

カーネル非線形直交相互部分空間法に基づく手形状識別

Hand Shape Recognition based on Kernel Orthogonal Mutual Subspace Method

大川 泰弘, 福井 和広

Yasuhiro Ohkawa, Kazuhiro Fukui

筑波大学大学院システム情報工学研究科

Graduate School of Systems and Information Engineering, University of Tsukuba

E-mail: ohkawa@viplab.is.tsukuba.ac.jp, kfukui@cs.tsukuba.ac.jp

Abstract

本論文は、複数視点画像を用いて高精度に手形状を識別する手法を提案する。近年、3次元物体を識別する研究が盛んに行われている。3次元物体は、視点や照明により見え方が大きく変動するため、高精度に識別することは難しい。特に、ハンドジェスチャ(以下、手形状)識別は、人により形状や大きさが変動するため、難しい問題である。そこで、提案手法では複数視点画像から位置不変特徴である高次局所自己相関特徴を抽出し、これにカーネル非線形直交相互部分空間法を適用して識別を行う。提案手法を実装したシステムを用いて評価実験を行い、類似した形状を含む10カテゴリの手形状を高精度でリアルタイム識別できることを確認した。

1 はじめに

近年、手形状識別を行う研究が盛んに行われている[1][2]が、手形状は図1のように視点により見え方が大きく変動するため、高精度に識別することは難しい。

図2は図1に示すように手形状を回転させながら撮影した複数視点画像をその固有空間に射影したものである。3つのカテゴリ分布が複雑に絡み合い、単一視点画像のみを用いた単純な方法では識別困難な問題であることがわかる。このような識別問題に対しては複数視点画像を用いることが有効であると考えられる。

複数視点から得られた物体の見え方のパターン分布は低次元の線形部分空間で近似可能であることが知られている[6][7]。物体の見え方が、その3次元形状を反映していると考え、線形部分空間同士の類似度を測ることで、3次元形状同士の類似度を測ることができる。二つの線形部分空間の類似度は、両者なす複数の正準角により規定される。この部分空間同士のなす正準角を基に識別を行う手法が相互部分空間法(Mutual Subspace Method:MSM)[8]である。

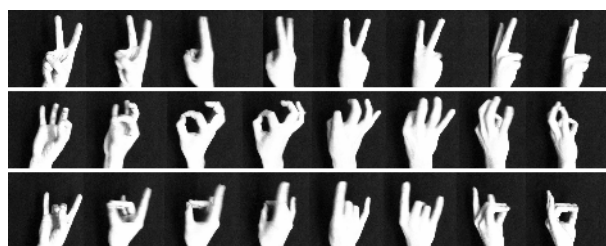


図1 手形状の多視点画像: 上からチョキ, OK, キツネ

MSMを用いることで2つのパターン分布の類似度を測ることが可能となるが、他のカテゴリとの関係が考慮されていないため、MSMの識別性能は必ずしも最良とは言えない。これに対して直交相互部分空間法(Orthogonal Mutual Subspace Method: OMSM)[11]では、直交化行列を用いた線形変換により、部分空間同士のなす角度を広げることで、カテゴリ間の差異を強調し識別率の向上を図る。

OMSMにより単一視点画像を用いる方法に比べて高い識別性能が実現できるが、依然として次のような問題が残されている。手形状のような複雑な3次元形状を有する物体の見え方のパターン分布は、一般に非線形構造を持つため、線形部分空間で他カテゴリと重なり無く近似することは難しい。このような場合には、パターンを高次元(あるいは無限次元)の特徴空間 \mathcal{F} へ非線形写像し、写像されたパターンに対してOMSMを適用する、カーネル非線形直交相互部分空間法(Kernel Orthogonal Mutual Subspace Method: KOMSM)[14]が有効である。

KOMSMを用いることにより、非線形構造を持つパターン分布に対しても高い識別性能を実現できるが、学習時には全学習パターン数の3乗、識別時に全学習パターン数の2乗に比例した計算量が必要となる。したがって学習パターン数の多い問題や、リアルタイムシステムの場合、KOMSMを適用することが困難になる。

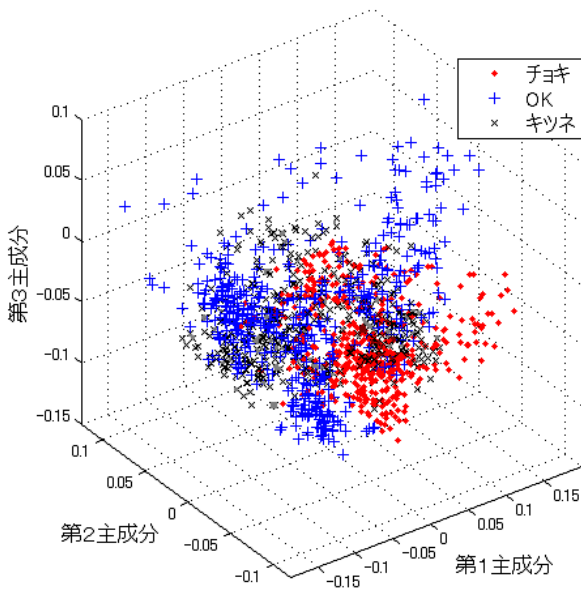


図2 手形状の分布

計算量の問題を解決するためには、学習パターン数を削減すれば良いが、学習パターン数を削減すると識別率が低下する。識別率を下げずに学習パターン数を削減する手法として、スパースカーネル主成分分析 [17] や学習パターンを K-平均法で求まるクラスター中心で近似する手法 [15] などが提案されている。本論文では、それらの中でもカーネル MSM [9] に対して最も有効であると示された K-平均法による削減法を適用する。

更に本論文では識別特徴についても検討を行った。画像認識では濃淡画像を直接ベクトル化したものを識別特徴とする場合が多いが、多視点画像の場合には、見え方が大きく変わるために各カテゴリ分布が大きく広がってしまう。このために精度が高い辞書を作成するためにはそれに見合った大量の学習データが必要となり、これにより計算困難となる。そこで手形状の位置に依らない不変特徴を用いることで、パターン分布をコンパクト化し、必要な学習データを削減する。位置不変特徴としては高次局所自己相関特徴 (HLAC) [3] を用いる。

本論文では第2章において、カーネル非線形直交相互部分空間法について述べる。第3章では、カーネル非線形直交相互部分空間法を用いた手形状識別アルゴリズムについて述べる。第4章では、提案手法の有効性を実験により示し、考察を行う。第5章はまとめである。

2 カーネル非線形直交相互部分空間法

本章では、複数視点画像を用いた手形状識別に用いるカーネル非線形直交相互部分空間法について述べる。

まず相互部分空間法とそれに直交化による特徴抽出を付加した直交相互部分空間法について述べ、次にカーネル非線形直交相互部分空間法への非線形拡張について述べる。

2.1 相互部分空間法

相互部分空間法は学習パターンと入力パターンの双方を主成分分析により部分空間で近似し、部分空間同士のなす正準角に基づき、識別を行う手法である。正準角とは、ベクトル間のなす角度を拡張したものである。M次元部分空間 P と N次元部分空間 Q との間には M個 (便宜上、 $M \leq N$) の正準角が定義できる。部分空間 P の第 i 正規直交基底ベクトルを p_i 、部分空間 Q の第 j 正規直交基底ベクトルを q_j とすると、第 n 正準角 θ_n の余弦 $\cos\theta_n$ は、次式で示す $M \times N$ 行列 D の第 n 番目に大きい特異値として求めることができる。

$$d_{i,j} = (p_i \cdot q_j)$$

$$D = (d_{i,j}), (i = 1 \sim M, j = 1 \sim N)$$

(1)

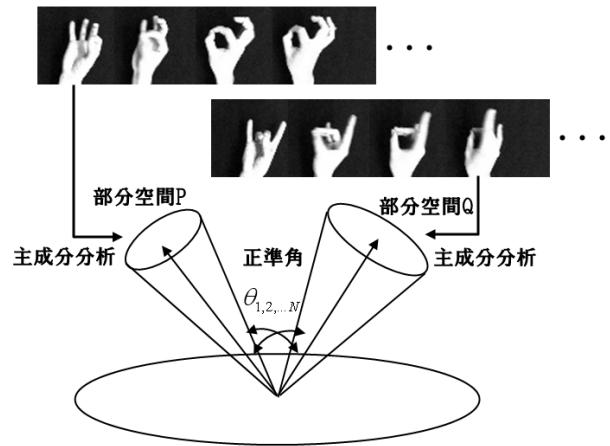


図3 相互部分空間法

2.2 直交相互部分空間法

相互部分空間法は他のカテゴリとの関係が考慮されていないため、識別に関しては必ずしも最良とは言えない。直交相互部分空間法 [11] では直交化行列 O を用いた線形変換により、部分空間同士のなす角度を広げることによって、カテゴリ間の差異を強調し識別率の向上を図る。直交化行列 O は以下の手順で求める。

各カテゴリの学習用部分空間への射影行列 $P_k (k = 1 \sim L)$ 、その L 個の平均を P とする。

$$P = \frac{1}{L} \sum_{k=1}^L P_k$$

(2)

P は L 個の学習部分空間の分布を表現する行列となる。直交化行列 O は、P の固有値をすべて 1 にする白色化

変換を表す行列として, 次式で与えられる.

$$\mathbf{O} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{B}^T \quad (3)$$

ただし, $\mathbf{\Lambda}^{-\frac{1}{2}}$ は \mathbf{P} の固有値の平方根の逆数を並べた対角行列, \mathbf{B} は \mathbf{P} の固有ベクトルを列に並べた行列である.

2.3 カーネル非線形直交相互部分空間法

KOMSM では, 入力空間のパターン x を非線形写像 Ψ により, 極めて高次元 (あるいは無限次元) の特徴空間 \mathcal{F} に写像し, 写像された $\Psi(x)$ に対して, 先に述べた直交相互部分空間法を適用する.

特徴空間 \mathcal{F} 上で直交相互部分空間法を構築するためには, 特徴空間上において各カテゴリ分布を部分空間 (非線形カテゴリ部分空間) で表し, それらを直交化するための直交化行列 (カーネル直交化行列 \mathbf{O}_Ψ と呼ぶ) を求める必要がある.

\mathbf{O}_Ψ の計算過程において非線形部分空間の基底ベクトル同士あるいは入力ベクトルの写像 $\Psi(x)$ と基底ベクトルの内積計算が必要となるが, 特徴空間の次元が極めて高いために計算困難 (無限次元空間では不可能) となる. そこで特徴空間での内積 ($\Psi(x) \cdot \Psi(y)$) をカーネル関数 $k(x, y)$ に置き換えるカーネルトリックと呼ばれる技法を用いる. 代表的なカーネル関数にガウシアンカーネルがある.

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (4)$$

具体的にカテゴリ数を r , 各カテゴリ非線形部分空間の次元を d とするとき, 非線形写像 $\Psi(x)$ のカーネル直交化行列 \mathbf{O}_Ψ により直交変換された $\chi(\Psi(x))$ は次式で表される.

$$\chi(\Psi(x)) = (\chi_1, \chi_2, \dots, \chi_{r \times d})^T \quad (5)$$

$$\chi_i = (\Psi(x) \cdot \mathbf{O}_{\Psi_i}) \quad (i = 1 \sim r \times d) \quad (6)$$

ここで $\Psi(x)$ や \mathbf{O}_{Ψ_i} は陽には取り出すことができないが, 両者の内積 ($\Psi(x) \cdot \mathbf{O}_{\Psi_i}$) は計算可能なので, これにより直交化変換されたパターン $\chi(\Psi(x))$ は, $r \times d$ 次元の実体のあるベクトルとして求まる (詳細は文献 [14] を参照).

直交化変換された各カテゴリのパターン集合 $\{\chi(\Psi(x))\}$ に対し主成分分析を適用して部分空間 $\mathcal{P}_{\mathbf{O}_\Psi}$ を生成すると, これらはお互いに直交化された非線形カテゴリ部分空間となっている.

3 KOMSM による手形状識別

本章では, KOMSM を用いた手形状識別について述べる. まず画像から位置不変特徴を抽出する方法について概説し, 次に識別処理の流れと高速化について説明する.

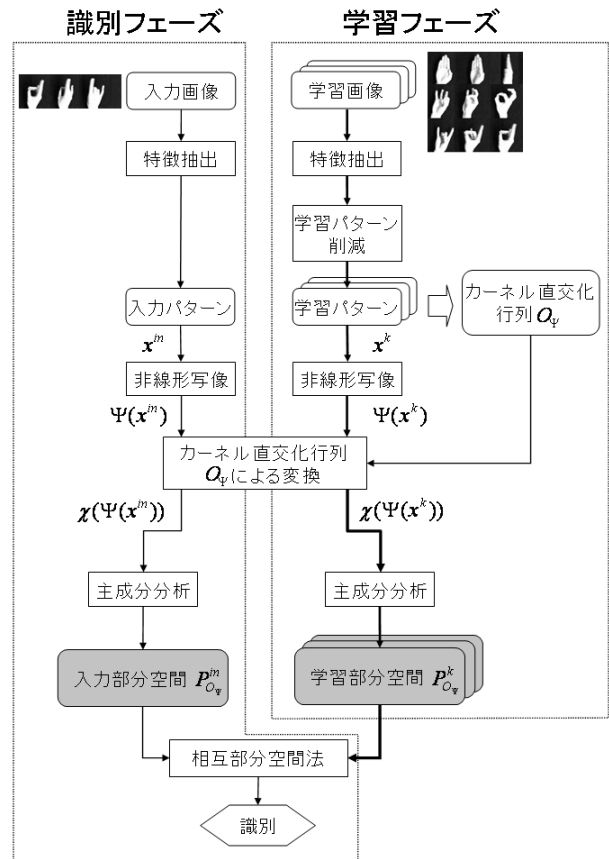


図 4 手形状識別の流れ

3.1 位置不変特徴の抽出

高次局所自己相関特徴 (HLAC) は, 加法性や位置不変性などのすぐれた性質を持ち, 顔認識や物体認識などで高い性能を示している. HLAC は何次元まで計算するかで得られる特徴ベクトルの次元数が違ってくるが, 本論文では 1 枚の濃淡画像から 35 次元の特徴ベクトルを抽出する.

HLAC を抽出する前処理として, 640×480 ピクセルの濃淡画像を 80×60 ピクセルに縮小し, ガウシアン・ラプラシアンフィルタを用いてエッジ抽出する. このエッジ画像に対し高次局所自己相関特徴を抽出する.

HLAC は, 識別対象の大きさや画像の解像度により性能が大きく変動ため, 画像ピラミッドを利用する手法 [4] や, マスクパターンのサイズを拡大する手法 [5][16] などが提案されている. 本論文では, マスクパターンのサイズを $3 \times 3, 5 \times 5, 9 \times 9, 17 \times 17$ とし, それぞれについて 35 次元の高次局所自己相関特徴を抽出する. したがって 1 枚の画像から 140 次元の特徴ベクトルが得られることになる.

3.2 識別処理の流れ

識別処理は図 4 に示すように, 大きく学習フェーズと識別フェーズに分けられる.

学習フェーズ

1. 学習画像から位置不変特徴 (HLAC) を抽出し, これを学習パターンとする.
2. KOMSM の高速化のために学習パターンの削減を行う.
3. 全カテゴリの学習パターンから, カーネル直交化行列 O_{Ψ} を計算する.
4. カテゴリ k に属するパターン $x_i^k (i = 1 \sim m)$ の非線形写像 $\Psi(x_i^k)$ をカーネル直交化行列 O_{Ψ} を用いて直交化変換する.
5. 直交化変換されたパターン集合 $\{\chi(\Psi(x^k))\}$ に対し, 主成分分析を適用し, 線形部分空間 $\mathcal{P}_{O_{\Psi}^k}$ を生成する.

識別フェーズ

1. 入力画像から位置不変特徴 (HLAC) を抽出し, これを入力パターンとする.
2. 入力パターンの集合 $x_j^{in} (j = 1 \sim n)$ の非線形写像 $\Psi(x_j^{in})$ を学習フェーズで求めたカーネル直交化行列 O_{Ψ} を用いて変換する.
3. 直交化変換された入力パターン集合 $\{\chi(\Psi(x^{in}))\}$ に対し, 主成分分析を適用し, 線形部分空間 $\mathcal{P}_{O_{\Psi}^{in}}$ を生成する.
4. 入力部分空間 $\mathcal{P}_{O_{\Psi}^{in}}$ と各カテゴリの部分空間 $\mathcal{P}_{O_{\Psi}^k}$ の正準角を式 (1) により計算する.
5. 全ての類似度の中で, しきい値以上で最も高い類似度に該当するカテゴリを入力パターン集合のカテゴリとして識別する.

3.3 KOMSM の高速化

カーネル非線形直交相互部分空間法は, 学習パターン数を d , カテゴリ数を r , 入力パターン数を n , とすると, 学習時には $O(d^3 r^3)$, 識別時には $O(d^2 r^2 n)$ の計算が必要となる. 図 5 に $n = 3$ とし, $d \times r$ を変化させたときの識別時間の変化を示す. 30fps でのリアルタイム識別を行うには, 特徴抽出の時間も考慮すると, $d \times r = 1200$ 程度にする必要があり, カテゴリ数 $r = 10$ の時, 各カテゴリの学習パターン数は $d = 120$ 程度に抑えなければならない. そこで, 学習パターン分布を K-平均法で求まる K 個のクラスタ中心で近似することにより, パターン分布を変化させずに学習パターンの削減を実現する.

4 実験および考察

本章では, まず複数視点画像の有効性の検証を行う. 次に学習パターン削減量と識別率の関係を評価し, どの程度までの削減が可能かを明らかにする.

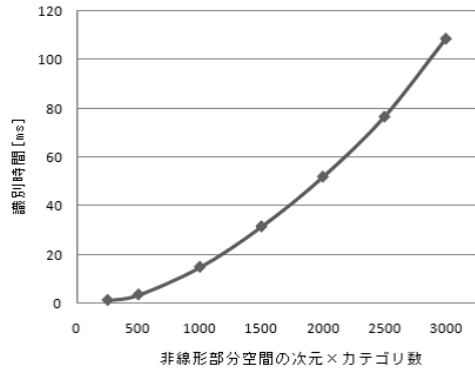


図 5 学習サンプル数と識別処理時間

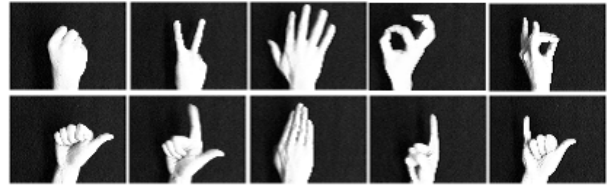


図 6 10 カテゴリの手形状: 左上から右へ, ゲー, チョキ, パー, OK, キツネ, 親指, ピストル, チョップ, 人指し, 親子

4.1 実験条件

手形状の撮影は, 図 7 に示すマルチカメラシステムで行った. 画像入力には IEEE1394 カメラ (Point Grey Research 社 Flea) に焦点距離 8mm のレンズを装着したものをを用いた. カメラは下面から約 45cm の高さに設置し, 左右のカメラ間隔は約 21cm とした. 図 6 は識別対象の 10 カテゴリの手形状である. 被験者数は実験 1, 2 では 1 名, 実験 3 では 8 名とした.

識別システムは C++ 言語を用いて実装した. エッジ抽出などの画像処理は OpenCV, 特異値問題などの行列演算は GNU Scientific Library を用いた. 識別に用いた CPU は Core 2 Duo 2.66GHz である.

4.2 複数視点有効性の検証 (実験 1)

複数視点画像を用いた識別の有効性について検証するために, 一定速度で手を回転させながら撮影した. ただし問題を単純化するために図 7 の中央カメラのみを用いた. 評価では入力画像数を 1 枚から 17 枚まで変化させて, 識別率がどのように変化するかを調べた. ここで入力画像の枚数が多いほど, 多くの視点の画像を用いた識別になっている. 図 8 に 1 秒間に撮影された 30 枚の手形状画像を示す.

学習パターンは以下の手順で生成した. 各カテゴリ毎に収集した 500 枚の学習画像に対して HLAC 特徴抽出を行った後, これらに K-平均法を適用して K 個のクラスタ中心を求めた. 各クラスタ中心を学習パターン

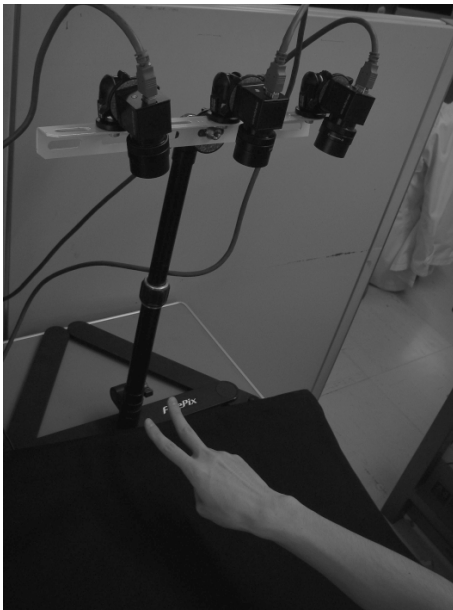


図7 マルチカメラシステム

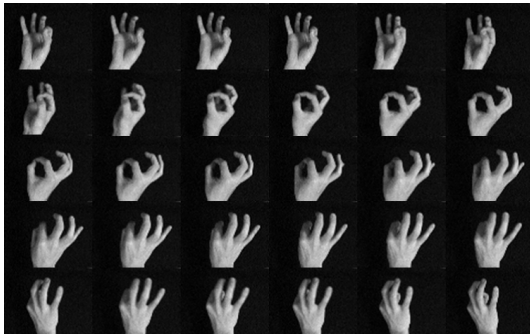


図8 1秒間に撮影された30枚の手形状画像

とする。

予備実験により, 学習パターン数 $K=80$, 非線形カテゴリ部分空間の次元 $d=80$, カーネル関数はガウシアンカーネル ($\sigma^2=0.1$) に設定した。非線形入力部分空間の生成には, 入力パターンを全て使ったので, その次元は入力パターン数を同じとなっている。

学習画像とは異なる画像に対して, 各カテゴリ 500 回, 計 5000 回の識別を行った。実験結果を表 1 に示す。この表により, 入力パターン数, つまり視点数が増えるにつれ, 識別率が確実に上昇していることがわかる。これにより KOMSM が複数視点画像の情報をうまく引き出せていることが確認できた。また 15 枚を撮影する時間, つまり 0.5 秒程度の時間遅れで, 実用的な識別性能が実現できることも併せて確認できた。

4.3 学習パターン削減の検討 (実験 2)

K-平均法による学習パターン削減量と識別率の関係性を調べる。この実験では 3 台のカメラを全て使う。これにより手を回転させることなく多視点画像が確実に

表 1 入力パターン数と識別率

入力パターン数	識別率 [%]
1	91.4
3	94.8
5	96.4
7	96.8
9	98.2
11	98.6
13	99.0
15	99.6
17	99.8



左カメラ 中央カメラ 右カメラ

図9 3台のカメラから同時に入力される画像

得られる。また入力としてはリアルタイム性を重視して, 各カメラから 1 枚ずつ, 図 9 に示す合計 3 枚の画像を用いる。

学習パターンは各カテゴリ毎に 1500 枚 (各カメラ 500 枚) の学習画像の HLAC 特徴に対して K-平均法を適用して求めた。カーネル関数としてはガウシアンカーネル ($\sigma^2=0.1$) を用いた。非線形カテゴリ部分空間は学習パターンを全て使って生成したので, その次元は学習パターン数を同じとなっている。

学習画像とは異なる画像に対して各カテゴリ 500 回, 計 5000 回の識別を行う。学習パターン数を 10 から 200 まで変化させ, 識別率と識別時間を計測した結果を表 2 に示す。学習パターンを削減するほど識別率は低下するが, 識別速度は向上している。学習パターン数を 80 より増やしても識別率の向上は見られないことから, 80

表 2 学習パターン数と識別率

学習パターン数	識別率 [%]	識別時間 [ms]
10	85.7	0.26
20	94.2	0.46
40	98.1	0.96
60	99.2	2.54
80	99.5	8.81
100	99.5	12.85
120	99.5	18.23
200	99.5	45.84

個のクラスタ中心でパターン分布を十分に近似できていると考えられる。学習パターン数が80個の時、識別時間は8.81msであるので特徴抽出などにかかる時間を考慮しても十分にリアルタイム識別が可能であると考えられる。

4.4 他人の手形状による識別 (実験3)

一般的な使用環境を想定し、他人の手形状画像から作成した学習パターンを用いて、識別がどの程度可能であるかを検証する。8人から10カテゴリの手形状画像を撮影し、Leave-one-out法により評価を行う。カメラは中央カメラのみを用いて一定速度で手を回転させながら撮影した。学習パターンは各カテゴリ毎に7000枚(各人物1000枚)の学習画像から先の実験と同様に生成した。

予備実験により、学習パターン数 $K=180$ 、非線形カテゴリ部分空間の次元 $d=140$ 、カーネル関数にはガウシアンカーネル($\sigma^2=0.04$)に設定した。入力パターン数は32として、これらを全て使って32次元、非線形入力部分空間を生成した。

実験を行った結果、8人の平均識別率は91.8(%)、最良の人物では97.2(%)となった。これにより、異なる人物の手形状画像から学習したシステムでも識別が可能であることがわかった。実験2,3に比べて識別率が下がっている原因は、学習に用いる人物やデータ数を小規模だった点が挙げられる。今後はデータ数を増やすと同時に、3台のカメラを全て使って識別率の向上を目指す。

5 まとめ

本論文では、複数視点から得られた画像に対し、高次局所自己相関特徴を抽出し、カーネル非線形直交相互部分空間法を基に識別を行う手法を提案した。提案手法により、見え方の変動の大きい3次元物体である手形状に対し、高精度に識別できることを実験により示した。今後の課題として、識別できる手形状を増やす、指文字、手話認識への応用などが考えられる。

参考文献

- [1] 上田悦子, 松本吉央, 今井正和, 小笠原司, “多視点シルエット画像を用いた手の形状推定”, 情報研報 CVIM-128-4, pp.25-31, 2001.
- [2] 東山和弘, 小野智司, 中山茂, “3次元テンプレートマッチングによる指文字認識”, 情報処理学会論文誌, vol.125-C, 2005.
- [3] 大津展之, “パターン認識における特徴抽出に関する数理的研究”, 電子技術総合研究所研究報告, vol.818, 1981.
- [4] T. Kurita, N. Otsu, T. Sato, “A face recognition method using higher order local autocorrelation and multivariate analysis”, Proc. International Conference on pattern Recognition, vol.2, pp.213-216, 1992.
- [5] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu, “Face recognition system using higher order local autocorrelation and multiscale intergration”, IEEE Trans. Pattern Anal. Mach. Intell., vol.18, No.10, pp.1024-1028, 1996.
- [6] エルツキ・オヤ著, (小川英光, 佐藤誠訳), “パターン認識と部分空間法”, 産業図書, 1986.
- [7] 村瀬洋, S.K.Nayar, “2次元照合による3次元物体認識-パラメトリック固有空間法-”, 信学論, J77-D-II,11, pp.2179-2187, 1994.
- [8] 前田賢一, 渡辺貞一, “局所的構造を導入したパターン・マッチング法”, 信学論, vol.J68-D, no.3, pp.345-352, 1985.
- [9] 前田英作, 村瀬洋, “カーネル非線形部分空間法によるパターン認識”, 信学論, vol.J82-D-II, no.4, pp.600-612, 1999.
- [10] 福井和広, 山口修, “部分空間法の理論拡張と物体認識への応用”, 情報処理学会論文誌, vol.46, No.SIG 15 (CVIM 12), pp.21-34, 2005.
- [11] 河原智一, 西山正志, 山口修, “直交相互部分空間法を用いた顔認識”, 情報研報 CVIM-151, pp.17-24, 2005.
- [12] Bernhard Scholkopf, Alexander Smola, Klaus-Robert Muller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”, Neural Computation, 10,5, pp. 1299-1319, 1998.
- [13] 坂野鋭, “パターン認識における主成分分析-顔画像認識を例として-”, 統計数理第49巻第1号, pp.23-42, 2001.
- [14] K. Fukui, O. Yamaguchi, “The Kernel Orthogonal Mutual Subspace Method and its Application to 3D Object Recognition”, ACCV07, pp.467-476, 2007.
- [15] 市野将嗣, 坂野鋭, 小松尚久, “クラスタリングを用いた核非線形相互部分空間法の処理量削減手法”, 信学論, vol.J87-D-II, No.8, pp.1622-1631, 2004.
- [16] 豊田崇弘, 長谷川修, “高次局所自己相関特徴の拡張”, 画像電子学会誌, Vol.34, No.4, pp.390-397, 2005.
- [17] M. Tipping, “Sparse kernel principal component analysis”, Proc. Conf. Advances in Neural Information Processing Systems 13, pp.633-639, 2001.