

ランドマークデータベースを用いたカメラ位置・姿勢推定による 実時間での複合現実型プレビジュアリゼーションの実現

Real-time MR pre-visualization by feature landmark based camera parameter estimation

武富 貴史† 佐藤 智和† 池田 聖† 横矢 直和†
Takafumi Taketomi Tomokazu Sato Sei Ikeda Naokazu Yokoya

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プレビジュアリゼーション (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、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。複合現実型プレビジュアリゼーション (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。しかし、従来提案されている MR-PreViz の手法 [1, 2] は、屋内での利用を想定しており、CG の位置合わせを実現するためにマーカを必要とするため、広域な屋外シーンを対象として、撮影現場における実時間での MR-PreViz 映像を作成することが困難であるという問題がある。これに対して、本報告では、撮影対象となるシーンに存在する自然特徴点を用いてあらかじめ画像データベースを構築しておき、画像に基づくカメラ位置・姿勢推定を実時間で行うことで、屋外環境において MR-PreViz 映像を実時間で生成する手法を提案する。

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 [3, 4] やランドマークデータベース [5] 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手法 [5] を採用する。ただし、従来手法 [5] 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、複雑な屋外環境においても利用が可能なランドマークデータベースによる手法 [5] を基礎として、以下の2つの改良を加えることで、カメラ位置・姿勢推定の計算コストを削減し、実時間での MR-PreViz 映像の生成を実現する。

- (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

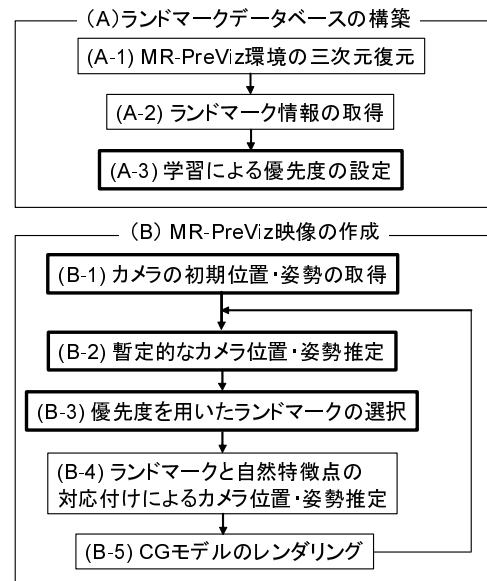


図 1: 全体の処理の流れ

- (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

また、手法 [5] では、カメラの初期位置・姿勢を手動で与える必要があり、ユーザの補助を必要としていたが、本研究ではこの点についても改善し、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] を導入することで、カメラの初期位置・姿勢推定処理を自動化する。

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 [5] 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 MR-PreViz 映像の作成処理 (B) からなる。以下、2 節では、オフラインでのデータベース構築処理 (A) について述べ、3 節では、オンラインでの MR-PreViz 映像作成処理 (B) について述べる。4 節では、プロトタイプシステムにより生成された MR-PreViz 映像を示し、提案手法の有効性について検証する。最後に5 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、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、推定に有効なランドマーク

† 奈良先端科学技術大学院大学
Nara Institute of Science and Technology (NAIST)

を効率的に選択するために、従来手法 [5] 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。また、初期フレームのカメラ位置・姿勢を推定するために、自然特徴点の固有スケールと SIFT 特徴ベクトルの情報を登録しておく。

2.1 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に、三次元位置が既知の少数の基準点 [7] または GPS から得られる絶対位置 [8] を利用して、動画全体での最適化処理を行う。

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) ランドマークの三次元位置、(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 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 (b-1) を作成し、データベースへ登録する。また、カメラの初期位置・姿勢を推定するために Harris-Laplacian [9]、SIFT-descriptor [10] を用いて、自然特徴点の固有スケール (b-2)、SIFT 特徴ベクトル (b-3) を算出し、データベースへ登録する。さらに、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-4)。

(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像に対して、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。

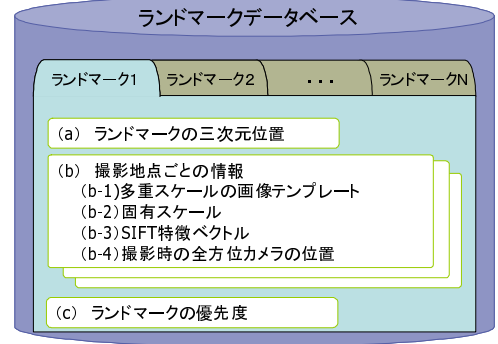


図 2: ランドマークデータベースの構成要素

3. 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。

3.1 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 (B-1)。初期フレームのカメラ位置・姿勢推定処理では、まず、入力画像中の自然特徴点の固有スケール、SIFT 特徴ベクトルを算出した後、これらをデータベース中の全てのランドマークの SIFT 特徴ベクトルと照合することで自然特徴点とランドマークを対応付ける。次に、ランドマークの観測地点の整合性を利用して、対応付けられたランドマークから誤対応を排除する。最後に、得られた対応関係を用いて、PnP 問題 [11] を解くことで 6 自由度のカメラ位置・姿勢を推定する。

上記の初期位置・姿勢推定処理では、ランドマークと自然特徴点の照合処理に数十秒を要するため、本研究では初期位置・姿勢の推定が完了するまでの間カメラを動かさないものとする。

3.2 暫定的なカメラ位置・姿勢の推定

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。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 [12] を用いて誤対応を排除した後、正しい対応関係を用いて、PnP 問題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。

3.3 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、

暫定的なカメラ位置から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。本研究では、ランドマークの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照合処理において除外し、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で高速化を行う。

3.4 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

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 (B-4)。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-1) 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。

3.5 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 (B-5)。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、画像上でのレンズ歪みも考慮する。

4. 実験

提案手法の有効性を示すために、まず、従来手法 [5] と計算コストの比較を行った。次に、プロトタイプシステムを用いて、CG キャラクタの合成実験を行った。ただし、現時点において構築したシステムでは、初期位置・姿勢の推定に多くの処理時間を必要とするため、後述する実験では、初期位置・姿勢の推定中は環境内に固定された三脚にカメラを設置しておき、初期位置・姿勢推定の完了後、カメラを三脚に設置した状態から逐次推定処理を開始する。

4.1 計算コストの比較

本実験では、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 (1260 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

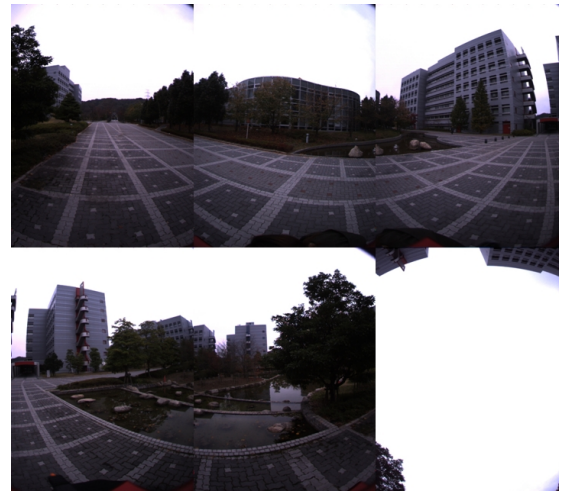


図 3: 全方位画像

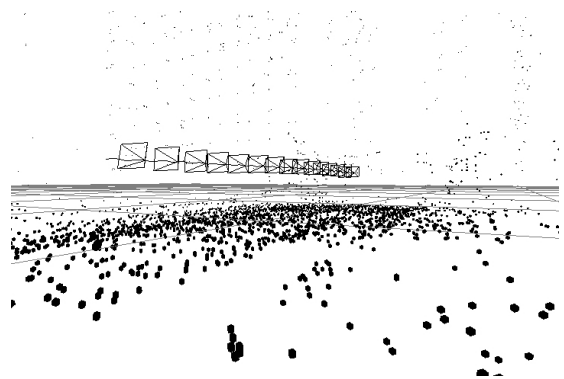


図 4: 三次元復元の結果

録されたランドマーク数は約 12000 個であり、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 (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、3 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 (1000 フレーム) を評価用として用いた。本実験で用いたパラメータを表 1 に示す。PC(Core 2 Extreme 2.93GHz, メモリ 2GB) を用いた場合においては、提案手法、従来手法 [5] とともに初期フレームのカメラ位置・姿勢推定の処理時間は約 63 秒であった。逐次推定の処理時間は、提案手法、従来手法 [5] でそれぞれ、1 フレーム当たり 66ms(15.1fps), 393ms(2.5fps) であり、提案手法は従来手法 [5] の約 1/6 の処理時間となった。また、従来手法 [5] 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。

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

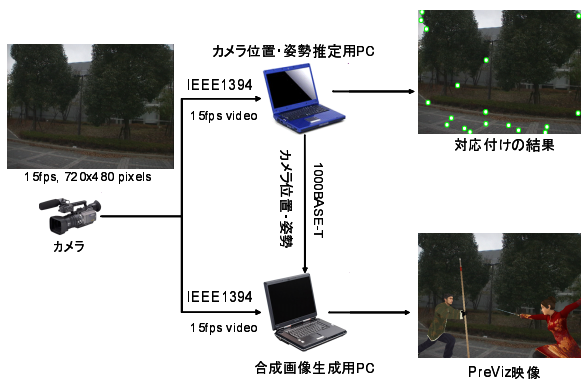


図 5: プロトタイプシステムの機器構成

た。このことから、提案手法では、処理の高速化が実現しただけでなく、従来手法 [5] 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。

4.2 プロトタイプシステムを用いた MR-PreViz 映像の生成

図 5 に示すプロトタイプシステムを用いて、CG キャラクタの合成実験を行った。同図に示すように、プロトタイプ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。プロトタイプシステムでは、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(PC1: Core 2 Extreme 2.93GHz, メモリ 2GB) と合成画像生成用 PC(PC2: Core 2 Duo 2GHz, メモリ 2GB) に分配する。次に、PC1 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 PC2 に送信する。PC2 では、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、ユーザに提示することで実時間での PreViz を実現する。

プロトタイプシステムを用いて、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、カメラ位置・姿勢の推定および CG モデルの合成を正しく行うことができた。プロトタイプシステムでは、1 秒間に約 15 フレームのカメラの位置・姿勢の推定と、それに基づく MR-PreViz 映像の提示が可能であった。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。

5. 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、実時間での MR-PreViz を実現する手法を提案した。また、MR-PreViz を実現するためのシステムを試作し、実際に実時間で 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、より高品位な MR-PreViz を実現するためには、CG キャラクタと現実物体との隠蔽関係の再現を行う必要がある。



620 フレーム目 640 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



620 フレーム目 640 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 MR-PreViz 画像

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] M. Shin, b.-s. Kim and J. Park: "AR storyboard: An augmented reality based storyboard authoring tool," Proc. ISMAR, pp. 98–99, 2005.
- [2] G. Thomas: "Mixed Reality Techniques for TV and their Application for On-Set and Pre-Visualization in Film Production," DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking, pp. 31–36, 2005.
- [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] 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.
- [6] 薄充孝, 中川知香, 佐藤智和, 横矢直和: "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", 日本バーチャルリアリティ学会論文誌, Vol. 13, No. 2, pp. 161–170, 2008.
- [7] 佐藤智和, 池田聖, 横矢直和: "複数動画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.
- [8] 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.
- [9] K. Mikolajczyk and C. Schmid: "Scale & affine invariant interest point detectors," Int. J. of Computer Vision, Vo. 60, No. 1, pp. 63–86, 2004.
- [10] D. G. Lowe: "Distinctive image features from scale-invariant keypoints," Int. J. of Computer Vision, Vo. 60, No. 2, pp. 91–100, 2004.
- [11] R. Klette, K. Schluns and A. Koschan Eds: "Computer vision: Three-dimensional data from image," Springer, 1998.
- [12] P. J. Rousseeuw: "Least Median of Squares Regression," J. of the American Statistical Association, 79, pp. 871–880, 1984.