基礎論文

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

薄 充孝[*1]　中川 知香[*1]　佐藤 智和[*1]　横矢 直和[*1]

Extrinsic Camera Parameter Estimation from a Still Image Based on Feature Landmark Databse

Mitsutaka Susuki[*1], Tomoka Nakagawa[*1], Tomokazu Sato[*1] and Naokazu Yokoya[*1]

Abstract – **In this paper, we propose a novel method that estimates camera position and posture from a single image using a feature landmark database. Conventionally, several kinds of camera parameter estimation methods are proposed in which pre-constructed database is used to estimate the camera parameters. In the most of these methods, they achieve fast and high accurate estimation by limiting searching range for the database using the assumption that camera motion for successive image frames is small. However, if the input is a single image, these approaches do not work because there is no good initial parameter to limit the searching range for the database. In this research, we gradually limits the searching range for the landmark database by using GPS position, SIFT distance, and consistency of camera position and posture. The validity of the proposed method has been shown through experiments for an outdoor environment.**

Keywords : **Extrinsic camera parameter estimation, Augmented Reality, Feature landmark database**

1　はじめに

近年，携帯電話によるヒューマンナビゲーションやユーザ位置に依存した情報配信が実用化されている．現在実用化されているナビゲーションシステムでは，携帯機器に内蔵されたGPSや複数の基地局からの電波強度を利用してユーザの位置を特定し，地図上での道案内を実現している．これらのナビゲーションシステムでは，一般に二次元の地図を利用してユーザの位置やガイド情報をディスプレイ上に提示しているが，ユーザにとって実環境と二次元地図上のガイド情報の関係を正しく把握することは必ずしも容易ではない．これに対して，拡張現実感技術を用いたナビゲーションシステムが複数提案されている [1, 2, 3, 4, 5]．これらのシステムでは，実環境を撮像するカメラの位置・姿勢を推定し，仮想環境と現実環境の位置合わせを行った上で，実環境の映像上にナビゲーション情報やガイド情報を重畳表示することで，ユーザに対する直観的なナビゲーションを実現している．

このような拡張現実感によるナビゲーションを実現するためのカメラ位置・姿勢推定手法として，GPSやジャイロなどのセンサを組み合わせて用いる手法やカメラから得られる画像を用いる手法などが提案されている．前者は，複数のセンサにより得られる計測値を組み合わせてカメラの位置・姿勢を推定することで比較的ロバストな推定を実現しているが，センサの実装コスト・サイズの問題から一般的な携帯電話で利用することは難しい．後者は，入力画像中の自然特徴点やエッジなどの自然特徴をあらかじめ構築されたデータベース内の情報と照合することで，特殊なセンサを用いることなくカメラ位置・姿勢を推定する．しかし，計算機リソースの乏しい携帯端末上において，広域環境を対象としたデータベースを保持することや，比較的計算コストの高い特徴照合の処理を行うことは困難である．これに対して，サーバ・クライアント型の枠組みを用い，サーバ上でデータベースと入力画像上の特徴を照合することでカメラ位置・姿勢推定を実現するアプローチが考えられるが，ネットワーク帯域の問題，サーバ側の計算負荷の問題，伝送遅延の問題から，動画像をサーバに伝送し処理することは現実的ではない．これに対して本研究では，以下のような枠組みで静止画像一枚からカメラ位置・姿勢推定を行うことで，ネットワーク帯域の問題およびサーバにおける計算負荷の問題を回避する．

(1) ユーザは携帯電話に内蔵されたカメラで静止画を撮影し，GPSや電波強度による位置情報と共に写真をサーバに送信する．

(2) サーバは事前に構築されている環境のデータベー

[*1] 奈良先端科学技術大学院大学　情報科学研究科
[*1] Nara Institute of Science and Technology

スと写真を照合し，カメラ位置・姿勢を推定する．

(3) サーバは受信した写真にナビゲーション情報および各種ガイド情報を合成し，ユーザに返送する．

上記の枠組みにより，現在既に市販されている携帯電話や，将来においても大勢を占めるであろう比較的安価で低機能な携帯端末上においても拡張現実感によるナビゲーションを実現することができる．ただし，静止画像一枚を対象とした場合には，従来手法の多くが用いてきた時系列情報を用いたデータベースの探索範囲の限定が行えないため，比較的広い空間範囲に対応するデータベースから，入力画像中の自然特徴と正しく対応づく対応点を探索する必要がある．

提案手法の処理の流れを図 1 に示す．本研究では，まず対象となる環境を全方位カメラで撮影し，オフライン処理で structure from motion 法 (以下，SFM 法) による三次元復元を行うことで，自然特徴点の三次元位置とその見え方の情報をランドマークとしてデータベースに登録する (以降，本論文では三次元位置と見え方の情報が既知の自然特徴点をランドマークと呼ぶ)．オンライン処理では，データベースに登録された膨大な数のランドマークから，入力画像中の自然特徴点に対応する正しいランドマークを検索するために，GPS または電波強度による位置情報，ランドマークの類似度，ランドマーク観測時のカメラ位置の整合性，を順に用いてカメラ位置・姿勢推定に利用するランドマークを段階的に絞り込み，最終的に誤対応を排除した上で 6 自由度でのカメラ位置・姿勢推定を行う．

以下，2 節では本研究に関連が深い画像を用いたカメラ位置・姿勢推定の従来手法について概観し，本手法の位置づけを述べる．3 節では，本研究で用いるランドマークデータベースの構成要素および作成方法について詳述し，続いて 4 節では，作成したランドマークデータベースを用いた静止画像一枚からのカメラ位置・姿勢推定手法について述べる．また，5 節では，実際に携帯電話に内蔵されたカメラで撮影した画像を用いたカメラ位置・姿勢推定実験について報告し，最後に，6 節でまとめと今後の課題について述べる．

2　関連研究

本節では，拡張現実感への応用が可能なカメラ位置・姿勢の推定手法を分類し，本研究の位置づけを述べる．画像を用いてカメラの位置・姿勢を推定する手法は，事前知識を用いない手法と，事前知識として各種のデータベースを用いる手法に大別することができる．

事前知識を用いない手法は SLAM(Simultaneous Localization and Mapping) と呼ばれ，動画像上の自然特徴点の動きから自然特徴点の三次元位置とカメラ位置・姿勢を同時に推定する．このようなアプローチを用いれば拡張現実環境をその場で簡単に構築できるため，近年盛んに研究されている [6, 7]．SLAM は，コンピュータビジョンの分野で研究されてきた SFM 法を，オンラインで実行可能なように拡張・改良したものであると言えるが，広域な環境において SLAM のようなオンライン推定の枠組みを用いた場合には誤差の蓄積が問題となり，何らかの手法を用いて蓄積した誤差を解消する必要がある．また，SLAM によって得られる自然特徴点の三次元位置およびカメラ位置・姿勢は相対的なものであるため，位置に依存した情報を利用するナビゲーションなどのアプリケーションにそのまま応用することはできない．これに対して，本研究では，3 節で述べる GPS または基準点を用いたオフラインでの最適化処理を行う SFM 法をデータベースの構築に用いることで誤差の累積の問題を解決し，また同時にランドマークの三次元位置を絶対座標で算出する．

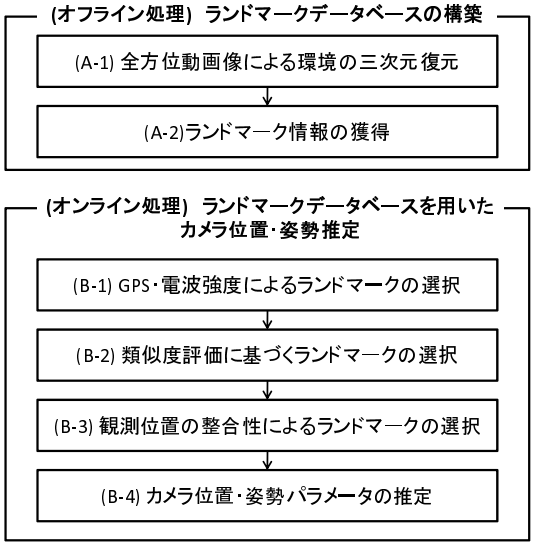

図 1　提案手法の処理の流れ
Fig. 1　Flow diagram of proposed method.

事前知識を用いてカメラ位置・姿勢推定を行う手法は，事前知識として用いるデータベースによって以下のように分類できる．

(a) 3 次元位置関係が既知の人工的なマーカを用いる手法 [8, 9, 10, 11]

(b) 環境中を事前に撮影した画像群とその撮影位置・姿勢情報などの付加情報から成る画像データベースを用いる手法 [12, 13, 14]

(c) 事前に作成した環境中の三次元モデルやランドマークなどを用いる手法 [15, 16, 17, 18, 19]

人工的なマーカを用いる手法 (a) では，環境内に配置した位置関係が既知の多数のマーカを利用すること

でカメラ位置･姿勢の推定を行う．これらの手法の多くは動画像を入力として想定しているが，大半の手法は静止画像一枚を入力とした場合にも利用できる．しかし，広域環境へのマーカの配置に多大な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り，特に屋外環境での利用は難しい．

画像データベースを用いる手法 (b) では，環境を事前に撮影した画像群とその撮影位置･姿勢情報などの付加情報からなるデータベースを利用する．岩佐ら [12] や Kourogi ら [13] は，全方位画像をデータベースに登録し，入力画像との見え方が近い位置・姿勢を探索す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．これらの手法は，入力として静止画像を扱うことができるが，データベース中で最も類似した画像の撮影位置をそのまま入力画像のカメラ位置として出力するため，精度の高いカメラ位置・姿勢が要求される拡張現実感への応用は難しい．これに対して，Cipolla ら [14] は画像データベースを用いたカメラ位置･姿勢推定の精度を高めるために，画像中に写るビルなどの輪郭や壁面上から縦・横方向の多数の平行線を用い，これらを用いてデータベース中の画像と入力画像間の平面射影変換行列を算出することで，入力画像と最も類似した画像からの相対的なカメラ位置・姿勢を推定する手法を提案している．この手法では，データベース構築時に，二次元地図上の建物の正面に相当する輪郭線を画像上で指定しデータベースを二次元地図と対応付けておくことで，静止画像一枚を入力とした二次元地図上での撮影位置と方位の推定を実現している．しかし，二次元地図を基礎としているため，建物の高さ方向に関する情報が与えられず，最終的にカメラ位置の高さ成分と姿勢の仰角成分を推定することができない．また，対象となる実環境に多数の平面や平行線が存在していることを前提としており，利用可能な環境が限定されるという問題がある．

三次元モデルやランドマークを用いる手法 (c) では，ワイヤフレームモデルなどによる環境の三次元モデルや，ランドマークから成る三次元情報を含むデータベースを事前に作成しておき，それらを入力画像上の自然特徴 (エッジや自然特徴点) と対応付けることでカメラの位置・姿勢を算出する．このようなアプローチを用いれば，マーカなどを用いることなく 6 自由度でカメラ位置・姿勢推定を推定できるため，マーカを用いないカメラ位置・姿勢推定手法として盛んに研究されている．このような従来手法の大半は動画像を扱うことを前提としており，一般的に時系列情報 (例えば前フレームのカメラ位置・姿勢の推定結果) を用いることで，利用するデータベースの範囲をフレーム毎に限定し，推定のロバスト性を高めている．しかし，推定開始時点における初期フレームでは時系列情報を用いることができないため，初期フレームにおけるカメラ位置・姿勢推定の問題 (初期化の問題) は，それ以降のフレームに対する推定の問題とは問題設定が異なる．本研究で扱う静止画像一枚からのカメラ位置・姿勢推定の問題は，このカメラ位置・姿勢の初期化の問題と本質的に同等の問題設定となる．

このような初期化の問題に対して，Skrypnyk ら [19] は，SIFT 特徴点を用いてデータベース構築用の画像群から初期フレームの画像に最も類似した画像を探索し，その画像取得時のカメラ位置・姿勢を初期フレームのカメラ位置・姿勢として与えている．この手法では，データベース構築時のカメラ位置から離れた場合において正しく初期化を行うことができない．また，この手法は比較的小規模な環境を対象としている (上記の手法では 20 枚程度の画像からデータベースを構築している) が，広域環境での利用を前提とした場合には，データベース中に見え方が類似したランドマークが多数登録されるため，自然特徴点の対応付けにおける誤対応率が過半数を上回り，正しいカメラ位置・姿勢を与えることが難しいという問題がある．一方，あらかじめ物体ごとに分類された特徴を用いて入力画像中に撮像されている物体を認識することで，撮影されている物体に対するカメラ位置・姿勢の初期化を行う手法 [20, 21] が提案されているが，類似した物体が多数存在し，また物体自体の分離が難しい広域環境に対するカメラ位置･姿勢推定に利用することは難しい．また，エッジ特徴を用い投票に基づいて初期化を行う手法 [22] が提案されているが，エッジのみから得られる特徴量は乏しく，この手法では傾斜角センサの利用が前提となっている．Reitmayr ら [23] は，屋外環境において GPS により計測された位置に対して，計測位置を中心とした一定の範囲内に複数の位置候補を生成し，各地点において実際に推定処理を行った上で信頼度の高い推定結果を初期値として採用している．この手法では，姿勢は磁気コンパスと傾斜角センサから得られる情報をそのまま用い，位置についても高さは GPS から得られるデータの平均値を用いることで，推定すべきパラメータの自由度を 2 に限定している．

本研究では，上記の動画像を対象としたカメラ位置・姿勢の初期化に関する従来研究ではあまり扱われてこなかった比較的広域で複雑な屋外環境を対象とし，データベース中に多数の類似したランドマークが存在する場合にも画像に基づいて 6 自由度のカメラ位置・姿勢推定を実現する手法を提案する．提案手法では，入力画像中に存在する自然特徴点と類似したランドマークを観測可能な空間中の位置・姿勢に投票を行うことで，誤対応を効果的に排除する．

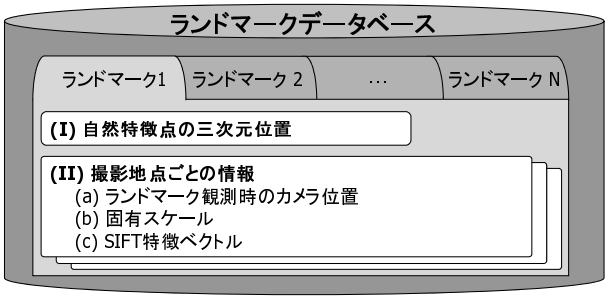

図 2　ランドマークデータベースの構成要素
Fig. 2　Elements of feature landmark database.

3　ランドマークデータベースの構築

本節では，オフラインでランドマークデータベースを構築する手法について述べる．図 1 に示したように，本研究では，まず撮影対象となる環境を全方位カメラを用いて撮影し，structure from motion 法によって自然特徴点の三次元位置と全方位カメラのカメラパラメータを算出する (A-1)．次に，三次元復元結果に基づきランドマークデータベースを構築する (A-2)．以下では，これらの処理について詳しく述べる．

3.1　全方位カメラによる環境の三次元位置復元

本研究では，対象となる環境を多眼型の全方位カメラを用いて移動撮影し，Harris オペレータ [24] を用いて動画像中から物体の角などの自然特徴点を抽出する．次に，抽出された自然特徴点を画像間で対応づけ，bundle adjustment[25] の枠組みで再投影誤差を最小化することで，自然特徴点の三次元位置と全方位カメラの位置・姿勢を推定する．ここで，実環境と自然特徴点の三次元位置の間の幾何学的な位置関係は，基準マーカと自然特徴点を同時に画像中で追跡する手法 [26] または GPS による位置計測情報と画像からの三次元復元情報を併用する手法 [27] を用いて求める．

3.2　ランドマーク情報の獲得

ランドマークデータベースの構成要素を図 2 に示す．ランドマークデータベースは (I) 自然特徴点の三次元位置と (II) 撮影地点ごとの情報から成る．ランドマークの見え方は撮影地点によって異なるので，本研究では各ランドマークに複数の撮影地点情報を登録する．撮影地点情報は，(II-a) ランドマーク観測時のカメラ位置，(II-b) 自然特徴点の固有スケール，(II-c)SIFT 特徴ベクトルから成る．本研究では，前節の三次元復元結果を，自然特徴点の三次元位置 (I) およびランドマーク観測時のカメラ位置 (II-a) としてそのまま利用する．また，自然特徴点の固有スケール (II-b)，SIFT 特徴ベクトル (II-c) の算出には，それぞれ Harris-Laplacian[28] および SIFT-descriptor[29] を用いる[1]．以下では，自然特徴点の固有スケール (II-b) および SIFT 特徴ベクトル (II-c) の算出手法について詳述する．

(自然特徴点の固有スケールの算出)

3.1 節の手法によって三次元位置が推定されたすべての画像特徴点に対して，特徴点の固有スケールを算出する．特徴点の固有スケールとは，特徴点周辺のテクスチャの局所構造から決定される特徴点固有のスケールであり，これを用いることで，物体とカメラ間の距離の変化による画像スケールの変化や画像解像度の変化が起こった場合にも，一定の空間領域に対応する画像の局所領域を切り出し，特徴点を正しく対応づけることができる．また，入力画像から検出した自然特徴点とそれに対応づけられたデータベース中のランドマークの固有スケールの比によって，4 節で述べるオンライン推定におけるカメラ位置とランドマークの間の距離を推定できる．これらについては 4 節で詳しく述べる．以下では固有スケールの算出手法について述べる．

ここではまず，カメラの姿勢変化による画像上の自然特徴点の見え方の違いをなくすためにレンズ歪みを排除した上で，カメラの投影中心を中心とする球面上に撮影した全方位画像を投影する．次に，球面画像上における自然特徴点の位置を中心として，スケール σ を変化させながら式 (1) で表される Laplacian-of-Gaussian(LoG) フィルタを適用し，極値をとるスケールを自然特徴点の固有スケールとする [28]．

$$f(r) = -\frac{r^2 - 2\sigma^2}{2\pi\sigma^6} exp(-\frac{r^2}{2\sigma^2}) \qquad (1)$$

ここで，スケール σ は Gaussian フィルタのサイズを表しており，r はオペレータ中心から注目画素までの距離を示す．図 3 は，同一のランドマークを異なる距離で撮影した 3 枚の画像に対して，LoG 値を算出した例である．同図から，LoG の極値に対応するスケールが，全ての画像上において同一の空間範囲に対応していることが分かる．本研究では，LoG の極値に対応したスケールを自然特徴点の固有スケール ω とすることで，自然特徴点までの撮影距離の違いにより画像スケールが異なっていても，空間的に同範囲の領域を決定する．

(SIFT 特徴ベクトルの生成)

上記の手法により算出された自然特徴点の固有スケー

[1] 文献 [29] では，画像の特徴を記述する SIFT-descriptor と共に，自然特徴点の位置と固有スケールを算出する SIFT-detector が提案されているが，本研究では三次元復元処理における自然特徴点の抽出に Harris 特徴点を用いているため，固有スケールの抽出にも Harris 特徴点に適した Harris-Laplacian を採用する．

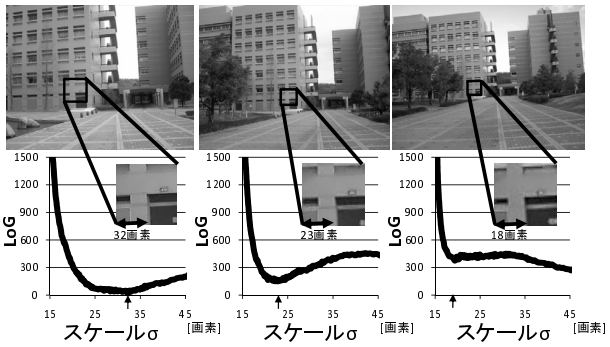

図 3　Harris-Laplacian による自然特徴点の固有スケールの算出
Fig. 3　Characteristic scale detection by Harris-Laplacian.

ルに対応する球面投影画像上の自然特徴点周辺の画像から，SIFT-descriptor[29] を用いて画像の回転に対して不変な特徴ベクトルを生成する．本研究では，Harris-Laplacian により算出された自然特徴点の固有スケール ω を用いて，自然特徴点の座標を中心とする半径 $k\omega(k$: 定数) の領域内の画像データから D 次元の SIFT 特徴ベクトル $\mathbf{f} = (v_1, \cdots, v_D)$ を生成する．

4　静止画像からのカメラ位置・姿勢推定

本節では，3 節で述べた手法により作成したランドマークデータベースを用いて，一枚の静止画像から撮影時のカメラ位置・姿勢を推定する方法について述べる．本手法では，まず GPS や携帯電話の電波強度を用いておおよそのカメラ位置を特定することでデータベースの選択を行う (B-1)．次に，SIFT を用いた類似度評価により入力画像上の自然特徴点と対応付くランドマークの候補を複数選択する (B-2)．更に，選択された各ランドマークが，それぞれ対応付いた入力画像上の自然特徴点と同じ見え方で観測できるカメラ位置・姿勢を算出し，投票によって 1 地点から最も多くのランドマーク候補を観測できるカメラ位置・姿勢の候補を決定する (B-3)．最後に，決定されたカメラ位置・姿勢の候補に投票したランドマーク群を用いて最終的に 6 自由度のカメラ位置・姿勢推定を行う (B-4)．以下では各処理について順に述べる．

4.1　GPS・電波強度によるランドマークの選択

現在，携帯電話に内蔵された GPS によって 10m から 100m 程度，複数の基地局からの電波強度を用いて 100m から 200m 程度の誤差を含んだ端末位置の特定が可能である．本研究では，あらかじめ多数の地点・地域においてランドマークデータベースが構築されていることを想定し，GPS または電波強度を用いて検索に用いるデータベースを選択する．

ここでは，ランドマーク観測時のカメラ位置を基準に，あらかじめランドマークデータベースが 100m×100m 程度の単位に分割されているものとし，まず GPS・電波強度によって得られる観測位置周辺に存在するデータベースを全て選択する．次に，選択されたデータベースに登録されているランドマークから，GPS・電波強度による計測地点を中心とした半径 γ[m] 内の領域に，ランドマーク観測時のカメラ位置が存在するものを全て，以降の処理で用いるランドマークとして選択する．半径 γ は計測誤差および携帯機器から最も近いランドマーク観測時のカメラ位置までの距離を考慮して決定する必要がある．

ただし，ここで述べた手法は大規模なデータベースを扱うことを想定した場合に必要となるものであり，本論文では，上記のランドマークの選択について，実装および実験による検証は行わない．

4.2　類似度評価に基づくランドマークの選択

入力画像上で検出された自然特徴点と見え方が類似したランドマークをデータベースから選択する．ここではまず，入力画像上の自然特徴点を Harris オペレータによって抽出し，抽出された各自然特徴点に対する SIFT 特徴ベクトル $\mathbf{f}' = (v'_1, \cdots, v'_D)$ をデータベース構築時と同様の手法によって算出する．次に，以下の式を用いて，入力画像上の自然特徴点とランドマークの類似度 S を算出する．

$$S = |\mathbf{f}' - \mathbf{f}|^2 = \sum_{d=1}^{D} (v'_d - v_d)^2 \qquad (2)$$

最後に，各自然特徴点に対して算出された類似度 S を昇順に並び替え，S が一定の閾値以下の上位 α 個のランドマークを自然特徴点と対応づける．これにより，画像上の各自然特徴点との類似度の高い複数のランドマークをデータベース中から選択する．

4.3　観測位置の整合性によるランドマークの選択

前節の処理で対応付けられた自然特徴点とランドマークの組み合わせには，自然特徴点と真に対応するランドマーク以外の誤対応が多数存在する．提案手法では，このような誤対応を排除するために，入力画像が環境中の単一の位置・姿勢で撮影されているという事実に着目し，前節の処理で選択されたランドマークを最も多く観測可能なカメラ位置・姿勢を投票によって算出する．またこれにより，投票値が最大となったカメラ位置・姿勢以外に投票したランドマークを排除する．

まず，GPS または電波強度から得られる誤差を含む端末位置 (g_x, g_y, g_z) を中心とする一定範囲の領域を，図 4 に示すように，地面に対して水平方向に格子状に分割し，それぞれの格子に水平方向の姿勢の回転に対応する l 個（360 度を $360/l$ 度ずつ分割）の投票箱を設置する．ここでは，世界座標系における $(2h+1)\times(2h+1)$

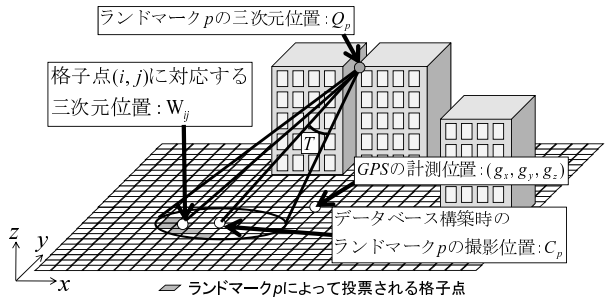

図 4　1 つのランドマークからの投票例
Fig. 4　Voting from a landmark.

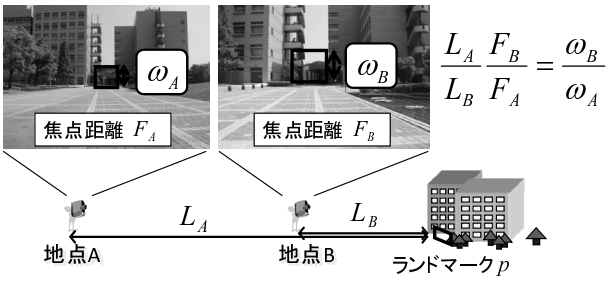

図 5　自然特徴点の固有スケールとランドマーク撮影距離の関係
Fig. 5　Relationship between characteristic scale and camera-landmark distance.

個の格子点の xy 座標 $\mathbf{w}_{ij}(-h \leq i \leq h, -h \leq j \leq h)$ を以下のように定義する．

$$\mathbf{w}_{ij} = \begin{pmatrix} w_i \\ w_j \end{pmatrix} = \begin{pmatrix} g_x + iL \\ g_y + jL \end{pmatrix} \tag{3}$$

ただし，L は格子間隔を表す．

次に，前節の処理で選択された各ランドマークが，それぞれ対応付いた入力画像上の自然特徴点と同じ見え方で観測できるカメラ位置・姿勢を算出し，投票を行う．ここでは，まず自然特徴点の固有スケールを用いることで，カメラと各ランドマークの間の距離を算出する．図 5 に示すように，ランドマーク p との間の距離が異なる 2 つの地点 A, B において撮影された 2 枚の画像上で，同一のランドマーク p に対応する自然特徴点の固有スケールがそれぞれ ω_A, ω_B であるとする．ただし，以下では簡単のために，地点 A, 地点 B において用いられたカメラの画素サイズ (1 画素の物理的な大きさ) の違いはあらかじめ正規化されているものとする．このとき，固有スケール ω_A, ω_B に対応する実空間中での範囲は同一である．物体の画像上での大きさは，物体とカメラ間の距離に反比例し，カメラの焦点距離に比例するため，地点 A, B とランドマーク p の間の距離をそれぞれ L_A, L_B, 地点 A, B におけるカメラの焦点距離をそれぞれ F_A, F_B とおけば，$L_A F_B / L_B F_A = \omega_B / \omega_A$ が成り立つ．

よって，図 4 に示すように，ランドマーク p の三次元位置を $\mathbf{Q}_p$, ランドマーク p を撮影したデータベース構築時のカメラ位置を $\mathbf{C}_p$ とし，地点 $\mathbf{C}_p$ の画像上におけるランドマーク p の固有スケールを ω_p，ランドマーク p に対応づけられた入力画像上の自然特徴点 q の固有スケールを ω_q とすれば，以下の式によって，ランドマーク p と入力画像を撮影したカメラ位置の間の距離 L_p を算出することができる．

$$L_p = |\mathbf{C}_p - \mathbf{Q}_p| \frac{\omega_p F_{MD}}{\omega_q F_{DB}} \tag{4}$$

ただし，F_{DB} はデータベース構築時のカメラの焦点距離を, F_{MD} は携帯端末に内蔵されたカメラの焦点距離を表わす．

ここでは，前節において選択された各ランドマーク p に対して算出された距離 L_p を用いて，以下に示す条件 (1), (2) を同時に満たすすべての格子点 $\mathbf{w}_{ij}$ の m 番目の投票箱に投票する．ただし，以下の条件における $\mathbf{W}_{ij}$ は, 格子点 $\mathbf{w}_{ij}$ に $\mathbf{C_p}$ の高さ成分を与えた三次元位置 (w_i, w_j, c_z) を表わす．

条件 (1) 格子点とランドマークの間の距離が式 (4) によって算出される距離 L_p に一致すること．すなわち，$1-\alpha < \frac{|\mathbf{W}_{ij}-\mathbf{Q}_p|}{L_p q} < 1+\alpha$ (α:定数).

条件 (2) 格子点からランドマーク p への視線方向がデータベース構築時と一致すること．すなわち，$\frac{(\mathbf{Q}_p-\mathbf{C}_p)\cdot(\mathbf{Q}_p-\mathbf{W}_{ij})}{|\mathbf{Q}_p-\mathbf{C}_p||\mathbf{Q}_p-\mathbf{W}_{ij}|} \geq \cos(T)$ (T:定数).

なお，m は以下の式により決定する．$m = [\theta l / 2\pi]$．ただし，$[a]$ は a を超えない最大の整数を表し，世界座標系における x 軸に対するベクトル $\mathbf{W}_{ij} - \mathbf{Q}_p$ の水平面上での回転角を θ[ラジアン] とする．以上の処理により得られた投票結果に対して，投票数が最大となる位置・姿勢に投票したランドマークを抽出し，次節で述べるカメラパラメータ推定処理に用いる．

4.4　カメラ位置・姿勢パラメータの推定

投票により抽出されたランドマークと各ランドマークに対応けられた入力画像上の自然特徴点を用いて 6 自由度のカメラの位置・姿勢を推定する．ここでは，各ランドマークを画像上に投影した座標と，各ランドマークに対応する自然特徴点の画像座標の間の二乗距離 (再投影誤差) の総和を最小化することで 6 自由度のカメラ位置・姿勢を推定する．ただし，前節の投票結果には誤対応が含まれるため，ここでは投票結果からランダムに自然特徴点・ランドマークの組を繰り返し抽出し，LMedS 基準 [30] を最小とする自然特徴点・ランドマークの組を抽出することで，誤対応を含まない自然特徴点・ランドマークの組からカメラ位置・姿勢を選択する．

ここで，最終的に得られるカメラ位置･姿勢パラメータに対して，再投影誤差の平均値が R 画素を上回る

図 6　データベースの構築に用いた全方位型マルチカメラシステム Ladybug と入力画像の一部
Fig. 6　Omni-directional multi-camera system and sampled frames of input video sequences for database construction.

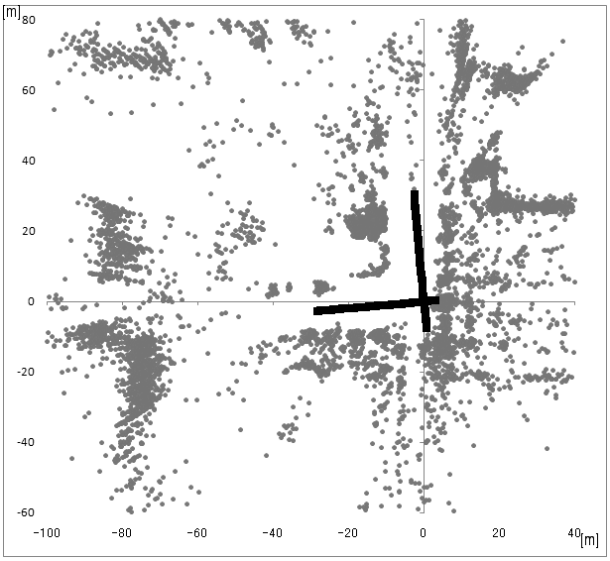

図 7　データベースに登録された自然特徴点と全方位カメラの位置
Fig. 7　Position of registered landmarks and omni-directional camera for database construction.

場合には，システムは推定結果を失敗と判定する．この場合には，ユーザは入力画像を撮り直し，再度カメラ位置・姿勢推定を実行する必要がある．

5　実験

提案手法の有効性を示すために，屋外環境においてランドマークデータベースを構築し，提案手法によるカメラ位置・姿勢推定の精度と推定成功率を検証した．

5.1　データベースの構築

本実験ではまず，図 6 左に示す全方位型マルチカメラシステム (PointGreyResearch 社製 Ladybug) を用いて，実験対象となる環境内の 2 本の経路上を移動しながら撮影し，ランドマークデータベースを構築した．Ladybug は水平方向に 5 つ，上方向に 1 つの合計 6 つのカメラユニットから構成されており，各カメラユニットはそれぞれ 768 × 1024 画素の解像度で動画像を同期撮影できる．ここでは，図 6 右に示す 6 枚の画像を含む，合計 7,200 枚 (1,200 フレーム) の撮影画像を入力として用い，3.1 節で述べた手法を用いて，動画像上の自然特徴点の動きから自然特徴点の三次元位置と全方位カメラのカメラパラメータを復元した．また，撮影した環境中において自然特徴点をトータルステーションで計測し，その三次元座標を全方位動画像のキーフレームで手動で指定することで，トータルステーションでの計測時に設定した実環境中の世界座標系に対するランドマークの三次元位置と全方位カメラの位置を得た [26]．

次に，得られた三次元復元結果を用いて 3.2 節で述べた手法によりデータベースを構築した．本実験では，SIFT による特徴ベクトルの次元数 D を 128 次元とした．図 7 に，データベースに登録されたランドマークの三次元位置と全方位カメラの位置の関係を，地表に対する上面図として示す．同図中の太い実線は全方位カメラのカメラパスを，点群はランドマークの位置を表している．本実験では，データベースに約 12,500 個のランドマークが登録され，各ランドマークに対して平均 7.5 地点の撮影地点ごとの情報が登録された．

5.2　カメラ位置・姿勢の推定

市販の GPS・カメラ付き携帯電話 (Casio 社製 GzOne W42CA) を用いて撮影した静止画像を用い，提案手法によるカメラ位置・姿勢の推定結果と真値を比較することで推定精度を評価する．カメラ位置・姿勢の真値は，あらかじめ環境内の自然特徴点をトータルステーションで計測し，入力画像上で手動で位置を指定した上で再投影誤差の最小化によりカメラ位置・姿勢を算出することで作成した．本実験ではサーバ・クライアント型システムは構築せず，携帯電話による画像撮影後に PC 上に画像を転送した上でカメラ位置・姿勢推定処理を行った．また，用いたデータベースの規模が比較的小さいため，4.1 節で述べたデータベースの選択は行わず，今回は登録されたすべてのランドマークを用いて実験を行った．なお，携帯電話に内蔵されたカメラの内部パラメータは Tsai の手法 [31] によってあらかじめ校正した．

まず，図 8 に示すように，データベース構築時の全方位カメラの撮影経路周辺において，5m 間隔の格子点上 ($6 \times 6 = 36$ 地点) で，異なる 2 方向 (方向 1, 方向 2) に対して 640 × 480 画素の 72 枚の静止画像を撮影した．本実験では，72 枚の画像のうち，カメラ位置・姿勢の真値を作成することが可能な 65 枚の画像を入力画像として用いた．次に，表 1 に示すパラメータを用いて各画像に対するカメラ位置・姿勢推定を行った．

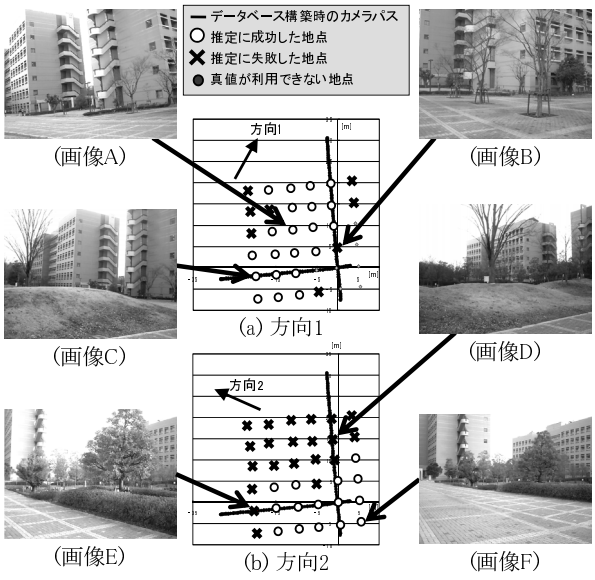

図 8　入力画像の撮影地点，撮影画像例と各地点における推定成功・失敗の判定結果
Fig. 8　Positions of input images, examples of input images and result of success/fail judgement.

表 1　カメラ位置･姿勢の推定に用いたパラメータ
Table 1　Given parameters for estimation.

条件 1 の閾値 α	0.2
閾値 T(度)	10
再投影誤差の閾値 R（画素）	5.0
投票箱の設計パラメータ (h, L, l)	(80, 0.5m, 72)

ただし，本実験では，投票空間の設定に GPS による計測値は用いず，上記の格子点の重心位置を中心とする 80m × 80m の範囲を投票空間として用いた．

図 8 に，4.4 節で述べた再投影誤差による基準に基づきシステムが判定した，撮影地点ごとの成功・失敗の結果を示す．また，表 2 に，方向 1 および方向 2 におけるカメラ位置・姿勢推定の成功率，推定精度および再投影誤差の平均と標準偏差を示す．画像中に人工物を主に捉えることができる方向 1 に関しては，多くの入力画像で推定に成功した．図 8 の画像 C に対する位置の投票結果を一部拡大したものを図 9(a) 左に，姿勢に関する投票結果を同図右に示す．ただし，位置の投票結果は，位置 2 次元・姿勢 1 次元から成る三次元の投票空間 (x,y,θ) に対して，姿勢方向に対する投票値を各位置でそれぞれ累計したものを表わしており，濃度値が濃いほど投票数が多い．また，姿勢の推定結果は，真値を通過する $x\theta$ 平面上の投票結果を表わしている．この例では，投票数が最大となった位置が真値から約 3.5m 離れているが，4.4 節で述べた手法により誤対応を排除した上でカメラ位置・姿勢を推定することで，最終的なカメラ位置・姿勢の推定誤差はそれぞれ約 1.2m, 0.7 度となった．方向 1 に対して成功

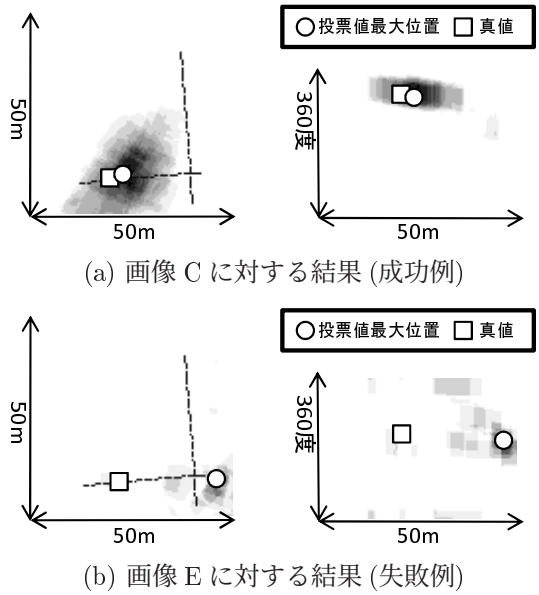

図 9　推定成功・失敗地点における投票結果の例
Fig. 9　Examples of voting results.

と判定された全ての画像に対する推定誤差の平均は，位置誤差が 1.4m，姿勢誤差が 1.4 度であり，これは静止画像による拡張現実感ヒューマンナビゲーションを行うために問題ない精度であると考えられる．

これに対して方向 2 では，撮影された半数以上の画像上で画像中に自然物が大きく写り込んでおり，方向 1 に比べ推定成功率が低い．図 9(b) は，画像 C 撮影時と同一の地点から方向 2 を撮影した画像 E に対する投票結果である．この例では，投票値が分散したため投票値が最大となった位置が真値から大きく離れており，最終的なカメラ位置・姿勢推定処理において，再投影誤差が閾値 $R = 5.0$ 画素を上回った．このためシステムは推定結果を失敗と判定している．方向 2 に対して推定が成功と判定された全ての画像に対する推定誤差の平均は，位置誤差が 6.8m，姿勢誤差が 3.9 度であり，方向 1 に比べて推定精度が低い．この原因として，多くの地点でカメラ位置・姿勢推定に用いられた建物上のランドマークとカメラ位置の間の距離が，方向 1 の場合と比較して，撮影場所により 3 から 5 倍程度離れていること，また自然物による遮蔽によりカメラ位置・姿勢の推定に利用できるランドマーク数が少ないことが挙げられる．

以上のことから，ランドマークを多数検出可能な建物などの人工物を捉えた画像に対して，提案手法は比較的精度の高いカメラ位置・姿勢を実現できることが分かる．しかし，自然物が画像上の大半の領域を占める入力画像に対しては，ランドマークと自然特徴点が正しく対応付かず，投票値が分散するためにカメラ位置・姿勢推定が失敗または推定精度が低下することが分かる．

表 2　方向 1・方向 2 におけるカメラ位置・姿勢推定の成功率，推定精度および再投影誤差
Table 2　Success rate and accuracy of estimated results.

	方向 1	方向 2
システムの判定による推定成功率 (%)	72.4	41.7
平均位置誤差 (m)	1.4	6.8
位置誤差の標準偏差 (m)	2.5	9.1
平均姿勢誤差 (度)	1.4	3.9
姿勢誤差の標準偏差 (度)	2.0	4.5
平均再投影誤差 (画素)	2.0	2.0
再投影誤差の標準偏差 (画素)	0.9	1.1

6　まとめ

本論文では，サーバ・クライアント方式による携帯電話上での拡張現実感への応用を想定し，事前に構築したランドマークデータベースを用いる静止画像一枚からの新たなカメラ位置・姿勢推定手法を提案した．本手法では，静止画像一枚からのカメラ位置・姿勢推定が可能であるため，現在普及しているカメラ付き携帯電話をそのまま用いることができるという特長を持つ．実験により，人工物を入力画像中に十分捉えている場合には，データベース構築時のカメラ位置から離れた地点においても，静止画像へのナビゲーション情報を重畳するために問題ないと考えられる精度でカメラ位置・姿勢推定を行えることを確認した．しかし，樹木などによりランドマークの大半が遮蔽されてしまう場合には，現状の手法ではランドマークを正しく対応づけることが難しいことを確認した．

今後は，環境内に類似したランドマークが存在しない特徴的なランドマークを優先的に用いることで推定のロバスト性を高める手法を開発する．また，同一地点から異なる環境条件で撮影された複数の画像をデータベースの構築時に用いることで推定のロバスト性を高める手法についても検討する．

謝辞 本研究は，総務省戦略的情報通信研究開発推進制度 (SCOPE) により実施したものである．

参考文献

[1] T. Höllerer, S. Feiner and J. Pavlik: “Situated documentaries: Embedding multimedia presentations in the real world,” Proc. Int. Symp. on Wearable Computers, pp. 79–86, 1999.

[2] 前田真希, 小川剛史, 清川清, 竹村治雄: “ウェアラブル拡張現実感のための赤外マーカのステレオ計測と姿勢センサを用いた位置・姿勢推定”, 日本バーチャルリアリティ学会論文誌, Vol. 10, No. 4, pp. 459–466, 2005.

[3] 天目隆平, 神原誠之, 横矢直和: “「平城宮跡ナビ」マルチメディアコンテンツを利用したモバイル型観光案内システム”, 第 1 回デジタルコンテンツシンポジウム講演予稿集, No. S3-6, 2005.

[4] 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.

[5] M. Kourogi, N. Sakata, T. Okuma and T. Kurata: “Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/Self-contained sensor system,” Proc. Int. Conf. on Artificial Reality and Telexistence, pp. 1310–1321, 2005.

[6] G. Reitmayr, E. Eade and T. Drummond: “Semi-automatic annotations in unknown environments,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 67–70, 2007.

[7] G. Klein and D. Murray: “Parallel tracking and mapping for small ar workspace,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 225–234, 2007.

[8] 齊藤滋, 谷川智洋, 廣瀬通孝: “コード化した模様を内装に施した室内における位置同定システム”, 信学技報, MVE2006-1, 2006.

[9] 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”, 信学技報, ITS2003-76, 2004.

[10] 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.

[11] D. Wagner and D. Schmalstieg: “First steps towards handheld augmented reality,” Proc. Int. Symp. on Wearable Computers, pp. 21–23, 2003.

[12] 岩佐英彦, 粟飯原述宏, 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”, 信学論 (D-II), Vol. J84-D-II, No. 2, pp. 310–320, 2001.

[13] M. Kourogi, T. Kurata, K. Sakaue and Y. Muraoka: “A panorama-based technique for annotation overlay and its real-time implementation,” Proc. Int. Conf. on Multimedia and Expo (ICME 2000), pp. 657–660, 2000.

[14] R. Cipolla, D. Robertson and B. Tordoff: “Image-based localization,” Proc. Int. Conf. Virtual Systems and Multimedia, pp. 22–29, 2004.

[15] L. Vacchetti, V. Lepetit and P. Fua: “Stable real-time 3D tracking using online and offline information,” Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, No. 10, pp. 1385–1391, 2004.

[16] 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.

[17] E. Rosten and T. Drummond: “Fusing points and lines for high performance tracking,” Proc. Int. Conf. on Computer Vision, pp. 1508–1515, 2005.

[18] 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.

[19] 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.

[20] G. Blasko and P. Fua: “Real-time 3D object recognition for automatic tracker initialization,” Proc. Int. Symp. on Augmented Reality, pp. 175–176, 2001.

[21] V. Lepetit and P. Fua: “Keypoint recognition using randomized trees,” Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 9, pp. 1465–

1479, 2006.
[22] D. Kotake, K. Satoh, S. Uchiyama and H. Yamamoto: “A fast initialization method for edge-based registration using an inclination constraint,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 239–248, 2007.
[23] G. Reitmayr and T. Drummond: “Initialization for visual tracking in urban environments,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 161–160, 2007.
[24] C. Harris and M. Stephens: “A combined corner and edge detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.
[25] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: “Bundle adjustment a modern synthesis,” Proc. Int. Workshop on Vision Algorithms, pp. 298–372, 1999.
[26] 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.
[27] 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.
[28] K. Mikolajczyk and C. Schmid: “Scale & affine invariant interest point detectors,” Int. Journal of Computer Vision, Vo. 60, No. 1, pp. 63–86, 2004.
[29] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int. Journal of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.
[30] P. J. Rousseeuw: “Least median of squares regression,” J. of American Statistical Association, Vol. 79, pp. 871–880, 1984.
[31] R. Y. Tsai: “An efficient and accurate camera calibration technique for 3D machine vision,” Proc. Computer Vision and Pattern Recognition, pp. 364–374, 1986.

(2007 年 12 月 10 日受付)

［著者紹介］

薄 充孝

2005 年神戸大・工・電気電子工卒．2007 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．現在，トヨタ自動車株式会社勤務．修士 (工学)．

中川 知香

2004 年熊本電波高専・専攻科・制御情報システム工学専攻卒．2006 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．現在，富士ゼロックス株式会社勤務．修士 (工学)．

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教．コンピュータビジョン，画像からの三次元復元の研究に従事．博士 (工学)．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 年マッギル大・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．工博．電子情報通信学会，情報処理学会各フェロー．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．