

見えの変形を学習させた分類木に基づく高精度実時間三次元手指姿勢推定

○宮本翔 藤本光一¹ 松尾直志 島田伸敬 白井良明 (立命館大学)

Real-Time and Precise 3-D Hand Posture Estimation Based on Decision Tree Trained with Variations of Appearances

*S. Miyamoto, K. Fujimoto¹, T. Matsuo, N. Shimada and Y. Shirai (Ritsumeikan University)

Abstract— We propose a method for estimating 3-D hand postures from 2-D images in real-time. The estimation is based on finding the best matched posture from typical postures whose appearances are learned in advance. For high accuracy, conventional methods require high computational cost for comparing an input with many typical postures. In our method, a tree is automatically generated and trained with typical postures and their variations. Efficient search with the tree brings about real-time estimation. We show the effectiveness of our method by some experimental results.

Key Words: Hand posture estimation, Image recognition, Real-time, 3D-posture, Hierarchical

1 はじめに

ヒューマンインタフェースの一つとしてジェスチャ認識が注目されている。道具や物体を操作の様子や手話のような複雑な手の形と動きを推定することは、単純な見えの照合では非常に困難である。手指姿勢推定問題に対してこれまで提案されてきた手法を大別すると三次元モデルによるものと二次元の見えによるものの二種類に分けられる。

前者の手法は画像から得られた特徴に符合するよう手の三次元モデルの形状パラメータを変化させて手形状を見つける手法である^{1, 2, 3, 4, 5}。モデルを用いることで精度の高い姿勢推定が行うことができるが、指が他の指や掌によって隠れる自己隠蔽を含む姿勢などの推定には一般に頑健でない。これを解決するために多視点の複数カメラで得られた情報を用いる手法^{6, 7}も提案されている。

後者の手法は手の様々な既知の姿勢での見えを二次元情報として登録しておき、入力と類似度の高いものを探索し手の三次元形状を推定するアプローチである。見え自体を登録するので自己隠蔽の問題を解消できる^{9, 10}が、精度を高めるには非常に多くの姿勢を登録する必要があり、単純な全探索ではリアルタイムの推定はできない。手指姿勢は連続的にしか変化しないことを利用して探索範囲を限定し高速化する手法が提案されている^{8, 9}が、これらは自由度の制限や大型の計算機を用いて実時間を実現している。星野らは実測画像とデータグローブによって学習を行い、輪郭の高次局所自己相関関数による特徴量と Kohonen の自己組織化マップ¹⁴を使った多階層データベースを用いることで実時間推定を可能にしている¹¹。

以上の手法を踏まえて本研究では、CGによって学習を行い、時間的連続性ではなく見えで探索範囲を限定し、見えに基づいた三次元モデルを作成する手法を提案する。関節角度を粗く変化させて三次元CGモデルを作成し、輪郭形状の凹凸部分に着目した特徴量を作り、これを基にした分類木を用いる。作成した分類木をもとに、三次元CGモデルの変形を学習した分類木と新たな三次元CGモデルを作成する。新たな三次元CGモデルは見えに従って量子化間隔が狭くなるた

め、高精度な推定を行える。また、提案法で構成した分類木は各枝分かれでの判断にかかる計算量が小さいので非常に効率的に姿勢候補群を絞り込める。実験では入力画像が約86000個の手指形状クラスのどの形状であるかの識別を行った。分類木によって約400個まで候補を絞り込むことができる。見えを学習した分類木では末端ノードあたりの姿勢数が増加するが、高精度な推定が行える。Intel Core i7 2.66GHzのマシンで秒間30フレーム以上の処理が可能なシステムを構成できた。

2 処理の概要

提案システムの処理の概要は以下の通りである。

1. カメラから得た入力画像から色情報を基に手領域を抽出する。
2. 得られた手領域の輪郭点から手形状の特徴を計算し、粗い特徴による分類木を用いて姿勢の絞り込みを行う(絞り込み)。
3. 絞り込みによって得られた姿勢候補のそれぞれと入力とを今井らの手法¹⁰で比較し、最も合致するものを推定結果とする(詳細照合)。

3 姿勢推定

3.1 典型姿勢の作成

手指姿勢は連続的に変化するが、その全ての見えを学習するのは不可能なので関節角度を量子化してその組み合わせで定まる姿勢について学習を行う。この量子化された関節角度を持つ姿勢を典型姿勢と呼ぶ。

手指の自由度は非常に高いが、通常のジェスチャを行うときはいくつかの関節は互いに連動して動くので、これを用いて手のモデルを単純化する。制約条件と典型姿勢の選択は藤本らの手法¹³を用いているのでここでは概要について述べる。詳細は¹³を参照されたい。

Fig. 1中の関節角度について、以下の関係を仮定する。

$$\theta_{f1} = \theta_{f2} = \theta_{f3} = \theta_f$$

$$(f = \text{thumb, index, middle, ring, pinky})$$

$$\theta_{o,\text{index}} = \theta_{o,\text{middle}} = 0^\circ$$

$$\theta_{o,\text{ring}} = -4^\circ, \theta_{o,\text{pinky}} = -14^\circ$$

¹現所属:東芝ソリューション株式会社

独立なパラメータは以下の通りである。

$\theta_{wrist1}, \theta_{wrist2}$ 画面内回転を除いた手首回転

$\theta_{thumb}, \theta_{index}, \theta_{middle}, \theta_{ring}, \theta_{pinky}$ 指の曲げ具合

上記のパラメータについて見えの形状変化への寄与が大きいものと小さいものと異なる間隔で量子化する。 $\theta_{thumb}, \theta_{index}, \theta_{middle}, \theta_{ring}, \theta_{pinky}$ については4段階に θ_{wrist1} は12段階に、 θ_{wrist2} は7段階に量子化する。曲げについてはどの指も同じ量子化間隔で角度を取るのので、Table1では $\theta_{thumb}, \theta_{index}, \theta_{middle}, \theta_{ring}, \theta_{pinky}$ をまとめて θ_{bend} と表す。合計で86016 ($= 4^5 \times 12 \times 7$)の典型姿勢を生成する。

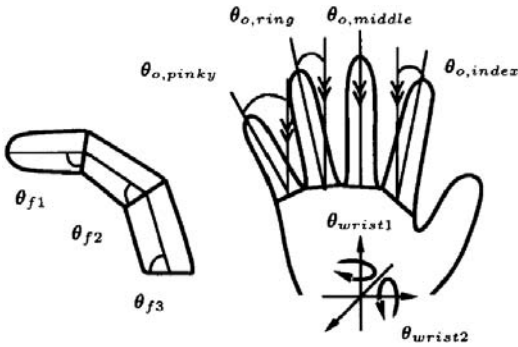


Fig. 1: Freedom of posture.

Table 1: Parameters of typical postures.

θ_{bend}	$8^\circ, 26^\circ, 48^\circ, 75^\circ$
θ_{wrist1}	$-150^\circ, -120^\circ, -90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 180^\circ$
θ_{wrist2}	$-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ$

3.2 木構造生成

3.1節で生成した典型姿勢のそれぞれについて特徴量(4節で後述)を求め、その値をもとに分類木を生成する。

まず各典型姿勢を凸数で分類(0から5個)して1段目の子ノードを生成する。得られた凸数2~5に対応する子ノードのそれぞれについて、凸間向きベクトル(4.3節で後述)が混合ガウス分布に従うと仮定して教師なしクラスタリング¹²⁾を適用し分類する。3段目以降についてもそれぞれ別の特徴量について分類して木構造を生成する。各階層におけるクラスタ数はMDL基準により決定する。

3.3 見えの変形の学習

本手法においては三次元手指姿勢は典型姿勢で量子化されており(3.1節)、典型姿勢はある範囲の手指姿勢を代表する姿勢である。しかし3.2節の分類木では典型姿勢そのものの特徴しか考慮されておらず、ある典型姿勢とそれを微小変形した姿勢が異なるグループに分類され、本来対応すべき典型姿勢が絞り込みによって除外される恐れがある。

そこで提案法では各典型姿勢についてそれを微小変形した姿勢を多数生成し、それらが分類木のどのノードに分類されるかをあらかじめ学習しておくことで絞り込みによる除外を回避する。

具体的には各典型姿勢 \hat{p}_t について以下の処理を行い学習する。まず、 \hat{p}_t を微小変形して得られる姿勢の集

合 U_t を同値関係 $p \equiv q \Leftrightarrow H(p) = H(q)$ で分類する。ここで $H(p)$ は姿勢 p を3.2節の木構造で分類した結果を表す。得られた同値類を要素数についての降順で番号付けして $A_{t,0}, A_{t,1}, \dots, A_{t,N-1}$ とする。微小変形のとり得る全ての分類先を考慮すると詳細照合における比較対象が多くなるため、ここでは稀にしか現れない分類先を省いて高速化する。全分類先に対する割合を τ とし、採用する同値類の数 N_t を下記のように定める。

$$N_t = \min \left\{ n \in [0, N-1] \mid \tau < \frac{\sum_{k=0}^n |A_{t,k}|}{\sum_{k=0}^{N-1} |A_{t,k}|} \right\} \quad (1)$$

こうして得られた同値類 $A_{t,0}, A_{t,1}, \dots, A_{t,N_t}$ のそれぞれについて属する姿勢のなかで平均的な関節角度をもつ姿勢を、その同値類の代表姿勢 $\hat{p}_{t,n}$ として記憶しておく。

3.4 木構造を用いた姿勢探索

典型姿勢全ての集合を見えを元にして階層的に分類すればFig.2のような木構造が得られる。このようなノードの親子関係が包含関係に対応した木構造が与えられれば、ある入力画像に対する姿勢推定は以下のような手順で行える。

1. 背景差分し輪郭を抽出する。輪郭点の座標を集めたベクトル v を求める。
2. 変数 n を木の根ノードで初期化し深さ L を0とする。
3. ノード n の子ノード集合 C_n から特徴 F^L についてのマハラノビス距離が最小になるノード m を求める。

$$m = \operatorname{argmin}_{k \in C_n} (F^L(v) - \mu_k)^T \Sigma_k^{-1} (F^L(v) - \mu_k) \quad (2)$$

4. m が末端でなければ、 $n = m$ とし、深さ L をひとつ増やして3に戻る。
5. m が末端であれば m に属する典型姿勢 \hat{p}_t について今井らの手法¹⁰⁾による距離 $D(v, t)$ を最小化する典型姿勢 $\hat{p}_{\tilde{t}}$ を求める。

$$T(m) = \{t \mid \exists n \leq N_t, \exists p \in A_{t,n} \text{ s.t. } m = H(p)\} \\ \tilde{t} = \operatorname{argmin}_{t \in T(m)} D(v, t) \quad (3)$$

6. $A_{\tilde{t},n}$ の平均的な代表姿勢 $\hat{p}_{t,n}$ を推定結果とする。

ここで特徴 F^L は木構造の各深さ L について定義され、各ノード k について平均 μ_k と共分散行列 Σ_k が学習されているものとしている。

$$\mu_k = \mathbb{E}_{v \in k} [F^L(v)] \quad (4)$$

$$\Sigma_k = \mathbb{E}_{v \in k} \left[(F^L(v) - \mu_k) (F^L(v) - \mu_k)^T \right] \quad (5)$$

木構造を用いて探索範囲を絞り込んでいるため、従来手法で探索する姿勢数は末端ノードに所属する姿勢の数で抑えられる。

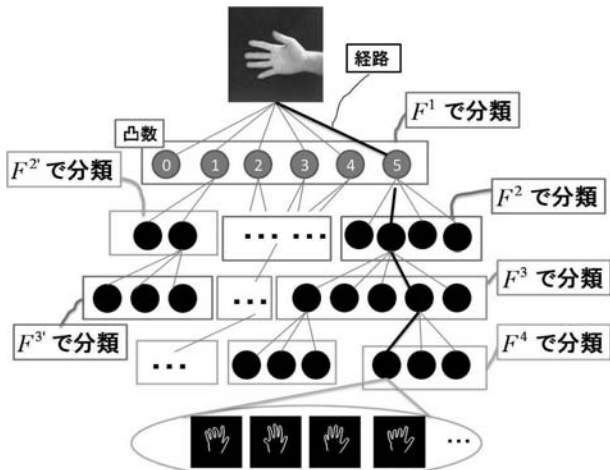


Fig. 2: Narrowing by the tree structure.

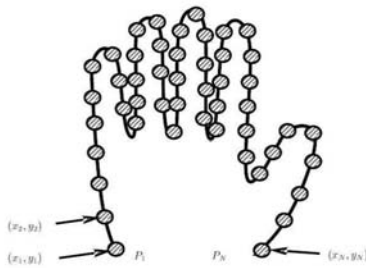


Fig. 3: Contour feature.

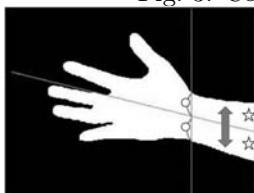


Fig. 4: Detection of the arm axis.

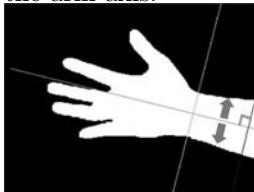


Fig. 5: Scanning the width of arm.

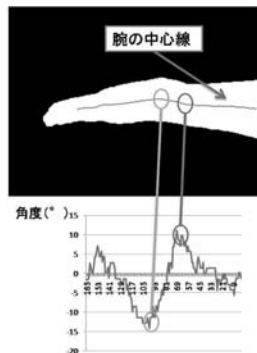


Fig. 6: Detection of the wrist.

効率が悪い. ここでは粗い特徴でありながら姿勢の大きな見えをよく表す輪郭の凹凸情報を用いて姿勢を分類する. 凸は指先によって輪郭が大きく変化する部分で, 指先にあたる凸を基準に姿勢を分類する.

4.1 輪郭特徴の抽出

CG モデルの輪郭形状から凸を検出する. まず $N(=128)$ 個の点を輪郭上に等間隔にサンプルする (Fig. 3). サンプル点の座標 $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ を重心 (g_x, g_y) と手領域の面積 A で正規化した特徴ベクトル f (式 (6), (7)) を形状特徴として用いる. ここでは面積によって正規化することで, 輪郭特徴が手の画面上の大きさに依存しないようにしている.

$$r_n = \frac{1}{\sqrt{A}} \begin{bmatrix} x_n - g_x \\ y_n - g_y \end{bmatrix} \quad (6)$$

$$f = [r_1^T, \dots, r_n^T, \dots, r_N^T]^T \quad (7)$$

輪郭サンプル点の始点, 終点 $(x_1, y_1), (x_N, y_N)$ は手首位置を基準にとる. 手首は腕からの手になるときの太さと曲率変化で見つける.

1. 入力画像の縁から腕を見つけ腕の軸を検出する (Fig. 4). 腕の軸から垂直に腕の太さを計算し急に太くなったところを手首の候補とする (Fig. 5).
2. 腕の太さから腕の中心を出し, その曲率が変化した所 (Fig. 6) を腕の候補とする.

これに過去の手首位置も考慮し手首を検出する. 詳細は割愛する. 手首位置の始点を (x_1, y_1) , 終点 (x_N, y_N) する.

4.2 凸の検出

凸は突起度合い $\rho_f(p)$ を基にし検出する. 凸らしさを Fig. 7 のように形状が尖っている度合いで評価する. そこで輪郭のサンプル p 番目を中心として $2I + 1$ 個のサンプル点を含む区間を用い突起度を評価し, 凸を検出する. 突起度合いを (1) 区間の端点距離 $\|r_{p-I} - r_{p+I}\|$ と (2) 区間の中間輪郭サンプル点 r_p と区間端点の中点 $\frac{1}{2}(r_{p-I} + r_{p+I})$ との距離の比 (式 (8)) で定義する.

$$\rho_f(p) = \frac{d \cdot \|r_p - \frac{1}{2}(r_{p-I} + r_{p+I})\|}{\|r_{p-I} - r_{p+I}\|} \quad (8)$$

$$d = \begin{cases} 1, & ((r_{p+I} - r_{p-I}) \times (r_p - r_{p-I})) \geq 0 \\ -1, & ((r_{p+I} - r_{p-I}) \times (r_p - r_{p-I})) < 0 \end{cases} \quad (9)$$

$a \times b$ は 2 次元ベクトル a, b 間の外積である.

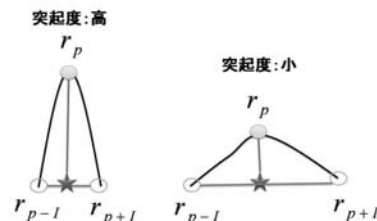


Fig. 7: Degree of convex.

p の付近が指先のような突起であると $\rho_f(p)$ の値が大きくなるので $\rho_f(p)$ が p に関して極大かつ閾値 θ 以

M を典型姿勢数, H は各末端ノードの平均姿勢数とすると, 一つの入力を識別するのに必要な計算量は全数探索では $O(C_1 M)$, 提案手法では $O(C_1 H + C_2 \log(M/H))$ となり計算量を典型姿勢数の対数オーダーに減らすことができる. ここで C_1 は今井らの手法における一つの姿勢との照合にかかる計算量, C_2 は木構造の各ノードとの照合にかかる計算量である.

4 姿勢分類に用いる特徴

姿勢の分類において, 入力の細かいノイズにも左右されにくい粗い特徴を用いれば多数の子ノード数が生成されることが避けられる. しかしほとんど全ての姿勢に共通する特徴では探索範囲が絞り込めず分類する

上となるような点 p を集め, $\rho_f(p)$ が大きいものから最大5つ採用して凸とする(式(10)).

$$F^1(f) = \min \{|S_p(\theta)|, 5\} \quad (10)$$

$$S_p(\theta) = \left\{ p \mid \theta \leq \rho_f(p) = \max_{|k-p| \leq I} \rho_f(k) \right\} \quad (11)$$

ここで, $|S_p(\theta)|$ は $\rho_f(p)$ の極大点の集合 $S_p(\theta)$ の要素数を表すものとしている. 輪郭上に等間隔にサンプルされた点を用いて凸を検出するためには姿勢ごとの平均サンプル点間隔を同じにしないと姿勢ごとに同じ感度の凸を検出することが難しい.

あるモデル m の見えの輪郭点系列を f_m とする. このモデルの輪郭長さ $l(f_m)$ (式(12))と全てのモデルに渡る平均輪郭長さ $E[l(f_m)]$ から, この形状特徴 f_m で使用する区間のサンプリング点数 I を決定する(式(13)).

$$l(f_m) = \sum_{n=2}^N \|r_n - r_{n-1}\| \quad (12)$$

$$I = \frac{E[l(f_m)]}{l(f_m)} \cdot I_v \quad (13)$$

I_v は輪郭が平均的な長さの場合に用いるサンプル点数である. I_v が小さいとノイズに左右されやすく不必要な小突起を検出する恐れがあり, I_v が大きいと間隔の狭い2つの突起が1つの突起と見なされる恐れがある. 今回の実験では I_v を上記のような悪影響が小さくなった7とした.

4.3 凸間特徴 (凸数が2以上の時)

凸間向き ($L=2$ の時)(Fig. 8), 凸間での最凹距離 ($L=3$ の時)(Fig. 9), 凸間の距離 ($L=4$ の時)(Fig. 10)の特徴を使用して木構造を作る. 凸間の情報なので凸が2つ以上の時に扱う. 凸で分類された姿勢の輪郭特徴は面積で正規化しているため輪郭特徴とオリジナル輪郭のスケールが変わる. 距離の特徴を分類に使用するにはスケールを揃える必要がある. そこで凸数と凸間向きによって似た姿勢をみつめた後, 距離に基づく特徴を使用する. そして凸間の凹み, 距離を用いて姿勢を分類する.

凸のサンプル点番号を昇順にソートしたものを p_1, p_2, \dots, p_K とし, 対応するベクトルをそれぞれ P_1, P_2, \dots, P_K ($2 \leq K \leq 5$) とする. P_n の番号は p_n であるので $P_n = r_{p_n}$ となる.

1. 凸間向き

凸間向きのベクトル $F^2(f)$ を式(15)で定義する. e_1 ($\|e_1\|=1$) を手首を横切る方向の単位ベクトル, e_2 ($\|e_2\|=1$) を e_1 と直交し手首から手のひらへ向かう方向の単位ベクトル, R は e_1 を基準とした座標系に写す回転行列(14)とする.

$$R = \begin{bmatrix} e_1^T \\ e_2^T \end{bmatrix} \quad (14)$$

$$F^2(f) = \begin{bmatrix} \frac{1}{\|P_2 - P_1\|} R(P_2 - P_1) \\ \frac{1}{\|P_3 - P_2\|} R(P_3 - P_2) \\ \vdots \\ \frac{1}{\|P_k - P_{k-1}\|} R(P_k - P_{k-1}) \end{bmatrix} \quad (15)$$

2. 凸間の最凹距離

2つの凸 (P_k と P_{k+1}) 間の部分輪郭(サンプル点番号 p_k から p_{k+1}) について考える. 凸間を結んだ直線と, 部分輪郭上の任意の点 r との距離 $l_k(r)$ は式(16)で表せる. 部分輪郭上での l_k の最大値を \tilde{l}_k とし, これを全ての凸について集めたベクトルを $F^3(f)$ とする(式(18)).

$$l_k^2(r) = \|r - P_k\|^2 - \frac{(r - P_k) \cdot (P_{k+1} - P_k)}{\|P_{k+1} - P_k\|^2} \quad (16)$$

$$\tilde{l}_k^2 = \max_{p_k \leq m \leq p_{k+1}} l_k^2(r_m) \quad (17)$$

$$F^3(f) = [\tilde{l}_1, \dots, \tilde{l}_K]^T \quad (18)$$

3. 凸間の距離

$F^4(f)$ は凸間の距離を用いて式(19)で定義される.

$$F^4(f) = [\|P_1 - P_2\|, \dots, \|P_K - P_{K-1}\|]^T \quad (19)$$

4.4 凸位置特徴 (凸数が1の時)

凸が1つの場合 ($F^1 = 1$) は, 手首から凸までの距離 ($L=2$)(Fig. 11)と手首基準の凸の角度 ($L=3$)(Fig. 12)を使用し3段階の木構造を作る. 手首基準の凸の位置で姿勢を分類する.

1. 手首から凸までの距離

手首の端点中点 w から凸 P_1 までの距離を $F'^2(f)$ とする(式(20)).

$$F'^2(f) = \|P_1 - w\| \quad (20)$$

2. 手首に対する凸の向き

手首に対する凸の向きを表すベクトル(式(21))で $F'^3(f)$ を定義する.

$$F'^3(f) = \frac{1}{\|w - P_1\|} R(w - P_1) \quad (21)$$

4.5 凸数検出あいまいさの解決

凸数で姿勢を分類するので似た姿勢であっても, 凸の数を間違えると正しく姿勢を分類することができない. $\rho_f(p)$ が閾値に近い値になるような輪郭の場合, 凸数 $F^1(f)$ の値は不安定になり得るためである. ここでは以下を考慮することであいまいさの悪影響を緩和する.

- $\rho_f(p)$ が閾値より小さい時, 通常凸でないが僅差であれば凸である場合を考慮する.
- $\rho_f(p)$ が閾値より大きい時, 通常凸であるが僅差であれば凸でない場合を考慮する.

今回の手法では木構造の各ノードは閾値を $\theta_0 = 0.5$ としあらかじめ作成しておく(認識時は閾値 $\theta_0 = 0.5$ を用いて凸を検出する). 許容範囲を $S = 0.1$ とし以下の条件の閾値変化で凸数が変化するとき, 一つの典型姿勢に対して複数の F^1, \dots, F^4 を作る.

式(10), 式(11)において θ の代わりに $\theta_+ = \theta_0 + S$ 及び $\theta_- = \theta_0 - S$ を使って F^{1+}, F^{1-} をそれぞれ求め

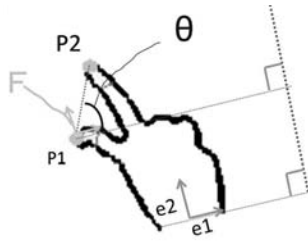


Fig. 8: Direction between convexes.

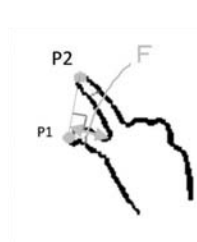


Fig. 9: Depth of concave.

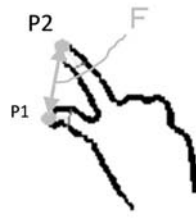


Fig. 10: Distance between convexes.



Fig. 11: Distance from wrist.

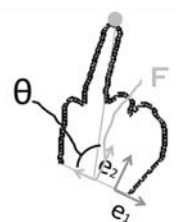


Fig. 12: Angle of convex.

る. $F^{1-} \neq F^{1+}$ のとき, $F^{1-} - F^{1+}$ 個だけ突起度合いに誤差を考慮して凸になる場合と凸にならない場合の2通り考える. 計 $2^{(F^{1-} - F^{1+})}$ 個の凸特徴を作る. 以下の手順で凸数で誤差のある姿勢の特徴を作り, その姿勢を新しくノードの所属姿勢に追加する.

1. $2^{(F^{1-} - F^{1+})}$ 個の凸の組み合わせについてそれぞれ F^1, \dots, F^4 まで計算.
2. 典型姿勢のみで学習させた階層構造に対してどのノードに属するかを計算.
3. (2) で計算されたノードに属する姿勢として典型姿勢を追加 (Fig. 14).

Fig. 13 では小指, 薬指, 人差し指についてそれぞれ凸になる場合とならない場合を考慮した $2^3 = 8$ 個のパターンを示す.

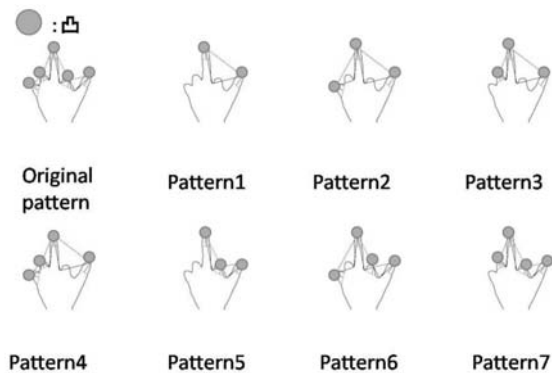


Fig. 13: Ambiguity of number of convexes.

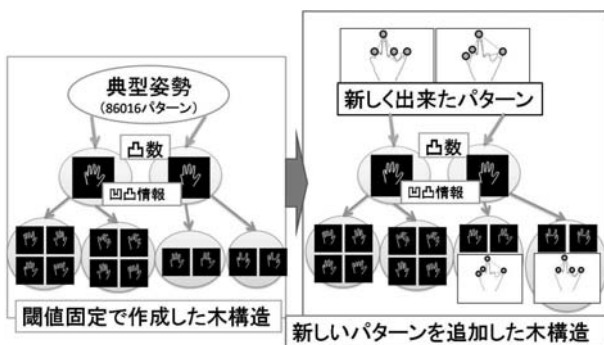


Fig. 14: Introduction of convex ambiguity.

4.6 動きの連続性を考慮した動的計画法による姿勢のフィルタリング

人間の手指姿勢変形は時間的に連続なので, ある時点で入力との合致度があまり高くないと判定された姿勢であっても, その後の形状変化を考慮すると最適であった, ということが起こり得る. そこで藤本らの手法¹³⁾を用い, 過去フレームの評価値から遷移可能な手形状の連続変化の系列を抽出し, その連続変化で現在の手形状を評価した. 詳細は¹³⁾を参照されたい. 連続性によってあらかじめ探索範囲を限定するのは素早い動きに対して正しい姿勢候補を見落とす原因となるので, ここでは木構造による分類では連続性を考慮せず, 詳細照合においてのみ探索範囲を制限している.

5 実験

5.1 木構造生成

3次元手指形状モデルをつかって生成した 86016 姿勢の CG 画像を用いて本手法に基づいて探索のための木構造を生成した. その分類能力の指標を Table 2 に示す. 第1層のノード (凸数 F^1 (0 から 5) で分類) ごとに, その下に所属する総姿勢数, 末端ノード (F^4 で分類された第4層のノード) の個数, 末端ノード中の平均所属姿勢数 (さらなる分類に必要な詳細照合¹⁰⁾ の回数に等しい) を分類能力の指標として示した. 平均所属姿勢数は微小変形を考慮せず, 典型姿勢のみを用いた場合のものである. 入力シルエット特徴を本手法に基づく木構造で分類を行うと, 詳細照合 (今井らの手法¹⁰⁾ の必要な候補を, たとえば凸数が3の姿勢の場合には平均的に 165 姿勢まで減らすことができる. これらより, ほとんどの末端ノードで 100-500 個程度 (全姿勢数の約 0.1-0.5%) にまで詳細照合の必要は姿勢候補の個数を減らせていることが分かる.

次に微小変形を考慮し, 3.3節で導入した τ のいくつかの値について平均所属姿勢数を求めた結果が Table 3 である.

5.2 姿勢推定

人の手の画像計 1170 枚に対してそれぞれ単一フレームでの推定実験を行った. 推定結果の一部を Fig. 15 に示す. 実際の人の手を入力とした場合, 正しい角度が未知であるため認識結果が「正しい」かどうかを主観的に判断したところ入力の 76.50% が正しく認識された.

学習に用いた CG モデルを用いた実験では 100% の認識精度が得られた. また客観評価のため学習に用いた姿勢とは少し異なる関節角度をもつ CG を多数生成して認識実験を行った結果が Table 4 である. ここでは, 典型姿勢からの関節角度の最大変化幅を量子化間隔の $1/4$ として実験している. $0.9 < \tau$ であれば十分な認識精度といえる.

Table 2: End nodes for each convex number.

凸数	総姿勢数	末端ノード数	平均姿勢数
0	2561	1	2561
1	18476	40	462
2	37954	75	500
3	36976	224	165
4	16707	209	81
5	3410	38	89

Table 3: Number of typical postures in end nodes.

τ	0.5	0.7	0.9	1.0
平均姿勢数	710.7	1319.5	2923.2	9336.2

Table 4: Recognition result for CG.

τ	0.5	0.7	0.9	1.0
認識率	70.3%	84.5%	96.0%	99.8%

5.3 計算時間

約1000フレーム、Core i7 2.66GHz 8コアでデータベースとの照合を並列処理させ実験した。全数(86016通り)照合した場合1フレーム平均509ms、提案手法で微小変形を学習しない($\tau = 0$)場合、平均464姿勢/frameの探索で推定解が見つかり、平均5.88msecで処理でき170fpsが実現できた。微小変形を学習した場合の結果はTable 4である。 $\tau = 0.9$ とすれば30fps超で高い精度の認識が行える。動きの連続性を考慮して推定解を減らすとさらに高速に姿勢を推定することができる。

6 おわりに

本研究では凸数と凸間の特徴空間に階層構造を導入した。階層構造内の分類は計算量が小さくなるよう設計されているため大型計算機を使用せず手姿勢推定を実時間に行えるようになった。約86000姿勢の手形状を類似見えて約800クラスに分類し効率化することで、30fps超で高精度な推定を行える。

提案法はCGについては非常に良好な認識精度が得られているものの、人間の手指を撮影した実画像に関しては微妙に異なる推定結果も見られる。これは関節角度の量子化間隔の粗さによるものと思われる。適切な関節角度の選択が今後の課題である。

参考文献

- 1) 亀田, 美濃, 池田: シルエット画像からの関節物体の姿勢推定法, 電子情報通信学会論文誌, vol.J79-D-II, no. 1, pp. 26-35, Jan.1996.
- 2) S. Uk Lee, I. Cohen: 3D Hand Reconstruction from a Monocular View, Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04), Volume 3 - Volume 03 2004, pp. 310 - 313, Cambridge, Aug.2004
- 3) J. M. Rehg, T. Kanade: Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking, In Third European Conference on Computer Vision (ECCV'94), pp. 35-46, Springer-Verlag, May.1994.

Table 5: Time for recognition.

τ	0.5	0.7	0.9	1.0
平均処理時間 [msec]	12.6	16.9	27.8	76.7

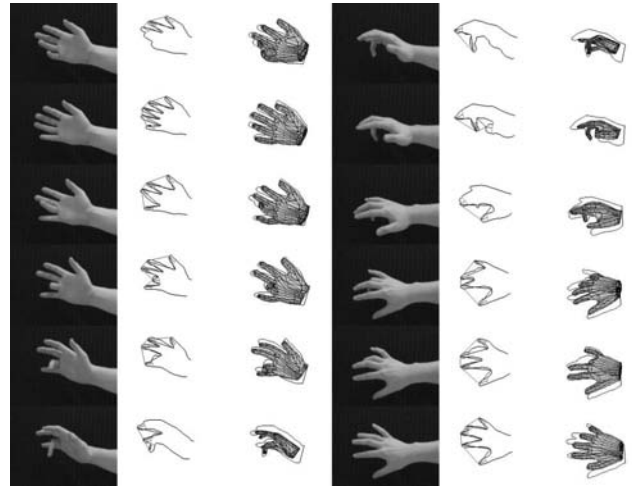


Fig. 15: Recognition result.

- 4) D. G. Lowe: Fitting Parameterized Three-Dimensional Models to Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, no. 5, pp. 441 - 450, May.1991
- 5) 藤木, 陳, 吉本, 有田, 谷口: 手の制約知識を用いた逆運動学解法による実時間三次元手形状推定, 画像の認識・理解シンポジウム MIRU2005, pp. 1380-1387, 兵庫, Jul.2005
- 6) W. CHEN, R. FUJIKI, D. ARITA and R. TANIGUCHI: Real-time 3D Hand Shape Estimation using Multiple Cameras, The 13th Japan-Korea Joint Workshop on Frontiers of Computer Vision, Busan, pp. 15-20, Korea, Jan.2007,
- 7) E. Ueda, Y. Matsumoto, M. Imai and T. Ogasawara: Hand Pose Estimation for Vision-based Human Interface, IEEE Transactions on Industrial Electronics, Vol. 50, No. 4, pp. 676-684, Aug.2003
- 8) 島田, 白井, 久野: 確率に基づく探索と照合を用いた画像からの手指の三次元姿勢推定, 電子情報通信学会論文誌 D-II J79-D-II/7, pp. 1210-1217, Jul.1996
- 9) B. Stenger, A. Thayananthan, P. H. S. Torr, R. Cipolla: Model-Based Hand Tracking Using a Hierarchical Bayesian Filter, IEEE T. PAMI, Vol. 28, No. 9, pp. 1372-1384, Sep.2006.
- 10) 今井, 島田, 白井: 輪郭の変形の学習による3-D手指姿勢の認識, 電子情報通信学会論文誌 D-II, Vol. J88, No. 8 pp. 1643-1651, Aug.2005
- 11) K.Hoshino and M.Tomida: "3D hand pose estimation using a single camera for unspecified users," Journal of Robotics and Mechatronics, 21, 6, pp.749-757, 2009
- 12) Charles A. Bouman: CLUSTER: An Unsupervised Algorithm for Modeling Gaussian Mixtures, <http://www.ece.purdue.edu/bouman>
- 13) 藤本, 島田, 白井: 見えの学習と動きの連続性を考慮した並列探索による高速3次元手指姿勢推定システム, 電子情報通信学会2008年総合大会 情報・システムソサイエティ誌 2008年総合大会特別号 学生ポスターセッション, pp.123, Mar.2008
- 14) Kohonen, T., The neural phonetic typewriter, IEEE computer, 21, 3 (1988), 11-22