

輪郭部分特徴の階層構造学習による三次元手指姿勢推定の高速化

藤本 光一[†] 松尾 直志[†] 島田 伸敬[†] 白井 良明[†]

[†] 立命館大学大学院理工学研究科 〒525-8577 滋賀県草津市野路東 1-1-1

E-mail: †{fujimoto,matsuo}@i.ci.ritsumeai.ac.jp, ††{shimada,shirai}@ci.ritsumeai.ac.jp

あらまし 本論文では実時間での二次元の見えに基づく三次元手指姿勢の推定方法について述べる．二次元画像からの姿勢を推定ではあらかじめ定めた典型姿勢について二次元の見えを学習しておき，入力に一致する見えを探索することで三次元姿勢を求める．十分な精度を得るには典型姿勢を多数学習しなければならないため従来法では探索に要する計算量が大きく，実時間処理には大型の計算機が必要であった．我々は輪郭形状の凸情報を基にした粗い情報に着目した見えの階層構造を用いて手指の見えを効率的に探索する手法を提案する．見えの階層構造を凸特徴で典型的な見えの集合をクラスタリングすることにより自動的に生成する．階層構造は詳細な照合をすべき見えの数ができるだけ少なくなるように設計されているため手姿勢推定を大型な計算機を使用しなくても実時間で行える．システムを実装して本手法の有効性を示す．

キーワード 画像認識， 手指姿勢推定， 実時間， 三次元形状， 階層構造， クラスタリング

High Speed 3-D Hand Posture Measurement by Tree-based Learning Contour Features

Koichi FUJIMOTO[†], Tadashi MATSUO[†], Nobutaka SHIMADA[†], and Yoshiaki SHIRAI[†]

[†] School of Information Science and Engineering, Ritsumeikan University Nojihigashi 1-1-1, Kusatsushi, Shiga, 525-8577 Japan

E-mail: †{fujimoto,matsuo}@i.ci.ritsumeai.ac.jp, ††{shimada,shirai}@ci.ritsumeai.ac.jp

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 matched posture from typical postures whose appearances are learned in advance. To achieve high accuracy, conventional methods requires the high computational cost for comparing an input with many typical postures. In our method, we perform the efficient hierarchical search by using the tree generated from convex features of typical appearances. The tree is automatically generated by clustering typical postures with convex features, where the features are selected in order for the hierarchical search to effectively narrow the set of candidate postures. By using our method, one can estimate postures by an ordinary PC with sufficient accuracy in real-time. We show the effectiveness of our method by some experimental results.

Key words Image recognition, Hand posture estimation, Real-time, 3D-posture, Hierarchical, Clustering

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．バーチャルリアリティ環境において特殊な手袋等を用いずに手指の状態を入力できるよう，カメラ画像だけから手指の三次元形状を認識することが求められている．またコンピュータの操作や手話による対話などにも直観的なジェスチャインタフェースが有効な場合がある．コンピュータの進歩により，簡単なジェスチャに限れば実時間で認識できるようになったが，道具や物体を操作する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，背景と手指領域の峻別や，手指と

カメラ位置の関係による見え方の変化，ジェスチャ自体に揺らぎがあることなどから，単純な見えの照合では非常に困難である．この問題に対処するために，これまで大きく分けると 3D-model-based と 2D-appearance-based の二種類のアプローチが提案されてきた．

前者の手法は画像からの局所的な特徴を抽出し，その特徴に合うような手の三次元モデルの形状パラメタを変形させ特徴が合致する手形状を見つけるアプローチである．実画像とモデルをシルエットでマッチングしてパラメタを推定する手法 [1] [2] [3] [4] や手の可動範囲や依存関係の制約知識を利用した逆運動学解法を用いて手領域内エッジ情報を利用した手法 [5] が提案されている．これ

らの手法は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、一般的な性質として指が他の指や掌によって隠れる姿勢などの推定にはロバスト性に欠ける。それらを解決するために多視点の複数カメラで得られた画像からセルフオクルージョンの少なく手形状の情報量が最も多いカメラ画像を使用し認識する手法 [6] が提案されている。また二次元画像ではなく三次元空間中でマッチングを行う手法として多視点のカメラを用いて得られる手シルエット画像から視体積交差法により生成された手の三次元形状と三次元手指モデルが一致するようにパラメタ推定を行う手法 [7] が提案されている。

後者の手法は二次元における手の様々な姿勢での見えを登録しておき、入力と類似度の高いものを選択し手の三次元形状を推定するアプローチである。このアプローチには様々な手の姿勢を事前にデータベースに登録しておくためセルフオクルージョンの問題に対処することができる [9] [10] [11]。この手法では手の内部自由度が高いため、探索する対象姿勢の数が非常に多くなり、推定精度と速度の要求仕様が相反する。

人間の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「見え」のマッチングを高速化する手法が提案されている。Stenger ら [9] はあらかじめ学習しておいた頻度分布を基に姿勢空間を階層構造に分割しておきマッチング時にはベイズ推定の枠組みで姿勢空間を探索した。また島田ら [8] は探索する姿勢を、直前時刻の画像に対して推定された姿勢候補の近傍に限定して計算量を削減した。谷本ら [10] は自己組織化マップを用いて関節角度の近い姿勢グループのクラスタリングを行い、階層構造を導入した上で関節角度の連続性から探索範囲を限定した。しかしこれら動きの連続性を考慮した場合でも自由度の多い手指の動きを追跡するには大量の候補を残さなければならず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。また手の動作が遅いときには時間連続性を用いるとより安定な認識結果が得られるが逆に手の動作が速いときにはかえって正しい姿勢が認識できなくなることや、一度手の表裏の認識を誤ると誤った判断が長時間解消されないことがある。

本論文では見えで探索範囲を限定する手法を提案する。3DCG モデルから生成した輪郭形状の突起数と凹凸部分に着目して階層的に特徴のクラスタリングを繰り返すことにより探索木を構成した。凹凸部分特徴は輪郭の凸部分間向き、各凹部分の最凹距離、凸部分間の距離を使用する。提案法で構成した探索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クラスを絞り込める。見えで探索範囲を限定するものとして富田ら [12] は手の主要領域を含む矩形を求め、その縦横比や手との位置関係を元に手指姿勢候補を限定する手法を提案した。ただしこの手法は手の形が正面を向いている場合は効率的に探索範囲を限定することができるが、セルフオクルージョンが起こるような姿勢の時は指の屈

伸度合いによる形状比率変化をとらえることが難しく、大きく候補姿勢を減らすことは難しいと考えられる。また使う特徴として、輪郭に加えて爪の位置 [13]、手領域の内部エッジ [5] に着目したものが存在するが照明条件や低解像度の時に弱く輪郭だけの方が抽出するのが容易なので我々の研究では輪郭情報だけで手姿勢を認識することを目的とする。実験では入力画像が約 86000 個の手指形状クラスなどの形状であるかの識別を行った。探索木によって約 400 個まで候補を絞り込むことができる。Intel Core i7 2.66GHz のマシンで秒間 150 フレーム以上の処理が可能なシステムを構成できた。

2. 姿勢推定

二次元の見えを元にした姿勢推定ではあらかじめ様々な姿勢について画像を学習しておき、入力画像に類似の学習データを探索することで姿勢推定を行う。しかし手指姿勢は自由度が高いため十分な精度で関節角度を得るためには非常に多くの姿勢について探索を行わなければならない。本論文では木構造を用いて探索範囲の効率的絞り込む手法を提案する。

2.1 典型姿勢の作成

手の自由度は大きいですが、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。そのため手のモデルを関節の連動の制約を用いて単純化する。姿勢の自由度と関節角度を図 1、表 1 に示す。

各指には三つの関節があり、一つの指は全ての関節が連動して曲がる。ここで $\theta_{f1} = \theta_{f2} = \theta_{f3} = \theta_f$ と仮定する。 $(f = \text{thumb}, \text{index}, \text{middle}, \text{ring}, \text{pinky})$

だから各指のパラメータを $\theta_{\text{thumb}}, \theta_{\text{index}}, \theta_{\text{middle}}, \theta_{\text{ring}}, \theta_{\text{pinky}}$ で定義する。

指には曲げと垂直な方向の指の軸の周りに回転するもう一つの自由度がある。この指の開きの動きを「open」と定義する。親指以外の指の open の関節角度を $\theta_{o,\text{index}}, \theta_{o,\text{middle}}, \theta_{o,\text{ring}}, \theta_{o,\text{pinky}}$ とする。

人は自然な動作を行うときには open に関して各指を連動させて動かす。したがって独立変数 θ_{open} を用いて視線な動作の際には open の各関節角度が連動することが多いので、 $\theta_{o,\text{index}} = \theta_{\text{open}}, \theta_{o,\text{middle}} = 0^\circ, \theta_{o,\text{ring}} = -\theta_{\text{open}} - 4^\circ, \theta_{o,\text{pinky}} = -2\theta_{\text{open}} - 14^\circ$ と仮定する。 $\theta_{\text{open}} = 0^\circ$ のとき自然に open が開いた状態になる。上述の制約によってモデルの姿勢は 8 自由度 ($\theta_{\text{thumb}}, \theta_{\text{index}}, \theta_{\text{middle}}, \theta_{\text{ring}}, \theta_{\text{pinky}}, \theta_{\text{open}}, \theta_{\text{wrist1}}, \theta_{\text{wrist2}}$) で表される。

典型姿勢 (モデル形状) は量子化された姿勢パラメータの組み合わせで定義される。ジェスチャをするときには open の自由度よりも曲げの自由度の方が重要となるので典型姿勢を定義するときには $\theta_{\text{open}} = 0^\circ$ に固定しておく。残り 7 個のパラメータについて手指の形状の変化が大きい突起と小さいときで異なる間隔で量子化する。 $\theta_{\text{thumb}}, \theta_{\text{index}}, \theta_{\text{middle}}, \theta_{\text{ring}}, \theta_{\text{pinky}}$ 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
 で, 表 1 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.

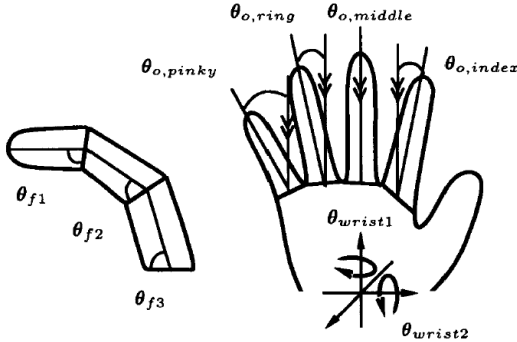


図 1 姿勢の自由度

表 1 典型姿勢のパラメータ

θ_{bend}	8	26	48	75			
θ_{open}	0						
θ_{wrist1}	-150	-120	-90	-60	-30		
	0	30	60	90	120	180	
θ_{wrist2}	-90	-60	-30	0	30	60	90

2.2 木構造生成

1 段階目は凸数 (0 から 5 個) に対応した子を持つ, 2 段階目以降は各特徴が類似した候補群を混合ガウス分布に基づく教師無しクラスタリング [14] を階層的に繰り返すことで探索木を構成する. 各階層におけるクラスタ数は MDL 最小基準により選択する. 各段階初期クラスタを 10 にした.

2.3 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図 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.

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

(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 に属する姿勢について今井らの手法 [11] を用い輪郭の細かな変化を評価する. 入力 v に最も合致する姿勢を探索し, 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.

$$\mu_k = \mathbb{E}_{v \in k} [F^L(v)] \quad (2)$$

$$\Sigma_k = \mathbb{E}_{v \in k} [(F^L(v) - \mu_k)(F^L(v) - \mu_k)^T] \quad (3)$$

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.

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式 (4), 提案手法では式 (5) になり計算量を対数的に減らすことができる. $C1$ を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, $C2$ を木構造の各ノードとの照合にかかる計算量, M を典型姿勢数, H は各末端ノードの平均姿勢数とする.

$$O(C1 * M) \quad (4)$$

$$O(C1 * H + C2 * \log(M/H)) \quad (5)$$

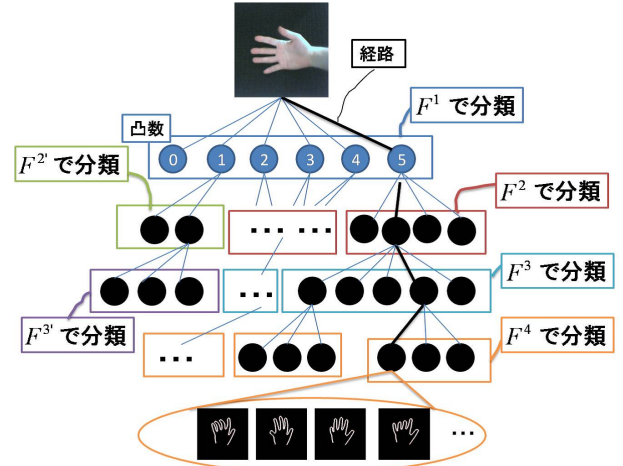


図 2 木構造を用いた探索範囲の絞り込み

3. 姿勢分類に用いる特徴

姿勢の分類において, 多くの姿勢に共通する粗い特徴を用いれば入力画像の細かいノイズにも左右されず, 大まかに分類できるので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効率が悪い. ここでは粗い特徴でありながら姿勢の大ま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. 3.1 に輪郭特徴の抽出, 3.2 に凸

の検出, 3.3 に凸が 2 個以上の時の特徴, 3.4 に凸が 1 つの時の特徴, 3.5 に凸数検出あいまいさの解決, 3.6 に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について説明する.

3.1 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 (図 3). 手領域の面積を A と定義する. 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. 輪郭特徴を面積によって正規化し, 輪郭特徴を個人差や画像の見えによる手の大きさの違いに依存させないようにする.

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

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

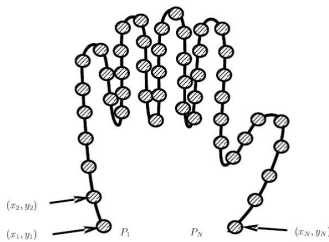


図 3 輪郭特徴

輪郭サンプル点の始点, 終点 $(x_1, y_1), (x_N, y_N)$ は手首位置になる. 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) 太さでの検出

入力画像から手首を検出するには, 画像の縁から腕を見つけ腕の軸を検出する (図 4). 腕の軸から垂直に腕が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 (図 5).

(2) 曲率変化での検出

腕の太さから腕の中心を出し, その曲率が変化した所 (図 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.

3.2 凸の検出

凸は突起度合い $\rho_f(p)$ を基にし検出する. より凸らしいというのは細長く尖っている (図 3.2) とする. そこで輪郭のサンプル I 番目の点から $N - I$ 番目の点まで $\pm I$ サンプル点の区間を用い突起度を評価し, 凸を検出

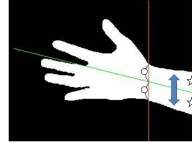


図 4 腕軸検出

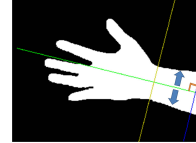


図 5 太さ変化

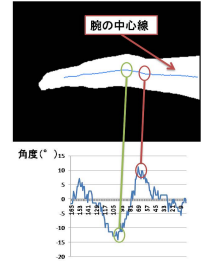


図 6 腕中心曲率変化

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.

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

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

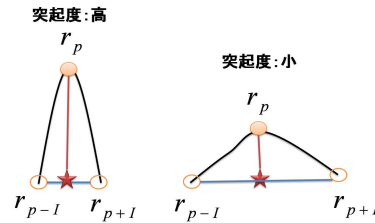


図 7 突起度合い概要

正面手輪郭画像の時, 区間 (5 サンプル) の中点と輪郭上の中点の距離を示したものを図 8, 突起度合いを図 3.2 に示す. 図 3.2 の値の大きい山 5 つと値の小さな谷が 4 つあり, これらが指の凸と凹に対応している. 指先のような突起の時に区間の中点と輪郭上の中点の距離がおおきくなる. 凸は突起度合いが多い順に閾値 θ 以上かつサンプリング間隔前後の突起度が最大 (式 (11)) なものを最大 5 つ凸とする (式 (10)).

$$F^1(f) = \begin{cases} |S_p(\theta)| & (0 \leq |S_p(\theta)| \leq 5) \\ 5 & (5 < |S_p(\theta)|) \end{cases} \quad (10)$$

$$S_p(\theta) = \{p \mid \rho_f(p) \geq \max_{k \in [p-I, p+I]} \rho_f(k) \text{ かつ } \rho_f(p) \geq \theta\} \quad (11)$$

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.

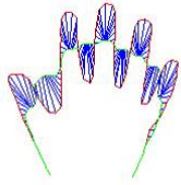


図 8 突起度合いイメージ

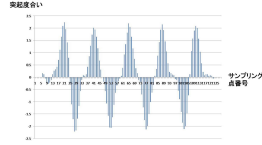


図 9 突起度合い

あるモデルの形状特徴 f_m とする．全ての典型姿勢の平均輪郭長さ ave (式 (13))，輪郭長さ l_{all} (式 (12)) よりある典型姿勢での区間で使用するサンプリング点の数 I は式 (14) で求められる．平均輪郭区間を Iv をかけた数とする． Iv が小さいと，小さい突起と大きい突起のどちらの突起が大きいかが評価するのが難しく， Iv が大きいと，2つの突起に区間がかぶり突起度の小さい方が無視される．今回の実験では Iv を上記の2点を影響が小さくなった7とした．

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

$$ave = E_f[c(f_m)] \quad (13)$$

$$I = \frac{ave}{c(f_m)} \times Iv \quad (14)$$

3.3 凸間特徴 (凸数が2以上の時)

凸間向き ($L=2$ の時) (図 10)，凸間での最凹距離 ($L=3$ の時) (図 11)，凸間の距離 ($L=4$ の時) (図 12) 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，ベクトルを小さい順にソートしたものを $p_1, p_2, \dots, p_K, P_1, P_2, \dots, P_K (2 \leq K \leq 5)$ とする． $P_n = r_{p_n}$ となる．

(1) 凸間向き

凸間向きのベクトル $F^2(f)$ は以下の式 (16) で計算される． $e_1(|e_1|=1)$ を手首を横切るベクトル， $e_2(|e_2|=1)$ を e_1 と直行し手首から手のひらへのベクトル， R は手首を横切る方向を基準とした座標系に写す回転行列式 (15) とする．

$$R = [e_1, e_2]^T \quad (15)$$

$$F^2(f) = \begin{bmatrix} \frac{1}{\|P_2 - P_1\|} R(P_2 - P_1) \\ \frac{1}{\|P_3 - P_2\|} R(P_3 - P_2) \\ \dots \\ \frac{1}{\|P_k - P_{k-1}\|} R(P_k - P_{k-1}) \end{bmatrix}^T \quad (16)$$

(2) 凸間の最凹距離

凸間のベクトル $P_k - P_{k+1}$ から凸間の輪郭ベクトル ($P_k + 1$ から $P_{k+1} - 1$) の任意の輪郭ベクトル C_v までの距離 l_k は式 (17) で表される． l_k が最大になる距離を $F^3(f)$ (式 (18)) とする．

$$l_k^2 = \max_{c \in [P_k+1, P_{k+1}-1]} \{ \|C_v - P_k\|^2 - \frac{1}{\|P_{k+1} - P_k\|^2} \{ (C_v - P_k) \cdot (P_{k+1} - P_k) \}^2 \} \quad (17)$$

$$F^3(f) = [l_1, \dots, l_k] \quad (18)$$

(3) 凸間の距離

凸間の距離 $F^4(f)$ は式 (19) で計算される．

$$F^4(f) = [\|P_1 - P_2\|, \dots, \|P_k - P_{k+1}\|] \quad (19)$$

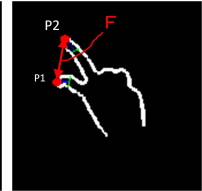
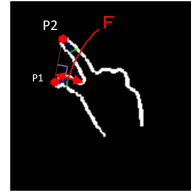
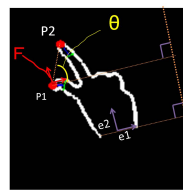


図 10 凸間向き 図 11 凸間の凹距離 図 12 凸間の距離

3.4 凸位置特徴 (凸数が1の時)

凸が1つの時，手首から凸までの距離 ($L=2$) (図 13) と手首基準の凸の角度 ($L=3$) (図 14) 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) 凸距離

手首の端点中点 w から凸のベクトル P_1 に対して凸距離を $F^2(f)$ とする (式 (20)) ．

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

(2) 凸角度

凸間向きのベクトル $F^3(f)$ は以下の式 (21) で計算される．

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

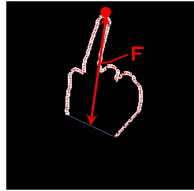


図 13 凸距離

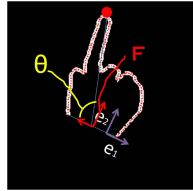


図 14 凸角度

3.5 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。ここで以下を考慮することであいまいさを解消できると考えた。

- 通常凸でないが $\rho_f(p)$ が閾値より少し小さい時、凸である場合を考慮する。
- 通常凸であるが $\rho_f(p)$ が閾値より少し大きい時、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 $\theta_0=0.5$ であらかじめ作成しておく（認識時は閾値 $\theta_0=0.5$ で凸を検出する。）誤差範囲 $S=0.1$ とし以下の条件の閾値変化で凸数に誤差が出るとき、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 θ_+ (式 (22)), θ_- (式 (23)) を使って F^{1+}, F^{1-} を求める。

$$\theta_+ = \theta_0 + S \quad (22)$$

$$\theta_- = \theta_0 - S \quad (23)$$

$F^{1-} \neq F^{1+}$ 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を作り 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 (図 16)。

図 15 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。

3.6 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

単一フレームの認識ではそのフレームの見えに一致度の高い姿勢が選択される。しかし過去の手形状変化と合致するとは限らず、過去にあまり評価の高くない姿勢群

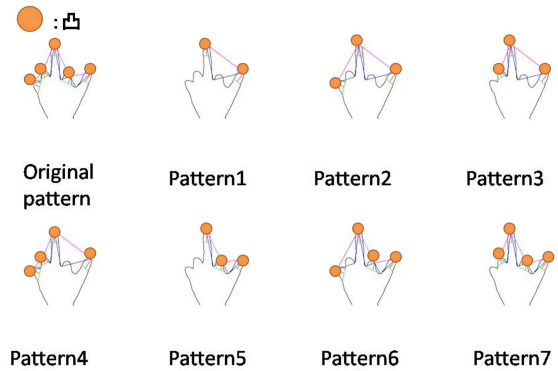


図 15 凸数誤差例

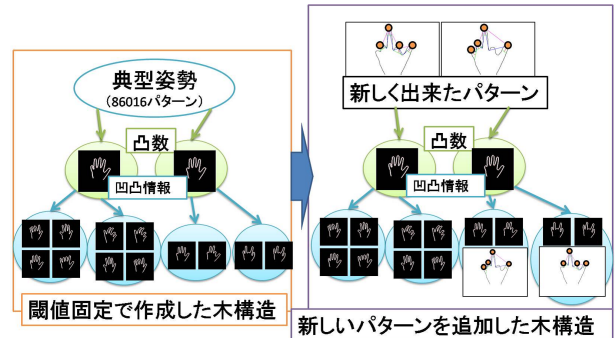


図 16 新しいパターンを追加した木構造

であっても単一フレームでは一致度が一番高くなることもある。

そこで藤本ら [15] の手法を用い、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。過去 5 フレーム間、上位 5 候補の姿勢から遷移可能な姿勢（関節角度が量子化 1 段階以内 (表 1)) を辿る。連続変化の系列ごとに照合度を足した評価値が最も高い連続変化を採用し各自由度ごとに関節角度の平均をとる。その平均の関節角度を認識結果とする。詳細は割愛する。図 17 では t フレームにおいては第 1 候補が手の裏側の姿勢と誤認識してしまうが、連続的に姿勢を迎れないので t フレームにおいて第 2 候補からの系列が採用される。



図 17 動きの連続性を考慮した姿勢の評価

4. 実験

4.1 木構造生成

86016 姿勢を分類したときの凸数ごとの姿勢数，末端ノード数と末端ノードにおける平均所属姿勢数を示したものを表 2 に示す．凸数が 1 以上の姿勢の場合，410 姿勢まで階層構造を辿ることで限定することができる．各凸数 (1,2,3,4,5) ごとの末端ノードの所属する姿勢数を図 18 から 22 に示す．

表 2 凸数ごとの木構造データ

凸数	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

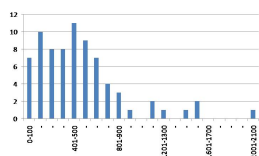


図 18 凸 1:平均 462 姿勢



図 23 認識結果例

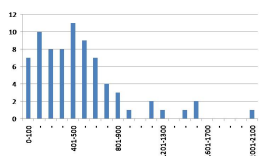


図 19 凸 2:平均 500 姿勢

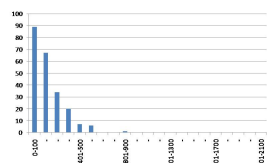


図 20 凸 3:平均 165 姿勢

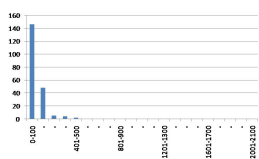


図 21 凸 4:平均 89 姿勢

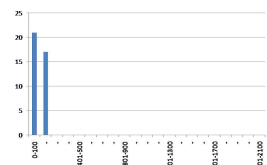


図 22 凸 5:平均 81 姿勢

4.2 姿勢推定

推定結果の例を図 23 に示す．人の手の画像，画像計 1170 枚に対して単一フレームによる評価による実験を行った結果 76.50% の入力为正しく認識された．

4.3 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，提案手法では平均 464 姿勢/frame，の探索で推定解が見つかり，平均 5.88ms (図 24) で処理でき 170fps が実現できた．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．

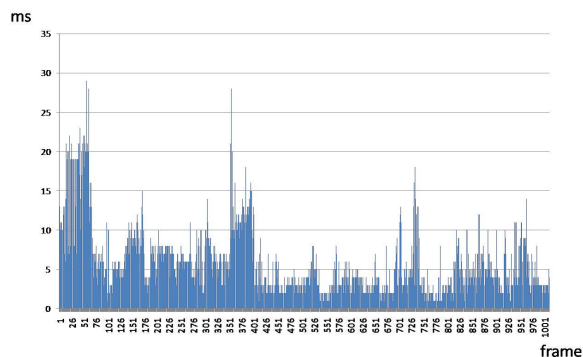


図 24 認識時間

5. おわりに

本研究では凸数と凸間の特徴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、認識の効率化させ 170fps を実現した。

文 献

- [1] 亀田 能成, 美濃 導彦, 池田 克夫 “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] Sung Uk Lee, Isaac 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] James M. Rehg and Takeo 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.
- [4] David 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] 藤木隆司, 陳維英, 吉本廣雅, 有田大作, 谷口倫一郎: “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] Weiyang CHEN, Ryuji FUJIKI, Daisaku ARITA and Rin-ichiro 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] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai and Tsukasa Ogasawara. , “Hand Pose Estimation for Vision-based Human Interface”, IEEE Transactions on Industrial Electronics. Vol. 50, No. 4, pp. 676–684, Aug.2003
- [8] 島田伸敬, 白井良明, 久野義徳: “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] 谷本 貴頌, 星野 聖, 廣池 敦, “自己増殖型 SOM を用いた画像データベースからの手指形状の実時間推定”, 情報処理学会論文誌 コンピュータビジョンとイメージメディア, Vol. 47, No. SIG 5, pp. 100–107, Mar.2006
- [11] 今井章博, 島田伸敬, 白井良明, “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
- [12] 富田元将, 谷本貴頌, 小渡悟, 星野聖, “情報機器インターフェースへの応用を目的とした単眼カメラによる手指動作推定”, 電子情報通信学会技術研究報告. EID 電子ディスプレイ, 108(235), pp. 1–4, Oct.2008
- [13] 玉城絵美, 味八木崇, 曆本純一, “インタラクティブシステムのための高精度な 3 次元ハンドジェスチャ認識手法”, 情報処理学会論文誌, Vol. 5, No. 2, pp. 229–239, Feb.2010.
- [14] Charles A. Bouman, “CLUSTER: An Unsupervised Algorithm for Modeling Gaussian Mixtures”,

<http://www.ece.purdue.edu/bouman>

- [15] 藤本 光一, 島田 伸敬, 白井 良明, “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, 北九州, pp.123, Mar.2008