

法線ヒストグラムを用いたモデルベースの手指姿勢推定

Model-Based estimation of human hand postures using normal histograms

村井 隼人 (中央大) ○ 正 多田 充徳 (産総研) 増山 岳人 (中央大) 正 梅田 和昇 (中央大)

Hayato MURAI (Chuo University, murai@sensor.mech.chuo-u.ac.jp),
Mitsunori TADA (Digital Human Research Center, AIST, m.tada@aist.go.jp),
Gakuto MASUYAMA (Chuo University, masuyama@mech.chuo-u.ac.jp),
Kazunori UMEDA (Chuo University, umeda@mech.chuo-u.ac.jp)

In this paper, we propose a method for estimating the pose of a human hand using normal histograms. The method is composed coast to fine of three steps. First, within the database which consist of hand model postures in a reduced degree of freedom space about human hand postures in a daily life, screening of postures is conducted using dispersion of point clouds. Second, the similarity between normal histograms generated from the input posture and the database posture is evaluated with Chi Square. Third, optimization for minimizing the evaluation function is conducted in the reduced parameter space. In the experiment, the hand model was used as input postures. The result of the experiment shows that the maximum estimation error was 21[%] using this method.

Key Words: normal histoglams, model-based, pose estimation, principal component

1 序論

近年、デジタルカメラやスマートフォンなどの、手で握って操作する製品の仮想的なエルゴノミック性評価が試みられている。製品を操作する際の手指運動動作は、主にモーションキャプチャを利用して計測される。しかし、モーションキャプチャ自体は非常に高価な機器であり、キャリブレーション(校正)や後処理(リンクモデルのラベリング)に手間がかかるという課題がある。

このような課題に対して、より安価な計測デバイスを使用した、非侵襲の手指姿勢推定に関する研究が多くなされている [1]。代表的な手法は、カラーカメラなどを用いた二次元画像による手指姿勢推定と、距離画像センサなどを用いた距離画像による手指姿勢推定である。

二次元画像を用いた手指姿勢推定は、高速な計測と画像処理が特徴である。一方で、肌色検出や手指の輪郭による手指形状の抽出を行う必要がある。そのため、姿勢推定精度が形状抽出精度に依存する課題が挙げられる。

このような理由から、近年では、距離画像センサを用いた手指姿勢推定手法が多く利用されている。特に、手モデルを姿勢推定に用いた、連続的な手指運動姿勢推定の手法が多く提案されている。一方で、手モデルには多くの自由度があり、多次元空間内での手指姿勢推定は、推定のプロセスが長くなる問題がある。そこで Schröder ら [2] は、低次元空間内での手指姿勢推定手法を提案した。手モデルと計測された姿勢の三次元的な整合性は、非剛体 ICP (Iterative Closest Point) アルゴリズムを用いた点群の位置合わせによって評価される。これにより、指の変形を考慮した姿勢推定が可能となる。しかし、高い位置合わせ精度が求められるため、コンピュータ上での計算コストが増大し、GPU の処理に依存しなければならない課題がある。

そこで本稿では、手モデルと手指三次元形状の法線情報の分布(以降、法線ヒストグラムと呼称)を用いることで、非剛体 ICP を必要としない、計算コストの低減を目的とした三次元手指姿勢推定の手法を提案する。手指運動姿勢を低次元化した主成分空間(以降、低次元化空間と呼称)内で姿勢推定を行うことで、姿勢推定の計算量削減を目的とする。

2 手指姿勢推定手法

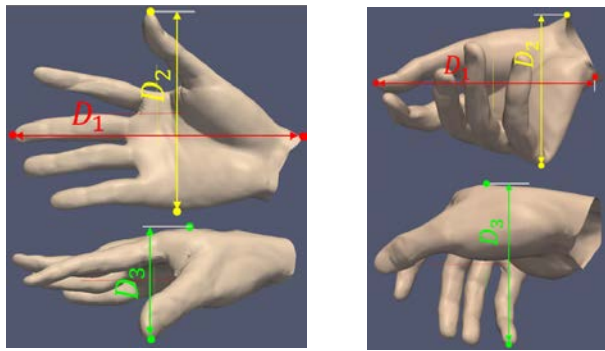
本研究の手指姿勢推定手法は、入力手指姿勢と低次元化空間内の代表的な手指点群(以降、データベースと呼称)との比較と、最適化による低次元化空間内での手指姿勢推定により構成される。本手法を三段階のステップで分けると、以下のような流れになる。

- Step1 点群データのばらつきを用いた、データベース内の類似度の低い手指点群データの除去
- Step2 法線ヒストグラムを用いた、データベース内の類似度の高い手指点群データのマッチング(低次元化空間内での最適化の初期値決定)
- Step3 法線ヒストグラムを用いた、低次元化空間内での最適化による類似度の高い手指点群データの探索(最適解の探索)

2.1 データベース内の類似度の低い手指点群データの除去

低次元化空間から生成されるデータベースに対して、入力手指姿勢とは明らかに異なる姿勢の除去(スクリーニング)を行う。手指点群データベースに関しては、3 章にて説明する。

まずは、入力手指点群データとデータベース内点群データの初期的な位置合わせを行う。点群データは三次元座標情報を持つので、点群のばらつきを三つの主軸方向に分けることが出来る。それぞれの主軸方向のばらつきを、 D_1 、 D_2 、 D_3 と定義する。例として、開いた手と閉じた手の主軸方向を、Fig. 1 に示す。



(a) Opened hand posture (b) Closed hand posture

Fig. 1 Distribution of hand point cloud

次に、求められた点群データ主軸沿いのばらつきを基準として、手指点群データの手指姿勢としての類似度を評価する。これにより、データベース内の類似度の低い手指姿勢を削減する。ここで、それぞれ手指点群データ主軸沿いの点群のばらつきを、関数 f_{D_1} , f_{D_2} , f_{D_3} で定義する。手指点群のスクリーニングには、式 (1) に示す評価値 S_d を用いる。姿勢類似度の評価は、 $1 - S_d$ とする。

$$S_d(P, Q) = \sqrt{\left(\frac{f_{D_2}(P)}{f_{D_2}(Q)} - \frac{f_{D_2}(Q)}{f_{D_2}(P)}\right)^2 + \left(\frac{f_{D_3}(P)}{f_{D_3}(Q)} - \frac{f_{D_3}(Q)}{f_{D_3}(P)}\right)^2} \quad (1)$$

ただし、入力手指姿勢を P 、データベース内手指姿勢を Q とする。

2.2 データベース内の類似度の高い手指点群データのマッチング

低次元化空間内の手指姿勢から生成した手指点群データベースと入力手指姿勢の姿勢類似度を、法線ヒストグラムにより評価する。

本手法で用いられる法線ヒストグラムは、Horn らによって提案された Extended Gaussian Image (EGI) [3] を基にした、物体の三次元形状の表現方法である。Fig. 2 に、球面上に表現した法線ヒストグラムと、三次元平面で表現した法線ヒストグラムの例を示す。法線ヒストグラムは、法線の方向を仰角 ($-90[\text{deg}] \sim 90[\text{deg}]$) と方位角 ($-180[\text{deg}] \sim 180[\text{deg}]$) で表す。

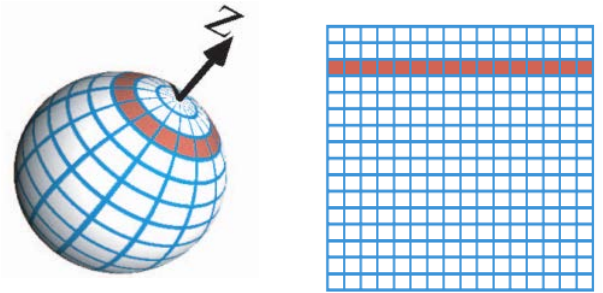
法線ヒストグラムによる類似度評価には、カイ二乗を用いる。カイ二乗の値が 0 に近いほど、互いのヒストグラムは類似していると言える。法線ヒストグラム同士のカイ二乗の値 S_h を、式 (2) に示す。

$$S_h(P, Q) = \sum_i \sum_j \frac{(g(P)_{ij} - g(Q)_{ij})^2}{g(P)_{ij} + g(Q)_{ij}} \quad (2)$$

ただし、 $g(P)$ は推定対象となる手指姿勢 P の法線ヒストグラムを、 $g(Q)$ は手指点群データベース内手指姿勢 Q の法線ヒストグラムを表わす。

2.3 最適化による手指姿勢推定

2.2 節では、入力手指姿勢と最も類似度の高いデータベース内手指姿勢が求められる。この手指姿勢を初期値として、低次元化空間内での最適化を行う。最適化関数は法線ヒストグラムの誤差二乗和とし、最適化パラメータは低次元化空間内での主成分得点とする。



(a) Normal histogram described on a sphere

(b) 2D normal histogram

Fig. 2 Normal histogram description [4]

3 手指点群データベース

本研究では、日常生活における手指運動姿勢のみに限定して、手指姿勢データベースを構築している。この制約により、姿勢推定の候補が削減され、姿勢マッチングの回数が単純に少ない姿勢推定を行うことが可能である。

生成の手法は、多田らによって提案された手法 [5] に準拠する。また、本研究では、産業技術総合研究所にて開発された『Dhaiba-Hand』 [6] を手モデルとして使用する。

多田らは、作業療法における手の動的な使用形態の分類 [7] から、代表的な二十八動作を選定した手指運動動作の主成分分析を行った。実験では、四名の被験者を対象にして、動作中の手指運動をモーションキャプチャ (VICON MX40: Vicon Motion Systems Ltd.) を用いて計測した。主成分分析の結果、上位六主成分までで、説明率は 80% を超えることが分かった。これにより、25DoF 以上の自由度を持つ手指運動姿勢を、低次元化した空間内で表わすことが可能となった。

手指の自由度を減らすことで、主成分空間内での類似姿勢探索に費やす計算コストを減らすことが可能となる。そこで、説明率が 60% を超える上位二主成分空間から、手指点群データベースを生成する。上位二主成分空間に、計測した手指運動姿勢をプロットした散布図と、本研究で使用する手指点群データベースを、Fig. 3 に示す。動的な手指運動姿勢を計測したため、散布図上では同じ動作に属する手指姿勢が線状で現れている。また、灰色の楕円は、主成分分析結果の $\pm 2\sigma$ 楕円を表わしている。

4 提案手法検証実験

4.1 手指点群データベースの生成

まず、本実験で用いる手指点群データベースの生成を行った。ここでは、上位二主成分空間における $\pm 2\sigma$ 楕円上の手指姿勢を八姿勢、空間中心の手指姿勢を一姿勢、計九姿勢を生成した。

これにより、上位二主成分空間内での姿勢を万遍なく網羅した手指点群データベースとなる。生成した手指点群データベース姿勢は、Fig. 3 に示した通りである。

4.2 入力手指点群データの生成

本手法は、将来的に距離画像センサで計測した手指姿勢の点群データに対して姿勢の推定を行う。そこで本稿では、姿勢推定の精度を評価するために、データベース内の姿勢が既知である点群データに対して推定を行う。ただし、距離画像センサで計測した点群データを想定し、入力点群データに対して粗雑処理を加える

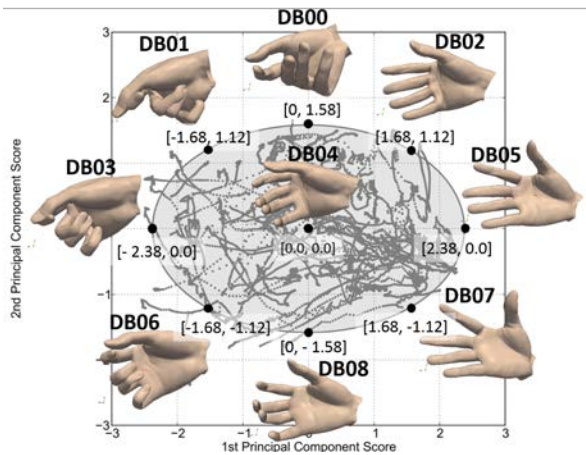


Fig. 3 Distribution of hand postures within first and second principal component space

ことで、疑似的な計測データを生成する。

4.2.1 姿勢の選定

推定対象とする姿勢は、生成した手指点群データベースの中から四姿勢を選定した。選定した姿勢は、低次元化空間内での主軸と $\pm 2\sigma$ 楕円が重なる点に生成された手指姿勢とした。つまり、DB00, DB03, DB05, DB08 と同様の手指姿勢を、入力手指姿勢とする。それぞれ、Input Posture(IP)00, IP03, IP05, IP08 と呼称する。

4.2.2 粗雑処理

Kinect for Windows v2(以下, Kinect) を用いて, Kinect から約 0.5[m] の位置で手指掌側の三次元形状を計測した場合, およその点群数は 6,000 点程度である。掌側, 背側を合わせると, 約 12,000 点になると考えられる。そこで, 入力点群データを, ランダムに 12,000 点程度に削減した。

4.3 データベースのスクリーニングの検証実験

式 (1) により求められた類似度を降順に並べた結果を, Fig. 4 に示す。スクリーニングを行う姿勢は, 降順の類似度の差分により求める。差分が最大となる類似度以下のデータベース姿勢を, 姿勢推定対象から除外する。Fig. 4 棒グラフ内の黒色バーは, スクリーニングによってデータベースから除外される姿勢を表わしている。よって, 36 姿勢中 9 姿勢が除外されていることが分かる。つまり, 次ステップの姿勢推定に用いられるデータベース手指姿勢が 1/4 削減されるため, 計算時間を 3/4 に削減できることが分かる。また, 結果より, Fig. 3 に示したデータベース手指姿勢の位置関係を見ると, 入力手指姿勢の最近傍には存在しない手指姿勢が, スクリーニングによって除外される姿勢に選ばれている。つまり, 推定対象姿勢とは明らかに異なる姿勢のみを除外できていることが分かる。

4.4 最適化による手指姿勢推定の初期値決定の検証実験

入力手指点群データとデータベース内手指点群データの初期位置姿勢合わせは, 点群の各主軸方向を一致させることで行われる。初期位置姿勢合わせが行われたデータベース手指点群データと入力手指点群データの法線ヒストグラムを導出し, その類似度を式 (2) により比較した。例として, 初期位置姿勢合わせ後の手指点群データベースの DB05 点群データと導出された法線を Fig. 5 に, 法線ヒストグラムを Fig. 6 に示す。ただし, 検証実験で用いる法

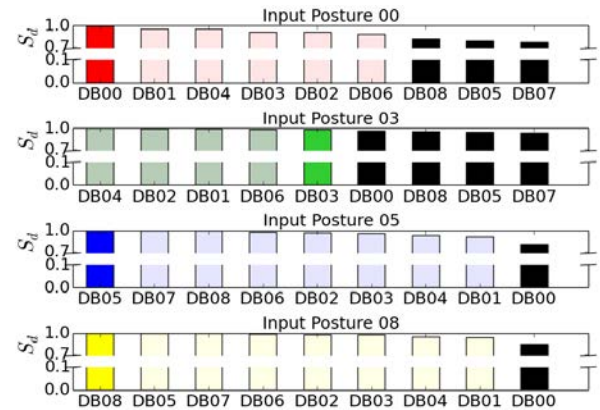


Fig. 4 Result of hand posture screening

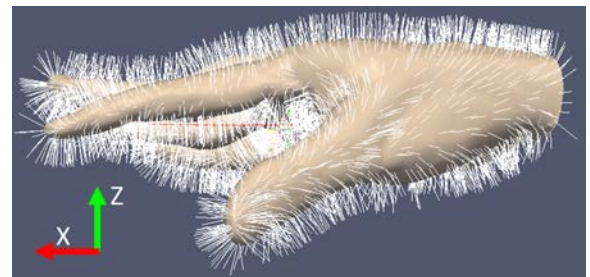


Fig. 5 DB05 with meshes and normals

線ヒストグラムの解像度は, 40×80 とした。Fig. 5 より, 手指の掌側と背側には Z 軸方向を向く法線が多く存在していると見て取れる。また, Fig. 6 を見ると, 法線ヒストグラムの上下に分布が広がっており, Z 軸方向 (仰角最大, 最小方向) に法線が多く存在していることが見て取れる。よって, 法線情報がヒストグラムとして正しく出力されていることが分かる。

本稿では, 法線ヒストグラムに対して, 標準偏差 $\sigma = 1$ のガウシアンフィルタによりヒストグラムの平滑化を行う。また, 法線ヒストグラムは正規化されている。

代表的な手指点群データベースの中で, 入力手指点群データとの類似度が最も高くなった手指姿勢を, Fig. 7 に示す。ここでは, Step2 における低次元化空間内での姿勢推定結果 (IP~) と, 予想される推定結果のデータベース姿勢 (生成元の姿勢) (DB~) を表わしている。

結果を見ると, 手指点群データベースの中で IP05 及び IP08 と最も類似した姿勢は, 点群データの生成元である DB05 と DB08 であると推定された。一方, IP00 と IP03 に関しては, 生成元である DB00 と DB03 が最も類似した姿勢であるという推定結果にはならなかった。しかし, Fig. 3 に示したデータベース手指姿勢の位置関係を見ると, 生成元である DB00 と DB03 は, DB04 の最近傍に存在する姿勢であることが分かる。よって, 推定結果としては, 生成元である DB00 と DB03 に次いで, 類似した姿勢が推定されることが分かった。

4.5 最適化による低次元化空間内での手指姿勢推定

最適化の手法には, 比較的計算効率が良いとされる BFGS 法を用いる。最適化による低次元化空間内での姿勢推定の結果と, それぞれの入力手指姿勢に対する全関節角度の平均誤差率を, Table 1 に示す。Step2 により推定された初期点を Initial に, Step3 に

Table 1 Principal component score and joint angle error of estimated posture in Step3

Input posture		IP00		IP03		IP05		IP08	
Estimation step		Initial	Final	Initial	Final	Initial	Final	Initial	Final
Principal component score	The 1st axis	0.00	-0.38	0.00	0.00	2.38	2.38	0.00	0.05
	The 2nd axis	0.00	0.36	0.00	0.00	0.00	0.00	-1.58	-1.69
Joint Angle Error[%]		13.46		21.14		0.00		5.78	

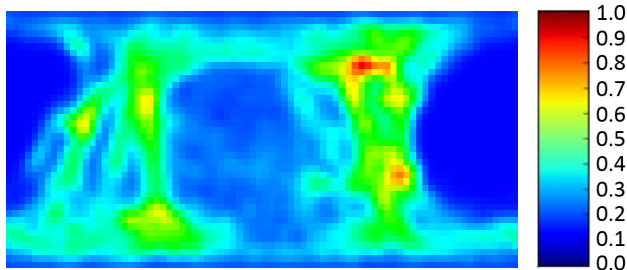


Fig. 6 Normal histogram of DB05

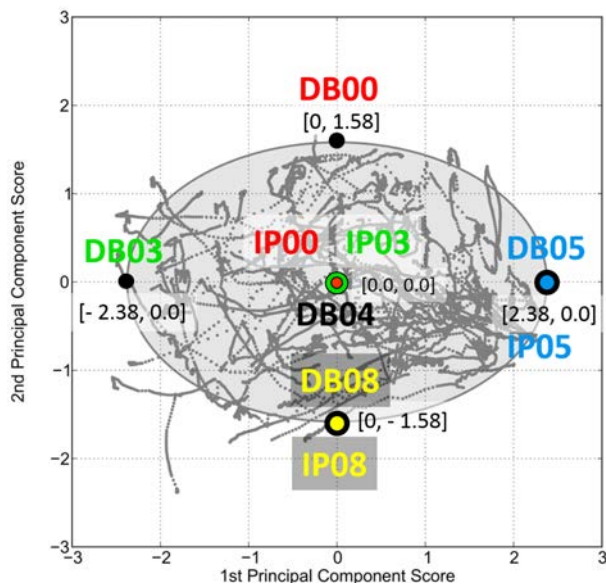


Fig. 7 Estimated hand postures in Step2

5 結論

本論文では、法線ヒストグラムと低次元化空間を用いて、計算コストの低減を目的としたモデルベースの手指姿勢推定手法を提案した。実験では、ただか 40 × 80 の情報量での細かい手指姿勢の推定を行い、計算コストの低減を図った。また、低次元化空間内での姿勢推定では、非剛体 ICP を必要としないアルゴリズムで、探索範囲を狭めた最適化による細かい手指姿勢の推定を可能とした。結果より、法線ヒストグラムを用いることで、低次元化空間内の予想される姿勢と近い姿勢を推定することが出来た。姿勢推定誤差は、最大で約 21[%] となり、法線ヒストグラムを用いることで関節角度レベルでの姿勢推定誤差を求めることが可能となった。

姿勢推定時間に関して、Step2 の一対一の姿勢類似度評価では、類似度評価時間の平均は 0.63[s] となった、また、Step3 では、入力手指姿勢四姿勢の平均姿勢推定時間は 688[s] となった。これは、数万のランドマーク頂点を持つ DhaibaHand 手モデルの SSD を点群データへ変換する際に、高い計算コストを費やしているためである。点群データへの平均変換処理時間は 5.47[s] であった。コンパイラ型言語による姿勢推定環境を構築した上で、手モデルメッシュの簡略化や、姿勢の変形をあらかじめ法線ヒストグラムによってデータベース化を行うことで、姿勢推定の高速化が可能である。

今後は、最適化のパラメータを調整し、局所最適解に陥らないロバストな姿勢推定手法の確立を目指す。また、実際に距離画像センサを用いた姿勢推定を行い、実用的な姿勢推定システムの構築を行っていく。

参考文献

- [1] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Comput. Vis. Image Underst.*, vol. 108, no. 1-2, pp. 52-73, 2007.
- [2] M. Schröder and M. Botsch, "Online Adaptive PCA for Inverse Kinematics Hand Tracking," *Proc. of Vision, Modeling and Visualization*, pp. 111-118, 2014.
- [3] B. K. P. Horn, "Extended Gaussian Images," *Proc. IEEE*, vol. 72, no. 12, pp. 1671-1686, 1984.
- [4] A. Makadia, A. I. Patterson, and K. Daniilidis, "Fully automatic registration of 3D point clouds," *Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1297-1304, 2006.
- [5] M. Tada, S. Masuda, Y. Endo, and H. Takemura, "Principal Component Analysis of Hand Postures in Daily Life," *日本機械学会ロボティクス・メカトロニクス 2014 講演論文集*, pp. 3A1-V05, 2014.
- [6] M. Kouchi et al., "An Analysis of Hand Measurements for Obtaining Representative Japanese Hand Models," *SAE 2005 Transactions Journal of Passenger Cars - Electronic and Electrical Systems*, 2005-01-2734, 2005.
- [7] 鎌倉矩子, "手のかたち 手のうごき", 医歯薬出版株式会社, 1989.

より推定された最適解を Final に示す。

IP00 は、Sep2 における初期値の決定が、低次元化空間内における原点とされている。そのため、Step3 の最適化では、生成元の姿勢である DB00 を表す低次元化空間内の座標に向けて収束していると分かる。

IP03 は、IP00 に比べて、生成元の点群データである DB03 の座標に向けて収束していない。ここで、低次元化空間内での DB03 の座標位置を見てみると、 $\pm 2\sigma$ 楕円の第一主軸沿いに存在している。そのため、初期点から最適解に陥ると予想される座標が DB00 よりも離れており、最適化の途中で局所解に陥ってしまったと考えられる。

IP05, IP08 に関しては、Step2 における初期点決定の時点で、生成元である DB05, DB08 の座標位置に解が求まっている。そのため、最適解が初期点近くに存在すると推測される。このような理由から、Step3 における最適化の結果は、その初期点の周辺の座標位置に存在する最適解に陥ったと考えられる。