

パーツ識別に基づくモデル照合を用いた手指の位置姿勢推定に関する研究

情報科学研究科 情報科学専攻 H11401M 有賀治樹

指導教員：橋本学 教授

1. はじめに

ビジョンベースの人物の動作認識は古くからの課題であり、現在でも多くの研究がなされている。その中でも手指の位置姿勢推定は、Human Computer Interaction (HCI) における、さまざまなアプリケーションの開発のために必要不可欠な技術である。この分野の課題として、高精度な位置姿勢の推定が挙げられる。

従来の手指の位置姿勢推定手法として、任意の位置姿勢パラメータを持つ手指のモデルと観測データの整合性評価に基づいて位置姿勢を推定する生成ベースの手法 [1] や、決定木ベースの識別器を用いた手指のパーツ識別に基づいて関節位置を推定する識別ベースの手法 [2] が提案されている。近年では、生成ベースの手法と識別ベースの手法を組み合わせた統合ベースの手法 [3] が提案され、推定精度や素早い動きへの耐性の面で高い性能を示している。しかしながら、手指のモデルと観測データの整合性を評価する際に手全体の整合性のみを使用するため、手指のモデルと観測データ間で局所的な誤対応が起こっていた場合にも高い整合性評価値が得られる可能性があり、推定精度の低下が懸念される。

そこで本研究では、手指のモデルと観測データ間の局所的な誤対応を低減した高精度な手指の位置姿勢推定を目的とする。提案手法では、手全体の大局的な情報だけでなく、事前にパーツ識別された手指の局所領域の情報を用いて手指のモデルと観測データの整合性を評価することによって、局所領域の誤対応を低減する。

2. 局所領域の整合性を考慮した位置姿勢推定

2.1 提案アルゴリズムの概要

図 1 に提案アルゴリズムの流れを示す。

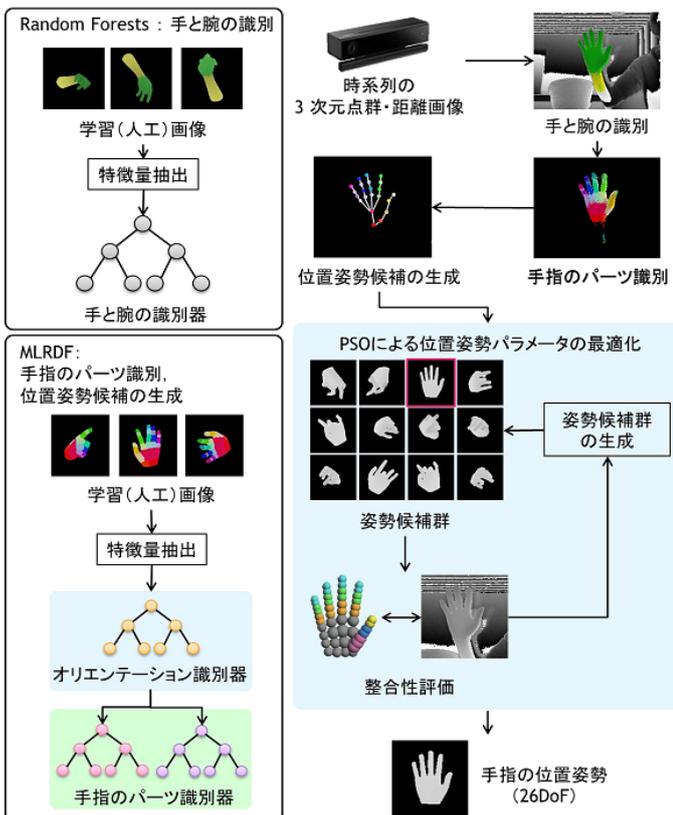


図 1 提案アルゴリズムの流れ

まず、レンジセンサから取得した距離画像の各画素を Random Forests を用いて手指と腕に分類し、手指を表す画素を手指のパーツに識別する。この際、Random Forests を階層的に構築することによって識別性能を向上させる Multi-layered Randomized Decision Forests を用いて高精度な手指のパーツ識別を実現する。また、セルフオクルージョンに頑健な投票ベースの Regression Forests を用いて単一フレームから関節位置を推定し、位置姿勢の候補とする。最後に、位置姿勢の候補と前フレームで推定された位置姿勢をもとに最適化問題の解法である Particle Swarm Optimization (PSO) を用いて手指の位置姿勢を推定する。

2.2 手指のパーツ識別

Multi-layered Randomized Decision Forests (MLRDF) [2] を用いることによって、高精度なパーツ識別を実現する。MLRDF は、Random Forests を階層化することによって、識別性能を向上させ、効率的な学習を実現している。

提案手法では、上位の階層に手全体のオリエンテーションのクラスを識別するための識別器を設定し、下位の階層に各画素を手指のパーツに識別する識別器を構築する。MLRDF を用いた手指のパーツ識別の流れを図 2 に示す。

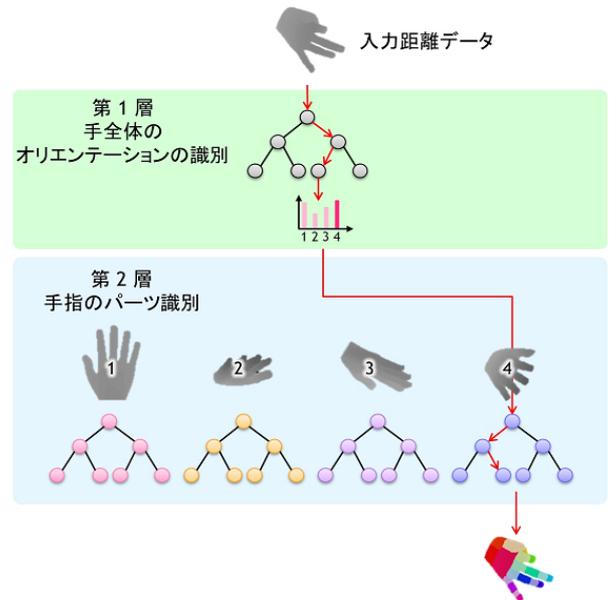


図 2 手指のパーツ識別の流れ

第 1 層では、入力距離画像から選択された N 画素を手全体のオリエンテーションの識別器に入力することによって得られた N 個のクラス尤度の和を最終的なクラス尤度として、最も尤度の高いクラスを識別結果とする。第 2 層では、第 1 層で識別されるクラス数の識別器を構築しておき、第 1 層の識別結果から第 2 層で使用する識別器を選択する。選択された識別器に入力距離画像の各画素を入力することによって、各画素を手指のパーツに識別する。

2.3 組み合わせ最適化に基づく位置姿勢推定

提案手法では、手指の位置姿勢推定を位置姿勢パラメータの組み合わせ最適化問題と捉え、手指のモデルと観測データの整合性評価に基づいて位置姿勢パラメータを最適化する。

前フレームで推定された位置姿勢と推定された位置姿勢候補に基づいて位置姿勢仮説群の初期姿勢を設定し、これらの

位置姿勢を手指のモデルで表現したデータと観測データ（距離画像）の整合性を評価し、最適化手法を用いて位置姿勢仮説群を更新し、最終的に最良の位置姿勢仮説を出力する。提案手法では、最適化過程に Particle Swarm Optimization (PSO) [4] を適用する。従来手法 [3] の目的関数をベースに、手指のパーツ識別の結果を目的関数に反映することによって、高精度な手指の位置姿勢推定を実現する。PSO における最適化で使用する目的関数を式 (1) に示す。

$$E(\Theta) = E_{sim} + E_{const} + E_{lim} \quad (1)$$

目的関数は、3つの項から成り立っており、 E_{sim} は手指のモデルと観測データの相違度、 E_{const} は物理的な制約に基づく姿勢の評価値、 E_{lim} は手指の運動学的な可動範囲に基づく姿勢の評価値を示す。

E_{sim} は、入力された3次元の点群と手指のモデルとの相違度を示しており、手指のモデルとして図3の球の集合で構成された手指のモデルを使用する。

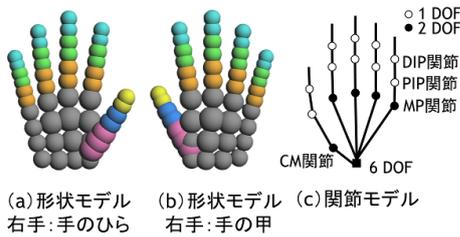


図3 球の集合としてあらわされる手指のモデル

47個の球体から構成された手指のモデルであり、 $M(\Theta) = \mathbf{s}_i$, $\{i = 1, 2, \dots, 47\}$, $\mathbf{s}_i = (\mathbf{c}(\Theta), r)$ として表現される。 \mathbf{s}_i は、各球を示しており、中心座標 \mathbf{c} と半径 r の情報を持っている。 E_{sim} は式 (2) によって算出される。

$$E_{sim}(\Theta) = \sum_{n=1}^N D(\mathbf{p}_n, \mathbf{s}_{near}) \quad (2)$$

\mathbf{p} は入力された3次元点群の1点を示しており、 N は3次元点の数を示している。 D は、式 (3) によって算出される。入力された3次元点とこれに最も近い球の中心座標のノルムから、その球の半径を引いた値を示している。

$$D(\mathbf{p}, \mathbf{s}) = (\|\mathbf{p} - \mathbf{c}\|) - r \quad (3)$$

従来手法 [3] は以上の処理によって“すべての球”の中で3次元点との距離が最も近い球との距離を計測しているのに対して、提案手法では“3次元点が表す手指のパーツ情報と対応する球”の中から距離が最も近い球との距離を計測する。従来手法では、単純に3次元点から最も近い球までの距離を計測していたが、この場合、実際には対応していない点同士（例：人差し指の指先を表す点と親指の指先を表すモデル球の中心点）から距離値を計測することがあり、対応していない点同士の距離が小さい場合にも評価値が小さくなることによって、誤推定が懸念される。そこで提案手法では、入力された3次元点に対して、手指のパーツ識別で得られた識別結果を付与し、対応するモデル球との距離を計測することによって、従来手法の問題点を解決する。

3. 実験結果と考察

レンジセンサ (Kinect v2) から取得した実データ (15秒間, 380フレーム, 約25fps) を用いて手指の位置姿勢推定の実験をおこなった。実データの各フレームで推定した関節の座標と正解データの座標の誤差を算出することによ

て、推定性能を評価した。また、従来手法 [3] の目的関数と PSO を使用した比較手法の実験結果と比較する。実験結果を図4に示す。

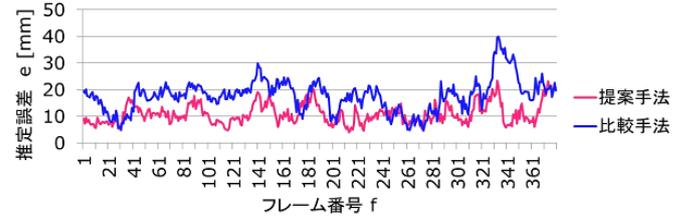


図4 推定誤差の比較実験の結果

横軸がフレーム番号、縦軸が正解データと比較した推定誤差を示す。このグラフの推定誤差は、すべての関節位置の誤差の平均値である。すべてのフレームの平均誤差は、提案手法が10.9mm、比較手法が17.8mmであり、提案手法が比較手法よりも約7mm精度が高いことを確認した。

図5に実データに対する手指と腕の識別結果、手指のパーツ識別結果、位置姿勢の推定結果の例を示す。

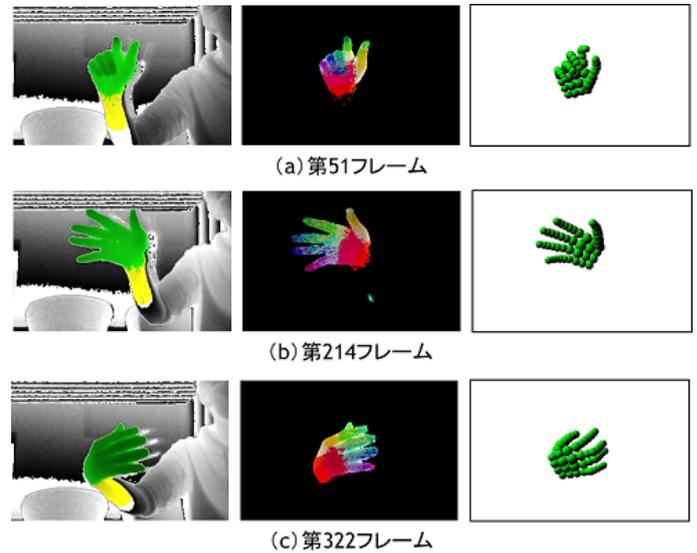


図5 実データに対する手指と腕の識別結果、手指のパーツ識別結果、位置姿勢の推定結果の例

複数の手指の姿勢において、正しく位置姿勢を推定できたことを確認した。

4. おわりに

手全体の大局的な情報だけでなく、事前にパーツ識別された手指の局所領域の情報を用いて、手指のモデルと観測データの整合性を評価することによって、モデルと観測データ間の誤対応を低減し、高精度な手指の位置姿勢を実現した。実データを用いた実験の推定精度が10.9mmであり、比較手法よりも約7mm精度が向上したことを確認した。

参考文献

- [1] Oikonomidis et al., Efficient Model-based 3D Tracking of Hand Articulations using Kinect, BMVC, pp.101.1-101.11, 2011.
- [2] Keskin et al., Hand Pose Estimation and Hand Shape Classification Using Multi-layered Randomized Decision Forests, ECCV, pp.852-863, 2012.
- [3] Qian et al., Realtime and Robust Hand Tracking from Depth, CVPR, pp.1133-1144, 2014.
- [4] Kennedy et al., Swarm intelligence, Morgan Kaufmann Publishers, 2001.