

## 基礎論文

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・姿勢推定大江 統子<sup>\*1</sup> 佐藤 智和<sup>\*1</sup> 横矢 直和<sup>\*1</sup>

Camera Position and Posture Estimation

Based on Feature Landmark Database for Geometric Registration

Motoko Oe<sup>\*1</sup>, Tomokazu Sato<sup>\*1</sup> and Naokazu Yokoya<sup>\*1</sup>

**Abstract** – In the field of wearable augmented reality, absolute position and posture of user's camera are necessary to overlay CGs on real images. In this research field, two kinds of approaches have been investigated; sensor-based and vision-based methods. However, both kinds of conventional methods have a problem that it was not easy to use these methods in large and variable environments; usable environment is limited by selection of sensors in the sensor based method, and many artificial markers must be allocated in large environment in the vision based method to prevent accumulative errors. This paper proposes a novel vision-based camera position and posture estimation method which does not require sensors and artificial markers. In the proposed method, feature landmark database is constructed in advance by using large number of natural feature points in omni-directional image sequences. Next, by detecting feature landmarks from images captured by wearable camera, camera position and posture are estimated in every frame without accumulative errors.

**Keywords** : camera position and posture estimation, natural features, landmark database, wearable augmented reality

## 1 はじめに

近年、カメラによって撮影された画像にCGを重畳表示することで、現実環境の映像を拡張する拡張現実感(AR)に関する研究が盛んである。このような拡張現実感の研究においては、ユーザの眼前に設置された小型のディスプレイ上にCGを重畳表示するビデオスルー型のシステムが多数開発されている。このようなビデオスルー型拡張現実感システムにおいて、CGを幾何学的に正しい位置に重畳合成するためには、カメラと撮影対象の間の位置・姿勢の関係を正確に推定することが必要となる。特に、近年盛んに研究されているウェアラブル拡張現実感の分野では、環境に注釈やナビゲーション情報を表示するためにカメラの絶対位置・姿勢を推定することが必要となる。このようなウェアラブル拡張現実感を実現するためのカメラ位置・姿勢の推定手法としては、従来、GPSや磁気センサを用いる手法[1, 2, 3, 4, 5]、入力画像に撮像されるマーカ(ランドマーク)や画像特徴点を用いる手法[6, 7, 8, 9, 10, 11, 12]、それらのハイブリッド手法[13, 14, 15]など多くの手法が提案されている。

センサ類を用いるカメラ位置・姿勢推定手法は、用いるセンサの組み合わせによって、屋外で利用可能な手法と屋内で利用可能な手法に大別される。屋外において一般的に利用されるセンサは、GPSとジャイロまたは慣性航法センサの組み合わせである[1, 2, 3]。これらを用いる手法では、GPSによって得られる位置情報とジャイロや小型慣性航法センサによって得られる姿勢情報をカメラの位置・姿勢パラメータとして利用する。しかし、GPSを用いるために屋内では利用できないという問題や、センサの計測誤差が大きいためカメラ位置・姿勢の推定精度が悪いという問題がある。屋内環境で利用可能な手法としては、赤外線ビーコンや磁気センサを用いる手法が提案されている[4, 5]。しかし、これらの手法は環境に何らかの装置を設置することを前提としており、広域な環境においてカメラ位置・姿勢計測を実現するためには装置の設置に膨大なコストがかかるという問題がある。このように、センサのみを用いる手法では、利用する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ために、様々な環境でカメラの位置・姿勢を推定することは難しい。また、センサの組み合わせによってシステムが重装備となり、装着するユーザの負担になるという問題がある。

\*1奈良先端科学技術大学院大学 情報科学研究科

\*1Graduate School of Information Science, Nara Institute of Science and Technology

一方、カメラからの入力画像を用いる手法としては、ARToolkit[6]などの人工的なマーカを用いる手法[7, 8]が提案されている。これらの手法では、マーカが画像内に撮像されていれば、安定にかつ精度よくカメラの位置・姿勢を推定することができる。しかし、広範囲を移動するカメラの位置・姿勢を推定するためには、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[11, 12]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[11]。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・姿勢推定を実現しているが、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撮影対象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[12]。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発生するために、小物体、小領域への適用にとどまっている。

また、センサ類とランドマークを用いるハイブリッド手法についても近年盛んに研究されている[13, 14, 15]。しかし、従来提案されている手法では、センサ類によって得られるカメラ位置・姿勢を、画像内のランドマークを使って補正するというアプローチを採っているために、センサ類の組み合わせによって利用可能な範囲が限定されるという問題や、センサ類とカメラの同期を取ることが難しいという問題が残されている。

本論文では、移動を伴って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事前にデータベース化し、これを用いてユーザが装着するカメラの位置・姿勢を逐次推定する手法を提案する。本手法の基本的なアプローチは、従来手法[11, 12]と同様、事前に行う環境情報のデータベース化(モデル化)と、これを用いたオンラインでのカメラ位置・姿勢推定から成る。ただし、本研究では、データベースの作成に基準点と自然特徴点の併用による全方位画像系列からの三次元復元手法を利用することで、誤差の累積しない広域のランドマーク情報を自動で構築する。

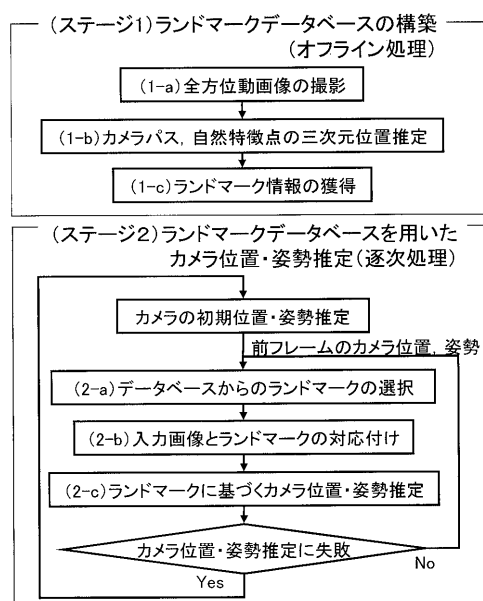


図1 提案手法のフローチャート  
Fig. 1 Flow diagram of proposed method.

加えて、ユーザの位置・姿勢に応じて利用するランドマークをデータベースから自動選択することで、広域環境においても高精度なカメラ位置・姿勢推定を実現する。また、本論文では、処理の高速化による実時間処理の実現や試作システムの作成については保留し、推定精度の向上を主題とすることで、多様な広域環境におけるビジョンによるカメラ位置・姿勢推定の実現可能性を定量的評価実験によって検証する。

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、structure-from-motionによって環境の粗な三次元復元を行う。また、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、ウェアラブルユーザのカメラ位置・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築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てユーザが装着するカメラの位置・姿勢を逐次推定する手法について述べる。4章では、屋外・屋内環境を対象としたデータベースの作成とカメラ位置・姿勢の復元実験について報告する。実験では、推定されたカメラ位置・姿勢に基づくMatch Moveおよび定量的評価によって、本手法の

ウェアラブル拡張現実感システムへの応用可能性を検証する。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作成方法(ステージ1)について詳述する。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、全方位動画像上の自然特徴点追跡による三次元復元を行い、自然特徴点の三次元位置と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に示すように、複数のランドマークによって構成される。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

本手法では次章で述べるステージ2において、ランドマーク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に、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、環境に固定した世界座標系で保持する。以降、本論文における世界座標系は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、各ランドマークに複数の撮影地点ごとの情報を保持する。

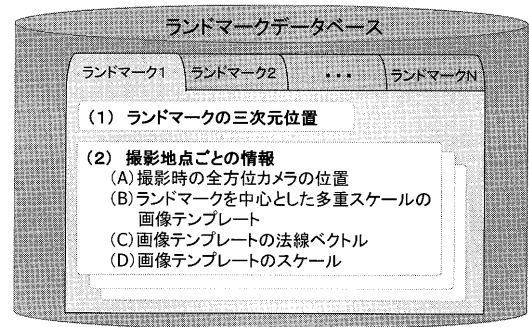


図2 ランドマークデータベースの構成要素  
Fig.2 Elements of feature landmark database.

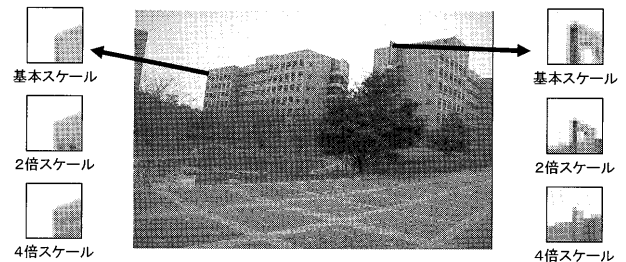


図3 スケールの異なる画像テンプレート  
Fig.3 Multi-scale image templates.

- (A) 撮影時の全方位カメラの位置: 全方位カメラの位置を世界座標系で保持する。入力画像と対応付けるランドマークの選択に用いられる。
- (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: 図3に示すように、撮影地点ごとに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。入力画像中からランドマークを探索するために用いられる。
- (C) 画像テンプレートの法線ベクトル: 図4に示すような、実空間中における、画像テンプレートの面の法線方向を表す。画像テンプレートを選択する基準として用いられる。また、入力画像からランドマーク照合用の画像パターンを作成する際に用いられる。
- (D) 画像テンプレートの基本スケール: 多重スケールの画像テンプレートの中で、基本スケールとなる画像テンプレートの実空間中での大きさを表す。入力画像のランドマーク照合用画像パターンのスケールを決定するために用いられる。

### 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、全方位動画像からの三次元復元手法[16]を利用してランドマーク情報を獲得し、自然特徴点による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、全方位カメラ

を用いて対象となる環境を移動しながら撮影し、撮影された動画像を用いて三次元復元を行うことで、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 (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、全方位動画像からランドマークデータベースを作成する手順について詳述する。

### 2.2.1 全方位動画像からの環境の三次元復元

ここでは、全方位型マルチカメラシステムを用いた三次元復元手法 [16] を用いて、全方位カメラの外部パラメータと自然特徴点の三次元位置を推定する。この手法では、まずデータベース化の対象となる環境を撮影した全方位動画像から少数のキーフレームを選択し、キーフレーム内に存在する自然特徴点から6点以上の任意の特徴点を基準点として選択する。次に、これらの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する。最後に、全方位動画像上で基準点と Harris オペレータ [17] によって検出される多数の自然特徴点を自動追跡し、撮影された全方位動画像全体での再投影誤差の和 [16] を最小化することで、累積的な推定誤差を含まない全方位カメラのカメラパラメータおよびランドマークの三次元座標を推定する。これにより、複雑で広範囲な環境を撮影した場合でも、全方位動画像のカメラパスと自然特徴点の三次元位置を基準点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

三次元復元によって得られた情報を用いてランドマークデータベースを自動構築する。以下では、ランドマークデータベースのそれぞれの構成要素について、作成方法を述べる。

#### (1) ランドマークの三次元位置

三次元復元によって得られた各自然特徴点の三次元座標を対応するランドマークの三次元位置として登録する。三次元座標が推定されなかった自然特徴点については、データベースに登録せず破棄する。

#### (2) 撮影地点ごとの情報

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、撮影地点ごとにランドマーク情報を登録する。以下に、撮影地点ごとに登録される要素の作成方法について詳述する。

- (A) 撮影時の全方位カメラの位置: 三次元復元手法によって推定されたカメラ外部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登録する。
- (B) ランドマークを中心とした多重スケールの画像

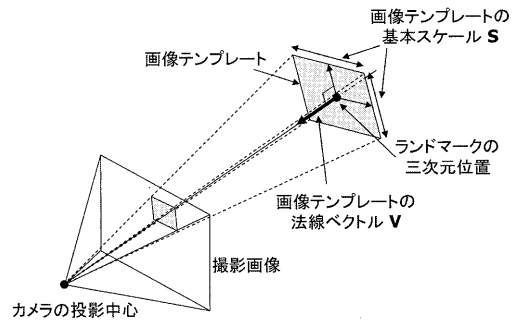


図4 ランドマークとその画像テンプレート  
Fig.4 Landmark and its image template.

テンプレート: 画像撮影時のカメラ姿勢の変化によるランドマークの画像上での見え方の変化を補正するために、図4に示す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、画像上でのランドマーク位置周辺のパターンを投影し、多重スケールの画像テンプレートを作成する。ここではまず、図4に示すような法線ベクトル  $V$  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  $XY$  平面と面の横軸が平行

設定された投影面を、等間隔な  $N \times N$  の格子に分割する。最後に、各格子点の三次元座標を撮影画像面上に投影し、各格子点の輝度値を決定することで基本スケールの画像テンプレートを作成・登録する。また同様に、スケールが2倍、4倍の画像テンプレートをデータベースに登録する。

- (C) 画像テンプレートの法線ベクトル: 図4に示すように、ランドマークからカメラの投影中心への単位ベクトルを画像テンプレートの法線ベクトルとして登録する。
- (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1画素を撮影画像上に投影した大きさが撮影画像の1画素のサイズと等しくなるように、基本スケール画像テンプレートの世界座標系での大きさ  $S \times S$  を決定する。これにより、作成される基本スケール画像テンプレートの解像度と撮影画像の解像度を一致させ、必要以上に高解像な画像テンプレートの作成を防ぐ。ここでは、算出された大きさ  $S \times S$  を基本スケールとしてデータベースに登録する。

### 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・姿勢の逐次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(ステージ2)の流れは先に示した図1の通りである。ただし、本論文では初期カメラ位置・姿勢は何らかの手法によって取得できることを前提としている。本手法では、まず前章で作成されたランドマークデータベースから、入力画像との対応付けに利用するランドマークを選択する(図1中2-a)。次に、入力画像上で自然特徴点を検出し、検出された自然特徴点と選択されたランドマークの画像テンプレートを照合することでランドマークデータベースと自然特徴点の対応付けを行う(2-b)。最後に、得られた対応関係を用いてPnP問題[18]を解くことでカメラの位置・姿勢推定を行う(2-c)。また、これらの処理を繰り返すことで、カメラの位置・姿勢を逐次推定する。ただし、処理(2-c)においてカメラ位置・姿勢が正しく推定されない場合には、再度カメラの初期位置・姿勢推定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して、以下の3つの条件を全て満たすものを選択する。

- (条件1) ランドマークが入力画像の視野範囲内に存在
- (条件2) 画像テンプレートの法線ベクトルとランドマークと入力画像の投影中心を結ぶベクトルの成す角が閾値以下
- (条件3) 入力画像面上において既に選択された近接するランドマークとの距離が閾値以上

ただし、後段のカメラパラメータの推定処理(2-c)で用いるPnP問題の解法では、十分な数のランドマークが与えられれば、推定精度がおおむね上限に達するため、選択されるランドマークの個数に上限を設ける。条件1~3を全て満たすランドマークの数が、あらかじめ設定した上限個数を上回る場合には、カメラの移動による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カメラ位置と入力画像撮影時のカメラ位置の距離が小さい順にランドマークを選択する。ただし、ここでは

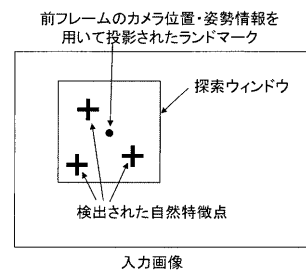


図5 ランドマークと対応付ける自然特徴点の検出  
Fig.5 Detection of image features.

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・姿勢を近似的に用いる。

#### 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、自動で対応付ける手法について述べる。ここではまず、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、ランドマークデータベースと自然特徴点の対応付けを行う。以下に、それぞれの処理手順を詳述する。

##### 3.2.1 対応付けに用いる自然特徴点の検出

図5に示すように、まず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ィンドウ内で、データベース作成時と同じHarrisオペレータ[17]を用いて自然特徴点を検出する。これによっ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 テンプレート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入力画像内の自然特徴点ごとに、その点を中心とした画像パターンを作成し、正規化相互相関法によって照合する。それぞれの自然特徴点の画像パターンは、ランドマークの画像テンプレート作成と同様に、実空間中において自然特徴点を中心とする正方形の面上に入力画像を投影することで作成する。ここで、自然特徴点を中心とする面の法線として、照合を行うランドマークの画像テンプレートの法線を用い、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する。

また、ランドマークの画像テンプレートと同様に多重スケールの画像パターンを作成する。ここでは、ランドマーク、自然特徴点双方に対して、基本、2倍、4倍スケールの画像パターンを連結した画像パターンを作成し、双方の多重スケールテンプレートを正規化相互相関により一度に照合する。これにより、最も相関の高い画像パターンを持つ自然特徴点を一意に選択することで、データベースに登録されているランドマークの三次元座標と入力画像上の二次元座標を対応付ける。なお、相関値が閾値以下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

ここでは、推定するカメラ位置・姿勢を用いてランドマークの三次元座標を画像上に投影した座標  $(u, v)$  と、ランドマークに対応づけられた自然特徴点の画像上での検出座標  $(u', v')$  の自乗距離の和  $E = \sum\{(u - u')^2 + (v - v')^2\}$  を再投影誤差として定義し、これを最小化することでカメラ位置・姿勢の推定を行う。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、ここでは線形解法を用いてまずカメラ位置・姿勢の初期値を求め [18]、補正を行った後に最急降下法によって再投影誤差  $E$  を最小化する。また、RANSAC [19] による誤対応の排除を行い、正しい対応関係のみを用いて再投影誤差を最小化することでカメラ位置・姿勢を精度良く推定する。

### 3.4 カメラ位置・姿勢推定の失敗判定

前節で述べたカメラ位置・姿勢の推定処理において、線形演算によるカメラ位置・姿勢の初期値の算出 [18] には6点以上のランドマークが必要となる。また、テンプレートマッチングによるランドマークの対応付けにおいて、誤対応が過半数を上回る場合には、原理的に誤対応を排除することが不可能となるため、カメラ位置・姿勢の推定処理において、再投影誤差が十分に最小化されない。そこで、本手法では、RANSAC によって得られる対応点数が6点を下回る場合、または再投影誤差が十分に最小化されない場合、カメラ位置・姿勢の推定が失敗したとみなし、図1に示したカメラ位置・姿勢の初期推定から処理を行う。

## 4 実験

提案手法の有効性を検証するために、屋外・屋内環境においてランドマークデータベースを構築し、これらを用いたカメラ位置・姿勢推定を行った。また、推定されたカメラパラメータに基づきCGを重畳表示することで、ウェアラブル拡張現実感システムへの応用可能性を検証し、加えて推定結果に関する定量的な評価実験を行った。ただし、計算量の問題から、現状では試作したシステムをビデオレートで作動させること

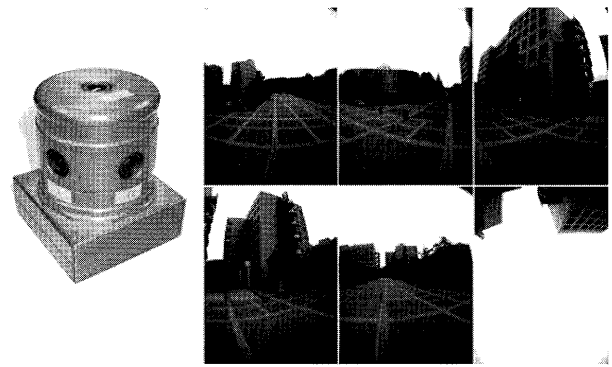


図6 全方位型マルチカメラシステム Ladybug と撮影された全方位画像

Fig. 6 Omni-directional multi-camera system Ladybug and images taken by Ladybug.

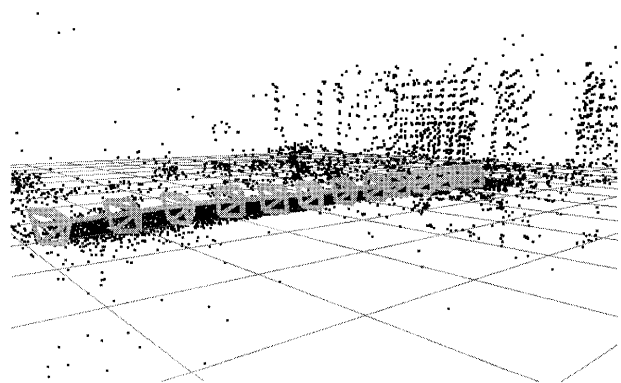


図7 屋外環境において推定されたカメラパスと自然特徴点の三次元位置

Fig. 7 Estimated camera path and 3-D coordinates of natural feature points.

が難しいために、本実験では、カメラを動かしながら撮影した動画をオフラインで逐次処理することで、ステージ2のカメラ位置・姿勢推定処理を行った。

### 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右に示すような本学の屋外環境を複数の動画像系列 (1250 フレーム × 6 台, 15fps) として、約 80m の距離を歩きながら撮影した。次に、2.2.1 項で述べた手法 [16] によって全方位カメラの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。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。本実験では、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15×15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

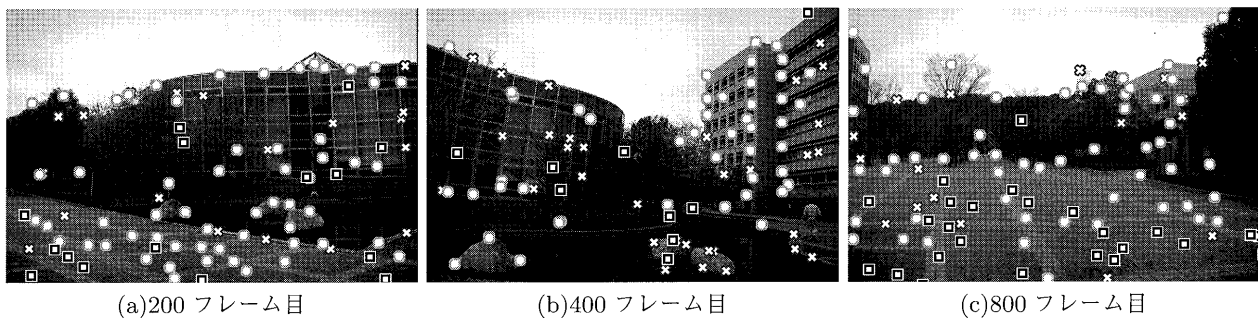


図8 屋外環境におけるカメラ位置・姿勢推定に用いられたランドマーク  
 Fig. 8 Landmarks for camera position and posture estimation (outdoor).

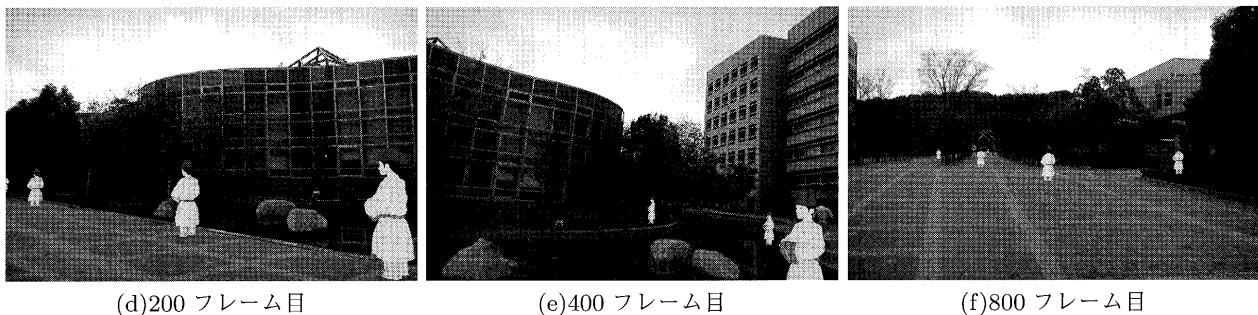


図9 屋外環境におけるカメラ位置・姿勢推定結果を用いた match move  
 Fig. 9 Match move based on estimated camera position and posture (outdoor).

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。

次に、広角レンズ (SONY VCL-HG0758) を取り付けたビデオカメラ (SONY DSR-DP-150) を用いて 1000 フレームの単眼動画像 (解像度 720×480 画素, プログレッシブスキャン, 15fps) を撮影し, 3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, 本実験ではカメラ内部パラメータをあらかじめ Tsai の手法 [20] によって校正し, 撮影中は内部パラメータを固定した。また, 初期フレームのカメラ位置・姿勢は既知とし, あらかじめ手で与えた。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の上限は 100 個, 入力画像中の対応付け候補の探索範囲は 120×60 画素, RANSAC による繰り返し回数は 500 回とした。また, ここでは 3.1 節で述べたランドマーク選択の閾値として, (条件 2) の法線ベクトルの成す角の閾値を 15 度, (条件 3) の画像上での距離の閾値を 10 画素とした。

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<sup>1</sup>。

本実験において, 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 秒であり, 1000 フレームの入力に対してカメラパラメータ推定処理が途中で失敗することはなかった。1 フレームの処理に対する平均処理時間の内訳は, ランドマークの選択処理 (2-a) に 0.46 秒, 入力画像とランドマークの対応付け処理 (2-b) に 2.18 秒, カメラの位置・姿勢推定処理 (2-c) に 0.19 秒であった。この結果から, 処理 (2-b) に多くの計算時間を費やしていることが分かる。この処理では, 入力画像上の特徴点周辺のパターンとランドマークのテンプレートを照合する処理において, 特徴点周辺の画素の 1 画素毎の幾何学的な変形操作が処理時間の大半を占める。このような処理においては, GPU 等のハードウェアを利用した高速な幾何変換手法を適用することや, 近似的なパターン類似度の判定を行うことで高速化が期待できる。また, その他の処理についても, 現時点では試作システムの実装面での高速化は行っておらず, 今後, 実装の改良やハードウェアの高速化によって, リアルタイム処理の実現が可能となる見込みである。

<sup>1</sup><http://yokoya.naist.jp/pub/movie/oe/outdoor.mpg>



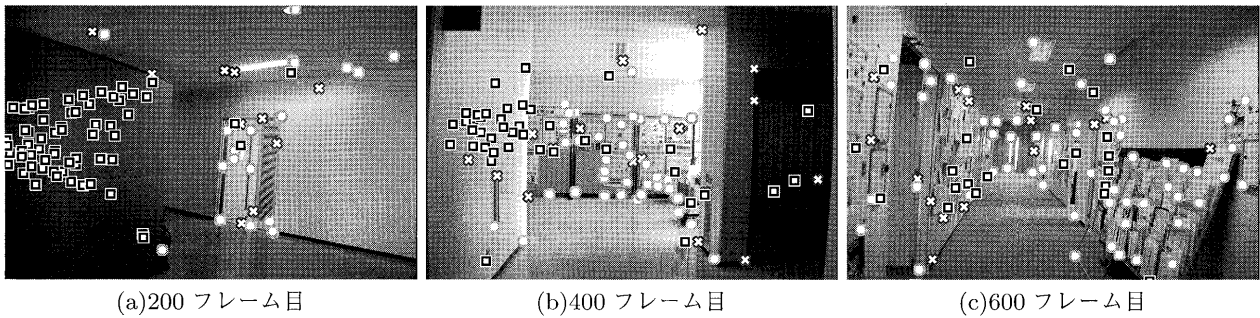


図 10 屋内環境におけるカメラ位置・姿勢推定に用いられたランドマーク  
 Fig. 10 Landmarks for camera position and posture estimation (indoor).

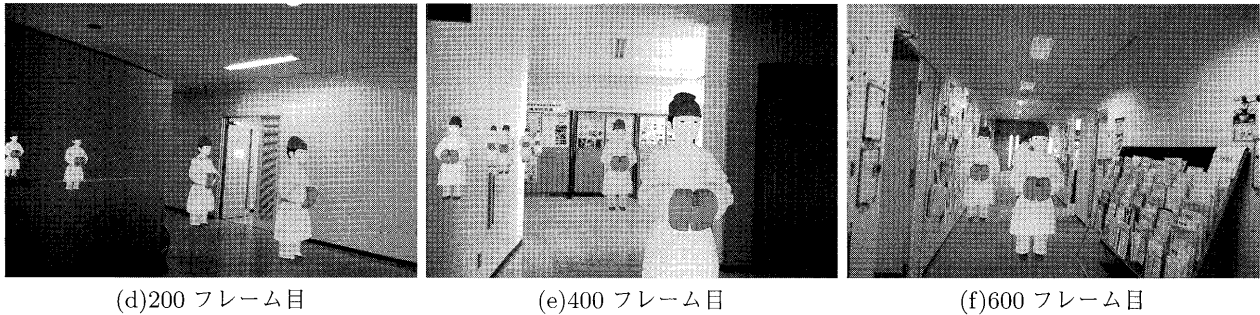


図 11 屋内環境におけるカメラ位置・姿勢推定結果を用いた match move  
 Fig. 11 Match move based on estimated camera position and posture (indoor).

#### 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。屋内環境におけるランドマークデータベース構築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。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・姿勢推定では、900 フレームの単眼動画をフレーム毎に逐次推定した。本実験では、処理 (2-a) におけるデータベースから一度に選択するランドマークの数を 100 個、処理 (2-b) における入力画像中の対応付け候補の探索範囲を  $60 \times 60$  画素、処理 (2-c) における RANSAC による繰り返し回数を 500 回と設定した。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。本実験においても、入力画像の全てのフレームの入力に対してカメラパラメータ推定処理が途中で失敗することはなかった。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 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。図中では、遮蔽物を考慮せずに CG を重畳表示しているために、CG と現実環境の前後関係が一部

誤っているが、屋外実験と同様に、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・姿勢推定結果には大きな推定誤差が含まれていないことを確認した<sup>2</sup>。

#### 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の位置を手動で対応付け、PnP 問題を解くことでカメラ外部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。

屋外実験において、初期フレームから 1000 フレーム目まで推定されたカメラ位置とその正解データを図 12 に示す。図中の横軸、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。また、逐次推定に用いたカメラの進行方向は、同図中において X 軸の正の方向である。同図から、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。また、ランドマークデータベース作成時の全方位カメラの位置か

<sup>2</sup><http://yokoya.naist.jp/pub/movie/oe/indoor.mpg>



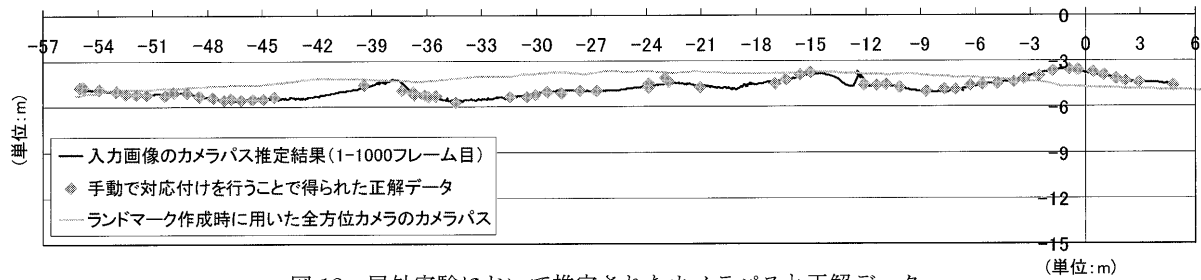


図 12 屋外実験において推定されたカメラパスと正解データ  
 Fig. 12 Estimated camera path and its ground truth (outdoor).

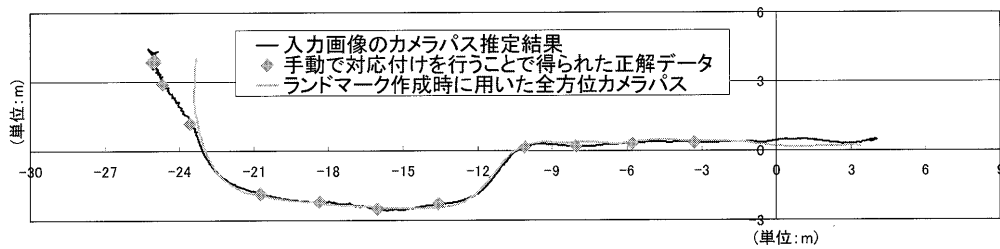


図 13 屋内実験において推定されたカメラパスと正解データ  
 Fig. 13 Estimated camera path and its ground truth (indoor).

ら数メートル離れた位置においても、大きな誤差を含むことなくカメラ位置が推定されていることが確認できる。屋外実験におけるカメラパラメータの推定結果において、1000 フレーム中で正解データを作成した 66 地点でのカメラの位置推定誤差は平均 220mm、カメラ光軸角度の推定誤差は平均 0.37 度であった。これらの結果から、今回用いたカメラでは視点位置から 14m 以上離れた地点に CG を重畳表示すれば、投影誤差を 1 画素以内に収めることができ、幾何学的に正確な位置合わせを実現可能である。

同様に屋内実験の結果を図 13 に示す。同図中においても、逐次推定に用いたカメラの進行方向は X 軸の正の方向である。同図から、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。本実験で得られた 900 フレームの画像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。

これらの結果から、本手法は、中景から遠景に CG を重畳する場合や、CG の位置合わせに画素単位の精度を要求しない注釈提示・ナビゲーション等の分野において、十分な精度でカメラ位置・姿勢を推定できることが分かる。ただし、ランドマークの撮影に用いた全方位カメラの移動経路に対して、ユーザのカメラ位置が大きく離れる場合には、入力画像中からのランドマークの検出精度の低下や、これに起因するカメラ位置・姿勢の推定精度の低下が起これると考えられる。このような問題に対応するためには、ユーザが移動することが予見される範囲内において全方位画像系列を複

数の経路で撮影することが必要である。

## 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に対応付けることによりカメラの絶対位置・姿勢をフレームごとに逐次推定する、ウェアラブル拡張現実感の幾何学的位置合わせのためのカメラ位置・姿勢推定手法を提案した。本手法では、まず、環境を全方位動画像として撮影し、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選択し、ウェアラブルユーザが装着するカメラの位置・姿勢を逐次推定する。これにより、ユーザに装着の負担を強いる複雑なセンサシステムや、景観を損なう多数の人工的マーカを利用することなく、広範囲の環境においてカメラの位置・姿勢パラメータを精度良く推定することができる。

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、ウェアラブル拡張現実感システムへの応用の可能性を示した。定量的評価実験では、手動で作成した正解データと比較することによって、累積的な誤差が発生しない高精度なカメラパラメータ推定が実現できることを確認した。ただし、本論文は、複雑な広域環境での利用が可能な高精度なカメラ位置・姿勢推定手法の確立を主題としたため、実時間処理の実現は保留されている。

今後、実時間でのカメラ位置・姿勢推定を実現するた

めには、実装面の改良と共に、より高速なランドマークと入力画像の対応付け手法の検討が必要である。また、本論文では初期カメラ位置・姿勢推定手法については既知としたが、実際には広範囲を撮影した画像データベース等を利用することで初期カメラ位置・姿勢を推定することが必要である。

謝辞 本研究は文部省科学研究費・基盤(B)16300057の助成を受けて行った。

#### 参考文献

- [1] T. H. S. Feiner, B. MacIntyre and A. Webster: "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment," Proc. Int. Symp. on Wearable Computers, pp. 74-81, 1997.
- [2] 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.
- [3] 神原誠之, 横矢直和: "RTK-GPS と慣性航法装置を併用したハイブリッドセンサによる屋外型拡張現実感", 電子情報通信学会 技術研究報告, PRMU2004-160, pp. 37-38, 2005.
- [4] D. Hallaway, T. Höllerer and S. Feiner: "Coarse, Inexpensive, Infrared Tracking for Wearable Computing," Proc. IEEE Int. Symp. on Wearable Computers, pp. 69-78, 2003.
- [5] R. Tenmoku, M. Kanbara and N. Yokoya: "A Wearable Augmented Reality System Using Positioning Infrastructures and a Pedometer," Proc. IEEE Int. Symp. on Wearable Computers, pp. 110-117, 2003.
- [6] H. Kato and H. Billinghurst: "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System," Proc. IEEE/ACM Int. Workshop on Augmented Reality, pp. 85-94, 1999.
- [7] 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, 2004.
- [8] L. Naimark and E. Foxlin: "Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-inertial Self-tracker," Proc. IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp. 27-36, 2002.
- [9] 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.
- [10] U. Neumann and S. You: "Natural Feature Tracking for Augmented-reality," IEEE Trans. on Multimedia, Vol. 1, No. 1, pp. 53-64, 1999.
- [11] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: "Fully Automated and Stable Registration for Augmented Reality Applications," Proc. Int. Symp. on Mixed and Augmented Reality, pp. 93-102, 2003.
- [12] I. Gordon 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.
- [13] Y. Kameda, T. Takemasa and Y. Ohta: "Outdoor See-through Vision Utilizing Surveillance Cameras," Proc. Int. Symp. on Mixed and Augmented Reality, pp. 151-160, 2004.
- [14] K. Satoh, M. Anabuki, H. Yamamoto and H. Tamura: "A Hybrid Registration Method for Outdoor Augmented Reality," Proc. Int. Symp. on Augmented Reality, pp. 67-76, 2001.
- [15] 興梠正克, 蔵田武志, 坂上勝彦: "ウェアラブルシステムのための画像と複数センサのデータ統合による位置・方位取得手法", 第8回画像センシングシンポジウム (SSII2002), pp. 439-444, 2002.
- [16] 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.
- [17] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- [18] R. Klette, K. Schluns and A. Koschian Eds.: Computer Vision: Three-dimensional Data from Image, Springer, 1998.
- [19] 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.
- [20] 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.

(2005年4月15日受付)

#### [著者紹介]

##### 大江 統子



2003年同志社・工・知識工卒。2005年奈良先端科学技術大学院大学情報科学研究科修了。修士(工学)。現在、日本IBM勤務。

##### 佐藤 智和



1999年阪府大・工・情報工卒。2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。現在、同大情報科学研究科助手。コンピュータビジョンの研究に従事。2001年電子情報通信学会学術奨励賞受賞。電子情報通信学会、情報処理学会、IEEE各会員。

##### 横矢 直和 (正会員)



1974年阪大・基礎工・情報工卒。1979年同大学院博士後期課程了。同年電子技術総合研究所入所。以来、画像処理ソフトウェア、画像データベース、コンピュータビジョンの研究に従事。1986~87年マツギル大・知能機械研究センター客員教授。1992年奈良先端科学技術大学院大学・情報科学センター教授。現在、同大情報科学研究科教授。1990年情報処理学会論文賞受賞。2005年情報処理学会フェロー。工博。電子情報通信学会、情報処理学会、人工知能学会、日本認知科学会、映像情報メディア学会、IEEE各会員。