

論 文

輪郭の変形の学習による 3-D 手指姿勢の認識

今井 章博^{†a)} 島田 伸敬^{††b)} 白井 良明^{††c)}

3-D Hand Posture Recognition by Learning Contour Variation

Akihiro IMAI^{†a)}, Nobutaka SHIMADA^{††b)}, and Yoshiaki SHIRAI^{††c)}

あらまし 本論文では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。従来の方法では、入力の見え（輪郭）と三次元手指モデルから生成された見えとの照合により姿勢を認識していた。しかし、この方法は三次元姿勢や視点の変化に伴って見えが変化する場合には認識精度が低下した。この問題は起こり得る見えをすべて登録しておけば解決できるが、手の自由度は大きいためにそれは難しい。本研究では、三次元手指モデルから生成された CG によって典型的な見えに対して起こり得る見えの変形を作成し、特徴空間上で見えの分布を求めておく。ただし、見えが複数のクラスを形成している場合は、各クラスごとに分布を求める。輪郭上のサンプル点の輪郭線に沿ったずれに対する制約部分空間を作成し、その空間上で入力の見えと手指モデルの見えとのマハラノビス距離を用いて類似度を評価する。実験結果によって本手法の有効性を示す。

キーワード 手指姿勢, 三次元形状モデル, 関節角度推定, 輪郭画像, 制約部分空間

1. ま え が き

1.1 従来の方法

最近ではマウスやキーボードなどのこれまでの入力デバイスに代わるものとして、ジェスチャによるヒューマンインタフェースが注目を集めている。ジェスチャを認識できれば、仮想世界や実世界での物体のマニピュレーション、指文字や手話などへの広い応用が可能である。ジェスチャ認識に対して今まで 3D model-based と 2D appearance-based の 2 種類のアプローチが提案されてきた。

前者の手法は、画像から局所的な特徴を抽出して、その特徴に対して三次元形状モデルを当てはめている [1]。この方法は、最小二乗基準に基づいた高精度な姿勢の推定を試みているが、セルフオクルージョンなどによって見えが多様に変化する手の姿勢に対してはロバスト性に欠ける。また、画像の局所的な特徴では

なく多視点画像から作成されたボクセルに対する三次元モデルの当てはめにより姿勢を推定する方法が提案されている [2]。しかしセルフオクルージョンが起こるとボクセルが正しく作成されず、認識がうまく行わない。

後者の方法では、記録されている対象物体の様々な姿勢の二次元の見えから入力の手指の見えと最も照合するものが選び出される [3]。これは特徴の抽出を行わず、直接入力の見えと記録されているモデルの見えを比較するのでセルフオクルージョンに対してロバストである。見えの画像を主成分分析 (PCA) で圧縮すれば、計算時間の短縮が可能である [4]。しかし、これらの方法は入力を少数のパターンに分類したのみで、三次元情報の抽出は行っていない。Black ら [5] はこのアプローチを二次元の位置と方向の推定まで拡張したが、三次元までは至っていない。

3D model-based と 2-D appearance-based の方法との橋渡しをする考え方として、“Estimation by Synthesis (ES)” [6], [7] がある。これは、起こり得る見えを三次元形状モデルから生成し、その見えと入力の見えとの照合を行う。しかし、入力画像が与えられるたびに複数の三次元モデルから見えを作成するので、実時間処理を行うには計算量が大きすぎる。

見えと姿勢パラメータとの間の関係を学習しておけ

[†] 大阪大学大学院工学研究科電子制御機械工学専攻, 吹田市
Dept. of Computer-Controlled Mechanical Systems, Osaka
University, 2-1 Yamadaoka, Suita-shi, 565-0871 Japan

^{††} 立命館大学情報理工学部知能情報学科, 草津市
Dept. of Human and Computer Intelligence, Ritumeikan
University, 1-1-1 Nojihigashi, Kusatsu-shi, 525-8577 Japan

a) E-mail: imai@cv.mech.eng.osaka-u.ac.jp

b) E-mail: shimada@ci.ritsumeiji.ac.jp

c) E-mail: shirai@ci.ritsumeiji.ac.jp

ば高速に姿勢を推定することができる [8]. また, あらかじめモデルの見えを登録しておくことによって高速にしたアルゴリズムが提案されている [9], [10]. その推定方法を図 1 に示す. このシステムでは各モデル (典型姿勢) に対してオフラインで三次元モデルから CG 画像を作成しておく. オンラインでは, 入力の見えと十分に照合する見えを選び出す. 認識結果はその見えに対応する典型姿勢である. この方法は, モデルの見えと入力の見えとの類似度を評価しているので, 入力とモデルとが多少異なっても認識を行うことができ, 手指形状の個人差が認識に与える影響を軽減できる.

輪郭点の座標値を手指の特徴とする方法 [9] では, 画像に比べて少ないデータ量で手指の特徴を表現できる. しかし, この方法はしばしば認識を誤る. 誤認識の例

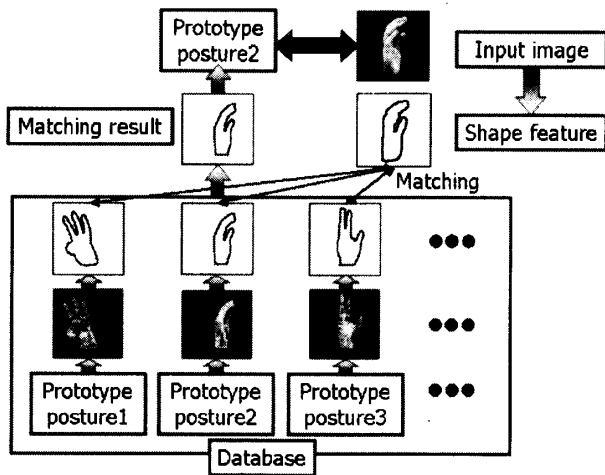


図 1 姿勢推定方法
Fig. 1 Method of posture estimation.

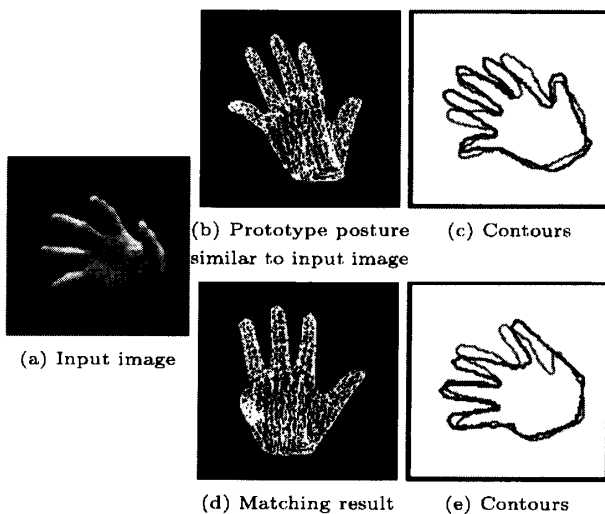


図 2 誤った認識の例
Fig. 2 Example of mismatch.

を図 2 に示す. 入力画像 (a) は典型姿勢 (b) (5 本の指が見えている) と照合すべきであるが実際は典型姿勢 (d) (4 本の指しか見えていない) と誤って照合した. (c), (e) は入力の輪郭 (薄い曲線) と典型姿勢の輪郭 (濃い曲線) を重ね合わせた図である. 二つの輪郭の類似度は輪郭点のずれの総和に基づいている. (a) と (b) は人指し指と薬指の輪郭が離れているので, 指が同じ本数であるにもかかわらず (a) と (b) の輪郭の類似度は小さくなる. 一方, (a) と (d) はもともと指の本数が異なるにもかかわらず, 輪郭のずれはあまり大きくない. 両者の差はわずかではあるが後者の方が類似度が大きくなるので (d) が選ばれた.

本研究ではこの不十分な照合基準の問題解決のために, 各典型姿勢に対して, 少し姿勢が変化したときに起こり得る輪郭の変形を学習させておく. 学習した輪郭の変形を考慮して照合すれば適切な典型姿勢が選ばれる.

2. 本手法の概要

輪郭の変形を学習して高次元特徴空間上の任意の姿勢の見えの分布を求める. まず図 3 のように可能な関節角度空間を等間隔に分割する. 分割されたそれぞれをクラスと呼ぶ. 関節角度が一定範囲に収まるようにクラスを認識の対象とする. またクラスの中央の姿勢を典型姿勢とする. 各クラスに属する姿勢 (バリエーション) から手指の CG モデルを作成し, その見え (学習サンプル) の分布を求める. 入力の見えは変形の分布を考慮して距離が最も近いクラスに分類される (注1).

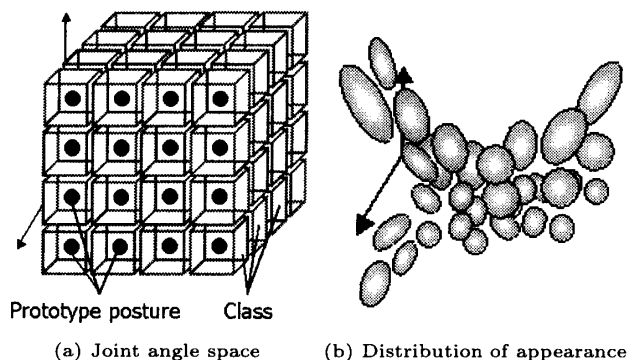


図 3 Locally-Compressed Feature Manifold モデル
Fig. 3 Locally-Compressed Feature Manifold model.

(注1): 見えの変化に応じてクラスを定めると, 一つのクラスがカバーする関節角度の範囲はクラスごとに異なるので, 関節角度の一定変化ごとにクラスを定めている.

ただし、各クラスの学習サンプルの分布はただ一つのクラスを形成しているとは限らない。複数のクラスを形成している場合には、各クラスごとに上述と同様の方法で学習サンプルを学習して、クラスの学習サンプルの分布を複数の分布モデルで表現する。このようにクラスごとにバリエーションを作成して、その見えを学習させた分布モデルを Locally-Compressed Feature Manifold (LCFM) モデルと呼ぶ。

3. 輪郭の変形を考慮した認識

本章ではまず 3.1, 3.2 で典型姿勢、バリエーションの作成について述べる。次に説明を簡潔にするため、3.3 で、クラス内で学習サンプルがクラスを一つだけ形成すると仮定した場合の学習について、3.4 でそのときの照合について述べる。その仮定が成立しない場合もあるので、3.5 でクラス内の学習サンプルが複数のクラスを形成している場合の学習と照合方法を示す。最後に 3.6 で輪郭点の対応が輪郭線に沿ってずれていた場合にずれの影響を除去して学習と照合を行う方法を述べる。

3.1 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自由度を減らす。

図 4 に手指モデルの関節角度を示す。 θ_{f1} , θ_{f2} , θ_{f3} の値は比例すると近似して、各指の曲げの自由度を一つの変数 (θ_{thumb} , θ_{index} , θ_{middle} , θ_{ring} , θ_{pinky}) で表す。また、比例係数を 1 で近似して、 $\theta_{f1} = \theta_{f2} = \theta_{f3}$ とする。

指には曲げと垂直な方向の軸の周りに回転するもう

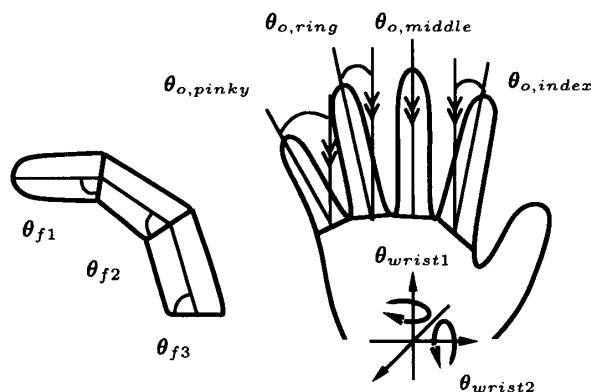


図 4 手指モデルの自由度
Fig. 4 Degrees of freedom of hand model.

一つの自由度 (open と呼ぶ) がある。open に関しても変数 θ_{open} を用いて $\theta_{o,index} = \theta_{open}$, $\theta_{o,middle} = 0^\circ$, $\theta_{o,ring} = -\theta_{open} - 4^\circ$, $\theta_{o,pinky} = 2\theta_{open} - 14^\circ$ と仮定する。

モデルがもっている手指の回転の自由度は手首を中心とした θ_{wrist1} と θ_{wrist2} の二つだけである。画像平面内での回転については照合の際に入力を正規化することにより対応している。

上述の制約によってモデルの姿勢は 8 自由度 (θ_{thumb} , θ_{index} , θ_{middle} , θ_{ring} , θ_{pinky} , θ_{open} , θ_{wrist1} , θ_{wrist2}) で表される。本研究の目的のためには 8 自由度で十分であるし、解剖学から手指の機構を調べても上述の制約条件は近似として妥当である。

典型姿勢は量子化された姿勢パラメータの組合せで定義される。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 θ_{open} の値は 0° に固定しておく。残りの 7 個のパラメータを形状変化の度合に応じて量子化する。 θ_{thumb} , θ_{index} , θ_{middle} , θ_{ring} , θ_{pinky} は 4 段階に、 θ_{wrist1} は 12 段階に、 θ_{wrist2} は 7 段階に量子化する (表 1, 曲げについてはどの指も同じ方法で量子化するので、表中では θ_{thumb} , θ_{index} , θ_{middle} , θ_{ring} , θ_{pinky} をまとめて θ_{bend} と表す)。合計で 86016 ($4^5 \times 12 \times 7$) 個の典型姿勢が定義される。典型姿勢から作成した CG モデルの例を図 5 に示す。

3.2 バリエーションの作成

それぞれのクラスに対して輪郭の変形を学習させるためにモデルの 8 個のすべてのパラメータ (open を含む) を典型姿勢の値から少し変化させてバリエーションを作成する。指の曲げと手首回りの回転の関節角度の変化量は 5 段階に量子化して、open については 3 段階に量子化する。このときある二つのクラスの境界上にある姿勢は両方のクラスのバリエーションに含まれるように変化量を設定する。各パラメータの典型姿勢の値からの変化量を表 2 に示す。ここで α はクラスによって異なる値をとり、境界上の姿勢を近傍として含むように値を調整する。典型姿勢の姿勢のパラメー

表 1 典型姿勢のパラメータ (単位は $[\circ]$)
Table 1 Parameters of prototype postures.
(unit is $[\circ]$)

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

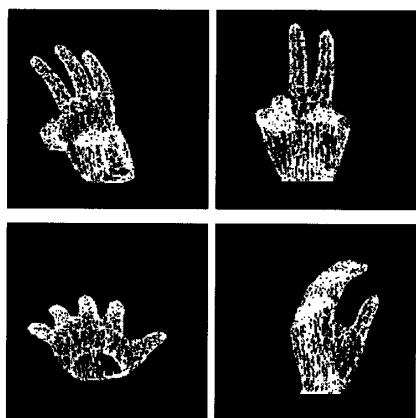
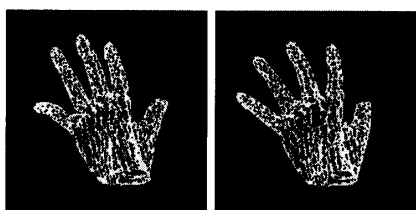


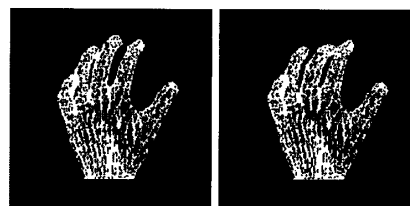
図5 典型姿勢
Fig. 5 Prototype postures.

表2 バリエーションの姿勢の変化量 (単位は [°])
Table 2 Change from prototype posture. (unit is [°])

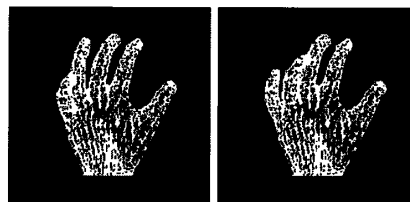
θ_{bend}	-2α	$-\alpha$	0	α	2α
θ_{open}	—	-6	0	6	—
θ_{wrist1}	-15	-7.5	0	7.5	15
θ_{wrist2}	-15	-7.5	0	7.5	15



(a) Prototype posture (b) Variation
図6 典型姿勢とバリエーション (1)
Fig. 6 Prototype posture and variation. (1)



(a) Prototype posture (b) Variation1



(c) Variation2 (d) Variation3

図7 典型姿勢とバリエーション (2)
Fig. 7 Prototype posture and variations. (2)

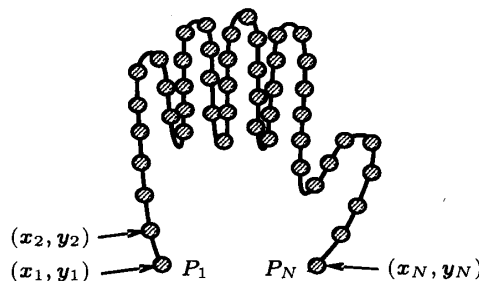


図8 形状特徴
Fig. 8 Shape feature.

$$\mathbf{x} = \frac{1}{\sqrt{A}}(x_1 - x_g, x_2 - x_g, \dots, x_N - x_g, y_1 - y_g, y_2 - y_g, \dots, y_N - y_g) \quad (1)$$

タの値 $8^\circ, 26^\circ, 48^\circ, 75^\circ$ に対して設定した α の値はそれぞれ 4, 5, 6, 7.5 である。バリエーションは各クラスごとに約 200,000 ($5^7 \times 3$) 個生成する。

図6(a)は図2(b)の典型姿勢のCGモデルである。図6(b)は図2(a)と最も似ているバリエーションから作成したCGモデルである。図7は別の典型姿勢のバリエーションのCGモデルである。図7の例では(b), (c), (d)の姿勢は典型姿勢(a)と少し異なるだけだが、隠ぺいのために輪郭は大きく変化している。

それぞれのバリエーションのCGモデルから以下に示す形状特徴を計算する。まず $N (= 128)$ 個の点をCGの輪郭上に等間隔にサンプルする(図8)。(x_1, y_1), (x_2, y_2), ..., (x_N, y_N) をサンプル点の座標, (x_g, y_g) を典型姿勢の手の領域の重心の座標, A を手領域の面積とする。ベクトル

を形状特徴として用いる。

3.3 学習方法

形状特徴の次元は $2N (= 256)$ なので、すべてのクラスのすべての学習サンプルの総データ量は大きくなる。ゆえに、学習サンプルをPCA (Principle Component Analysis) で圧縮する[11]。まずすべての典型姿勢の形状特徴から $N' (= 50)$ 次元の固有空間を作成する。この固有空間を「クラス間固有空間」と定義する。バリエーションの形状特徴をクラス間固有空間に投影して次元を N' に圧縮する。

クラス間固有空間上でクラスごとに学習サンプルの平均と共分散行列を求める。また各クラスで学習サンプルの N' 次元の固有空間を計算する。この固有空間を「クラス内固有空間」と定義する。 $\Sigma_i, \lambda_{ij}, \mathbf{e}_{ij}$ をそれぞれクラス i の学習サンプルの共分散行列, クラス内固有空間の j 番目に大きい固有値, j 番目の固有

ベクトルとする。

3.4 照合度

入力の手指領域は明るさに対するしきい値処理によって抽出することができる。文献 [12] の方法により、典型姿勢と入力の両方で輪郭上の特徴点（指先など）を抽出し、それが最も一致するように位置合わせをし、入力の輪郭上で典型姿勢の P_1, P_N (図 8 参照) に最も距離が近い点を P_1, P_N とする。

クラス間固有空間に投影した入力の形状特徴を \mathbf{y} とし、クラス i の学習サンプルの平均を $\bar{\mathbf{y}}_i$ とする。またクラス i のクラス内固有空間に投影した入力の形状特徴の第 j 成分を $g_{ij} (= (\mathbf{y} - \bar{\mathbf{y}}_i) \cdot \mathbf{e}_{ij})$ とする。クラス i の照合度はマハラノビス距離

$$\begin{aligned} d_i^2 &= (\mathbf{y} - \bar{\mathbf{y}}_i)^T \Sigma_i^{-1} (\mathbf{y} - \bar{\mathbf{y}}_i) \\ &= \sum_{j=1}^{N'} \lambda_{ij}^{-1} g_{ij}^2 \end{aligned} \quad (2)$$

である。照合結果は d_i が最小となるクラスである。マハラノビス距離を用いることによって、クラス内固有空間上で入力の輪郭が大きく変形し得る場合でも正しく照合することができる。

しかしながら学習サンプルの数が十分でないために、いくつかの分散（固有値）の値が非常に小さくなってマハラノビス距離が大きくなりすぎる可能性がある。したがって、十分たくさんの学習サンプルがあればすべての分散が定数 σ_0^2 よりも大きくなるという仮定のもとで、 σ_0^2 よりも値が小さい分散を σ_0^2 で置き換える。この σ_0^2 の値は実験的に 0.1 としている。 σ_0^2 よりも値が大きい固有値の数を r とする。式 (2) について $\lambda_{ij} (j = r + 1, \dots, N')$ を σ_0^2 で置き換えると修正したマハラノビス距離は次のように表される。

$$\begin{aligned} d_i'^2 &= \sum_{j=1}^r \lambda_{ij}^{-1} g_{ij}^2 + \sum_{j=r+1}^{N'} (\sigma_0^2)^{-1} g_{ij}^2 \\ &= \sum_{j=1}^r (\lambda_{ij}^{-1} - (\sigma_0^2)^{-1}) g_{ij}^2 + (\sigma_0^2)^{-1} |\mathbf{y} - \bar{\mathbf{y}}_i|^2 \end{aligned} \quad (3)$$

この式の値は $r (< N')$ 個の固有値を用いることによって高速に計算できる。

3.5 バリエーションの分割

前節までで、一つのクラス内の学習サンプルは正規分布に基づいて単調に分布しているという仮定のもと

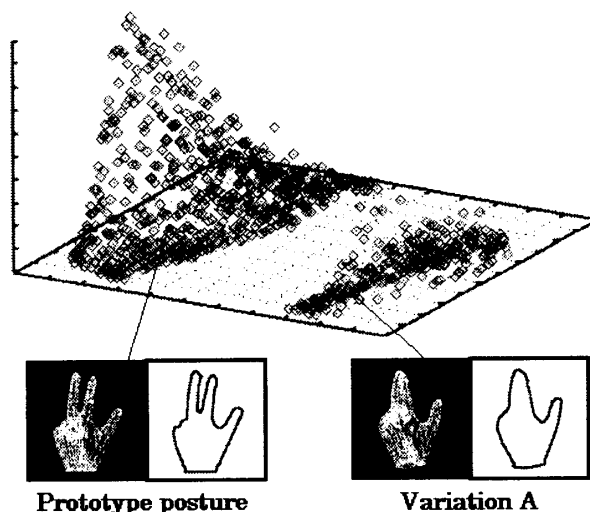


図 9 バリエーションの形状特徴の分布
Fig. 9 Distribution of shape feature variations.

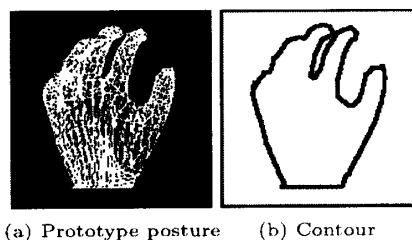
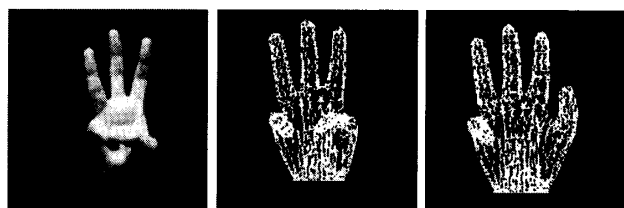


図 10 誤ってマッチした典型姿勢
Fig. 10 Mismatched prototype posture.

での輪郭の変形の学習と照合について述べた。しかし実際の学習サンプルは単調には分布していない場合がある。図 9 にあるクラスの学習サンプルの分布を示す。グラフの底面の横軸、縦軸はそれぞれこのクラスのクラス内固有空間の第一、第二主成分を表し、高さ軸は学習サンプルの密度を表している。このクラスでは学習サンプルが 1 個所に集中して分布するのではなく複数個所に分かれて分布する。その理由はバリエーション A のように中指によって人指し指の一部が隠れられ、典型姿勢と比べて輪郭形状が大きく変形するからである。

このような分布を一つのクラスとみなして学習、照合を行うと誤認識が生じる。このクラスを一つのクラスとみなした場合バリエーション A とのマハラノビス距離は大きい ($d_1' = 12.0$)。バリエーション A とは姿勢が似ていない図 10 に示すモデルを典型姿勢とするクラスとのマハラノビス距離の方が小さい ($d_2' = 10.4$)。

正しく認識を行うために、各クラスごとにそのクラスの約 20 万個のバリエーションを分布に応じていく



(a) Input image (b) Prototype posture (c) Matching result similar to input image

図 11 誤認識の例
Fig. 11 Example of mismatch.

つかのクラスに分割する必要がある。分割は、学習サンプルに対して混合正規分布モデルを当てはめてクラスタリングすることによって行う [13]。ただし、ベクトルの次元が大きいと計算量が大きくなるので、クラス内固有空間に投影した学習サンプルの第 5 成分までの成分を用いてクラスタリングする。それによって輪郭形状が似ているバリエーションが同じグループに属するように分割される (図 9 の例では二つのクラスに分割された)。各クラスで構成される見えの分布が LCFM である。

分割後の各クラスで 3.3 と同様にそのクラスに属する学習サンプルの平均ベクトル、共分散行列、固有空間を計算しておく。

入力を与えられたとき入力と探索範囲内のクラス i に属するクラス j とのマハラノビス距離を d'_{ij} とすると、クラス i との距離 d''_i を以下で定義する。

$$d''_i = \min_j d'_{ij} \quad (4)$$

照合するクラス m は以下で表される。

$$m = \arg \min_i d''_i \quad (5)$$

認識結果はそのクラス m の典型姿勢とする。前述の例では $d''_1 = 7.2$, $d''_2 = 10.8$ になり、正しく認識が行われている。

3.6 サンプル点の不对応の問題の解決

3.5 の方法でも誤認識を起こしてしまうことがある (図 11)。(a) の入力画像と最も姿勢が似ている (b) を典型姿勢とするクラスとの距離 ($d''_3 = 21.8$) は、伸ばしている指の本数が入力と異なる (c) を典型姿勢とするクラスとの距離 ($d''_4 = 19.5$) より大きくなっている。

誤認識の原因は、輪郭同士でサンプル点の対応が正しくないためである。(a) と (c) の指の部分の輪郭を重ね合わせた模式図を図 12 に示す。実線が入力の輪郭で、破線がモデルの輪郭である。形状によらず輪郭

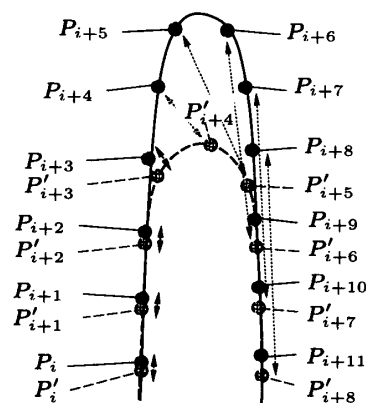


図 12 サンプル点の誤対応
Fig. 12 Mismatch of sample points.

上のサンプル点は同じ N 個なので、 P_{i+4} , P'_{i+4} 以降のサンプル点是对应を誤っている。このようにサンプル点が輪郭線に沿ってずれると、正しいクラスに対する距離が不当に大きくなり誤認識の原因になる。

全体を輪郭線方向にずらしながら対応付けを行って、最も良い評価となるものを採用する方法があるが、計算量が多くなる。また最も近い輪郭点同士を対応づける方法 (chamfer distance) もあるが、この方法では輪郭点同士の順番が保存されず、誤った対応に基づいて距離が計算されることがある。そこで、サンプル点がずれることにより生じる形状特徴変化の共分散行列をモデル化する。その共分散行列からサンプル点の輪郭線に沿ったずれに対する制約部分空間を作成しておく。その空間上で輪郭の変形の学習と入力の照合を行えばずれの影響を除去した照合が可能となる [14]。

3.6.1 ずれの共分散行列のモデル化

まず先見的な知識を用いてずれの共分散行列 (C とする) をモデル化する。サンプル点は輪郭の接線方向にずれると仮定する。サンプル点 P_i における輪郭の接線方向をクラス i の学習サンプルの平均のサンプル点の座標を用いて

$$\theta_i = \arctan \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}} \quad (j = 2, \dots, N - 1) \quad (6)$$

で与える。 $2N \times 2N$ 単位行列の (i, i) 成分, $(i, i+N)$ 成分, $(i+N, i)$ 成分, $(i+N, i+N)$ 成分をそれぞれ $\cos \theta_i$, $\sin \theta_i$, $-\sin \theta_i$, $\cos \theta_i$ で置き換えた行列を R_i と定義する。また $R = \sum_{i=1}^N R_i$ とする。接線方向は各サンプル点ごとに異なるのでサンプル点同士のずれの共分散をモデル化するのは難しい。そのため、ずれの方向を x 軸方向に統一した場合の共分散行列

$$C' = (R^{-1})^T C (R^{-1}) \quad (7)$$

をまずモデル化してからその C' を用いて

$$C = R^T C' R \quad (8)$$

を求める。

今、サンプル点 P_i が接線方向に距離 d ずれたとする。 $P_j (j = 2, \dots, i-1)$, $P_k (k = i+1, \dots, N)$ のずれの大きさは P_1 , P_N のずれの大きさが 0 になるように P_i と P_1 , P_i と P_N のそれぞれの間で線形補間することで求めることができる。つまり $P_j (j = 2, \dots, i-1)$, $P_k (k = i+1, \dots, N)$ のずれの大きさはそれぞれ $\frac{j-1}{i}d$, $(1 - \frac{k-i}{N-i})d$ となる。したがって P_i が接線方向に距離 d だけずれた影響により他のサンプル点がずれる場合の $2N \times 2N$ の共分散行列 C'_i の (j, k) 成分 $c'_{i,jk}$ は

$$\begin{aligned} c'_{i,jj} &= \left(\frac{j-1}{i}\right)^2 d^2 & (1 \leq j < i) \\ c'_{i,jk} &= c'_{i,kj} = \left(\frac{j-1}{i}\right) \left(\frac{k-i}{N-i}\right) d^2 & (1 \leq j < i \text{ かつ } k = i) \\ c'_{i,jj} &= \left(1 - \frac{k-i}{N-i}\right)^2 d^2 & (i \leq j \leq 2N) \\ c'_{i,jk} &= c'_{i,kj} = \left(1 - \frac{k-i}{N-i}\right) \left(\frac{j-1}{i}\right) d^2 & (i < j \leq 2N \text{ かつ } k = i) \end{aligned} \quad (9)$$

となる。 d の値は 0 以外の実数であれば得られる制約部分空間の基底は変化しないので $d = 1$ としている。 C'_i を用いて C' は次式で与えられる。

$$C' = \frac{1}{N-2} \sum_{i=2}^{N-1} C_i \quad (10)$$

C' から式 (8) によって C を計算し、更にクラス間固有空間に座標変換して (もとの特徴空間からクラス間固有空間への変換行列を A とする)

$$C'' = A^T C A \quad (11)$$

を求める。これがサンプル点の接線方向のずれの共分散行列である。

3.6.2 輪郭変形成分の抽出

C'' を固有値展開して得られる固有ベクトルを $\mathbf{e}_{ci} (i = 1, \dots, N')$ とする。値が大きい固有値に対応する固有ベクトルは、サンプル点のずれの影響による形状特徴変化を表している。学習サンプルの平均を中心とし、 $\mathbf{e}_{ci} (i = i_0, \dots, N')$ で張られる制約部分空

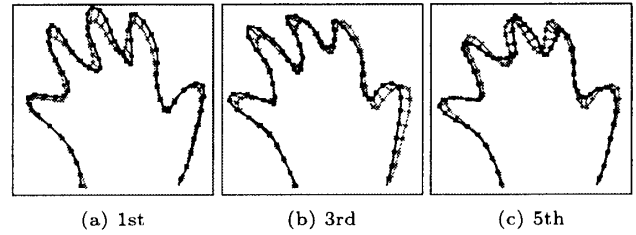


図 13 3 個の固有ベクトル
Fig. 13 Three eigenvector.

間を作成する (本研究では $i_0 = 12$ としている)。制約部分空間は値が小さい固有値に対応する固有ベクトルで張られ、形状特徴をこの空間に投影すれば、サンプル点のずれ以外の成分を抽出することができる。よって純粋に輪郭の変形によって生じた形状特徴変化のみを得る。

学習の結果得られたクラス固有ベクトルの例を図 13 に示す。灰色の輪郭は図 2 のクラスのあるクラスの平均の輪郭である。その輪郭にそれぞれの固有ベクトルを加えたのが黒い輪郭である。固有ベクトルを加える前と後のサンプル点を直線で結んである。図を見ると指の曲げなどの関節角度の変化による輪郭の変形が固有ベクトルとして学習されているのが分かる。

学習手順をまとめると以下ようになる。

1. 各クラスごとにバリエーションを作成し、学習サンプルをクラス間固有空間に投影して圧縮する。
2. 圧縮された学習サンプルをクラスごとにクラスタに分割する。
3. 各クラスの各クラスごとにサンプル点の輪郭線に沿ったずれに対する制約部分空間を作成し、学習サンプルを制約部分空間に投影する。
4. 各クラスごとに制約部分空間上で学習サンプルの分布を求める。

照合では学習と同様に入力の特徴をクラス間固有空間、制約部分空間、クラスの固有空間に投影して式 (3) のマハラノビス距離を計算する。図 11 の例について輪郭線方向のずれを除去した後の照合度は (c) が 8.6 で (b) は 5.7 になり (a) と (b) が正しく照合された。

4. 実験

4.1 照合結果

従来法で認識を誤っていた場合でも、各クラスで輪郭の変形を学習して照合度をマハラノビス距離で評価することにより正しく照合を行うことができる。図 2

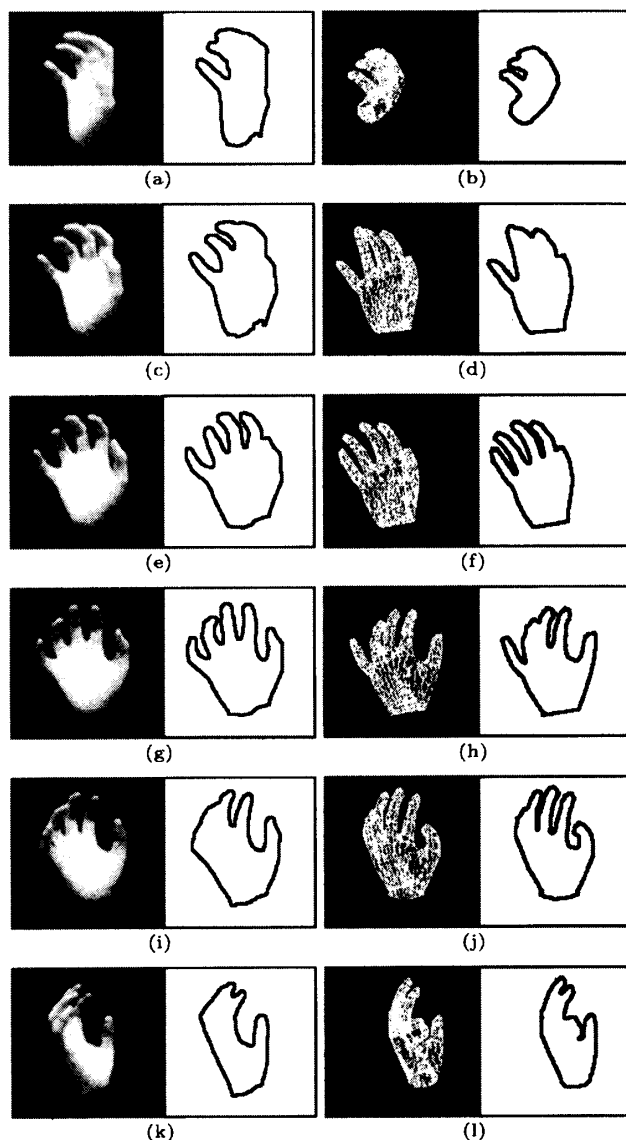


図 14 照合結果 (3.4 の照合方法)
Fig. 14 Matching result. (method described in 3.4)

の例では、(a) に対して (b), (d) とのマハラノビス距離はそれぞれ 9.4, 11.1 になり正しく照合が行われた。図 14 に時系列画像をオフラインで照合したときの照合結果を示す。左の画像は入力の手画像とその輪郭である。右の画像は照合の結果、全 (約 80,000 個) クラスの中から選択されたクラスの典型姿勢である。(c), (d), (e), (f), (g), (h), (i), (j) は正しく認識された結果で、(a), (b), (k), (l) では関節角度が少し異なるモデルが照合された。(c), (d), (g), (h), (i), (j) の例では、入力の手姿勢は典型姿勢と似ているが、入力の手輪郭形状はオクルージョンのために典型姿勢の輪郭形状と比べて大きく異なっている。このような例でも正しく認識することができた。(a), (b),



図 15 照合結果 (3.5 の照合方法)
Fig. 15 Matching result. (method described in 3.5)

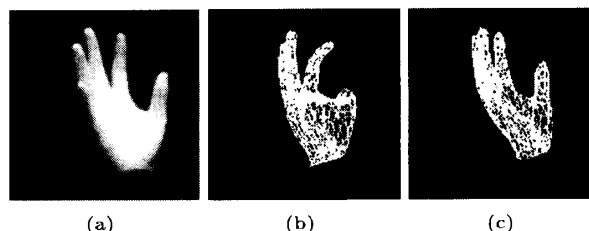


図 16 照合結果 (3.6 の照合方法)
Fig. 16 Matching result. (method described in 3.6)

(k), (l) では、入力と照合結果の姿勢が異なるにもかかわらず輪郭が似ているため、誤認識された。

また各クラスごとに輪郭の変形を考慮して照合することにより、隠ぺいが生じている入力に対して更に正確な照合を行うことができる。図 15 では、入力画像に対して、入力画像の関節角度が照合結果のクラスの範囲内となる正しいクラスが選択された。

サンプル点の不对応の問題を解決することによって正しく認識された例を図 16 に示す。(a) の入力画像に対し、3.5 の照合方法では (b) の典型姿勢が誤って選ばれた。それに対し、3.6 の照合方法では (c) の典型姿勢が選ばれた。

464 枚の入力画像に対して認識実験を行った結果、従来法 [9], 提案手法の認識率はそれぞれ 71.6%, 82.3% となった。認識率からも提案手法の有効性が確認できる。

4.2 計算時間

一つのクラス当りの照合にかかる計算時間は 0.72 ms である (CPU が Pentium III 1 GHz でメモリが 512 MByte)。時系列画像を認識する場合、1 フレームでの探索範囲は平均で 5,000 個のクラスに限定することができる [9]。計算時間は 1 フレーム当たり $0.72 \text{ ms} \times 5000 = 3.6 \text{ s}$ になる。更に、PC クラスタを用いるか [9], あるいは近い将来において PC 性能が向上すれば、実時間処理が可能になる。

1 クラス当りの学習にかかる平均時間は 173.0 秒である。CPU40 台で構成される PC クラスタを用いて全クラスの学習の計算を行った。

5. むすび

本研究では LCFM モデルを用いた手指輪郭の変形の学習と照合について述べた。学習では、クラスごとにバリエーションの見える分布を求め、LCFM を作成した。分布を考慮してマハラノビス距離で照合度を評価することにより、輪郭形状が変化しやすい入力画像でも正しく照合することができた。また輪郭上のサンプル点の誤対応の問題を解決した。実験によって本手法の有効性を示した。

今後の課題としては複雑背景下での手指輪郭を検出することである。

謝辞 本研究は文部科学省特別研究費補助 No. 15300058 の助成を受けて研究を行った。

本研究で用いている CG モデルを提供して下さった大阪大学大学院情報科学研究科の岸野文郎教授、北村喜文助教授に深く感謝致します。

文 献

- [1] 岩井儀雄, 八木康史, 谷内田正彦, “単眼動画画像からの手の 3 次元運動と位置の推定,” 信学論 (D-II), vol.J80-D-II, no.1, pp.44-55, Jan. 1997.
- [2] 藪上勝宏, 上田悦子, 松本吉央, 小笠原司, “PC クラスタを用いたリアルタイム手形状推定,” インタラクシオン 2004 論文集, pp.199-200, 2004.
- [3] X. Liu and K. Fujimura, “Hand gesture recognition using depth data,” Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition, pp.529-534, 2004.
- [4] K. Imagawa, R. Taniguchi, D. Arita, H. Matsuo, S. Lu, and S. Igi, “Appearance-based recognition of hand shapes for sign language in low resolution image,” ACCV2000, pp.943-948, 2000.
- [5] M.J. Black and A.D. Jepson, “Eigen tracking: Robust matching and tracking of articulated objects using a view-based representation,” Int. J. Comput. Vis., vol.26, no.1, pp.63-84, 1998.
- [6] 亀田能成, 美濃導彦, 池田克夫, “シルエット画像からの間接物体の姿勢推定法,” 信学論 (D-II), vol.J79-D-II, no.1, pp.26-35, Jan. 1996.
- [7] J. Cui and Z. Sun, “Visual hand motion capture for guiding a dexterous hand,” Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition, pp.729-734, 2004.
- [8] R. Rosales and S. Sclaroff, “Algorithms for inference in specialized maps for recovering 3D hand pose,” Proc. 5th Int. Conf. on Automatic Face and Gesture Recognition, pp.143-148, 2002.
- [9] N. Shimada, K. Kimura, and Y. Shirai, “Real-time 3-D hand posture estimation based on 2-D appearance retrieval using monocular camera,” Proc. Int. Workshop on RATFG-RTS, pp.23-30, 2001.

- [10] V. Athitsos and S. Sclaroff, “An appearance-based framework for 3D hand shape classification and camera viewpoint estimation,” Proc. 5th Int. Conf. on Automatic Face and Gesture Recognition, pp.40-45, 2002.
- [11] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, “Active shape models—Their training and application,” Comput. Vis. Image Underst., vol.61, no.1, pp.38-59, 1995.
- [12] N. Shimada, K. Kimura, Y. Shirai, and Y. Kuno, “Hand posture estimation by combining 2-D appearance-based and 3-D model-based approaches,” ICPR'00, pp.709-712, 2000.
- [13] C.A. Bouman, “Cluster: An unsupervised algorithm for modeling Gaussian mixtures,” <http://www.ece.purdue.edu/~bouman>
- [14] 福井和広, 山口修, 鈴木 薫, 前田賢一, “制約相互部分空間法を用いた環境変動にロバストな顔画像認識—照明変動の影響を抑える制約部分空間の学習,” 信学論 (D-II), vol.J82-D-II, no.4, pp.613-620, April 1999.

(平成 16 年 10 月 8 日受付, 17 年 2 月 7 日再受付)

今井 章博

2002 阪大・工・応用理工卒。2004 同大学院修士課程了。現在、同大学院後期課程在学中。単眼視画像列からの手指姿勢認識に関する研究に従事。



島田 伸敬 (正員)

1992 阪大・工・電子制御機械卒。1997 同大学院博士後期課程了。工博。同年同専攻助手。2001 同研究科研究連携推進室情報ネットワーク部門講師, 同研究科助教授を経て, 現在, 立命館大学情報理工学部知能情報学科助教授。コンピュータビジョン, ジェスチャ認識, ヒューマンインタフェース, インターネットソリューションの研究に従事。情報処理学会, IEEE 各会員。

白井 良明 (正員)

1964 名大・工・機械卒。1969 東大大学院博士課程了。同年電子技術総合研究所研究官, 研究室長, 部長となる。1988 大阪大学工学部教授。2005 立命館大学情報理工学部知能情報学科教授。2002~現在, 独立法人情報学研究所客員教授。知能ロボットに関する研究に従事。工博。1983, 1994 本会論文賞受賞。情報処理学会, 人工知能学会等各会員。