Environments, Vol.10, No.1, pp. 1–21, 2001.
- [10] D. Hallaway, T. Hollerer and S. Feiner: “Coarse, inexpensive, infrared tracking for wearable computing,” Proc. 7th IEEE Int. Symp. on Wearable Computers, pp. 69–78, 2003.
- [11] 加藤, M. Billinghamurst, 浅野, 橘: “マーカー追跡に基づく拡張現実感システムとそのキャリブレーション”, 日本バーチャルリアリティ学会論文誌, Vol.4, No.4, pp. 607–616, 1999.
- [12] 羽原, 町田, 清川, 竹村: “ウェアラブルPCのための画像マーカを用いた屋内位置検出手法とその評価”, 第7回日本VR学会大会論文集, Vol.102, No.216, pp. 65–70, 2002.
- [13] L. Naimark and E. Foxlin: “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker,” Proc. 1st IEEE Int. Symp. on Mixed and Augmented Reality, pp. 27–36, 2002.
- [14] 中里, 神原, 横矢: “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.
- [15] Y. Cho and U. Neumann: “Multi-ring color fiducial systems for scalable fiducial tracking augmented reality,” Proc. IEEE Int. Symp. on Virtual Reality, pp. 212–215, 1998.
- [16] J. Rekimoto: “Matrix: A realtime object identification and registration method for augmented reality,” Proc. 3rd on Asia Pacific Computer Human Interaction, pp. 63–68, 1998.
- [17] L. Chai, K. Nguyen, B. Hoff and T. Vincent: “An adaptive estimator for registration in augmented reality,” Proc. of IEEE Int. Workshop on Augmented Reality, pp. 23–32, 1999.

- [18] S. You and U. Neumann: “Fusion of vision and gyro tracking for robust augmented reality registration,” Proc. IEEE Conference on Virtual Reality, pp. 71–78, 2001.
- [19] 山本, 亀田, 大田: “複数の定点カメラを利用した屋外における複合現実型情報提示”, 第5回計測自動制御学会システムインテグレーション部門講演会, 2004.
- [20] 小竹, 佐藤, 内山, 山本: “傾斜角拘束を利用したハイブリッド位置合わせ手法”, 画像の認識・理解シンポジウム, pp. 241–248, 2006.
- [21] 佐藤, 内山, 山本: “UG+B 法: 主観及び客観視点カメラと姿勢センサを用いた位置合わせ手法”, 日本バーチャルリアリティ学会論文誌, Vol.10, No.3, pp. 391–400, 2005.
- [22] M. Pressigout and E. Marchand: “Real-time 3d model-based tracking: Combining edge and texture information,” Proc. IEEE Int. Conf. on Robotics and Automation, pp. 2726–2731, 2006.
- [23] L. Vacchetti, V. Lepetit and P. Fua: “Combining edge and texture information for real-time accurate 3d camera tracking,” Proc. 3rd IEEE Int. Symp. on Mixed and Augmented Reality, pp. 48–56, 2004.
- [24] E. Rosten and T. Drummond: “Fusing points and lines for high performance real-time tracking,” Proc. 10th IEEE Int. Conf. on Computer Vision, pp. 1508–1515, 2005.
- [25] R. Kumar and A. R. Hanson: “Robust methods for estimating pose and a sensitivity analysis,” Computer Vision, Graphics, and Image Processing. Image Understanding, Vol.21, No.3, pp. 313–342, 1994.
- [26] H. Najafi, Y. Genc and N. Navab: “Fusion of 3d and appearance models for fast object detection and pose estimation,” 7th Asian Conf. on Computer Vision, pp. 415–426, 2006.

- [27] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: “Fully automated and stable registration for augmented reality applications,” Proc. 2nd IEEE Int. Symp. on Mixed and Augmented Reality, pp. 93–102, 2003.
- [28] I. Gordon and D. G. Lowe: “Scene modeling, recognition and tracking with invariant image features,” Proc. 3rd IEEE Int. Symp. on Mixed and Augmented Reality, pp. 110–119, 2004.
- [29] 大江, 佐藤, 横矢: “幾何学的位置合わせのための自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定”, 日本バーチャルリアリティ学会論文誌, Vol.10, No.3, pp. 285–294, 2005.
- [30] B. Jiang, U. Neumann and S. You: “A robust hybrid tracking system for outdoor augmented reality,” Proc. IEEE Conf. on Virtual Reality, pp. 3–10, 2004.
- [31] G. Reitmayr and T. W. Drummond: “Going out: Robust model-based tracking for outdoor augmented reality,” Proc. 5th IEEE Int. Symp. on Mixed and Augmented Reality, pp. 109–118, 2006.
- [32] G. S. W. Klein and T. W. Drummond: “Tightly integrated sensor fusion for robust visual tracking,” Image and Vision Computing, Vol.22, pp. 769–776, 2004.
- [33] 神原, 横矢: “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, pp. 933–938, 2005.
- [34] S. Güven and S. Feiner: “Authoring 3D hypermedia for wearable augmented and virtual reality,” Proc. 7th IEEE Int. Symp. on Wearable Computers, pp. 118–126, 2003.
- [35] C. Randell, C. Djiallis and H. Muller: “Personal position measurement using dead reckoning,” Proc. 7th IEEE Int. Symp. on Wearable Computers,

pp. 166–175, 2003.

- [36] J. Kim, H. Jang, D. Hwang and C. Park: “A step, stride and heading determination for the pedestrian navigation system,” *Journal of Global Positioning Systems*, No.1-2, pp. 273–279, 2004.
- [37] T. Sato, S. Ikeda and N. Yokoya: “Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system,” *Proc. European Conf. on Computer Vision*, Vol.2, pp. 326–340, 2004.
- [38] R. Cipolla, D. Robertson and E. Boyer: “Photobuilder - 3D models of architectural scenes from uncalibrated images,” *Proc. IEEE Int. Conf. on Multimedia Computing and Systems*, vol.1, pp. 25–31, 1999.
- [39] 出口: “射影幾何学による PnP カメラ補正問題の統一的解法”, *情報シンポジウム*, Vol.90, pp. 41–50, 1990.
- [40] M. A. Fischler and R. C. Bolles: “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, Vol.24, pp. 381–395, 1981.