

三次元物体の姿勢検出における 2D-3D マッチング手法

Robust 2D-3D Matching for 3D Object Pose Detection

小林 達也† 加藤 晴久† 米山 暁夫†
Tatsuya Kobayashi Haruhisa Kato Akio Yoneyama

1. はじめに

近年 Augmented Reality(AR)の分野において、認識対象に仮想情報を重ね表示することで、アノテーションサービスや作業支援を実現する技術が注目されている。仮想情報の重ね表示には画面内の対象物に対する 6 自由度のカメラ姿勢を推定する必要があり、三次元物体を対象とした高速・高精度なカメラ姿勢推定技術が必要とされている。

三次元物体に対するカメラ姿勢推定では、入力画像内の二次元座標と、対応する三次元物体表面上の点(三次元座標)のマッチング(2D-3D マッチング)が必要である。従来文献[1]ではキーフレームを利用した 2D-3D マッチング手法が提案されているが、セルフオクルージョンの問題のためにマッチングの精度が三次元物体の姿勢に依存する問題点があった。

そこで本稿では、三次元物体の姿勢に依らず、高精度な対応点が取得可能な 2D-3D マッチング手法を提案する。

2. 関連研究

三次元物体に対するカメラ姿勢の推定を目的とした様々な手法の中で、多くの手法はオクルージョンや光源変化、視点の変化等に対する頑健性から、特徴点マッチングベースの手法を用いている。一般的に特徴点マッチングでは局所特徴量の計算に必要な計算量が問題になるが、事前に分類器を学習し、特徴点マッチングをクラス分類の問題に置き換える高速な手法が提案されている [2]。

特徴点マッチングを用いた従来のカメラ姿勢推定手法では、対象とする三次元物体が写った複数枚の参照画像と三次元モデルを用意し、予め参照画像内の特徴点の抽出とモデルを利用した逆投影による三次元座標の計算を行い、データベースとして記録する。この処理はオフラインで実施しておき、オンライン処理では入力画像から検出した特徴点と、記録した(三次元座標が既知である)特徴点の間でマッチングを行い、得られた 2D-3D の対応点からカメラ姿勢を算出する。このようなキーフレームベースの手法は、複数の参照画像を用意することで視点の変化に対して頑健性を持つため、特にカメラ姿勢の初期値計算に用いられている [1]。しかし 2D-3D マッチングの処理の間に入力-参照画像間の 2D-2D マッチングを挟むため、入力画像中の対象物の姿勢が参照画像内の対象物の姿勢のいずれからも離れているような場合に、マッチング率が劣化する問題点がある。

3. 提案する 2D-3D マッチング手法

本稿が提案する手法は、入力画像内の特徴点と三次元モデルの表面上の点を直接的にマッチングすることで、

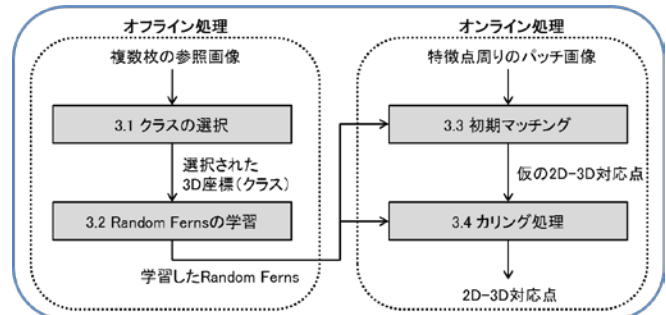


図1 提案する2D-3Dマッチング手法のフロー

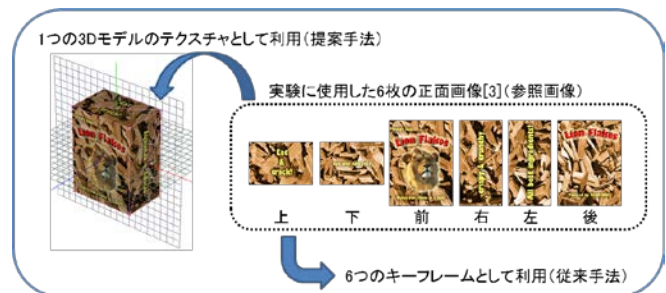


図2 参照画像の利用方法の違い

対象物の姿勢に依存しない 2D-3D マッチングを実現する。提案手法のフローを図 1 に示す。モデルの表面上の三次元座標をクラスに持つ Random Ferns[2]をオフラインで学習し、学習した分類器を用いて直接的な 2D-3D マッチングを行う。以下に各ブロックの処理を説明する。

3.1. クラスの選択

オフライン処理ではまず分類器のクラスとして扱う三次元座標の選択を行う。図 2 に示すようにテクスチャマッピングによって参照画像から 3D モデルを構築し、ランダムなカメラパラメータでモデルの投影を行ってモデル画像を生成する。モデル画像から特徴点を検出し、それを三次元座標に逆投影する。以上の処理を繰り返し行い、高頻度で特徴点として検出された三次元座標を H 個選択してクラスとして扱う。

3.2. Random Ferns の学習

提案手法では、特徴点を中心とした 32×32 のパッチ画像を入力として、最も確率の高い三次元座標を割り当てる Random Ferns[2]を学習する。学習には[2]と同様の手法を用いるが、特徴点の代わりに三次元座標をクラスとして持つため、学習画像の生成には 3.1 節と同じ手法を用い、パッチ画像はモデル画像内で三次元座標が対応する点を特定し、取得する必要がある。

† KDDI 研究所 KDDI R&D Laboratories Inc.

3.3. 初期マッチング

オフライン学習を行うことで、オンライン処理ではパッチ画像を分類器に入力することで高速にクラス分類を行うことができる。初期マッチングでは各 Fern の分類結果 F_k から確率密度の対数和が最大のクラスを選択する。

3.4. カリング処理

提案手法の分類器は対象物の表面を覆う三次元座標をクラスとするため、初期マッチングのみを行う場合でもマッチング精度が対象物の姿勢に依存しない 2D-3D マッチングが可能である。しかし初期マッチング後の対応点の中には、表と裏の関係にあり実際には同時に見えることのない三次元座標が含まれている。そこで初期マッチング後の対応点を用いて 3D モデルの面に対してカメラ画像に写っている/いないの判定を行い、学習した確率密度を修正する手法を提案する。修正後の各 Fern のリーフノード F_k ($k=1, \dots, M$) が持つ確率密度は以下の式で得られる。

$$P(F_k = l | C = c_i) = \begin{cases} \frac{N_{l,c_i} + 1}{N_{c_i} + K'}, & \text{if } \frac{N_f}{N_F} > T_f, \\ 0, & \text{otherwise.} \end{cases}$$

ここで Random Ferns は M 本の深さ S の決定木 (Fern) で構成され、 N_{l,c_i} は学習時にクラス c_i ($i=1, \dots, H$) のパッチ画像がリーフノード l ($l=1, \dots, K=2^S$) に分類された回数、 N_{c_i} はクラス c_i のパッチ画像の総数である。 N_f は対応点に含まれる三次元座標のうち、ポリゴン面 f 上の三次元座標の数、 N_F はポリゴン面 f 上の三次元座標の総数、提案手法は初期マッチングに含まれる三次元座標の割合が閾値 T_f 以下のポリゴンを写っていないと推定し、確率分布を修正することでマッチング率を改善する。 T_f の値は各モデルでモデル画像を使用したテストから最適の値を選択して使用する。

4. 評価実験

評価実験では [1] の検出技術に含まれるキーフレームベースの 2D-3D マッチング手法と提案手法で比較を行った。実験には図 2 の画像 (FlakesBox) [3] の他に、市販の菓子箱を撮影した正面画像を用い、従来手法は 6 面を直接キーフレームとして、提案手法は同じ画像をモデルのテクスチャとして利用した。クラスの選択はランダムなカメラパラメータで生成した 100,000 枚のモデル画像を用いて行った。クラスは $H=120$ に設定し、3D モデルの 6 面に対して均等に 20 点ずつ割り振った。Random Ferns のパラメータは $M=30$, $S=11$ に設定し、学習には別に生成した 10,000 枚の画像を使用した。従来手法は正面画像を直接キーフレームとしたため、 $H=20$ の分類器を 6 つ構築し、常に最適なキーフレーム選択を行うと仮定して毎回のテストで最もマッチング率の高い分類器を選択した。

テスト画像は別に生成した 1,000 枚のモデル画像を使用した。学習した 120 点の三次元座標をモデル画像に投影し、モデル画像に映っている三次元座標を中心としたパッチ画像に対してそれぞれの手法でマッチングを行い、正しくパッチ画像を元の三次元座標とマッチングできた割合 (マッチング率) を比較した。

図 3 に提案手法の T_f とマッチング率の関係のグラフを示す。 $T_f=0$ は初期マッチングのみ行うことを意味する。

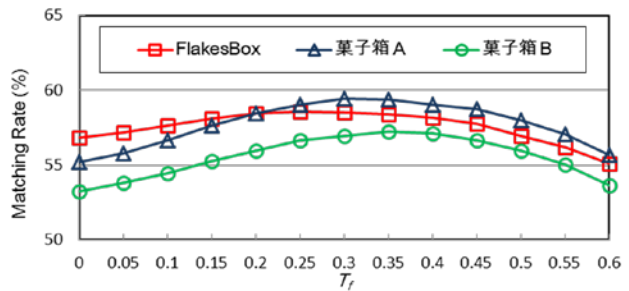


図3 T_f とマッチング率の関係

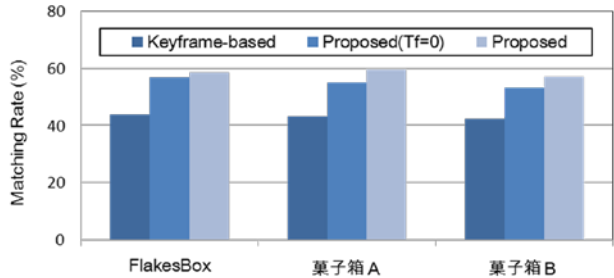


図4 従来手法と提案手法のマッチング率の比較

各モデルについて概ね $T_f=0.3$ の時、つまり 3D モデルの各面について、30%の三次元座標が初期マッチングで対応付けられた場合に、モデル面が見えていると判断し、裏面の三次元座標を対応点から除外することで、最も良いマッチング率が得られることが分かった。 T_f をさらに高めた場合、実際には写っている面を写っていないと判定する頻度が増加し、マッチング率が劣化することを確認した。

図 4 に従来手法と提案手法のマッチング率を比較した結果を示す。各モデルでテクスチャの違いからマッチング率自体には若干の差があるものの、共通して提案手法のマッチング率の方が約 15 ポイント優れていることを確認した。特に複数面が大きく写る入力画像において提案手法の顕著な優位性が確認された。従来手法は 6 面のそれぞれをキーフレームとしたため、このような画像に対してはどのキーフレームを用いても最大 1/2 あるいは 1/3 までのパッチ画像しかマッチングすることができないが、提案手法ではこの問題が発生せず、対象物の姿勢に依らず高精度なマッチングが可能である。

5. まとめ

本稿では、三次元物体に対するカメラ姿勢推定のための新しい 2D-3D マッチング手法を提案した。3D モデル表面の三次元座標をクラスとして学習した Random Ferns を用いたことで対象物の姿勢に依らない直接的なマッチングが可能になり、従来のキーフレームベースの手法のマッチング率を約 15 ポイント改善できることを確認した。

参考文献

- [1] Y. Park, V. Lepetit and W. Woo, "Extended Keyframe Detection with Stable Tracking for Multiple 3D Object Tracking," *Visualization and Computer Graphics, IEEE Transactions on*, vol.17, no.11, pp.1728-1735, Nov. 2011.
- [2] M. Ozuysal, M. Calonder, V. Lepetit and P. Fua, "Fast Keypoint Recognition Using Random Ferns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.32, no.3, pp.448-461, March 2010.
- [3] <https://ar.qualcomm.at/qdenvet/>