

# 優先度情報を付加した自然特徴点ランドマークデータベースに基づく 実時間でのカメラ位置・姿勢推定

武富 貴史<sup>†</sup> 佐藤 智和<sup>†</sup> 横矢 直和<sup>†</sup>

<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: †{takafumi-t,tomoka-s,yokoya}@is.naist.jp

あらまし カメラ位置・姿勢を推定する技術は、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が可能である。従来提案されているランドマークデータベースを用いたカメラ位置・姿勢推定手法は、データベース構築の人的コストが低く、広域な環境を対象とした場合にもカメラ位置・姿勢推定の誤差が累積しないという特長を持つ。しかし、従来手法ではデータベース中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処理に多くの計算コストを必要とするため、実時間でのカメラ位置・姿勢の推定が困難であるという問題がある。そこで、本研究では連続フレーム間でのランドマークの追跡とランドマークへの優先度情報の付加により、対応付け処理に用いられるランドマーク数と照合される自然特徴点数を削減することで、実時間でのカメラ位置・姿勢推定処理を実現する。実験では、仮想物体の実時間合成の結果を示し、提案手法の有効性を確認する。

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

## Real-time camera parameter estimation using a feature landmark database with priorities of landmarks

Takafumi TAKETOMI<sup>†</sup>, Tomokazu SATO<sup>†</sup>, and Naokazu YOKOYA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science, Nara Institute of Science and Technology  
Takayama 8916-5, Ikoma, Nara, 630-0192 Japan  
E-mail: †{takafumi-t,tomoka-s,yokoya}@is.naist.jp

**Abstract** In the field of augmented reality, many kinds of image based camera parameter estimation methods have been proposed to achieve accurate geometric registration. To achieve geometric registration between real and virtual worlds, some kinds of databases that contain knowledge of the target environment are usually used. As one of such approaches, a landmark based camera parameter estimation method has been proposed. In this method, database construction cost is lower than other approaches and the method can be used in various environments. However, computational cost is expensive and thus it cannot work in real-time. To achieve real-time camera parameter estimation, priorities are associated with landmarks by using a previously captured video sequences. According to these priorities, unreliable landmarks are efficiently discarded to realize fast and robust camera parameter estimation. To demonstrate the effectiveness of the proposed method, the robustness and the computational cost is compared between the proposed and conventional methods.

**Key words** camera parameter estimation, landmark database, natural features, augmented reality

### 1. はじめに

近年、実世界を撮影した画像中に CG などによって表現される仮想世界を位置合わせして描画することで情報を付加する拡張現実感技術に関する研究が盛んに行われている。拡張現実感において、CG などを所定の位置に描画するためには、一般にカメラの位置・姿勢を推定することが必要となり、従来から様々な手法が提案されてきたが、本研究では特に、実世界と仮想世界の位置合わせの精度を高めやすい画像に基づくカメラ位

置・姿勢推定手法に着目する。

画像からカメラ位置・姿勢を推定する手法は、環境の事前知識を用いる手法 [1] ~ [6] と環境の事前知識を用いない手法 [7] ~ [9] に大別できる。前者には、環境の事前知識として、画像データベースを用いる手法 [1], [2] や環境の三次元モデルを用いる手法 [3], [4], 自然特徴点による三次元点群を用いる手法 [5], [6] などがある。画像データベースを用いる手法 [1], [2] では、事前に環境内を撮影した画像とその撮影位置・姿勢をデータベースに登録しておき、入力画像とデータベース中の画像を対応

付けることでおよそのカメラの位置・姿勢を推定する．しかし，6自由度で正確なカメラ位置・姿勢を推定することは難しく，拡張現実感への応用は難しい．環境の三次元モデルを用いる手法 [3], [4] では，入力画像中のエッジや自然特徴点などをあらかじめ作成した三次元モデルと対応付けることによって，実時間でのカメラの位置・姿勢推定を実現している．しかし，一般に広域で複雑な屋外環境の三次元モデルを作成することは難しく，データベースの構築に多くの人的コストを必要とするという問題がある．自然特徴点からなる三次元点群を用いる手法 [5], [6] では，自然特徴点の三次元位置と自然特徴点周辺の局所的な画像情報をランドマークとしてデータベースへ登録しておき，ランドマークと入力画像中の自然特徴点を対応付けることでカメラ位置・姿勢の推定を行う．自然特徴点の三次元位置は structure-from-motion によって自動で推定されるため，データベース構築の人的コストが低く，広域で複雑な屋外環境を対象とした場合にも比較的容易にデータベースを作成することができる．しかし，データベースに登録されている多数のランドマークと画像上の特徴点を正しく対応付けるために多くの処理時間を必要とし，実時間でのカメラ位置・姿勢の推定が困難であるという問題が残されている．これに対して，あらかじめ特徴空間を KD-Tree によって分割しておくことで，高速に対応付け処理を行う手法 [10] が提案されているが，実時間での処理には到っておらず，また，広域な屋外環境のように高次元の特徴空間上で大量のデータを扱う場合には，このようなアプローチによる探索処理の高速化を行うことは難しい．

環境の事前知識を用いない手法 [7] ~ [9] では，一般に入力画像中の自然特徴点を追跡することにより，カメラ位置・姿勢の推定と環境のモデル化を同時に行う SLAM (Simultaneous Localization and Mapping) と呼ばれるアプローチが用いられる．これらの手法では，事前にデータベースを構築する必要がなく，手軽に拡張現実感環境を構築できるという特長があるが，広域な環境を対象としたカメラの位置・姿勢推定では，カメラ位置・姿勢の推定誤差が蓄積するという問題があり，利用できる範囲が限定される．また，SLAM によるカメラ位置・姿勢の推定手法は，それ単体では現実環境に対する絶対的なカメラ位置・姿勢やスケール情報を取得できないため，拡張現実感技術を用いたナビゲーションのような実環境の位置に依存した情報の提示を必要とするアプリケーションにそのまま用いることはできない．

本研究では，画像からの位置・姿勢推定の問題に対して，ランドマークデータベースを用いる手法に着目し，従来手法 [6] で問題となっていたカメラ位置・姿勢推定処理の計算コストを低減することで実時間でのカメラ位置・姿勢推定を実現する．従来手法 [6] では，データベース中より選択した  $N$  個のランドマークと入力画像中より抽出した平均  $F$  個の自然特徴点とを照合することで対応付け，カメラ位置・姿勢の推定を行っている．従来手法 [6] では，全体の処理時間の大半をこの対応付け処理が占めており，実時間での処理を達成するためには対応付け処理の高速化が必要である．従来手法 [6] の対応付け処理における計算コスト  $C_{prev}$  は，ランドマークと自然特徴点 1 個当

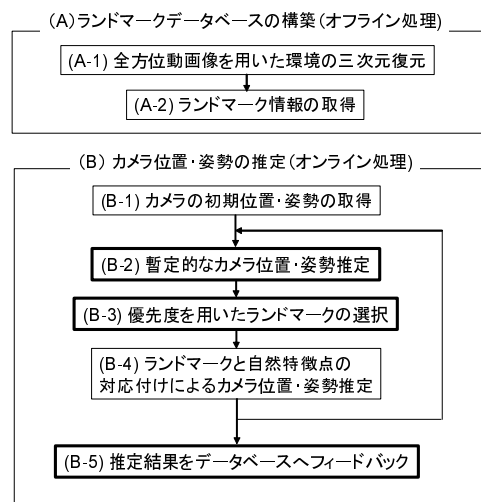


図 1 全体の処理の流れ

たりの照合コストを  $A$  とすると

$$C_{prev} = NFA \quad (1)$$

となる．照合コスト  $A$  には，撮影姿勢に依存しない画像パターンの作成コストと正規化相互相関によるテンプレートマッチングのコストが含まれる．本研究において，ランドマークと自然特徴点の照合コスト  $A$  を低減することは困難であり，対応付け処理の計算コストを低減するためには，対応点探索に用いるランドマークの数  $N$  とその照合に用いられる自然特徴点数  $F$  を削減する必要がある．そこで，本研究では，対応付け処理に用いられるランドマーク数  $N$  と照合される自然特徴点数  $F$  を削減するため，以下の 2 つの改良を加える．

- (1) 連続フレーム間でランドマークを追跡し暫定的なカメラの位置・姿勢を推定することで照合すべき自然特徴点数  $F$  を削減する
- (2) ランドマークへ優先度情報を付加することで，正しく対応づく可能性の高い少数のランドマークを選択し，対応付け処理に用いるランドマーク数  $N$  を削減する

提案手法の処理の流れを図 1 に示す．図中の太枠は，本研究で追加または改良した処理を表す．ランドマークデータベースを用いたカメラ位置・姿勢推定は，オフラインでのデータベースの構築処理 (A) とオンラインでのカメラ位置・姿勢の推定処理 (B) からなる．以下，2 節では，オフラインでのランドマークデータベース構築の処理について述べ，3 節では，オンラインでのカメラ位置・姿勢推定処理について述べる．4 節では従来手法 [6] と提案手法の対応付け処理の計算コストを比較し，5 節では実験によりロバスト性向上効果と処理速度向上効果を検証する．最後に，6 節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

本節では，オフラインでのランドマークデータベース構築処理 (A) について述べる．本研究では，3 節で述べるオンラインでのカメラ位置・姿勢の推定において，推定に有効なランドマークを効率的に選択するために，従来手法 [6] で用いられているデータベースに優先度情報を付加した新たなランドマークデータベースを用いる．ここでは，ランドマークデータベース

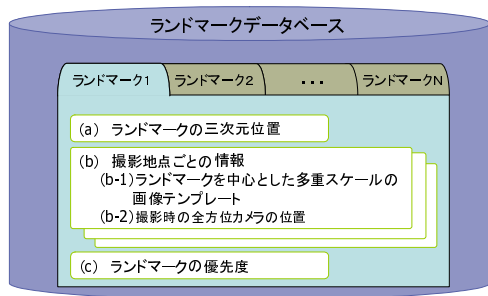


図 2 ランドマークデータベースの構成要素

を構築するために、まず、全方位画像を用いた環境の三次元復元を行う (A-1)。次に、環境の三次元復元結果を用いてランドマーク情報を取得する (A-2)。以下では、各処理について詳しく述べる。

### 2.1 全方位画像を用いた環境の三次元復元

本研究では、三次元位置が既知の少数の基準点や、GPS などの外部指標を用いた structure-from-motion 法 [11], [12] を用いることで、広域で複雑な環境においても蓄積誤差の生じない環境の三次元復元を行う。これらの手法では、まず、対象となる環境を移動しながら全方位カメラを用いて撮影する。次に、Harris オペレータ [13] によって検出された動画像中の自然特徴点を自動追跡することで、structure-from-motion によって自然特徴点の三次元位置と画像上の座標、および全方位カメラのカメラパラメータを推定する。最後に、三次元位置が既知の少数の基準点 [11] または GPS から得られる絶対位置 [12] を利用して、動画像全体での最適化処理を行うことで、カメラパラメータと自然特徴点の三次元位置の累積的な推定誤差を最小化する。このような手法を用いることで、広域で複雑な環境においても比較的容易に環境の三次元復元を行うことができる。

### 2.2 ランドマーク情報の取得

本研究で用いるランドマークデータベースには図 2 に示すように、(a) 自然特徴点の三次元位置、(b) 撮影地点ごとの情報、(c) ランドマークの優先度の情報が保持される。撮影地点ごとの情報は、(b-1) 多重スケールの画像テンプレートと (b-2) ランドマーク撮影時の全方位カメラの位置から成る。以下ではそれぞれの情報の取得方法について述べる。

(a) ランドマークの三次元位置：3 節で述べるオンラインでのカメラ位置・姿勢推定では、ランドマークの三次元位置と画像上の二次元座標を対応付けることによりカメラの位置・姿勢を推定するため、ランドマークの三次元位置情報が必要となる。ランドマークの三次元位置は 2.1 項で述べた処理によって得られるものであり、環境に固定された世界座標系で保持されている。(b) 撮影地点ごとの情報の取得：撮影地点によるランドマークの見え方の違いに対応するため、撮影地点ごとに多重スケールの画像テンプレートを作成し、データベースへ登録する。ここでは、世界座標系において、カメラの投影中心とランドマークの三次元位置を結ぶ直線に対して垂直な面に画像上での自然特徴点周辺のパターンを投影することで、カメラの撮影姿勢に依存しない画像テンプレートを作成し登録する (b-1)。また、オンラインでのカメラ位置・姿勢推定において、データベース中か

ら推定に用いるランドマークを選択するために、ランドマーク撮影時の全方位カメラの位置をデータベースへ登録する (b-2)。(c) ランドマークの優先度：3 節で述べるオンラインでのカメラ位置・姿勢の推定において、正しく対応づく可能性の高いランドマークをデータベース中から効率的に選択するために、ランドマークに優先度を付加する。優先度は、過去に同じ場所でユーザによって撮影された入力に対する推定結果を利用して、各ランドマークがカメラ位置・姿勢推定に利用される確率を算出することで決定する。ここで、ランドマークデータベースの構築時点においては優先度を算出することはできないため、全てのランドマークに対して同一の優先度を設定しておく。3 節で述べるオンライン処理によって、ユーザのカメラ位置・姿勢が推定された場合には、データベースからランドマーク  $i$  が選択された回数  $D_i$  とカメラ位置・姿勢推定の際に誤対応として排除されずに推定に用いられた回数  $E_i$  を用いて、ランドマーク  $i$  の優先度  $P_i$  を以下の式で更新する。

$$P_i = \frac{E_i}{D_i} \quad (2)$$

本研究では、データベースの管理者が、あらかじめデータベース構築時にいくつかの学習データを与えることで、一般ユーザが利用する前に優先度を更新しておくことを想定する。この優先度の更新処理については 3.2 節で詳しく述べる。

## 3. カメラ位置・姿勢の推定

本節ではオンラインでのカメラ位置・姿勢推定処理 (B) について述べる。本研究では、図 1 に示したように、まず初期フレームでのカメラ位置・姿勢を何らかの手法により取得し (B-1)、続いて逐次的なカメラ位置・姿勢推定処理 (B-2) ~ (B-4) を繰り返す。逐次推定では、まずカメラ位置・姿勢推定の高速化とロバスト性の向上のため、連続フレーム間で自然特徴点に対応付け、暫定的なカメラの位置・姿勢を推定することで、ランドマークの対応点探索範囲を限定する (B-2)。次に、優先度に基づいてランドマークを選択することで、推定に有効なランドマークを効率的にデータベース中から取り出す (B-3)。最後に、選択したランドマークと入力画像中の自然特徴点に対応付け、現フレームのカメラ位置・姿勢を推定する (B-4)。逐次推定の完了後に推定結果をデータベースへフィードバックすることで、データベース中の優先度を更新する (B-5)。以下では、処理 (B-2) ~ (B-5) について詳しく述べる。

### 3.1 暫定的なカメラ位置・姿勢の推定

本研究では、連続フレーム間でランドマークを追跡することにより、現フレームの暫定的なカメラ位置・姿勢を推定する (B-2)。暫定的なカメラ位置・姿勢の推定では、まず、前のフレームで自然特徴点と対応付けられ、誤対応として排除されずにカメラ位置・姿勢の推定に用いられたランドマークを選択し、現フレームにおける対応点を探索する。ランドマークの対応点の探索において、連続フレーム間ではパターンの変形や明るさの変化は微小であることから、前フレームにおけるランドマークの位置を中心とする  $T \times T$  画素の画像をテンプレートとして用い、ランドマークの前フレームの画像上の位置を中心とする一定のウィンドウ  $W_1$  内で SSD (Sum of Squared Differences)

を評価尺度としたテンプレートマッチングによりランドマークの追跡を行う。ただし、自然特徴点の対応付け処理はウィンドウ  $W_1$  内で Harris オペレータによって検出された自然特徴点のみを対象として行う。

次に、追跡した自然特徴点を用いて現フレームの暫定的なカメラ位置・姿勢を推定する。カメラ位置・姿勢の推定では、まず、LMedS 基準を用いて誤対応を排除する。次に、正しい対応関係を用いて、カメラ位置・姿勢の推定を行い、再投影誤差を最小化することによって現フレームの暫定的なカメラパラメータ  $\hat{M}$  を推定する。

### 3.2 優先度を用いたランドマークの選択

ランドマークの選択処理 (B-3) では、まず処理 (B-2) で推定された暫定的なカメラ位置・姿勢とランドマークが撮影された全方位カメラの位置を用いて、以下の幾何学的な条件を満たす画像テンプレートをデータベース中より選択する。

- ・ ランドマークが入力画像の画角内に存在する
- ・ ランドマーク撮影時のカメラ位置とランドマークを結ぶ直線と、暫定的なカメラ位置とランドマークを結ぶ直線の成す角が閾値以下
- ・ ランドマーク撮影時のカメラ位置と現フレームの暫定的なカメラ位置の間の距離が閾値以下

次に、選択されたランドマークの候補から優先度  $P_i$  の高い順にランドマークを  $N_{proj}$  個選択する。ただし、 $N_{proj}$  は最大  $N_{prior}$  個とし、 $N_{proj} = N_{prior} - N'_{track}$  とする。ここで、 $N'_{track}$  は処理 (B-2) で誤対応として排除されずに、暫定的なカメラ位置・姿勢の推定に用いられたランドマークの数である。

### 3.3 ランドマークと自然特徴点の対応付けによるカメラ位置・姿勢推定

ここでは、処理 (B-3) で選択した優先度の高いランドマークと入力画像中の自然特徴点を対応付け、カメラ位置・姿勢の推定を行う。ランドマークと自然特徴点の対応付け処理では、まず処理 (B-2) で推定した暫定的なカメラパラメータ  $\hat{M}$  を用いて、処理 (B-3) で選択されたランドマーク  $j(j = 1, 2, \dots, N)$  を以下の式により入力画像上へ投影する。

$$\begin{bmatrix} a_j u_j & a_j v_j & a_j \end{bmatrix}^T = \hat{M} \begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^T \quad (3)$$

ただし、 $(x_j, y_j, z_j)$  はランドマーク  $j$  の世界座標系での三次元位置、 $(u_j, v_j)$  はランドマーク  $j$  の画像上への投影位置、 $a_j$  はカメラ座標系でのランドマーク  $j$  の奥行きを表す。次に、投影したランドマーク  $j$  の画像上の位置  $(u_j, v_j)$  を中心とする一定ウィンドウ  $W_2$  内の自然特徴点を検出する。ただし、ここではウィンドウ  $W_2$  のサイズをウィンドウ  $W_1$  のサイズよりも小さく設定することでランドマークの画像上での探索範囲を小領域に限定する。最後にウィンドウ内で検出された全ての自然特徴点について、データベース構築時と同様に、カメラの撮影姿勢に依存しない画像テンプレートを作成し、作成した画像テンプレートとランドマークの画像テンプレートを正規化相互相関法によって照合する。本手法では相関値が最大となる自然特徴点をランドマークと対応付ける。

次に、ランドマークと対応付けられた自然特徴点と、処理

(B-2) で暫定的なカメラ位置・姿勢の推定に用いた自然特徴点を用いて、現フレームのカメラ位置・姿勢を推定する。現フレームのカメラ位置・姿勢は処理 (B-2) と同様に、LMedS 基準を用いて誤対応を排除した後に、再投影誤差を最小化することによって推定する (B-4)。

### 3.4 推定結果のフィードバックによる優先度の更新

カメラの位置・姿勢推定の完了後、ランドマークが利用された頻度の情報をデータベースへフィードバックすることで、データベース中の各ランドマークの優先度を更新する (B-5)。ここでは頻度の情報として、カメラ位置・姿勢推定の逐次処理において、ランドマーク  $i$  が処理 (B-3) でデータベース中から選択された回数  $D_{inew}$  と処理 (B-4) において LMedS 基準により誤対応として排除されずにカメラ位置・姿勢の推定に用いられた回数  $E_{inew}$  を利用することでデータベース中のランドマークの優先度を更新する。ランドマーク  $i$  の優先度  $P_i$  は、ユーザより新たに取得した頻度情報  $(D_{inew}, E_{inew})$  と、過去に同じ場所でカメラ位置・姿勢の推定を行ったユーザより得られている頻度情報  $(D_{iold}, E_{iold})$  を用いて、次式で更新する。

$$P_i = \frac{E_{iold} + E_{inew}}{D_{iold} + D_{inew}} \quad (4)$$

## 4. 対応付け処理の計算コストの比較

本節では、提案手法によるランドマークと自然特徴点の対応付けの理論的なコストの低減量について述べる。式 (1) に示したように、従来手法における対応付け処理の計算コストは  $C_{prev} = NFA$  であるが、本研究では照合される自然特徴点数  $F$  と対応付け処理に用いられるランドマーク数  $N$  を削減することで、対応付けの計算コスト  $C_{prev}$  を低減する。提案手法では、ランドマークの探索範囲を限定するために、新たに暫定的なカメラパラメータを推定することが必要となるため、ランドマーク対応付けの総コスト  $C_{new}$  は、

$$C_{new} = C_{track} + C_{proj} \quad (5)$$

となる。ただし、 $C_{track}$ 、 $C_{proj}$  はそれぞれ、暫定的なカメラパラメータの推定 (B-2)、暫定的なカメラパラメータを用いたランドマークの対応付け (B-4)、に必要な計算コストである。ここで、処理 (B-2) の計算コスト  $C_{track}$  は、追跡に用いるランドマーク数を  $N_{track}$  個とすると、

$$C_{track} = N_{track}FB + E_{LMedS} \quad (6)$$

となる。ただし、 $E_{LMedS}$  は誤対応の排除とカメラ位置・姿勢の推定にかかる計算コストであり、 $B$  は追跡処理においてランドマークと自然特徴点の 1 組のペアに対して SSD を算出する計算コストである。このコストは 1 節で述べた、幾何学的な変形や輝度値の正規化を必要とするランドマークの照合コスト  $A$  に対してはるかに小さく、また、追跡に用いるランドマーク数  $N_{track}$  は  $N$  に対して小さいため、提案手法で新たに導入した処理 (B-2) のオーバーヘッドは小さい。

処理 (B-4) では、処理 (B-2) で推定された暫定的なカメラ位置・姿勢を用いて画像上の探索範囲を限定することで、対応点候補として用いられる自然特徴点の平均数を  $S_2/S_1$  ( $S_1, S_2$  は

それぞれウィンドウ  $W_1, W_2$  の面積を表す) に、また、優先度に基づくランドマークの選択処理と、処理 (B-2) で追跡されたランドマークを対応付けるランドマークから除外することにより、ランドマークの数  $N$  を  $(N_{prior} - N'_{track})/N$  に削減できる ( $N_{prior} \leq N$ )。これにより、処理 (B-4) の計算コスト  $C_{proj}$  は、

$$C_{proj} = (N_{prior} - N'_{track}) \frac{S_2}{S_1} FA \quad (7)$$

となる。

ここで、従来の対応付け処理に対する高速化の効果 (処理コストの比) は以下の式で算出できるが、実際には各処理の繰り返しにおいて一定のオーバーヘッドが生じるため、5 節で述べる実験結果と完全には一致しない。

$$\frac{C_{new}}{C_{prev}} = \frac{C_{track} + C_{proj}}{C_{prev}} \quad (8)$$

## 5. 実 験

提案手法の有効性を示すために、本研究の基礎とした従来手法 [6] との比較により処理速度向上効果を検証した。データベースの構築には、全方位型マルチカメラシステム (Point Grey Research 社 Ladybug) で屋外環境 (約 75m) を歩きながら撮影した全方位動画画像を用いた。図 3 に、データベースの構築に用いた全方位動画画像の一部を示す。また、データベース構築時のカメラパスと復元されたランドマークの三次元位置を図 4 に示す。データベースに登録されたランドマーク数は約 12400 個であり、一つのランドマークにつき平均 8ヶ所の異なる地点で撮影された画像テンプレートが登録された。本実験ではビデオカメラ (SONY DSR-PD-150) を用いて撮影した 4 本の動画画像 (解像度 720×480 画素, プログレッシブスキャン, 15fps) に対してカメラの位置・姿勢推定を行った。そのうち、3 本の動画画像の推定結果を優先度設定のために用い、残りの 1 本の動画画像 (1000 フレーム) を評価用として用いた。カメラ位置・姿勢推定に用いた各パラメータを表 1 に示す。本実験では、提案手法、従来手法ともにカメラの内部パラメータは既知とし、初期フレームのカメラ位置・姿勢は手動で与えた。位置・姿勢推定精度の定量的な評価に用いるカメラ位置・姿勢の真値は、環境内の特徴点の三次元位置をトータルステーションを用いて測定し、画像上でそれらの特徴点を手動で対応付け、PnP 問題を解くことで算出した。ただし、再投影誤差が 1.5 画素以上であるフレームや自然物が入力画像中の大半を占めるなどの理由で、トータルステーションによって計測できる特徴点が少ないフレームについては評価対象にしなかった。

まず、データベースから選択するランドマークの最大数

表 1 実験で用いたパラメータ

	従来手法 [6]	提案手法
ランドマーク追跡処理の際の画像テンプレートサイズ	-	15×15画素
ランドマークの画像テンプレートサイズ	15×15画素	
ウィンドウ $W_1$ のサイズ	-	120×60画素
ウィンドウ $W_2$ のサイズ	120×60画素	20×20画素
学習データ	-	動画3本分の推定結果
優先度の初期値	-	0.5

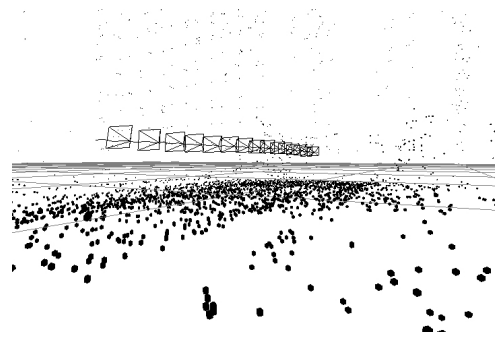


図 4 データベース構築時のカメラパスとランドマークの三次元位置  $N, N_{prior}$  を変化させ、以下の 4 つの手法に対する推定失敗率を比較した。

- (手法 A) 高速化処理無し (従来手法 [6])
- (手法 B) 提案手法において、暫定的なカメラ位置・姿勢の推定による探索範囲の限定のみを行った場合
- (手法 C) 提案手法において、優先度を用いたランドマークの選択のみを用いた場合
- (手法 D) 暫定的なカメラ位置・姿勢の推定による探索範囲の限定と優先度を用いたランドマークの選択を両方行った場合 (提案手法)

なお、本実験では、カメラ位置・姿勢の逐次推定の途中で、データベース中のランドマークと入力画像中の自然特徴点の対応関係が 6 組以上得られなかった場合にカメラ位置・姿勢の推定に失敗したとみなした。また、推定に失敗した場合には、手動で失敗したフレームのカメラ位置・姿勢を与えることにより、カメラ位置・姿勢の逐次推定処理を再開し、最終フレーム (1000 フレーム目) まで推定を完了させた上で推定失敗の回数を算出した。

図 5 にデータベースから選択するランドマーク数と推定に失敗したフレーム数の関係を示す。優先度情報を用いない手法 A、手法 B はデータベース中より選択するランドマーク数の減少に伴い、カメラ位置・姿勢の推定に失敗するフレームが発生するが、優先度を用いてデータベース中よりランドマークを選択する手法 C、手法 D ではランドマーク数を 30 個まで減らした場合においても最終フレームまで失敗せずにカメラ位置・姿勢を推定できている。このことから、優先度を用いることで、正しく対応づく可能性の高いランドマークを優先的に選択できていることが確認できる。

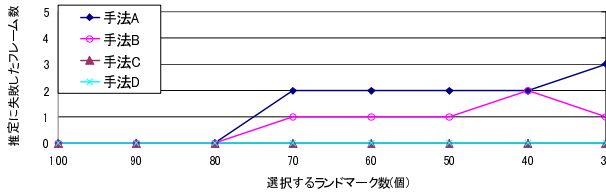


図5 ランドマーク数と推定に失敗したフレーム数の関係

次に、PC(Core 2 Extreme 2.93 GHz, メモリ 2 GB) を用いて、各手法の処理時間の比較を行った。処理時間の比較では、最終フレームまで失敗せずにカメラ位置・姿勢を推定できることを前提として、データベースより選択するランドマーク数を手法 A, 手法 B では 80 個, 手法 C, 手法 D では 30 個とした。このとき、それぞれの手法で推定されたカメラ位置・姿勢の精度を表 2 に示す。表 2 より、暫定的なカメラ位置・姿勢を用いた探索範囲の限定や優先度情報に基づくランドマークの選択を行うことによって、カメラ位置推定精度が向上していることが確認できる。これは、対応点探索範囲を小領域に限定したことと、優先度に基づいて、正しく対応づく可能性の高いランドマークを優先的にデータベース中より選択したことで、誤対応の割合が減少したためだと考えられる。カメラ位置・姿勢推定の各処理の処理時間と全体の処理時間を表 3 に示す。従来手法である手法 A と比較し、手法 B, 手法 C, 手法 D は、ランドマークと入力画像中の自然特徴点の対応付けの計算コストが削減されていることが確認できる。また、提案手法である手法 D では、全体の処理時間が平均 59ms となり、1 秒間に約 17 フレームの実時間でのカメラ位置・姿勢推定を実現することができる。表 3 から、手法 A に対する、手法 D のランドマークと入力画像中の自然特徴点の対応付けの計算コストは、従来手法 [6] の約 1/21 となり、大幅な高速化を実現できていることが分かる。ただし、この値は式 (7) に示した高速化効果の理論的な値である 1/48 に対して 2 倍程度となっているため、手法 D のランドマークと自然特徴点の対応付けコスト 15ms のうち、半分程度はオーバーヘッドであると考えられる。

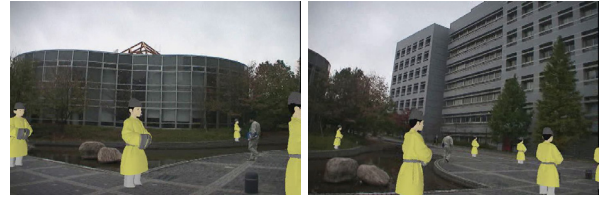
最後に、提案手法で推定されたカメラ位置・姿勢を用いて入力動画画像上に仮想物体を重畳表示した結果を図 6 に示す。生成された動画画像から仮想物体が位置ずれなく幾何学的に正しい位

表 2 推定精度の比較

	手法A	手法B	手法C	手法D
平均カメラ位置誤差 (mm)	360	257	231	256
カメラ位置誤差の標準偏差 (mm)	528	137	204	181
平均カメラ姿勢誤差 (度)	0.84	0.95	1.13	0.91
カメラ姿勢誤差の標準偏差 (度)	0.71	1.20	1.16	0.91

表 3 処理時間の比較

	処理時間 (ms)			
	手法A	手法B	手法C	手法D
暫定的なカメラ位置・姿勢推定	-	26	-	21
ランドマークの選択	12	3	2	1
ランドマークと自然特徴点の対応付け	316	51	131	15
誤対応の排除とカメラ位置・姿勢推定	61	16	16	17
画像の取得など	4	4	4	5
全体	392(2.5fps)	100(10fps)	153(6.5fps)	59(16.9fps)



(a)200 フレーム目

(b)400 フレーム目

図 6 カメラ位置・姿勢推定結果を用いた仮想物体の重畳表示位置に描画されていることを確認した。

## 6. まとめ

本稿では、自然特徴点の追跡による暫定的なカメラの位置・姿勢の推定による探索範囲の限定と優先度情報をランドマークへ付加することによって、カメラ位置・姿勢推定的高速化を図る手法を提案した。実験により、従来手法 [6] よりも精度よく、実時間でのカメラ位置・姿勢推定が実現できることを確認した。今後の課題として、データベースの更新による利用環境の変化への対応、テクスチャの少ないシーンでのロバスト性の向上などがあげられる。

謝辞 本研究の一部は、科学技術振興機構 (JST) の戦略的創造研究推進事業 (CREST) 「デジタルメディア作品の制作を支援する基盤技術」プログラムの支援による。

## 文献

- [1] R. Cipolla, D. Robertson and B. Tordoff: "Image-based localization," Proc. VSMM, pp. 22-29, 2004.
- [2] 岩佐英彦, 栗飯原述宏, 横矢直和, 竹村治雄: "全方位画像を用いた記憶に基づく位置推定", 信学論 (D-II), Vol. J84-D-II, No. 2, pp. 310-320, 2001.
- [3] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: "Fully automated and stable registration for augmented reality applications," Proc. ISMAR, pp. 93-102, 2003.
- [4] L. Vacchetti, V. Lepetit and P. Fua: "Combining edge and texture information for real-time accurate 3D camera tracking," Proc. ISMAR, pp. 48-57, 2004.
- [5] I. Skrypnik and D. G. Lowe: "Scene modelling, recognition and tracking with invariant image features," Proc. ISMAR, pp. 110-119, 2004.
- [6] 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", 日本パッチャリアリティ学会論文誌, Vol. 10, No. 3, pp. 285-294, 2005.
- [7] A. Davison, Y. G. Cid and N. Kita: "Real-time 3D slam with wide-angle vision," Proc. IFAC Symp. on Intelligent Autonomous Vehicles, 2004.
- [8] B. Williams, G. Klein and I. Reid: "Real-time SLAM relocation," Proc. ICCV, 2007.
- [9] G. Klein and D. Murray: "Parallel tracking and mapping for small AR workspaces," Proc. ISMAR, pp. 225-234, 2007.
- [10] J. S.Beis and D.G.Lowe: "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," Proc. CVPR, pp. 1000-1006, 1997.
- [11] 佐藤智和, 池田聖, 横矢直和: "複数動画画像からの全方位型マルチカメラシステムの位置・姿勢パラメータの推定", 信学論 (D-II), Vol. J88-D-II, No. 2, pp. 347-357, 2005.
- [12] S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: "Construction of feature landmark database using omnidirectional videos and GPS positions," Proc. 3DIM, pp. 249-256, 2007.
- [13] C. Harris and M. Stephens: "A combined corner and edge detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.