

# 複合現実感による映画制作支援のための 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

Real-time camera parameter estimation using a feature landmark database for MR pre-visualization

武富貴史, 佐藤智和, 池田聖, 横矢直和

Takafumi TAKETOMI, Tomokazu SATO, Sei IKEDA, Naokazu YOKOYA

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, {takafumi-t, tomoka-s, sei-i, yokoya}@is.naist.jp)

**Abstract :** This report describes a camera parameter estimation method for a pre-visualization tool for filmmaking using mixed reality (MR). In the pre-production of filmmaking, pre-visualization techniques are used for testing a camera work and acting. Usually a pre-visualization movie is created by computer graphics. On the other hand, a MR pre-visualization technique, which creates the pre-visualization movie using both real images and CG-rendered actors on site, has been proposed. By the MR pre-visualization, directors and actors can easily grasp the camera work and acting. To realize the MR pre-visualization, fast and accurate extrinsic camera parameter estimation method is necessary. Although many kinds of camera parameter estimation methods have been proposed, its computational cost is high in outdoor environments. To realize the real-time MR pre-visualization, the number of matching candidates are reduced by using priorities of landmarks that are determined from previously captured video sequences.

**Key Words:** *pre-visualization, camera parameter estimation, landmark database, mixed reality*

## 1. はじめに

近年、映画制作の初期段階において、CG を用いてシーンの見え方をシミュレートする Pre-visualization(PreViz) の利用が進んでいる。通常 PreViz は CG のみによって実現されるが、複合現実感技術 (MR) を用いることで、実写の背景を用いた PreViz 映像を撮影現場で作成することが可能となる。MR-PreViz は通常の PreViz と異なり実際の背景を用いるため、監督や俳優は、CG のみで作られた PreViz 映像よりも直感的に本番のカメラワークや演技を把握することができる。このような MR-PreViz を実現するためには、カメラの絶対位置・姿勢を実時間で高精度に推定する必要がある。従来から様々なカメラ位置・姿勢推定手法が提案されているが、本報告では PreViz での利用を想定し、実世界と仮想世界の画像上で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。

一般に、画像からカメラの絶対位置・姿勢を推定するためには、利用環境の事前知識が必要となる。利用環境の事前知識として、環境の三次元モデル [1, 2] やランドマークデータベース [3] を用いる手法が提案されているが、本研究では、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構築することができるランドマーク

データベースを用いるアプローチ [3] を採用する。ランドマークデータベースを用いる手法 [3] では、環境中に存在する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。ランドマークデータベースは structure-from-motion 法を利用することで自動で構築することができるが、従来手法 [3] では、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た。

本研究では、実時間処理を必要とする MR-PreViz での利用を想定し、従来提案されているランドマークデータベースによる手法 [3] を基礎として、以下の 2 つの改良を加えることで、複雑な屋外環境においても利用が可能な実時間でのカメラ位置・姿勢推定を実現する手法を提案する。

- (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を削減する

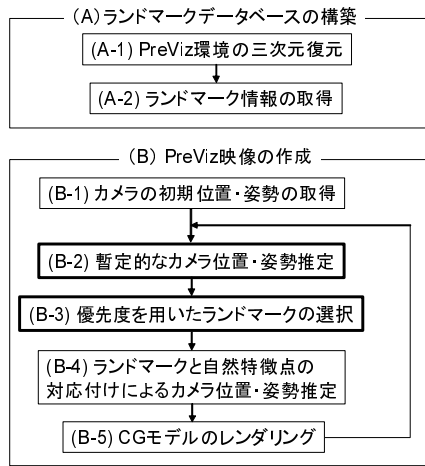


図 1: 全体の処理の流れ

(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を削減する

## 2. ランドマークデータベースを用いた実時間での

### CG モデルの位置合わせ

本研究では、ランドマークデータベースを用いたカメラ位置・姿勢推定によって、CG モデルの位置合わせを行い MR-PreViz 映像を作成する。提案手法の処理の流れを図 1 に示す。図中の太枠は、従来手法 [3] に対して、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 CG モデルの位置合わせは、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 PreViz 映像の作成処理 (B) からなる。

#### 2.1 ランドマークデータベースの構築

##### (A-1) MR-PreViz 環境の三次元復元

まず、MR-PreViz を実施する環境内を移動しながら全方位カメラを用いて撮影する。次に、動画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、および全方位カメラの外部パラメータを推定する。最後に、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ため、三次元位置が既知の少数の基準点 [4] または GPS から得られる絶対位置 [5] を利用して、動画全体での最適化処理を行う。

##### (A-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。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**: オンライン処理では、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 (A-1) 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。

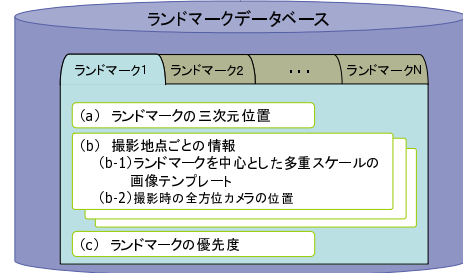


図 2: ランドマークデータベースの構成要素

(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、撮影地点ごとにカメラの撮影姿勢に依存しない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 (b-1)。また、オンライン処理でのランドマーク選択処理 (B-3) において、撮影可能な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**: オンライン処理でのカメラ位置・姿勢推定処理 (B-4) 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。ここでは、事前に MR-PreViz を実施する環境でランドマークの優先度を学習するための動画を撮影しておき、撮影された学習用の動画に対して、後述する処理 (B) と同様のカメラ位置・姿勢推定を行うことで、各ランドマークの優先度を決定する。

本研究では、ランドマーク  $i$  の優先度  $P_i$  として、ランドマーク  $i$  が推定に利用される確率  $P_i = E_i/D_i$  を用いる。ここで、 $D_i$ 、 $E_i$  は、それぞれ、学習データに対するカメラ位置・姿勢推定の処理結果について、ランドマーク  $i$  がデータベースから選択された回数とカメラ位置・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を表す。

#### 2.2 MR-PreViz 映像の作成

オンラインでの MR-PreViz 映像の作成処理 (B) について述べる。オンライン処理では、図 1 に示したように、まず初期フレームでのカメラ位置・姿勢を取得し (B-1)、続いて処理 (B-2)~(B-5) を繰り返す。以下では、各処理について詳しく述べる。

##### (B-1) カメラの初期位置・姿勢の取得

本研究では、初期フレームのカメラ位置・姿勢は、ランドマークデータベースを用いた静止画像からのカメラ位置・姿勢推定手法 [6] によって自動で取得することを想定している。ただし、現時点において構築したシステムでは手法 [6] を実装しておらず、後述する実験では、環境内に固定された三脚にカメラを設置した際の位置・姿勢を手動で与え、カメラを三脚に設置した状態から推定処理を開始することで、初期位置・姿勢を取得している。

##### (B-2) 暫定的なカメラ位置・姿勢の推定

本研究では、連続フレーム間でのランドマークの追跡により、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、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  $W_1$  内に存在する自然特徴点を対象として 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。次に、LMedS 基準 [7] を用いて誤対応の排除を行い、正しい対応関係を用いて、PnP 問題 [8] を解くことによって暫定的なカメラパラメータを推定する。

### (B-3) 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の幾何学的な関係を用いて、撮影可能なランドマークをデータベース中より選択する。次に、選択されたランドマークの候補から優先度の高い順にランドマークを  $N_{prior}$  個選択する。優先度を用いることによって、繰り返しパターンや自然物などの誤対応が生じやすいパターンを除外することができ、データベース中より選択するランドマークの個数を削減することができる。

### (B-4) ランドマークと自然特徴点の対応付によるカメラ位置・姿勢推定

処理 (B-3) で選択した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、暫定的なカメラパラメータを用いて、処理 (B-3) で選択されたランドマークを入力画像上へ投影する。次に、投影したランドマークの画像上の位置を中心とする一定ウィンドウ  $W_2$  内の全ての自然特徴点周辺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法によって照合する。ただし、ここでは、暫定的なカメラパラメータを用いることで、ウィンドウ  $W_2$  のサイズを処理 (B-2) よりも小さく設定することが可能となり、照合すべき自然特徴点数が削減される。

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パラメータを推定する。現フレームのカメラパラメータは処理 (B-2) と同様に、誤対応の排除後に、PnP 問題を解くことによって推定する。

### (B-5) CG モデルのレンダリング

推定された現フレームのカメラパラメータを用いて、事前に三次元位置や動きが与えられた役者の CG モデルをカメラで撮影した画像上にレンダリングすることで MR-PreViz 画像を生成する。ここで、CG モデルの三次元位置・姿勢・スケール等は、ランドマークと同一座標系で、あらかじめ定義しておく必要がある。また、CG をレンダリングする際には、カメラの内部パラメータを用いて、画像上での歪も考慮する。



図 3: 全方位画像

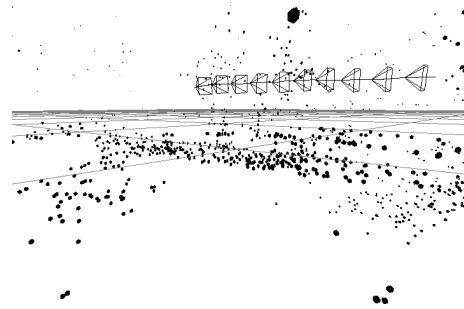


図 4: 三次元復元の結果

## 3. 実験

提案手法の有効性を示すために、まず、従来手法 [3] と計算コストの比較を行った。次に、試作システムを用いて、CG キャラクタの合成実験を行った。

本実験では、全方位型マルチカメラシステム (Point Grey Research 社 Ladybug2) で屋外環境 (約 10m) を歩きながら撮影した全方位動画 (200 フレーム) を用いてランドマークデータベースの構築を行った。図 3 に、データベースの構築に用いた全方位動画の一部を示す。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 664 個であり、平均 62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 3 本の動画 (解像度 720×480 画素, プログレッシブスキャン, 15fps) を取得し、そのうち、2 本の動画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 (300 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。

表 1: 実験で用いたパラメータ

	従来手法 [3]	提案手法
選択するランドマーク数 $N_{prior}$	60個	
自然特徴点追跡の際の探索範囲 $W_1$	-	120×60画素
ランドマークの対応点候補の探索範囲 $W_2$	120×60画素	20×20画素
学習データ	-	動画2本分の推定結果
優先度の初期値	-	0.5

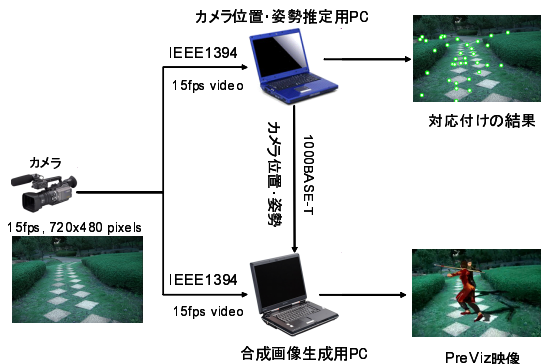


図 5: 試作システムの機器構成

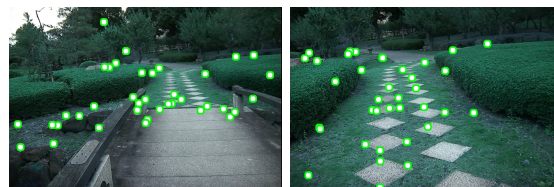
まず、PC(Core 2 Extreme 2.93GHz, メモリ 2GB) を用いて従来手法 [3] とカメラ位置・姿勢推定の処理時間の比較を行った。処理時間は、提案手法、従来手法 [3] でそれぞれ、1 フレーム当たり 68ms(14.7fps), 548ms(1.8fps) であり、従来手法 [3] の約 1/8 の処理時間となった。また、従来手法 [3] は、推定処理の途中でランドマークの対応付けに失敗し、最終フレームまで推定を完了することができなかった。このことから、提案手法では、高速化処理が実現しただけでなく、従来手法 [3] よりもロバストな対応付け処理が実現できていることが分かる。

次に、図 5 に示す試作システムを用いて、CG キャラクタの合成実験を行った。同図に示すように、試作システムはカメラ位置・姿勢推定と CG モデルのレンダリングの負荷を分散させるために 2 台の PC から構成されている。試作システムでは、まず、カメラで撮影された画像をカメラ位置・姿勢推定用 PC(Core 2 Extreme 2.93GHz, メモリ 2GB) と合成画像生成用 PC(Core 2 Duo 2GHz, メモリ 2GB) に分配する。次に、カメラ位置・姿勢推定用 PC でカメラ位置・姿勢の推定を行い、推定されたカメラパラメータを 1000BASE-T の有線 LAN 経由で合成画像生成用 PC に送信する。最後に、合成画像生成用 PC で、推定されたカメラパラメータを用いて CG キャラクタを実写画像に合成し提示することで実時間での PreViz を実現する。

試作システムを用いて、屋外環境で実写背景に CG キャラクタを合成した結果を図 6 に示す。図 6(a) 中の○印はランドマークと対応付けられた自然特徴点の位置を示す。提案手法では、図 6 に示すような、自然物が多く存在する複雑な環境でも、多くのランドマークが自然特徴点と対応付けられ、1 秒間に約 15 フレームのカメラの位置・姿勢を推定し、PreViz 映像を提示できることを確認した。なお、本手法では撮影から MR-PreViz 映像の提示までの間に、理論上、最低でも 1/15 秒以上の遅延が発生するが、主観的には MR-PreViz での利用において問題になるレベルの遅延は感じられなかった。

#### 4. まとめ

本報告では、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ランドマークへの優先度情報の付加によって、カメラ位置・姿勢推定の高速化を図り MR-PreViz を実現する手法を提案した。また、MR-PreViz を実現するため



60 フレーム目 220 フレーム目  
(a) ランドマークと対応づいた自然特徴点の位置



60 フレーム目 220 フレーム目  
(b) CG キャラクタの合成結果

図 6: プロトタイプシステムを用いた PreViz 画像

のシステムを試作し、実時間で CG キャラクタの合成ができることを確認した。

今後の課題として、より高度な PreViz を実現するため CG キャラクタと現実物体との隠蔽関係の再現などが挙げられる。

**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、総務省・戦略的情報通信研究開発推進制度 (SCOPE) の支援による。

#### 参考文献

- [1] V. Lepetit, L. Vacchetti, D. Thalmann and P. Fua: "Fully automated and stable registration for augmented reality applications," Proc. ISMAR, pp. 93-102, 2003.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] S. Ikeda, T. Sato, K. Yamaguchi and N. Yokoya: "Construction of feature landmark database using omnidirectional videos and GPS positions," Proc. Int. Conf. on 3DIM, pp. 249-256, 2007.
- [6] M. Susuki, T. Nakagawa, T. Sato and N. Yokoya: "Extrinsic camera parameter estimation from a still image based on feature landmark database," Proc. ACCV'07 Satellite Workshop on Multi-dimensional and Multi-view Image Processing, pp. 124-129, 2007.
- [7] P. J. Rousseeuw: "Least Median of Squares Regression," J. of the American Statistical Association, 79, pp. 871-880, 1984.
- [8] R. Klette, K. Schluns and A. Koschian Eds: "Computer vision: Three-dimensional data from image," Springer, 1998.