4Y-9

# 局所不変特徴量を用いた屋外 MR トラッキング法の高速化と安定化

樋下 航[†1]　天目 隆平[†2]　柴田 史久[†1]　田村 秀行[†1]
武富 貴史[†3]　佐藤 智和[†3]　横矢 直和[†3]
立命館大学 大学院理工学研究科[†1]　同 総合理工学研究機構[†2]　奈良先端科学技術大学院大学 情報科学研究科[†3]

## 1. はじめに

カメラトラッキングは複合現実感 (Mixed Reality; MR) の根幹を成す技術であり，動作速度や精度，安定性が要求される．武富ら [1] の手法では事前知識としてランドマークデータベース (以降 LDB) を利用し，屋外のような広域な環境下においても実時間で累積誤差の生じないトラッキングを実現している．しかし，自然特徴点の対応付けに要する計算コストが高く，一度トラッキングが破綻すると復帰できないといった問題があった．

本研究では，同手法の改良を試みる．ランドマークとカメラの 3 次元位置情報をもとに高速に記述される SIFT 特徴量 [2] を用いることで，対応付け処理の計算量を軽減し，トラッキング処理の高速化を図る．トラッキング破綻時には，画像中の全特徴点から得られた特徴量群に対して近似最近傍探索による対応付けを行い，実時間での復帰を実現する．また，事前に複数用意したキーフレームを利用して初期位置姿勢推定を行う．

## 2. トラッキングの高速化と安定化

### 2.1. LDB を用いたトラッキング

[1] の LDB 構築は，次の手順でオフライン的に実行される．(a)まず，利用する環境をカメラで撮影する．(b)次に，動画像から Structure-from-Motion により特徴点の 3 次元位置とカメラ位置姿勢を推定する．(c)最後に，特徴点の 3 次元位置や画像テンプレートなどをランドマーク情報として登録する．

一方，トラッキングは，次の手順で達成される (図 1)

- 前フレームで求めたカメラ位置姿勢から推定に用いるランドマークをデータベース中から選択する (T-1).
- 次に，入力画像から検出された特徴点とランドマークを対応付ける (T-2 ~ 4).
- 最後に，3D-2D の対応関係からカメラの位置姿勢を推定する (T-5).

計算コストが大きかったのは，対応付けに画像テンプレートを用いるためであり，また，そのテンプレート作成には前フレームでのカメラ位置姿勢が要るため，トラッキングが破綻すると復帰することができなかった．

本研究では，ランドマークの対応付けに SIFT 特徴量を利用し (T-4)，さらに破綻からの復帰処理 (T-6) を導入することで，トラッキングの高速化と安定化を図る．提案手法で用いる LDB には以下の情報を登録する．

(a) ランドマークの 3 次元位置
(b) SIFT 特徴量 (128 次元ベクトル)
(c) 固有スケール決定係数
(d) 登録時のカメラ位置

なお，スケールとは，各特徴点の画像上における特徴量を記述する領域の大きさとする (図 2)．また，特徴点の検出には高速で再現性の高い FAST 検出器を用いる．

"Improvement in Speed and Robustness for Outdoor Camera Tracking Using Local Invariant Feature"
†1 Graduate School of Science and Engineering, Ritsumeikan University
†2 Research Organization of Science and Engineering, Ritsumeikan University
†3 Nara Institute of Science and Technology

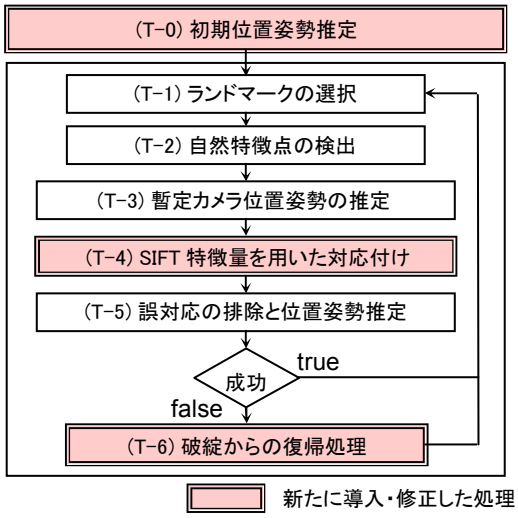


図 1　提案手法の流れ

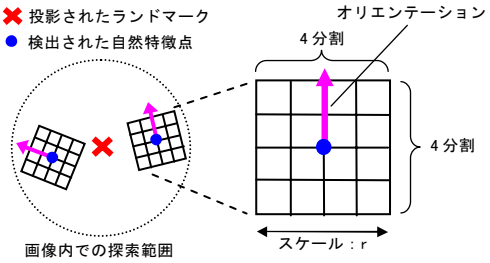


図 2　SIFT 特徴量記述領域

### 2.2. 対応付けの高速化

ランドマークの対応付けには画像の局所構造に不変，かつ照明変動にも頑健な SIFT 特徴量 [2] を採用する．しかし，一般的に SIFT は特徴量記述領域を定めるためのスケール算出にかかる計算量が多く実時間処理には適さない．そこで，ランドマークとカメラ間の 3 次元距離からスケールを一意に決定することで SIFT 特徴量を高速に記述し，対応付けに要する計算時間を削減する．

探索範囲内で検出された候補点に対し，スケール $r$ を

$$r = \frac{r' \times d'}{d} \tag{1}$$

によって計算する．ただし，$r'$ と $d'$ はそれぞれ LDB 構築時のスケール，カメラとランドマーク間の 3 次元距離であり，$d$ はトラッキング時のカメラとランドマークの距離を表す．$r' \times d'$ をスケール決定係数として登録しておき，内部パラメータは常に同じとする．スケールを求めた後は従来と同様に，勾配情報から求めたオリエンテーションの向きに記述領域を正規化し，128 次元の特徴量を記述する．今回の LDB 構築時には $r' = 24$ とした．

トラッキング時には，まず，暫定カメラ位置姿勢の推定によって対応付け候補の絞込みを行った後，残った候補点に対して上記の方法で SIFT 特徴量を記述し，データベースに登録してあるものと比較する．評価尺度は

SSD (Sum of Squared Differences) とし，この値が最小となる候補点を最終的な対応付け結果とする．最後に RANSAC を用いて誤対応を排除し，正しい対応関係から PnP 問題を解くことでカメラの位置姿勢を推定する．

### 2.3. 破綻からの復帰

トラッキングが破綻するとカメラの位置姿勢が未知となり，下記の 3 つの問題が生じる．破綻時のカメラ位置と光軸方向は破綻前と比べてその変化量が小さいと仮定することで，実時間での復帰を実現する．

まず，位置姿勢が未知の状態において，全てのランドマークを対応付けに用いると照合コストが増大してしまうため，ある程度数を絞り込む必要がある．そこで仮定より，対応付けに用いるランドマークは破綻する直前まで追跡されていたものに限定し，照合時間を削減する．

次に，特徴点ごとのスケールを定められないため，全特徴点に対して同一のスケールで特徴量を記述せざるを得ない．そこで仮定から，破綻前に各特徴点について求めたスケールの平均を用いて記述することとする．当然，適切なスケールとは異なる値をとり得るが，その差が小さければほぼ同じ特徴量を得られる傾向があるため，位置姿勢の計算に必要な分の対応関係は十分に取得できる．

最後に，ランドマーク 1 点につき画像内の全特徴点を対応付け候補として扱うため，その照合コストを抑える必要がある．そこで，得られた特徴量群から張られる特徴量空間内で近似最近傍探索による照合を行い，効率的に対応付けることで実時間での復帰を実現する．

### 2.4. キーフレームを用いた初期位置姿勢推定

初期フレームでは，前フレームのカメラ位置姿勢が与えられないため，トラッキングとは異なる方法で推定を行う必要がある．[1] では LDB 構築時のカメラパスから離れた位置においても高精度に推定可能な手法を用いているが，GPS を必要とし，推定に時間がかかるため，画像情報のみから高速に推定できる手法が望ましい．

本稿では，LDB 構築に用いる画像シーケンスから任意に選択された複数のキーフレームを利用することで，LDB 構築時のカメラパス付近であれば高速かつ高精度にカメラの初期位置姿勢を推定可能な手法を提案する．

推定処理 (T-0) は 2 つのステップからなる．最初のステップでは，入力画像と最も類似したキーフレームを検索し，対応付けの対象となるランドマーク数を絞り込むことにより，次のステップで特徴点を対応付ける際の照合時間や誤対応数を抑える．画像間の類似度 $S$ は

$$S = \sum_{i=1}^{L} \frac{1}{SSD(v_i, v')} \qquad (2)$$

から計算する．ただし，$L$ はキーフレーム内に存在する全ランドマーク数，$v$ は各ランドマークの SIFT 特徴量であり，$v'$ は最近傍と見なされた特徴点の特徴量を表す．

次のステップでは，最も類似度が高いと判断されたキーフレーム内に存在するランドマークと入力画像中の特徴点を，復帰処理と同様に特徴量空間内で近似最近傍探索によって対応付けることで初期位置姿勢を推定する．

**表 1　トラッキング平均処理時間 (ms)**

| 処理 | 従来手法 | 提案手法 |
|---|---|---|
| T-1~3 | 20.6 | 4.2 |
| T-4 | 35.1 | 3.6 |
| T-5 | 3.3 | 0.7 |
| 合計 | 59.0 | 8.5 |

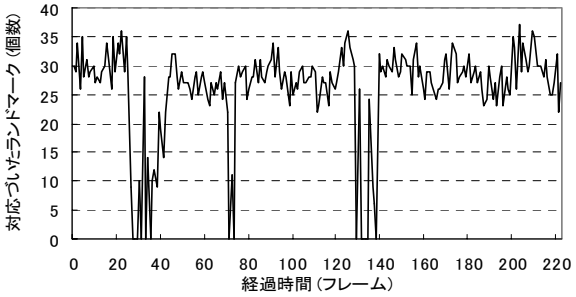


**図 3　対応付けに成功したランドマーク数**

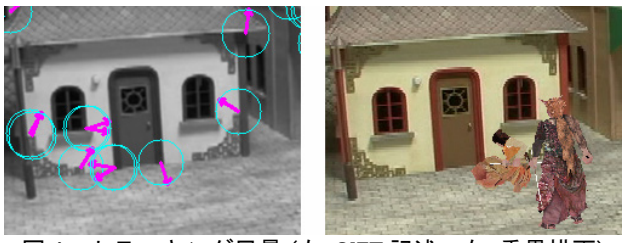

**図 4　トラッキング風景 (左: SIFT 記述，右: 重畳描画)**

## 3. 実験

提案手法の有効性を示すために，速度と安定性について従来手法との比較を行った．PC (CPU: Core 2 Extreme 2.8 GHz，メモリ: 4 GB)，ビデオカメラ (SONY DSR-PD170，720×480，プログレッシブ，15 fps) を使用し，カメラの内部パラメータは事前に求めた．400 フレームの動画像から LDB を構築し，そのうち見え方が異なる 10 フレーム分の画像を手動で選び，キーフレームとして用いた．まず初期位置姿勢の推定では，LDB 構築時のカメラパス付近であれば平均 45 ms で推定できることを確認した．次に，トラッキングの処理時間の比較を**表 1** に示す．対応付け処理 (T-4) の計算コストを抑えることで全体の速度が向上していることがわかる．**図 4 左**が SIFT 特徴量を記述する様子であり，矢印がオリエンテーション，円が記述領域に内接する円を表す．最後に，対応付けに成功したランドマーク数を**図 3** に示す．毎フレーム 60 個のランドマークを対応付けに用いている．30，70，140 フレーム周辺で十分な数のランドマークが対応付けられず破綻しているが，その後トラッキングに復帰できていることがわかる．実際に推定されたカメラの位置姿勢を用いて CG を重畳描画した (**図 4 右**)．

## 4. むすび

本稿では，高速なスケール計算によって記述される局所不変特徴量を用いて屋外MRトラッキング法の高速化と安定化を行う手法を提案した．本手法を実装し，従来手法よりも高速なカメラ位置姿勢推定が可能で，トラッキング破綻からも復帰できることを確認した．

**謝辞**　本研究は，JST の CREST「映画制作を支援する複合現実型可視化技術」の支援による．

### 参考文献

[1] 武富貴史, 他: “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”, 日本 VR 学会大会論文集, pp. 367 - 370, 2008.

[2] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” *Int J. Comput. Vision,* Vol. 60, No. 2, pp. 91 - 100, 2004.