

# 画像特徴点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

大江統子<sup>\*†</sup> 佐藤 智和<sup>†</sup> 横矢 直和<sup>†</sup>

† 奈良先端科学技術大学院大学 情報科学研究科 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{tomoka-s,yokoya}@is.naist.jp

あらまし 近年、動画像からのカメラの位置・姿勢推定手法が盛んに研究されており、拡張現実感やロボットナビゲーションに応用されている。これらの分野では、カメラの絶対位置・姿勢が必要とされるため、GPS や磁気センサを用いる手法や、カメラからの入力画像を用いる手法などが提案されている。しかし、センサ類を用いる手法ではカメラとセンサの同期をとることが難しい、利用可能な環境が制限されるなどの問題がある。また、カメラからの入力画像を用いる手法には、広範囲に人工的なランドマークを多数配置しなければカメラ位置の推定誤差が累積するという問題や、広範囲な屋外環境内での利用は難しいという問題がある。本研究では事前に撮影された動画像から検出された自然特徴点をランドマークとして用いることで、センサ類や人工的なマーカを必要としない、広範囲な環境におけるカメラ位置・姿勢推定手法を提案する。

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, 拡張現実感

## Camera Position and Posture Estimation Based on Feature Landmark Database

Motoko OE<sup>\*†</sup>, Tomokazu SATO<sup>†</sup>, and Naokazu YOKOYA<sup>†</sup>

† Nara Institute of Science and Technology 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan

E-mail: †{tomoka-s,yokoya}@is.naist.jp

**Abstract** Estimating camera position and posture can be applied to the fields of augmented reality and robot navigation. In these fields, to obtain absolute position and posture of the camera, sensor-based methods using GPS and magnetic sensors and vision-based methods using input images from the camera have been investigated. However, sensor-based methods are difficult to synchronize the camera and sensors accurately, and usable environments are limited according to selection of sensors. On the other hand, vision-based methods needs to allocate many artificial markers otherwise an estimation error will accumulate. Thus, it is difficult to use such methods in large and natural environments. This paper proposes a vision-based camera position and posture estimation method for large environments, which does not require sensors and artificial markers by detecting natural feature points from image sequences taken beforehand and using them as landmarks.

**Key words** camera position and posture estimation, natural features, landmark database, augmented reality

### 1. はじめに

動画像を用いたカメラ位置・姿勢推定は拡張現実感やロボットナビゲーションなどの分野に応用されている。これらの分野では、カメラの絶対位置・姿勢が必要とされ、GPS や磁気センサを用いる手法 [1][2] や、カメラからの入力画像を用いる手

法 [3][4][5][6][7][8][9] などが研究されている。

センサ類を用いる手法では、一般的に GPS やジャイロセンサが用いられており、広範囲な屋外で利用可能であるが、カメラとセンサの同期をとることが難しいという問題がある。また、センサの組み合わせによって利用できる環境が大きく異なることから、様々な環境でカメラ位置・姿勢を推定するためには、多数のセンサを組み合わせなければならず、システムが複雑になるという問題がある。一方、カメラからの入力画像を用いる

\* 現在, 日本 IBM 株式会社

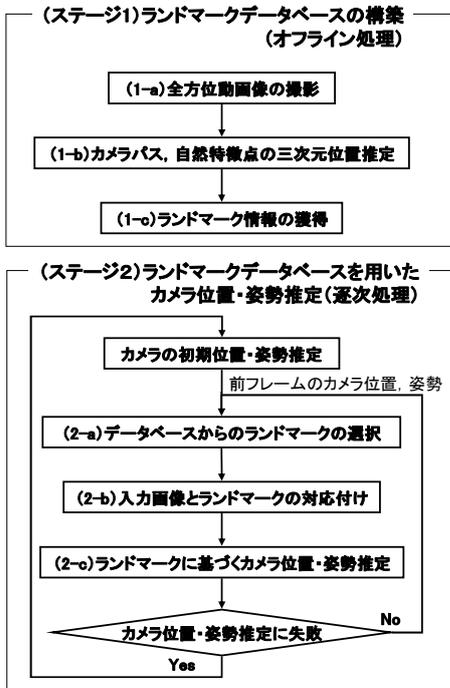


図 1 提案手法のフローチャート

手法としては, ARToolkit [3] などの人工的なマーカを用いる手法 [4] [7] や入力画像上の自然特徴点を用いる手法 [5], それらの手法のハイブリッド [6] などが従来研究されてきた. これらの手法では, マーカに基づいて位置・姿勢の推定を行うため, 人工的なマーカを広範囲に多数配置する必要がある, マーカの設置に多くの人的コストがかかるという問題や, マーカによって景観が損なわれるという問題がある.

このような問題を解決するために, 近年, 物体の三次元モデルを人工的なマーカの代わりに用いる手法 [8] [9] が提案され, 注目されている. Lepetit らは対象とする物体や環境の三次元モデルと特徴点情報を用いて三次元モデルと入力画像の対応関係を求めることで, カメラ位置・姿勢を推定する手法を提案している [8]. この手法では, 三次元モデルと前フレームのカメラ位置・姿勢情報を用いることで, 物体の見え方の変化やオクルージョンにロバストなカメラ位置・姿勢推定が可能であるが, 対象となる物体の三次元モデルを事前に人手で作成しておく必要があるため, 自然物を対象とした場合や広範囲な環境での利用は難しいという問題がある. また, Gordon らは小物体を撮影した複数枚の画像からセルフキャリブレーションによって特徴点の三次元復元を行い, スケールや回転に不変な SIFT 特徴量を用いて対応付けを行う手法を提案している [9]. この手法では特別な事前知識を必要としないが, 広範囲な環境への適用を考えた場合, 三次元復元の推定誤差が累積するという問題があり, 小物体, 小領域への適用にとどまっている.

本研究では事前に撮影された全方位動画画像から検出された自然特徴点をランドマークとして用いることで, センサ類や人工的なマーカを必要としない, 広範囲な環境におけるカメラ位置・姿勢推定手法を提案する. 提案手法は図 1 に示すように, オフライン処理によるランドマークデータベースの構築 (ステージ

1), とランドマークデータベースを用いたカメラ位置・姿勢の逐次推定 (ステージ 2), の二つのステージによって構成される. ステージ 1 では, 事前に環境内を全方位動画画像として撮影し, shape from motion によって環境の粗な三次元復元を行う. そして, 動画画像から検出された自然特徴点の画像テンプレートと, 三次元復元によって得られる自然特徴点の三次元位置を, ランドマーク情報としてランドマークデータベースに登録する. ステージ 2 では, これらのランドマークと入力画像上の自然特徴点との対応付けを行うことで, 累積誤差を含まないカメラ位置・姿勢を逐次的に推定する.

以降, 2 章では, オフライン処理で作成するランドマークデータベースの構成要素とランドマーク情報の獲得方法について述べる. 3 章では, 2 章で構築したランドマークデータベースを用いた逐次的なカメラ位置・姿勢推定手法について述べる. 4 章では, 実際に全方位カメラによって撮影された屋外全方位動画画像を用いた実験について報告する. 最後に, 5 章でまとめと今後の課題について述べる.

## 2. ランドマークデータベースの構築

本章では, 次章で述べるカメラ位置・姿勢の推定に必要なランドマークデータベースの構成方法 (ステージ 1) について詳述する. 本手法では, 環境内を撮影した動画画像から自然特徴点を検出し, ランドマークとして用いる. そこで, 本ステージでは, 図 1 に示すように, まず多数のランドマーク情報を効率的に取得するために, 広範囲を一度に撮影可能な全方位カメラを用いて環境内を移動しながら撮影する (1-a). 次に, 全方位動画画像上の自然特徴点追跡による三次元復元を行い, 全方位動画画像から検出された自然特徴点の三次元位置と動画画像のカメラパスを推定する. これにより, 自然特徴点の三次元位置, 自然特徴点撮影時のカメラ位置・姿勢情報を取得する (1-b). 最後に, 撮影された全方位動画画像とフレームごとのカメラ位置・姿勢, 自然特徴点の三次元位置情報に基づき, それぞれの自然特徴点ごとにランドマーク情報を自動的に取得し, ランドマークデータベースを構築する (1-c). 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, 次に各ランドマークの情報を取得するための手順について述べる.

### 2.1 ランドマークデータベースの構成要素

ランドマークデータベースは図 2 のように, 複数のランドマークによって構成される. これらのランドマークは次章で述べるカメラ位置・姿勢推定処理において, 入力画像中の特徴点と対応付けられ, カメラ位置・姿勢の推定 (ステージ 2) で用いられる. それぞれのランドマークは, 1 個の (1) ランドマークの三次元位置, 複数の (2) 撮影地点ごとの情報, から構成される. また, 撮影地点ごとの情報は, (A) 撮影時の全方位カメラの位置, (B) 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, (C) 画像テンプレートの法線ベクトル, (D) 画像テンプレートの基本スケールから成る. 以下にそれぞれの要素について詳述する.

#### (1) ランドマークの三次元位置

本研究では次章で述べるステージ 2 において, ランドマーク

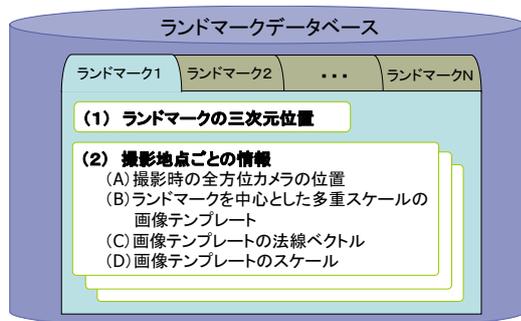


図2 ランドマークデータベースの構成要素

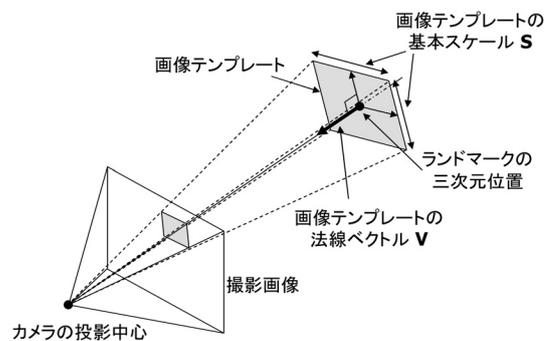


図3 ランドマークとその画像テンプレート

の三次元位置と画像上の二次元座標の対応からカメラパラメータを推定するため、ランドマークの三次元位置が必要となる。ランドマークの三次元位置は、次節で述べる環境の三次元復元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。世界座標系はX軸、Y軸が実環境における地面に対して水平、Z軸が地面に対して垂直な座標系であるとする。

## (2) 撮影地点ごとの情報

ランドマークと入力画像を対応付けるために必要な情報として、撮影地点ごとの情報を保持する。ランドマークの見え方は撮影地点によって異なるため、単一方向から撮影された画像のみでは高精度な対応付けは難しい。そこで、見え方の変化に対応するため、各ランドマークに対して複数の撮影地点ごとの情報を保持する。

(A) 撮影時の全方位カメラの位置: 入力画像と対応付けるランドマークを選択する際に用いる。全方位カメラの位置は世界座標系で保持されている。

(B) 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: 多重スケール画像テンプレートはステージ2において入力画像の中からのランドマークの探索に用いられる。これは、図3のように、世界座標系においてカメラの投影中心とランドマークを結ぶ直線に対して垂直な面に撮影画像を投影することで作成する。これにより、撮影カメラ姿勢やレンズ歪みの影響を受けない画像テンプレートを作成する。本研究では、図4のように、各撮影地点ごとに基本スケール（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）とその2倍、4倍のスケールで作成された多重スケールの画像テンプレートを保持する。また、作成する画像テンプレートの画素数は常に一定とし、スケールによって解像度が異なる画像テンプレートを作成する。

(C) 画像テンプレートの法線ベクトル: 図3に示すような、画像テンプレートの面の法線方向を表す。後述する画像テンプレートの選択や、入力画像の画像パターン作成に用いる。

(D) 画像テンプレートの基本スケール: 画像テンプレートのスケールとは、図3に示すような、画像テンプレートの実空間中での大きさであると定義する。画像テンプレートの基本スケールは、(B)で述べた多重スケールの画像テンプレートの中で、最小スケールの画像テンプレートの実空間中での大きさを表す。



図4 スケールの異なる画像テンプレート

本研究では、画像テンプレートの解像度がランドマーク作成時に撮影された全方位画像の解像度とほぼ等しくなるように基本スケールを決定する。

## 2.2 全方位カメラからの三次元復元によるランドマーク情報の獲得

本節では、全方位動画画像からの三次元復元を利用して、ランドマーク情報を獲得し、ランドマークデータベースを構築する方法について述べる。ランドマークデータベースの構成要素は先に図2に示した通りである。はじめに、全方位動画画像から環境の三次元復元を行うことによって、ランドマークの三次元位置(1)と撮影時の全方位カメラの位置・姿勢(A)を取得する。次に、三次元復元で得られた情報を用いて画像テンプレートの法線ベクトル(C)、画像テンプレートの基本スケール(D)を決定し、ランドマークを中心とした多重スケール画像テンプレート(B)を作成する。以下では、これらの作成方法について順に述べる。

### 2.2.1 全方位動画画像からの環境の三次元復元

ここでは、佐藤らの全方位マルチカメラシステムを用いた三次元復元手法[10]を用いて三次元位置が既知の少数の基準点とHarrisオペレータ[11]によって検出された多数の自然特徴点を全方位動画画像中で自動追跡し、カメラの外部パラメータを推定する。この手法では、まず基準点の三次元位置をトータルステーションと呼ばれる三次元測量機材を用いて計測し、少数のキーフレームの画像上で基準点を指定する。次に、全方位動画画像中の自然特徴点をフレーム毎に自動追跡し、動画全体での最適化処理を行うことで、カメラパラメータの累積的な推定誤差を最小化する。これにより、複雑で広範囲な環境を撮影した場合でも、全方位動画画像のカメラパスと自然特徴点の三次元位置を基準マーカによる絶対座標系で求めることができる。

### 2.2.2 ランドマーク情報の獲得

次に、三次元復元によって得られた情報を用いてランドマ

クの情報自動的に獲得し、ランドマークデータベースを構築する。2.2.1 項で述べた手法によって推定された全方位動画像のカメラパスと自然特徴点の三次元位置は、2.1 節で述べたランドマークの三次元位置 (1), ランドマーク撮影時の全方位カメラの位置 (A) として用いる。また、撮影された全方位画像とそのカメラ位置・姿勢, 自然特徴点のランドマークの三次元位置から画像テンプレートの法線ベクトル (C), 画像テンプレートの基本スケール (D) を決定し, それらを用いて多重スケール画像テンプレート (B) を作成する。以下に, それぞれの作成方法について述べる。

#### (1) ランドマークの三次元位置

本研究では, 環境の三次元復元によって三次元位置が推定された全方位画像上の自然特徴点をランドマークとして用いる。また, ランドマークの三次元位置として, 推定された自然特徴点の三次元位置を用いる。

#### (2) 撮影地点ごとの情報

(A) 撮影時の全方位カメラの位置: 2.2.1 項の手法によって復元されたカメラパラメータに基づき, ランドマーク撮影時の全方位動画像のカメラ位置を用いる。

(B) 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: 後述する画像テンプレートの法線ベクトルと基本スケールからランドマークを中心とした面を決定し, その面上に撮影画像を投影することで画像テンプレートを作成する。まず, 図 3 に示す法線ベクトル  $V$  と, 基本スケール  $S$  を求める。次に, 以下の条件を満たす正方形の面を, 基本スケールに対応する画像テンプレートの投影面として設定する。

- 中心がランドマークの三次元位置に一致
- 法線ベクトル  $V$  に対して垂直
- 世界座標系での大きさが  $S \times S$
- 世界座標系の  $XY$  平面と面の  $X$  軸が平行

画像テンプレートの画素数を  $N \times N$  個とした場合, この投影面を等間隔な格子で  $N \times N$  個に分割する。次に, 各格子点の三次元座標を撮影画像面上に投影し, 投影された画像上の画素値をその点の画素値とする。これを基本スケールの画像テンプレートとする。

同様に, スケールが 2 倍, 4 倍の画像テンプレートを作成し, 撮影地点ごとに保持する。しかし, このような手法で, スケールが 2 倍, 4 倍の画像テンプレートを作成すると, 撮影画像上で疎な点を用いるため, 投影された点の間に存在する画素の色情報は用いられない。従って, スケールが 2 倍, 4 倍の画像テンプレートを作成するときは, 撮影画像の代わりに撮影画像を  $1/2, 1/4$  に縮小した画像に三次元座標を投影し, その画素値を用いることで, 周辺画素の情報も考慮した色情報取得する。

(C) 画像テンプレートの法線ベクトル: 画像テンプレートの法線ベクトルはランドマークの三次元位置から撮影時の全方位カメラ位置への向きベクトルとする。

(D) 画像テンプレートの基本スケール: 本研究では, 画像テンプレートの 1 画素が, 撮影画像上の 1 画素に対応するように画像テンプレートの基本スケールを決定する。これにより, 作成

される画像テンプレートの解像度と撮影画像の解像度がほぼ等しくなるため, 必要以上に高解像な画像テンプレートが作成されることはない。

### 3. ランドマークデータベースを用いたカメラ位置・姿勢推定

本章では, 前章で述べた手法によって事前に作成されたランドマークデータベースを用いる, カメラ位置・姿勢推定手法について述べる。本手法のカメラ位置・姿勢推定処理の流れは先に示した図 1 の通りである。まず, 入力画像の初期カメラ位置・姿勢を推定する。ただし, 本稿では初期カメラ位置・姿勢は何らかの手法によって取得済みとする。次に, 前章で作成されたデータベースからランドマークを選択する (図 1 中 2-a)。ランドマークと入力画像の対応付けでは, まず, 入力画像から自然特徴点の検出を行い, 検出された自然特徴点とランドマークの画像テンプレートとのマッチングを行うことによって入力画像とランドマークの対応付けを行う (2-b)。最後に, それらの対応関係を用いて PnP 問題を解く事によってカメラ位置・姿勢推定を行う (2-c)。カメラ位置・姿勢推定に失敗していればカメラの初期位置・姿勢推定から, 成功していればランドマークの選択から処理を繰り返す。以下, それぞれの処理について述べる。

#### 3.1 ランドマークの選択

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選択する。また, ランドマークの見え方の変化に対応するため, 複数の画像テンプレートの中から入力画像と見え方が最も近いものを選択し, マッチングに使用する。本手法では, 対応付けに用いるランドマークとその画像テンプレートの組を以下の 4 つの条件によって選択する。ただし, ここでは, 現フレームの入力画像のカメラ位置・姿勢は未知であるため, 前フレームのカメラ位置・姿勢を代わりに用いる。

(条件 1) ランドマークが入力画像の視野範囲内に存在する: ランドマークの三次元位置を前フレームのカメラ位置・姿勢を用いて撮像面に投影する。これにより, 入力画像面上に投影されたランドマークのみを選択する。

(条件 2) ランドマーク撮影時のカメラ位置と入力画像撮影時のカメラ位置の距離が閾値以下である: ランドマークの撮影位置と前フレームのカメラ位置間の距離を算出し, 距離が閾値以下のランドマークを選択する。

(条件 3) ランドマークの画像テンプレートの法線ベクトルとランドマークから入力画像の撮影カメラ位置までのベクトルの成す角が閾値以下であり最小である: ランドマークの三次元位置と前フレームのカメラ位置を結ぶ直線と画像テンプレートの法線ベクトルとのなす角  $\theta$  が最小のものを選択する。直線と法線ベクトルのなす角  $\theta$  が閾値以上であれば, そのランドマークを用いない。

(条件 4) 既に選択されたランドマークと入力画像上で密接していない: まず, 入力画像を格子状に分割する。ランドマークの三次元位置を前フレームのカメラ位置・姿勢を用いて入力画像上に投影したとき, 同じ格子内に投影されるランドマークが既

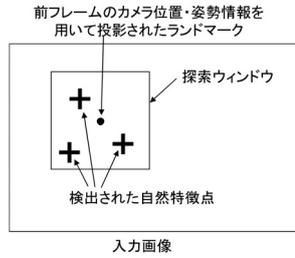


図5 ランドマークと対応付ける自然特徴点の検出

に選択されていた場合は、このランドマークを用いない。

はじめに、条件1を満たすランドマークの中から、条件2のランドマークの撮影位置と前フレームのカメラ位置間の距離が小さい順に、条件3を満たすランドマークを一定数選択する。選択されたランドマークの中から、条件3で述べる角度 $\theta$ が小さい順にランドマークを選択していき、条件4を満たすランドマークが一定数選択されるまで処理を繰り返す。

### 3.2 入力画像とランドマークの対応付け

本節では、データベースから選択されたランドマークを入力画像内で探索し、対応付ける手法について述べる。まず、ランドマークの三次元位置を前フレームのカメラ位置・姿勢を用いて現在の入力画像上に投影し、その周辺で検出された自然特徴点を、投影したランドマークと対応付ける候補とする。次に、検出された自然特徴点を中心とした画像パターンと、ランドマークの画像テンプレートとのパターンマッチングを行うことで、前節で選択されたランドマークとの対応付けを行う。以下に、それぞれの処理手順を詳述する。

#### 3.2.1 入力画像中の自然特徴点の検出

まず、図5に示すように、前フレームのカメラ位置・姿勢を用いてランドマークの三次元位置を入力画像上に投影する。投影された座標周辺にランドマークに対応する自然特徴点が存在すると仮定し、投影座標を中心とする一定サイズの探索ウインドウ内でランドマークデータベース作成時と同じHarrisオペレータ[11]により自然特徴点を検出する。本手法では、特徴量が閾値以上、かつ極大値である点を自然特徴点として検出する。このようにして検出された自然特徴点をランドマークと対応付ける候補とする。

#### 3.2.2 画像テンプレートとのマッチングによる対応付け

ここではまず、3.1節で述べた条件により選択された各ランドマークについて、対応候補点として検出された自然特徴点ごとに、その点を中心とした画像パターンを作成する。次に、これらの画像パターンとランドマークの画像テンプレートを正規化相互相関によって比較する。最も相関値が高い画像パターンを持つ自然特徴点周辺の全画素を候補として再度、画像パターンを作成し、ランドマークの画像テンプレートとの相関値を得る。このとき、最も相関値が高い点をランドマークと対応付ける。これらの詳細を以下で述べる。

それぞれの画像パターンはランドマークの画像テンプレートと同様に、自然特徴点を中心とした面上に入力画像を投影したものであり、前フレームのカメラ位置・姿勢、対応付けるラン

ドマークの三次元位置、画像パターンの法線ベクトルとスケール情報を用いて作成する。自然特徴点を中心とした面はランドマークの画像テンプレート作成時と同一の向きであるものとし、作成する画像パターンのスケールは対応付けるランドマークの画像テンプレートの基本スケールと同一のものと、その2倍、4倍の多重スケールとする。作成したスケールの異なる画像パターンのセットを多重スケール画像パターン $p$ とする。このようにして作成した多重スケール画像パターン $p$ とランドマークの多重スケール画像テンプレート $t$ から、以下に示す正規化相互相関により相関値を求め、まず画像テンプレートと最も相関の高い画像パターンを持つ自然特徴点を選択する。

$$C(t, p) = \frac{\sum_{j=0}^y \sum_{i=0}^x (t(i, j) - \bar{t})(p(i, j) - \bar{p})}{\sqrt{\sum_{j=0}^y \sum_{i=0}^x (t(i, j) - \bar{t})^2 \sum_{j=0}^y \sum_{i=0}^x (p(i, j) - \bar{p})^2}} \quad (1)$$

ただし、 $\bar{t}$ 、 $\bar{p}$ はそれぞれ $t$ 、 $p$ の画素値の平均であり、 $x$ 、 $y$ は画像パターン及び画像テンプレートの座標である。

入力画像上の自然特徴点がランドマークを投影した座標と同じ位置に検出されるとは限らないため、次に、相関値が最大となる自然特徴点周辺の全画素を対応付け候補として再度相関値を算出し、最終的に相関値が最も高い点をランドマークと対応付ける。これにより、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。なお、相関値が閾値以下の場合に対応付けは行わず、そのランドマークはカメラパラメータ推定に用いない。

### 3.3 ランドマークに基づくカメラ位置・姿勢推定

入力画像とランドマークの対応付けによって得られた二次元座標と三次元座標の対応関係を用いてカメラ位置・姿勢推定を行う。まず、RANSAC[12]により誤対応を排除する。そして、誤対応を除いた正しい対応関係のみを用いてカメラ位置・姿勢推定を行い、再投影誤差が最小となるカメラパラメータを最終的な推定結果とする。RANSAC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。
- (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。ここでは出口らの手法[13]を利用し、線形最小二乗法によってカメラパラメータの初期値を算出し、補正を行った後に、再投影誤差の非線形最小化を行う。このようにして得られたカメラパラメー



図 6 全方位型マルチカメラシステム Ladybug と撮影された全方位画像

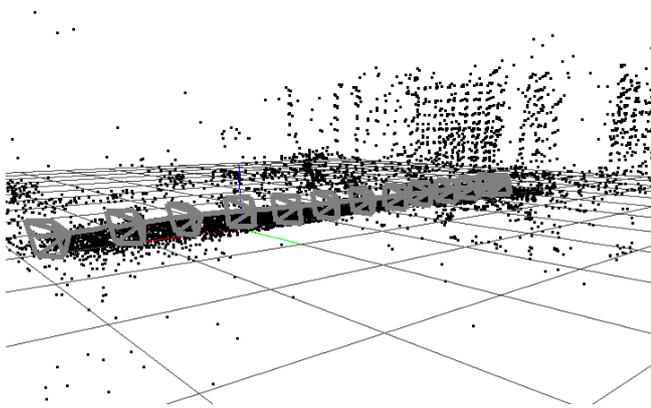


図 7 屋外環境において推定されたカメラパスと自然特徴点の三次元位置

タを最終的なカメラ位置・姿勢推定結果とする。

## 4. 実験

入力画像の自然特徴点とデータベースとの対応付けによるカメラ位置・姿勢推定手法の有効性を検証するために、屋外・屋内環境を撮影し、実験を行った。

### 4.1 屋外環境における実験

まず、図 6 左に示す全方位型マルチカメラシステム (Point Grey Research 社 Ladybug) を用いて図 6 右に示すような屋外環境を複数の系列 (1250 フレーム × 6 台) として撮影した。次に、2.2.1 項で述べた佐藤らの手法 [10] によってカメラパスと自然特徴点の三次元座標を推定し、ランドマークデータベースを構築した。ただし、データ量削減のため、三次元復元で得られた全フレームのデータに対し、10 フレームおきのデータを用いた。また、それぞれのランドマークにつき、3 つの  $15 \times 15$  画素の画像テンプレートからなる多重スケール画像テンプレートを撮影地点ごとに作成した。本実験で得られたランドマークの数は約 12400 個であり、一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。推定された全方位動画のカメラパスと自然特徴点の三次元位置を図 7 に示す。図中の曲線は全方位カメラシステムのカメラユニット 1 に関する推定されたカメラパスを、錐台は 100 フレームおきのカメラ位置・姿勢を表す。

次に、広角レンズ (SONY VCL-HG0758) を取り付けられたビデオカメラ (SONY DSR-DP-150) を用いて 1000 フレームの単眼動画 (解像度  $720 \times 480$  画素、プログレッシブスキャン、15fps) を撮影し、4 章で述べた手法による逐次的なカメラ位置・姿勢推定を行った。ただし、本実験では初期フレームのカメラ位置・姿勢はあらかじめ既知とし、手動で与えた。また、本実験においてデータベースから一度に選択するランドマークの数は最大 100 個、入力画像中の対応付け候補の探索範囲は  $120 \times 60$  画素、RANSAC による繰り返し回数は 500 回とした。逐次推定の結果、1 フレームの処理時間は PC (CPU: Pentium4 3GHz, Memory: 1.5GB) を用いて平均 2.8 秒であった。

入力画像との対応付けに用いられたランドマークの画像上の位置を図 8 に示す。図中、印はデータベースから選択されたランドマークのうち、入力画像と対応付けられなかったランドマーク、×印は RANSAC によって排除されたランドマーク、印は実際にカメラ位置・姿勢推定に用いられたランドマークである。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表示させた結果の一部を図 9 に示す。この実験により、CG で描かれた人物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・姿勢の推定結果に大きな誤差が含まれていないことを確認した。

4.2 屋内環境における実験

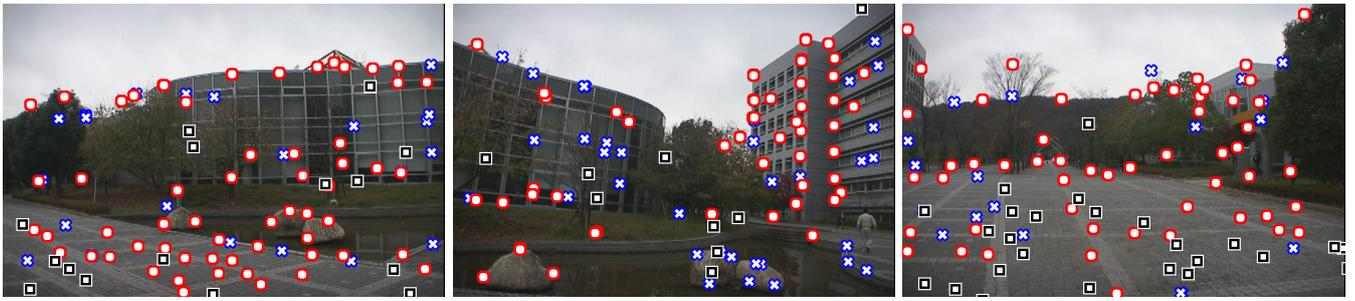
次に、屋外環境と同様の手順で屋内環境における実験を行った。屋内環境における実験で用いた全方位動画のフレーム数は 701 フレームであり、ランドマークデータベース作成には三次元復元で得られた全フレームのデータに対して 5 フレームおきのデータを用いた。また、屋外実験と同様に、それぞれのランドマークにつき 3 つの  $15 \times 15$  画素の画像テンプレートからなる多重スケール画像テンプレートを撮影地点ごとに作成した。実験で得られたランドマークの数は約 2100 個であり、一つのランドマークにつき平均 9.4ヶ所の異なる地点で撮影された画像テンプレートが登録された。

ランドマークデータベースを用いたカメラ位置・姿勢では、900 フレームの単眼動画をフレーム毎に逐次推定した。また、データベースから一度に選択するランドマークの数は最大 100 個、入力画像中の対応付け候補の探索範囲は  $60 \times 60$  画素、RANSAC による繰り返し回数は 500 回とした。逐次推定の結果、1 フレームの処理時間は屋外実験と同じマシンを用いて平均 2.5 秒であった。

最後に、入力画像との対応付けに用いられたランドマークの画像上の位置を図 10 に、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表示させた結果を図 11 に示す。本実験においても、CG で描かれた人物が入力画像と幾何学的に正しく合成されていることから、カメラ位置・姿勢推定結果に大きな推定誤差が含まれていないことを確認した。

4.3 定量的な評価実験

最後に、推定されたカメラ位置・姿勢を正解データと比較することで推定精度を評価した。本実験では、環境内の特徴点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の位置を手動で対応付け、PnP 問題を解くこ



(a)200 フレーム目

(b)400 フレーム目

(c)800 フレーム目

図 8 屋外環境におけるカメラ位置・姿勢推定に用いられたランドマーク

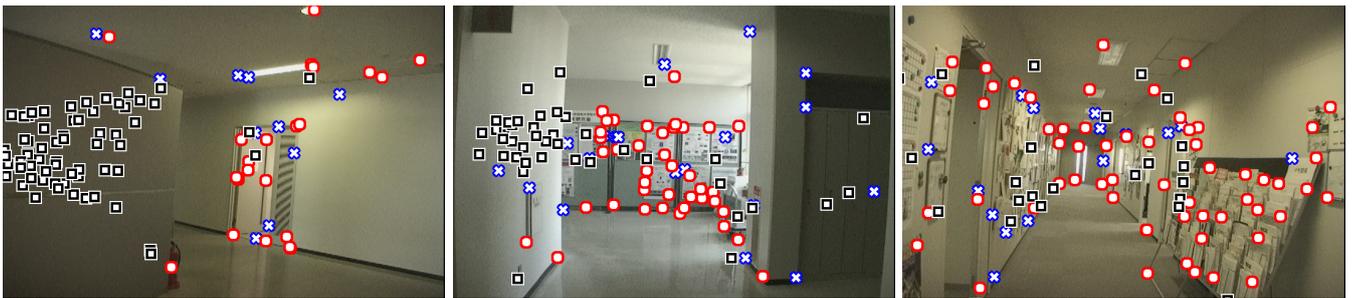


(d)200 フレーム目

(e)400 フレーム目

(f)800 フレーム目

図 9 屋外環境におけるカメラ位置・姿勢推定結果を用いた match move  
(<http://yokoya.naist.jp/pub/movie/oe/outdoor.mpg>)



(a)200 フレーム目

(b)400 フレーム目

(c)600 フレーム目

図 10 屋内環境におけるカメラ位置・姿勢推定に用いられたランドマーク



(d)200 フレーム目

(e)400 フレーム目

(f)600 フレーム目

図 11 屋内環境におけるカメラ位置・姿勢推定結果を用いた match move  
(<http://yokoya.naist.jp/pub/movie/motoko-o/indoor.mpg>)

とでカメラパラメータの正解データを 10 フレームおきに作成した。ただし、自然物が入力画像の大半を占めるなどの理由でトータルステーションによって計測できる特徴点が少ないフレームや、手動で対応付けを行った結果による再投影誤差が 1.5 画素以上であるフレームに関しては評価対象にできなかった。

屋外実験の結果に対しては、1000 フレーム中で正解データを作成した 66 地点におけるカメラの位置推定誤差は平均 220mm、

カメラの光軸角度の推定誤差は平均 0.37 度であった。図 12 に初期フレームから 1000 フレーム目までのカメラパラメータと正解データを示す。図 12 の横軸、縦軸はそれぞれランドマークデータベース構築時の世界座標系における X 軸、Y 軸に対応している。図 12 から、推定されたカメラ位置・姿勢に累積的な推定誤差が生じていないことを確認できる。

屋内実験では、屋外実験と同様に、環境内の特徴点を測定し、

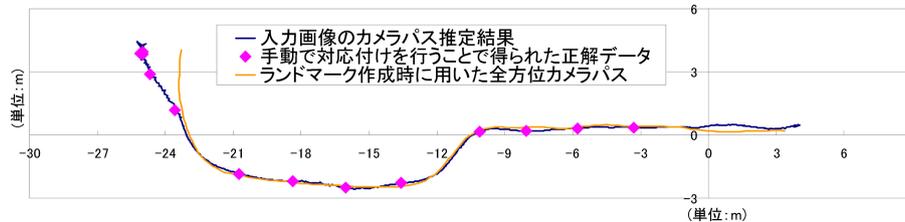
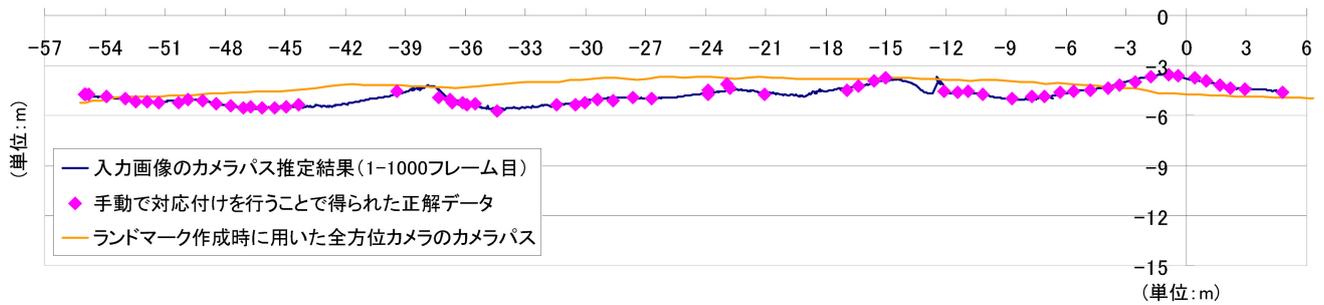


図 13 屋内実験において推定されたカメラパスと正解データ

各フレームの画像上でそれらの特徴点の位置を手動で対応付けることでカメラパラメータの正解データを 50 フレームおきに作成した。結果、900 フレーム中で正解データを作成した 14 地点におけるカメラの位置推定誤差は平均 52mm、カメラの光軸角度の推定誤差は平均 0.40 度であった。図 13 に初期フレームから 900 フレーム目までのカメラパラメータと正解データを示す。図 13 の横軸、縦軸はそれぞれランドマークデータベース構築時の世界座標系における X 軸、Y 軸に対応している。図 13 から、屋内環境においても推定されたカメラ位置・姿勢に累積的な推定誤差が生じていないことを確認できる。

## 5. ま と め

本稿では自然特徴点から成るランドマークデータベースを事前に構築し、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位置・姿勢をフレームごとに逐次推定する手法を提案した。本手法では、まず、環境を全方位動画として撮影し、三次元復元を行うことで特徴点ランドマークデータベースの構築を行う。次に、構築したランドマークから、カメラパラメータ推定に用いるランドマークを決定し、逐次単眼カメラからの動画のカメラ位置・姿勢推定を行う。実験では、推定されたカメラ位置・姿勢を用いて CG を重畳表示することで、カメラ位置・姿勢の推定に大きな推定誤差が含まれていないことを確認した。評価実験では、手動で作成した正解データと比較することによって、累積誤差が蓄積しないことを確認した。

今後の課題としては、カメラの初期位置・姿勢決定の自動化や逐次推定のリアルタイム化などが挙げられる。リアルタイム化には、より高速、高精度なデータベースからのランドマーク選択手法やランドマークと入力画像の自然特徴点の対応付け手法の検討が必要である。

## 文 献

- [1] A. Ronald, B. Hoff, H. Neely III and R. Sarfaty: "A motion-stabilized outdoor augmented reality system," Proc. IEEE Virtual Reality, pp. 252-259, 1999.
- [2] T. Höllerer, S. Feiner and J. Pavlik: "Situated documentaries: Embedding multimedia presentations in the real world," Proc. Int. Symp. on Wearable Computers (ISWC'99), pp. 79-86, 1999.
- [3] H. Kato and H. Billinghurst: "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," Proc. 2nd IEEE/ACM Int. Workshop on Augmented Reality (IWAR'99), pp. 85-94, 1999.
- [4] 羽原寿和, 町田貴史, 清川清, 竹村治雄: "ウェアラブル PC のための画像マーカを用いた広域屋内位置検出機構", 電子情報通信学会技術研究報告, ITS2003-76, 2003.
- [5] U. Neumann and S. You: "Natural feature tracking for augmented-reality," IEEE Transactions on Multimedia, Vol. 1, No. 1, pp. 53-64, 1999.
- [6] A. J. Davison, Y. G. Cid and N. Kita: "Real-time 3d slam with wide-angle vision," Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [7] L. Naimark and E. Foxlin: "Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker," Proc. 1st IEEE/ACM Int. Symp. on Mixed and Augmented Reality (ISMAR2002), pp. 27-36, 2002.
- [8] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: "Fully automated and stable registration for augmented reality applications," Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR2003), pp. 93-102, 2003.
- [9] I. Gordon and D. G. Lowe: "Scene modelling, recognition and tracking with invariant image features," Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR2004), pp. 110-119, 2004.
- [10] 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 (ECCV2004), Vol. 2, pp. 326-340, 2004.
- [11] C. Harris and M. Stephens: "A combined corner and edge detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- [12] M. A. Fischler and R. C. Bolles: "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Comm. of the ACM, Vol. 24, pp. 381-395, 1981.
- [13] 出口光一郎: "射影幾何学による P n P カメラ補正問題の統一的解法", 情処シンポジウム, Vol. 90, pp. 41-50, 1990.