

手指形状の個人差に応じて選択される参照データベースを持つ 装着型手指形状推定システム

A wearable hand pose estimation with reference data to be selected according to individual differences in hand shape

杉村 聡太[†] 星野 聖[†]
Sota Sugimura Kiyoshi Hoshino

1. はじめに

モーションキャプチャ技術は、幅広い分野で利用されている。例えば、映画で高精度な人体の CG アニメーションを作成することや、ゲームでキャラクタに人間の動作を模倣させることなどに利用されている。しかし、例えば光学式モーションキャプチャでは、腕、脚、体躯といった大きな身体部位の姿勢を高精度に検出できるが、それらと同時に「手指」のような小さな身体部位の動作や姿勢を検出することは難しい。これは、腕や脚などの大きな身体部位の動作範囲に比べて、手指などの小さな身体部位の動作が非常に小さく、さらに、自由度が高いため動作も複雑であることが原因である。したがって、体全体の動きと同時に手指の動作を検出するためには、別の技術を組み合わせることが必要になる。

手指の動作を検出する既存の技術の一つに、データグローブがある (例えば[1])。データグローブの手袋には、多数の歪ゲージが内蔵されており、歪ゲージの曲がり具合によって手指の関節角度を取得する仕組みである。しかし、手袋を手指に密着させるため、締め付けが強く、手指動作が拘束される。また、手指の大部分を手袋で覆うため、特にピアノや弦楽器の演奏など、指先を使って物体を操作する動作を阻害する恐れもある。

手指を拘束することなく、動作を検出する既存のシステムに Digits[2]がある。Digits は、手首 (掌側) に搭載した赤外線カメラと赤外線 LED、赤外線レーザを用いて、センサから各指までの距離と指先の位置を取得し、逆運動学に基づいて手指関節角度の推定を行うことができる。しかし、手指を伸ばす場合や、手指を使って他の物体を操作する場合、ユーザ自身の手指や他の物体によって指先が遮蔽されることがあり、正しく推定することができない可能性がある。また、装置が掌側にあるため、日常での動作 (物体の操作や歩行時の腕振り) の妨げになる恐れがある。

これらの問題に対処するため、著者らは手の甲にワイヤレス小型カメラを設置し、撮影された画像からユーザの手指形状を推定する手法を提案する。カメラを手の甲に設置することにより、ユーザは手指全体を拘束することなく、また、あらゆる作業を妨げることなく、最小限の拘束で手指形状の推定を行うことが可能になる [3]。提案システムでは、星野ら [4] の手法に基づき、事前に様々な姿勢の手指画像と手指関節角度を合わせてデータベースを構築し、輪郭線情報をもとに、入力画像に最も類似する画像を探索する手法を用いる。これにより、指先の情報を必要とせず手指形状推定を行うことが可能になる。しかし同手法では、ユーザの手指形状の個人差により、データベース構築に用いた手指画像とユーザの手指との輪郭線にずれが生じ、誤差が蓄積されることで推定精度が低下することが問題とさ

れていた。そこで、提案システムでは、事前に様々な手指形状の個人差を持つ手指画像を用いて複数のデータベースを構築し、ユーザの手指画像を用いて短時間でユーザの手指形状に最も近いデータベースを選択して手指形状推定を行うことに対応する。これにより、どんな手指形状のユーザに対しても、非常に短いキャリブレーション時間で高精度の手指形状推定を実現する。

2. システム構成

2.1 ハードウェア構成

ハードウェアは、ワイヤレスカメラ、固定用アーム、モバイルバッテリーによって構成される。ワイヤレスカメラで撮影した手指画像をコンピュータに送信し、手指形状推定を行う。図 1 に装着装置の外観を示す。ワイヤレスカメラは、小型 RGB カメラである 3rd Eye Electronics 社製の 5.8G Wireless Mini Camera -TE60A を用いた。画角は 90°、取得される画像の解像度は 720×480[pixel]である。カメラの位置は、図 2 に示すように手の甲、拇指側であり、手指の面に垂直に向けて設置する。カメラから手の甲までの距離は、約 120[mm]である。固定アームは、柔軟に変形できる可撓性ワイヤとリストサポータによって構成される。カメラとバッテリーとを繋ぐケーブルは、腕の長さ以上であり、バッテリーはモーメントを感じない体幹付近に装着することが可能である。バッテリーを除いた装着装置の重量は、約 75[g]である。

2.2 データベース構築

2.2.1 データベース構造

データベースは、手指関節角度と後述するシルエット情報、高次局所自己相関 [5] (以降、HLAC と略す)、尾根線情報を一つのデータセットとして構成される。様々な手指の見え方に対するデータセットを集めたものをデータベースとする。データベースの概念図を図 3 に示す。

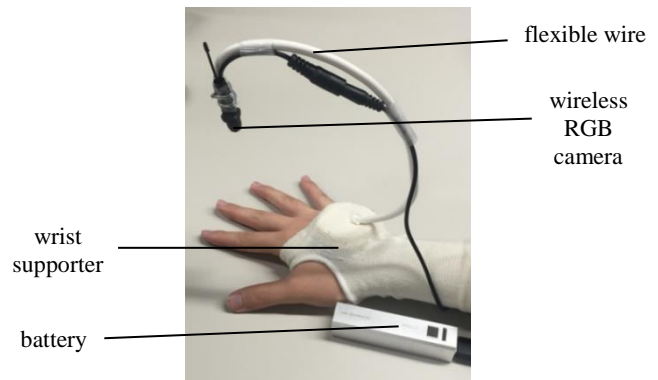


図 1. 装着装置の外観

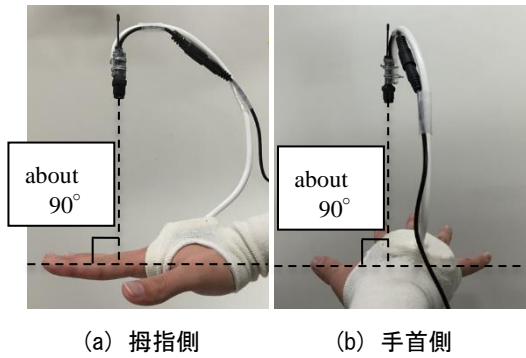


図 2. カメラ位置

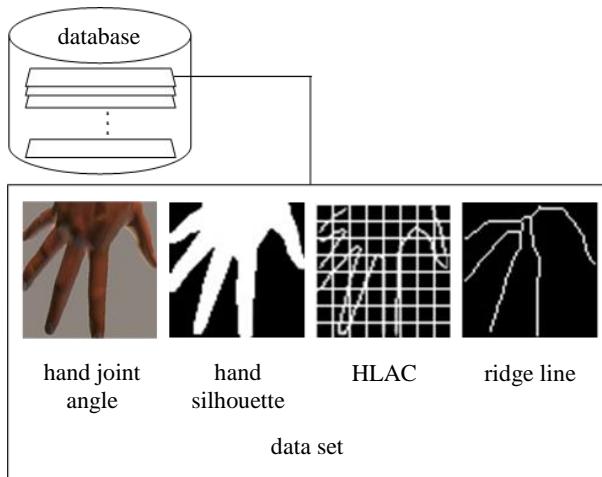


図 3. データベースの概念図

2.2.2 データベース構築用指画像の作成

データベース構築用の指画像は、3次元人間描画ソフト (Curious Labs 社製 Poser5) を用いて作成する。このソフトは、与えた手指関節角度に対応する手指形状 CG の画像を作成することができる。また、このソフトの手指形状 CG は、各指の第 1~3 関節のパーツと掌のパーツを合わせた 16 個のパーツで形成されており、各パーツを 3次元各方向にスケール変換することで、ユーザの手指形状の個人差を表現することができる。さらに、カメラパラメータを変更することで、ソフト内でのカメラの位置と姿勢を変更できるため、システム装着時に撮影されるユーザの手指の見え方をソフト内で再現することができる。

事前に、ソフトを用いて様々な特徴を持つ手指形状 CG を作成しておく。提案システムでは、まず標準的な形状の手指形状 CG (以降スタンダードモデルと略す) を用意し、個人差が生じやすい要素に関連するパーツを様々な比率でスケール変換することで、複数の手指形状 CG を作成した。個人差が生じやすい要素として、指の長さ、太さ、掌の幅、掌の下端から拇指の付け根位置の高さを選択した。指の長さを決めるパーツの比率を R_{length} 、太さを決めるパーツの比率を R_{width} 、掌の幅を決めるパーツの水平方向の比率を R_{palm} 、拇指の付け根位置の高さを決めるパーツの垂直方向の比率を R_{thumb} とし、それぞれ表 1 に示す 3 種類の比率を用意し、これらを組み合わせることで $3 \times 3 \times 3 \times 3 = 81$ 種類の手指形状 CG を作成した。作成した手指形状 CG の例を図 4 に示す。作成した手指形状 CG ごとに様々な手指関節

表 1. 各パーツにおけるスケール変換の比率

	小	標準	大
R_{length}	0.95	1.0	1.05
R_{width}	0.9	1.0	1.1
R_{palm}	0.95	1.0	1.05
R_{thumb}	0.9	1.0	1.1

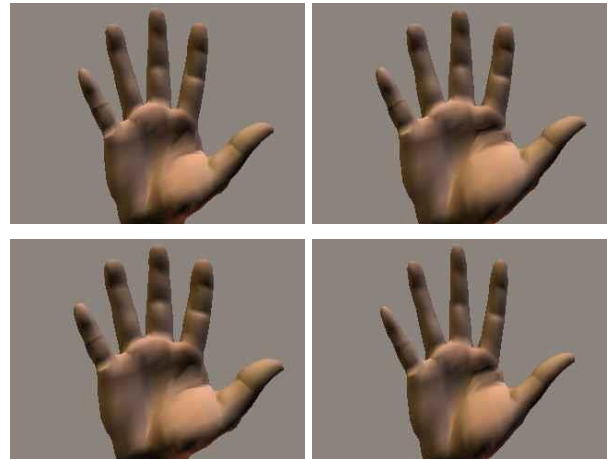


図 4. 手指形状 CG の例



図 5. 前処理の様子

角度を与え、レンダリングしてデータベース構築用指画像を作成する。画像作成時に与えた手指関節角度を、推定結果の出力用としてデータセットに保持する。

2.2.3 前処理

取得したすべての指画像に対して、まず、二値化によって背景分離を行う。次に、ラベリングを行い、面積が最も大きい領域を手指領域とする。最後に、手指領域画像を 64×64 [pixel] に縮小する。このときに得られた画像の各画素において、その画素が前景であるか背景であるかをシレット情報として保持する。入力画像と前処理によって得られる画像の例を図 5 に示す。

2.2.4 HLAC 特徴量

2.2.3 項で示した前処理によって得られた 64×64 [pixel] の画像から、手指領域の輪郭線を抽出する。図 6 のように、得られた輪郭線の画像を縦方向と横方向にそれぞれ 8 分割し、64 個のブロックに分ける。一つ一つのブロックに対して、図 7 に示す 25 種類の HLAC パターンがそれぞれいくつあるか数え上げ、それを HLAC 特徴量とする。一つのブロックに対して、25 次元の情報が得られるため、画像 1 枚あたり、 $25 \times 8 \times 8 = 1600$ 次元の特徴量を得る。

2.2.5 尾根線情報

手指姿勢の概形の情報を得るために、指や掌の中心を通り、手指形状の個人差の影響を受けにくい尾根線情報を用いる。2.2.3 項で示した前処理によって得られた 64×64[pixel]の画像に細線化処理を施し、細線化した画像の前景領域の座標を尾根線情報として取得する。細線化には、hilditch の手法[6]を用いる。図 8 に尾根線情報の例を示す。

2.3 推定手法

2.3.1 データベース選択

2.2.2 項で示したように、事前に用意した様々な特徴を持つ手指形状 CG は、スタンダードモデルの指の長さ、太さ、掌の幅、拇指の付け根位置の高さという四つの要素に関連するパーツを様々な比率でスケール変換することによって作成している。そこで、ユーザに対しても、四つの要素それぞれにおいてスタンダードモデルに対する比率を求め、得られた比率を事前に作成した手指形状 CG の比率と

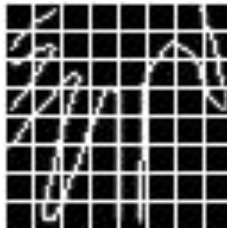


図 6. 8×8 ブロックに分割した手指領域輪郭線画像

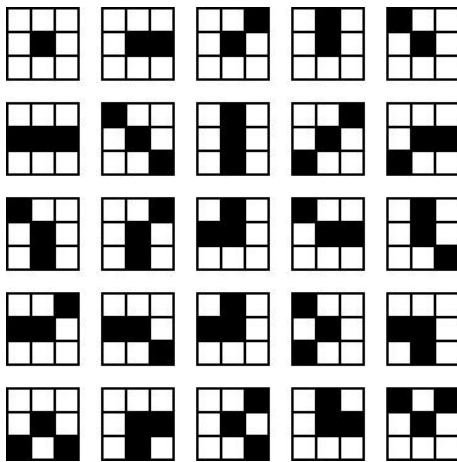


図 7. HLAC パターン



図 8. 尾根線情報の例

比較することで、ユーザの手指と最も類似した特徴を持つ手指形状 CG を選択する。各要素においてユーザのスタンダードモデルに対する比率に最も近い値の比率を求め、それらの組み合わせを持つ手指形状 CG をユーザの特徴を最もよく反映した手指形状 CG とし、選択した手指形状 CG に対応するデータベースを使用する。

ユーザの手指画像を用意し、画像から、全指の長さ、太さ、掌の幅、掌の下端からの拇指の付け根位置の高さの比率を算出するための情報を取得する。まず、指の長さの情報として、各関節間の長さをそれぞれ取得し、図 9 に示すように $L_1 \sim L_{14}$ とする。次に、指の太さの情報として、関節と関節の中心点における指の太さを取得し、図 10 に示すように $W_1 \sim W_{14}$ とする。また、掌の形状の情報として、四指の付け根と拇指の付け根の間の高さにおける掌の幅、中指の付け根から掌の下端までの垂直方向の距離を取得し、それぞれ P_{width} , P_{height} とする。拇指の付け根位置の情報として、拇指の付け根から中指の付け根までの水平方向の距離、拇指の付け根から掌の下端までの垂直方向の距離を取得し、それぞれ T_{width} , T_{height} とする。掌の形状、拇指の付け根位置の情報の例を図 11 に示す。

また、スタンダードモデルの画像を用意し、同様の情報を取得する。取得した値をそれぞれ、 $L'_1 \sim L'_{14}$, $W'_1 \sim W'_{14}$, P'_{width} , P'_{height} , T'_{width} , T'_{height} とする。次に、ユーザの各要素の情報とスタンダードモデルの各要素の情報を比較するために、掌の幅が等しくなるようにユーザの各要素の情報の値を正規化する。正規化した情報を $L''_1 \sim L''_{14}$, $W''_1 \sim W''_{14}$, P''_{width} , P''_{height} , T''_{width} , T''_{height} とする。こうして得られた正規化後のユーザの各要素の情報をスタンダードモデルの各要素の情報と比較し、個人差が生じやすい四つの要素それぞれにおいて、以下の式に基づき、ユーザのスタンダードモデルに対する比率を求める。

$$r_{length} = \frac{\sum_{i=1}^{14} \frac{L''_i}{L'_i}}{14} \quad (1)$$

$$r_{width} = \frac{\sum_{i=1}^{14} \frac{W''_i}{W'_i}}{14} \quad (2)$$

$$r_{palm} = \frac{\frac{P''_{width}}{P'_{width}} + \frac{T''_{width}}{T'_{width}}}{2 \frac{P''_{height}}{P'_{height}}} \quad (3)$$

$$r_{thumb} = \frac{T''_{height}}{T'_{height}} \quad (4)$$

- r_{length} : 指の長さにおけるユーザのスタンダードモデルに対する比率
- r_{width} : 指の太さにおけるユーザのスタンダードモデルに対する比率
- r_{palm} : 掌の幅におけるユーザのスタンダードモデルに対する比率
- r_{thumb} : 掌の下端から拇指の付け根位置の高さにおけるユーザのスタンダードモデルに対する比率

求めた比率と表 1 に示した様々な手指形状 CG の比率を要素ごとに比較し、最も類似する比率の組み合わせを持つ手指形状 CG を選択し、選択した手指形状 CG に対応するデータベースを使用する。

これらの工程は、ユーザの手指画像を撮影する作業を含めても、約 2 分間で行うことができる。

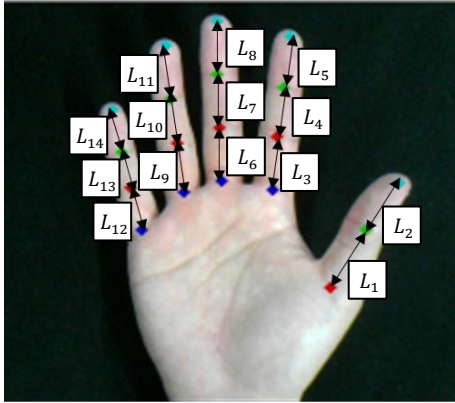


図 9. 指の長さの情報の例



図 10. 指の太さの情報の例

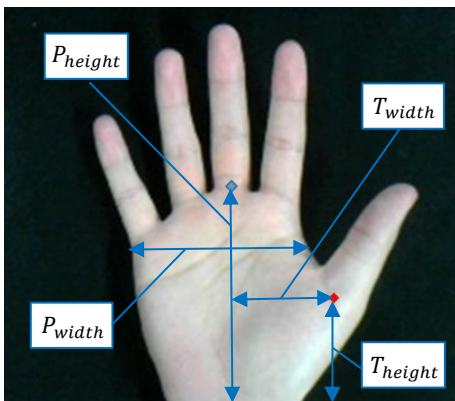


図 11. 掌の形状、拇指の付け根位置の情報の例

2.3.2 推定の流れ

ユーザは装置を装着し、入力画像からシルエット情報、HLAC 特徴量、尾根線情報を算出する。算出手法については、2.2.3～2.2.5 項に示したデータベース構築時と同様である。取得した特徴量をもとにデータベースを探索して、入力画像と最も類似するデータセットを推定結果とする。最後に、推定結果のデータセットが持つ手指関節角度を出力する。以下に、その詳細を説明する。

2.3.3 探索対象の絞り込み

入力画像と著しく異なる姿勢を持つデータセットを探索対象からはずすため、2 段階の探索対象の絞り込みを行う。2.2.3 項と 2.2.5 項で示した手法により取得したシルエット情報と尾根線情報を用いて、後述するペナルティを算出する。

まず、入力画像の尾根線情報とデータセットのシルエット情報をもとに、入力画像の尾根線領域がデータベース画像のシルエット領域からはみ出す面積を算出し、はみ出した面積の値[pixel]をペナルティ P_1 とする。図 12 にペナルティの例を示す。第 1 段階の絞り込みとして、 P_1 の値が以下の式で示す閾値を超えるデータセットを探索対象からはずす。

$$th_1 = C \cdot A_{input} \quad (5)$$

th_1 : 第 1 段階の絞り込みに関する閾値
 C : 定数
 A_{input} : 入力画像の尾根線領域の面積

第 1 段階の絞り込みをした後、探索対象として残ったデータセットに対して第 2 段階の絞り込みを行う。データセットの尾根線情報と入力画像のシルエット情報をもとに、データセットの尾根線領域が入力画像のシルエット領域からはみ出す面積を算出し、はみ出した面積の値[pixel]を 2 つ目のペナルティ P_2 とする。第 2 段階の絞り込みとして、 P_2 の値が以下の式で示す閾値 th_2 を超えるデータセットを探索対象からはずす。

$$th_2 = C_1 \cdot A_{dataset}^2 + C_2 \cdot A_{dataset} \quad (6)$$

th_2 : 第 2 段階の絞り込みに関する閾値
 C_1, C_2 : 定数
 $A_{dataset}$: データセットの尾根線領域の面積

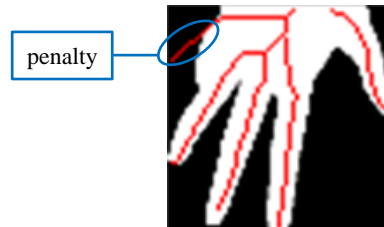


図 12. ペナルティの例

2.3.4 類似度計算

2.3.3 項に示した手法で探索対象の絞り込みを行った後、推定対象として残ったデータセットに対して類似度計算による探索を行う。類似度は、HLAC 特徴量と 2.3.3 項で示したペナルティ P_1 , P_2 を用いて、以下の式によって算出する。

$$E[i] = \sum_{k=1}^D (current_k - data[i]_k)^2 + C_3 \cdot P_1[i] + C_4 \cdot P_2[i] \quad (7)$$

- i : データベース番号
- $E[i]$: 入力画像とデータベース番号 i の類似度
- k : HLAC 特徴量の次元番号
- D : 画像一枚の HLAC 特徴量の次元数
- $current_k$: 入力画像の次元番号 k の HLAC 特徴量
- $data[i]_k$: データ番号 i の画像の次元番号 k の HLAC 特徴量
- C_3, C_4 : 定数
- $P_1[i]$: データ番号 i のシルエット情報と入力画像の尾根線情報から算出したペナルティ
- $P_2[i]$: データ番号 i の尾根線情報と入力画像のシルエット情報から算出したペナルティ

データベースを全て探索した結果、類似度計算によって得られた類似度が最小のデータセットを推定結果とし、推定結果のデータセットが持つ手指関節角度を出力する。

3. 評価実験

3.1 評価方法

提案システムの推定精度を検証するため、評価実験を行った。3名の被験者に対して、スタンダードモデルの画像によって構築されたデータベースを用いて推定を行った場合（以降、従来手法とする）と、2.3.1 項に示した手法により選択した被験者に類似した手指形状 CG の画像によって構築されたデータベースを用いて推定を行った場合（以降、提案手法とする）での出力される関節角度の推定精度を比較した。推定は 2.3.3, 2.3.4 項に示した手法により行った。右手に提案システム、左手にデータグローブ（Cyber Grove Systems 社製 CyberGrove II）を装着した状態で、左右の手で同じ動作を行い、データグローブで取得した値を真値として推定精度を検証した。なお、実験時の入力動作は全ての被験者で統一した。

全ての実験において、示指、中指、薬指、小指の各屈曲伸張動作、示指、中指、薬指、小指の内転外転動作、拇指動作を独立に動作させて作成した 20448 のデータセットを持つデータベースを使用した。また、実験では 2 章における変数をそれぞれ $C = 0.2$, $C_1 = 0.002$, $C_2 = 0.03$, $C_3 = C_4 = 800$ とした。実験で使用した計算機は Intel®Core™i5-2500 CPU 3.30GHz, 8.00GB RAM のデスクトップ PC である。

3.2 結果と考察

表 2, 3 に、実験によって得られた結果のうち、示指、中指、薬指、小指の四指の各関節を代表して、平均的な長さかつ関節の可動域が大きい示指の PIP 関節、四指とは異なる動作をする拇指の CM 関節における推定誤差の平均値、及び標準偏差の値を被験者間で比較したものを示す。表 2, 3 より、すべての被験者に対して、従来手法よりも提案手法の方が示指 PIP 関節、拇指 CM 関節の推定結果において、推定誤差の平均値が 0 に近く、標準偏差も小さいことがわかる。被験者平均では、推定誤差の平均値と標準偏差が、示指 PIP 関節で、従来手法では -3.66 ± 18.41 度、提案手法では -0.86 ± 17.67 度、拇指 CM 関節で、従来手法では 3.76 ± 11.76 度、提案手法では 0.98 ± 11.56 度となり、推定誤差の標準偏差は同程度であるが、平均値を約 2 度から 3 度低減することを実現した。また、提案手法では、従来手法に比べて被験者間の推定誤差の平均値のばらつきが少なくなっていることがわかる。このことから提案手法は、個人差対応に有効であることが示唆される。

また、提案システムの推定精度が直感的にわかるように、図 13 に入力画像と推定結果の手指画像を比較した様子を示す。さらに、推定結果のデータセットが持つ関節角度をロボットハンドの CG に出力し、ユーザの手指姿勢と比較した様子を図 14 にスナップショットにて示す。図 13 より、提案システムでは入力画像と類似する画像を高精度で探索できていることが直感的にわかる。また、図 14 より、ユーザの手指姿勢が CG のロボットハンドによって高い精度で再現されていることから、手指関節角度が高精度に推定されていることが示唆される。

画像が入力された後の、前処理から推定結果を出力するまでの処理速度は平均 60[fps]であった。

表 2 . 示指 PIP 関節の推定誤差 [deg]

		conventional method	Proposed method
subject A	average	-2.89	-2.02
	s.d.	15.42	15.04
subject B	average	-3.77	1.49
	s.d.	17.27	17.00
subject C	average	-4.31	-2.05
	s.d.	22.54	20.97

表 3. 拇指 CM 関節の推定誤差 [deg]

		conventional method	proposed method
subject A	average	4.95	1.08
	s.d.	11.48	11.37
subject B	average	4.31	1.18
	s.d.	10.31	9.85
subject C	average	2.03	0.68
	s.d.	13.49	13.46



図 13. 推定結果例

(左 : 入力画像, 右 : 推定結果の手指画像)

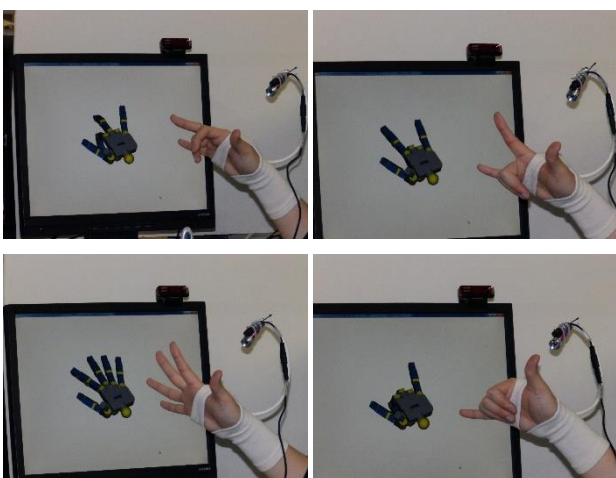


図 14. 推定結果のスナップショット

4. まとめ

モーションキャプチャにおいて、腕や脚、体躯といった大きな部位の動作と同時に手指などの小さな部位の動作を検出することは難しく、手指専用の計測技術を組み合わせる必要がある。これに関して本研究では、手の甲に小型カメラを設置し、撮影された手指画像の輪郭線情報を用いて手指形状推定を行うシステムの構築を行った。従来の輪郭線を情報源として、事前に用意したデータベースから入力画像に最も類似する画像を探索する手法では、ユーザの手指形状の個人差により、データベース構築に用いた手指画像とユーザの手指との輪郭線にずれが生じ、誤差が蓄積されることで推定精度が低下することが問題とされていた。そこで、事前に様々な手指形状の個人差を持つ手指画像を用いて複数のデータベースを構築し、ユーザの手指画像を用いて短時間でユーザの手指形状に最も近いデータベースを選択して手指形状推定を行う手法を提案した。被験者 3 名に対して、従来手法と提案手法の推定精度を検証する実験を行い、示指 PIP 関節と拇指 CM 関節における推定誤差の平均値と標準偏差を評価した結果、すべての被験者において、提案手法では従来手法に比べ推定誤差の平均値と標準偏差が低減し、推定精度の向上が見られた。被験者平均では、推定誤差の平均値と標準偏差が、示指 PIP 関節で、従来手法では -3.66 ± 18.41 度、提案手法では -0.86 ± 17.67 度、拇指 CM 関節で、従来手法では 3.76 ± 11.76 度、提案手法では 0.98 ± 11.56 度となり、推定誤差の平均値を 2 度から 3 度低減することを実現した。また、提案手法では従来手法に比べて、被験者間の推定誤差の平均値のばらつきが少なくなっており、ユーザの手指形状の個人差に対応することに有効であることが示唆された。提案システムの処理速度は平均 60[fps]であり、リアルタイムでの推定が可能であった。一連の結果は、提案システムの有効性を示していた。

参考文献

- [1] CyberGlove, <http://www.cyberglovesystems.com/cyberglove-ii/>
- [2] David Kim, Otmar Hilliges, Shahram Izadi, Alex Butler, Jiawen Chen, Iason Oikonomidis, Patrick Olivier, "Digits: Freehand 3D Interactions Anywhere Using a Wrist Worn Gloveless Sensor", UIST'12 Proceedings of the 25th annual ACM symposium on User interface software and technology, pp.167-176 (2012).
- [3] Kiyoshi Hoshino, "Ultraminiature camera-based hand pose estimation", Proc. 18th International Conference on Mechatronics Technology (ICMT), 18, 77, pp.1-8 (2014).
- [4] Kiyoshi Hoshino, Takanobu Tanimoto, "Real time search for similar hand images from database for robotic hand control", IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A, 10, pp.2514-2520 (2005).
- [5] Nobuyuki Otsu, Takio Kurita, "A new scheme for practical, flexible and intelligent vision systems", IAPR Workshop on Computer Vision, pp.431-435 (1998).
- [6] C. Judith Hilditch, "Linear Skeletons From Square Cupboards", Machine Intelligence, 6, pp.403-420 (1969).