

重み付き相関行列による局所部分空間法

山下 幸彦[†]

[†] 東京工業大学大学院理工学研究科 〒152-8552 東京都目黒区大岡山 2-12-1-S6-19

E-mail: [†yamasita@ide.titech.ac.jp](mailto:yamasita@ide.titech.ac.jp)

あらまし 飯島, Watanabe によって独立に提案された部分空間法は, 考え方が分かりやすいこと, 性能が高いこと, 実現が容易であることなどの理由により, 文字認識, 顔画像認識など極めて広い範囲で応用されている. 本稿では, 筆者らが提案してきた部分空間法に関係するパターン認識法として, 相対 Karhunen-Loève 変換法, カーネル Wiener フィルタ法を紹介する. また, 高次相関から計算される重み付き相関行列による局所部分空間法を提案する. そして, その有効性を計算機実験によって確認する.

キーワード 局所部分空間法, 重み付き相関行列, 部分空間法, 相対 Karhunen-Loève 変換法, カーネル Wiener フィルタ法

Local subspace method with weighted correlation matrix

Yukihiko YAMASHITA[†]

[†] Graduate School of Engineering and Science, Tokyo Institute of Technology O-okayama 2-12-1-S6-19, Meguro-ku, Tokyo, 152-8552 Japan

E-mail: [†yamasita@ide.titech.ac.jp](mailto:yamasita@ide.titech.ac.jp)

Abstract The subspace method originally proposed by Iijima and Watanabe is used for many areas in pattern recognition since its concept is simple, its performance is high, and its implementation is not complex. In this paper, we introduce the relative Karhunen-Loève method and the kernel Wiener method that are classification methods related to the subspace method, and are proposed by the authors. We propose a new local subspace method based on a weighted correlation matrix that can be calculated from the third or higher order correlations. We show its advantages by experimental results.

Key words Local subspace method, weighted correlation matrix, subspace method, relative Karhunen-Loève transform method, kernel Wiener filter method

1. はじめに

飯島 [1], [2], Watanabe [3] によって独立に提案された部分空間法は, 考え方が分かりやすいこと, 性能が高いこと, 実現が容易であることなどの理由により, 文字認識, 顔画像認識など極めて広い範囲で応用されている. また, この部分空間法を改良するために, 学習部分空間法, 相互部分空間法, 局所部分空間法, 非線型部分空間法などが提案されている.

基本的な部分空間法では, 認識に用いられる部分空間は主成分分析 (PCA), Karhunen-Loève 変換 (KLT), または, モード関数展開によって求められている. PCA の計算のためには, 固有値問題を解くことが必要となるが, パターンベクトルの次元が高い場合は計算が困難になる. また, 直接的にカテゴリーの分離に適した部分空間を求めているのではないという問題も存在する. 学習部分空間法は, カテゴリーを分離し, 認識に

適する部分空間を繰り返し法によって逐次的に求める手法である. そして, 単純な計算の繰り返しであるため, パターンベクトルの次元が高い空間でも計算が可能であるという特徴を持っている.

PCA ではカテゴリーの平均的な特徴を捉えている. 従って, 平均的でないパターンに対して認識率が劣化する. 局所部分空間法は, クラスタリング等により, カテゴリーをサブカテゴリーに分解し, そのサブカテゴリーを独立のカテゴリーとして部分空間法を適用するものである [4]~[6]. 平均的でない類似のパターンを集め, ひとつのカテゴリーとすることによって認識率を向上させることができる. しかしながら, このような手法には, クラスタリングアルゴリズムやその初期値によって, 結果が変わってしまうという問題が存在する. また, 適応局所部分空間法 [7] は, 入力パターンに近い学習パターンで相関行列を構成するものである. この手法では, 認識時に学習パター

ン数オーダーの計算が必要になるという問題がある。

相互部分空間法は、ひとつの対象に対して複数のパターンが得られる場合に、部分空間法を拡張したものである [8]~[10]. 入力パターンベクトルに対して部分空間を構成し、カテゴリを表す部分空間とのなす角 (複数) を評価することによって、パターン認識を行うものである。実際に、顔画像認識などに応用され、その有効性が確認されている。

サポートベクトルマシンのために導入されたカーネル法は、パターンベクトルを非線型写像によって Hilbert 空間に写像し、その Hilbert 空間で線形あるいは 2 次識別を行う手法である [11], [12]. そして、パターンベクトルの写像先の Hilbert 空間における内積をカーネル関数で与えることができれば、認識器を構成するための計算が、ヒルベルト空間でなく、学習パターン数次元のベクトル空間で実行できる [13]. このことによって、計算量を大幅に削減することができる。非線形部分空間法は、部分空間法にカーネル法を導入したものであり、部分空間はカーネル PCA (KPCA) によって計算される。非線型写像によって、入力パターンベクトルの空間で複雑な境界を表現することが可能で、認識率を向上させることができる。しかしながら、学習パターン数が非常に多くなることがあり、そのような場合は、計算量が大きくなりすぎ、実質的に実現不可能になる。

本稿では、筆者らが提案してきた部分空間法に關係するパターン認識法として、相対 Karhunen-Loève 変換法 (RKL 変換法) [14]~[16], カーネル Wiener フィルタ法 (KWF 法) [17], [18] を紹介すると共に、重み付き相関行列による局所部分空間法を提案する。

PCA により得られる正射影行列を、そのカテゴリの特徴を抽出するものと考え。このとき、RKL 変換法により得られる行列は、自カテゴリの特徴を抽出しながら、他カテゴリの特徴は抽出しないようにしたものである。そのため、カテゴリの分離に適した特徴を抽出できると考えられる。

KWF 法は、信号復元に使われる Wiener フィルタを、カーネル法を使って拡張し、パターン認識に応用したものである。カテゴリごとにそのカテゴリを代表する原信号を選び、入力パターンベクトルを観測信号と考える。識別は、その観測信号からのカテゴリを表す原信号への復元と考えることによって実現する。さらに、KWF に元のパターンベクトルの空間における距離に従った正則化法を導入している。

さらに、本稿では、入力パターンベクトルと学習パターンベクトルとの内積で、その学習パターンに重みを付けた相関行列による部分空間法を提案する。本手法では、この重み付き相関行列を使って PCA を計算し部分空間を求める。相関行列は、この重みによって入力パターン周辺の学習パターンの性質を強くを表していることになる。従って、本手法は局所部分空間法の一つと考えることができる。また、非線型な識別法ではあるが、クラスティングのような問題が生じることはなく、安定した認識結果が得られる。そして、高次相関を計算しておけば、認識時に学習パターン数のオーダの計算を不要にすることができる。最後に、計算機実験によって本手法の有効性を示す。

2. 相対 Karhunen-Loève 変換法 (RKL 変換法)

R^N を、パターンベクトルが存在する N 次元ベクトル空間とする。 R^N のベクトルは列ベクトルで表されているものとする。 R^N のベクトル f のノルムと転置したものを、それぞれ、 $\|f\|$ and f^T で表す。 R^N の 2 つのベクトル f, g の内積を、 $\langle f, g \rangle (= f^T g)$ によって表す。

部分空間法では、各カテゴリ $\Omega^{(c)}$ ($c = 1, 2, \dots, C$) に対して部分空間 $S^{(c)}$ を用意する。ここで、 C はカテゴリ数である。 $P^{(c)}$ を $S^{(c)}$ への正射影行列とする。未知パターンベクトル x の識別結果となるカテゴリの番号を $D(x)$ で表せば、

$$D(x) = \operatorname{argmax}_{c=1,2,\dots,C} \|P^{(c)}x\| \quad (1)$$

となる。

カテゴリ $\Omega^{(c)}$ に属しているパターンベクトル f に対する、母集団期待値を $E_{f \in \Omega^{(c)}}$ で表す。PCA によって得られる正射影行列 $P^{(c)}$ は、そのランクが一定であるという条件の元で、次の評価基準を最大にするものとして特徴づけることができる。

$$E_{f \in \Omega^{(c)}} \|P^{(c)}f\|^2 \quad (2)$$

$\Omega^{(c)}$ の相関行列 $R_{ff}^{(c)}$ は、

$$R_{ff}^{(c)} = E_{f \in \Omega^{(c)}} ff^T \quad (3)$$

によって定義される。 $R_{ff}^{(c)}$ の固有ベクトルを $\lambda_1^{(c)} \geq \lambda_2^{(c)} \geq \dots \geq \lambda_N^{(c)}$ で表し、それに対応する固有ベクトルを $\phi_1^{(c)}, \phi_2^{(c)}, \dots, \phi_N^{(c)}$ で表す。ここで、一般性を失うことなく、固有ベクトルは正規直交基底をなしているものと仮定することができる。 n を N 以下の自然数とする。このとき、PCA によって与えられる部分空間法のための n 次元部分空間 $S^{(c)}$ は、 $\phi_1^{(c)}, \phi_2^{(c)}, \dots, \phi_n^{(c)}$ によって張られる。

式 (2) を最大にする $P^{(c)}$ は、ランク一定の条件のもとで、

$$E_{f \in \Omega^{(c)}} \|X^{(c)}f - f\|^2 \quad (4)$$

を最小にする $X^{(c)}$ と本質的には同じものになる。すなわち、ランクが一定という条件の元で、平均 2 乗的にカテゴリのパターンを最良に近似できるものとなり、カテゴリの特徴を抽出していると考えることができる。

しかしながら、式 (4) には他カテゴリのことが考慮されていないため、この $P^{(c)}$ は他のカテゴリの特徴も抽出する可能性がある。従って、他のカテゴリのパターンベクトル g に対しても、 $\|P^{(c)}g\|$ の値が大きくなる可能性がある。この問題を解決するために、RKL 変換法が提案された。RKL 変換は、ランク一定の元で、

$$E_{f \in \Omega^{(c)}} \|X^{(c)}f - f\|^2 + \alpha E_{l \neq c} E_{g \in \Omega^{(l)}} \|X^{(c)}g\|^2 \quad (5)$$

を最小にする行列 $X^{(c)}$ として定義される。ここで、 $E_{l \neq c}$ は、 $\Omega^{(c)}$ 以外のカテゴリの出現確率による期待値であり、 α は正の定数 (パラメータ) である。

式 (5) の第 1 項は、 $X^{(c)}$ が抽出した $\Omega^{(c)}$ の特徴による近似誤差の 2 乗平均を表している。式 (5) の第 2 項は、 $X^{(c)}$ が抽出した $\Omega^{(c)}$ 以外のカテゴリの特徴量の 2 乗平均を表している。 $X^{(c)}$ が、自分のカテゴリの特徴は正確に抽出し、他カテゴリの特徴を抽出しないようにするためには、この 2 つの項を小さくすれば良い。従って、式 (5) では、その 2 項のどちらにどれだけ重点をおくか決めるためのパラメータ α を第 2 項にかけて和をとったものとした。

RKL 変換と式が類似したものに、ランク低減 WF が提案されている [19]。この場合、式 (5) の f と g は、それぞれ、同じ観測信号の信号成分と雑音成分を表すが、RKL 変換の場合は、異なるカテゴリの全く別のパターンベクトルを表している。

RKL 変換法によって、他カテゴリとの境界付近のパターンに対する認識率の向上が期待されたが、向上の幅はあまり大きくなかった。平均的な値、例えば、 $E_{f \in \Omega^{(c)}} \|X^{(c)} f\|^2$ と $E_{l \neq c} E_{g \in \Omega^{(l)}} \|X^{(c)} g\|^2$ の比などは向上している。しなしながら、誤認識するパターンは、カテゴリの平均を使った評価では捉えることが難しいため、認識率があまり改善しなかったものと考えられる。

3. カーネル Wiener フィルタ法 (KWF 法)

カーネル法では、非線型写像 Φ を使って、もとのパターンや信号が存在する空間から、特徴空間と呼ばれるヒルベルト空間に写像し、その空間で処理を行う手法である。このとき、

$$k(x, y) = \langle \Phi(x) \Phi(y) \rangle \quad (6)$$

を満たす、カーネル関数 $k(x, y)$ が与えられていれば、ヒルベルト空間での処理をカーネル関数を使った処理に書き換えることができる。PCA にカーネル法を適用した KPCA を使った部分空間法が提案され、その有効性が確認されている。

本稿では、カーネル Wiener フィルタを用いたパターン認識に関して説明する。Wiener フィルタは、劣化などを受けた観測信号から、原信号を推定するための線形復元フィルタである。原信号空間および観測画像空間から、それぞれの特徴空間への写像を、 Φ_S および Φ_O とおく。学習信号として、 L 個の原信号と観測信号の組 $\{(x_l, y_l)\}_{l=1}^L$ が与えられているとすれば、KWF は、

$$\frac{1}{L} \sum_{i=1}^L \|X \Phi_O(y_l) - \Phi_S(x_l)\|^2 \quad (7)$$

を最小にする線形作用素 X として与えられる。

これをパターン認識に応用する。カテゴリ数を C とし、各カテゴリ $\Omega^{(c)}$ ($c = 1, 2, \dots, C$) に対して、標準ベクトル s_c を定める。 L 個の学習パターンベクトル f_l ($l = 1, 2, \dots, L$) と、そのカテゴリ $\Omega_{c(l)}$ が与えられているときに、

$$\frac{1}{L} \sum_{i=1}^L \|X \Phi_O(f_l) - \Phi_S(s_{c(l)})\|^2 \quad (8)$$

を最小にする作用素 X を使ってパターン認識を行う。未知パターンベクトル x に対する認識結果は、

$$D(x) = \underset{c=1,2,\dots,C}{\operatorname{argmin}} \|Xx - s_c\| \quad (9)$$

によって与えられる。

正則化は、パターンの微小な変動に対して結果が大きく変化しないようにし、オーバーフィッティングを防ぐための有効な手段である。この正則化を特徴空間ではなく、元の空間の距離に従って行うことを考える。そのため、学習パターンベクトル f_l に仮想的な雑音 e_l が加算されるものとする。この e_l は f_l と独立で、 e_l は無相関であるものとする。このとき、

$$\frac{1}{L} \sum_{i=1}^L E_{e_l} \|X \Phi_O(f_l + e_l) - \Phi_S(s_{c(l)})\|^2 \quad (10)$$

を最小にする線形作用素 X を考える。ここで、 E_{e_l} は、 e_l に関する期待値を表している。

通常のカーネル法では、 e_l に対しても標準平均を計算する必要がある。しかしながら、その方法では計算量が膨大になるため、 e_l の大きさが小さいと考え、 Φ_O をその 1 次の Taylor 級数による近似に置き換え、式 (10) を評価する。 Φ_O を微分したものの内積は、カーネル関数を微分して得ることができる。このとき、 Φ_O の 1 次微分したものどうしの内積は、カーネル関数の 2 次微分になる。この項を導入すると、逆行行列を計算する行列の次数が正則化しない KWF のものよりも大幅に増加するため、省略して解を求めている。

表 1 に、13 個の 2 クラス標準データセット [20] に対する、KWF 法の認識実験の結果を示す。KWF 法には、元空間の距離に従った正則化法を導入している。パラメータは、5 fold のクロスバリデーションによって求めている。表の中で太文字は、SVM, AdaBoost Regression, KWF の中で誤認識率が最も低いもの示している。また、KWF における*は、優位水準 5% の t 検定で、SVM に対して優位にあるものを示している。表 1 より、KWF 法の有効性が確認できる。

4. 重み付き相関行列による局所部分空間法

はじめに、重み付き相関行列による局所部分空間法の最も簡単なものについて説明する。 x を入力パターンベクトルとする。カテゴリ $\Omega^{(c)}$ の重み付き相関行列 $R^{(c)}(x)$ を、

$$R^{(c)}(x) = E_{f \in \Omega^{(c)}} \langle x, f \rangle f f^T \quad (11)$$

で定義する。パターンベクトルがノルムで正規化されているとすれば、重み $\langle x, f \rangle$ は、 f と x の距離に近い方が大きくなる。従って、式 (11) の中で、平均される学習パターンに対する重みは、その学習パターンが入力パターンに近いほど大きくなる。カテゴリ $\Omega^{(c)}$ を表す部分空間 $S^{(c)}(x)$ は、 $R^{(c)}(x)$ から PCA によって得られた固有ベクトルで張られる空間とする。 $P^{(c)}(x)$ を部分空間 $S^{(c)}(x)$ への正射影行列とする。識別結果は、

$$D(x) = \underset{c=1,2,\dots,C}{\operatorname{argmax}} \|P^{(c)}(x)x\| \quad (12)$$

によって与えられる。 $P^{(c)}(x)$ は x に依存するため、全体としては非線型な識別法となる。

ベクトル f , 行列 R , 3 階のテンソル T , 4 階のテンソル U

表 1 Error rate of KWF-1 and KWF-2'

	Data set Name	error rate SVM	error rate AdaBoost Reg	error rate KWF'
1	banana	11.53 ± 0.66	10.85 ± 0.42	10.29 ± 0.39 *
2	breast cancer	26.04 ± 4.47	26.51 ± 4.47	24.45 ± 4.29 *
3	diabetis	23.53 ± 1.73	23.79 ± 1.80	22.97 ± 1.83 *
4	flare solar	32.43 ± 1.82	34.20 ± 2.18	33.18 ± 1.75
5	german	23.61 ± 2.07	24.34 ± 2.08	23.63 ± 2.24
6	heart	15.95 ± 3.26	16.47 ± 3.51	15.46 ± 3.30
7	image	2.96 ± 0.60	2.67 ± 0.61	2.69 ± 0.53
8	ringnorm	1.66 ± 0.12	1.58 ± 0.12	1.62 ± 0.10 *
9	splice	10.88 ± 0.66	9.50 ± 0.65	10.83 ± 0.67
10	thyroid	4.80 ± 2.19	4.55 ± 2.19	3.61 ± 2.05 *
11	titanic	22.42 ± 1.02	22.64 ± 1.20	22.00 ± 0.75 *
12	twonorm	2.96 ± 0.23	2.70 ± 0.24	2.34 ± 0.11 *
13	waveform	9.88 ± 0.43	9.79 ± 0.81	9.50 ± 0.35 *

* : superior to SVM by t-test of significance level 5%

に対して、第 i , (i, j) , (i, j, k) , and (i, j, k, l) 成分を、それぞれ、 f_i , R_{ij} , T_{ijk} , U_{ijkl} で表す。

認識時における $R^{(c)}(\mathbf{x})$ の計算を高速化するために、3 次相関テンソル $R_{fff}^{(c)}$ を、

$$(R_{fff}^{(c)})_{ijk} = E_{f \in \Omega(c)} f_i f_j f_k \quad (13)$$

によって与える。このとき、 $R^{(c)}(\mathbf{x})$ は

$$R^{(c)}(\mathbf{x})_{ij} = \sum_{k=1}^N (R_{fff}^{(c)})_{ijk} x_k \quad (14)$$

によって計算することができる。式 (14) を使えば、認識時に $R^{(c)}(\mathbf{x})$ を求めるために、学習パターンごとに計算する必要がなくなる。

計算量を削減したり、効率的な重みを構成するために、部分空間を張るためのパターンベクトルと、重みのためのパターンベクトルを分けることができる。また、部分空間法を行うために、前処理等を行うことも多い。それを一般的に表すために、2 つの変換 $p(\cdot)$ および $q(\cdot)$ を考える。前者を部分空間を構成するベクトルを生成する変換、後者を相関行列の重みのためのベクトルを生成するための変換とする。 N および M を、それぞれ、 $p(\cdot)$ および $q(\cdot)$ の像の次元とする。今回、扱う相関行列は、 $n = 1, 2$ に対して、

$$(R^{(c)}(\mathbf{x}) = E_{f \in \Omega(c)} (\langle \mathbf{q}(\mathbf{x}), \mathbf{q}(\mathbf{f}) \rangle + a) \mathbf{p}(\mathbf{f}) \mathbf{p}(\mathbf{f})^T \quad (15)$$

である。

$$\left(R_{p(f)p(f)}^{(c)} \right)_{ij} = E_{f \in \Omega(c)} p(\mathbf{f})_i p(\mathbf{f})_j \quad (16)$$

$$\left(R_{p(f)p(f)q(f)}^{(c)} \right)_{ijk} = E_{f \in \Omega(c)} p(\mathbf{f})_i p(\mathbf{f})_j q(\mathbf{f})_k \quad (17)$$

$$\left(R_{p(f)p(f)q(f)q(f)}^{(c)} \right)_{ijkl} = E_{f \in \Omega(c)} p(\mathbf{f})_i p(\mathbf{f})_j q(\mathbf{f})_k q(\mathbf{f})_l \quad (18)$$

とおく。式 (15) で $n = 1$ のときは、

$$R^{(c)}(\mathbf{x})_{ij} = \sum_{k=1}^M \left(R_{p(f)p(f)q(f)}^{(c)} \right)_{ijk} q(\mathbf{x})_k$$

$$+ a \left(R_{p(f)p(f)}^{(c)} \right)_{ij} \quad (19)$$

によって計算できる。同様に、式 (15) で $n = 2$ のときは、

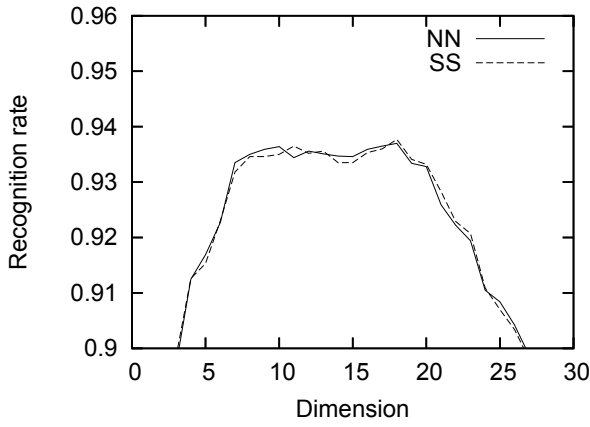
$$\begin{aligned} & R^{(c)}(\mathbf{x})_{ij} \\ &= \sum_{k=1}^M \sum_{l=1}^M \left(R_{p(f)p(f)q(f)q(f)}^{(c)} \right)_{ijkl} q(\mathbf{x})_k q(\mathbf{x})_l \\ &+ 2a \sum_{k=1}^M \left(R_{p(f)p(f)q(f)}^{(c)} \right)_{ijk} g(\mathbf{x})_k + a^2 \left(R_{p(f)p(f)}^{(c)} \right)_{ij} \end{aligned} \quad (20)$$

によって計算できる。 $q(\cdot)$ に、全カテゴリーの PCA などを適用し、写像して得られるベクトルの次元 M を小さくすれば、計算量を削減することができる。また、全カテゴリーに対するマハラノビス距離を使うなどすれば、パターンの分布を均一化させ、重みの作用が効果的になると考えられる。

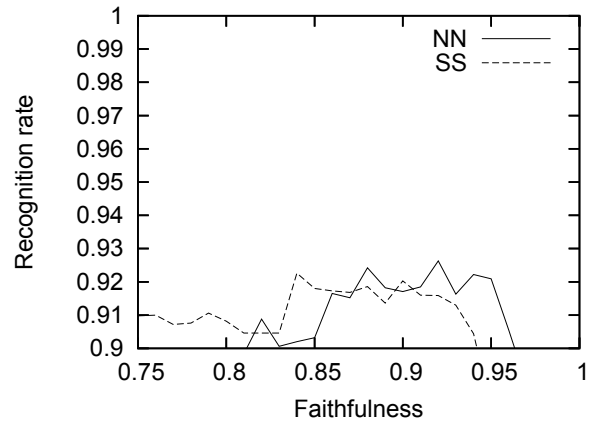
計算量に関して検討する。計算量のオーダーが L のとき、 $O(L)$ と表す。提案手法で、認識時の処理において計算量が多い部分は、重み付き相関行列の計算と固有ベクトルの計算である。重み付き相関行列のための計算量は、 $n = 1, 2$ のとき、それぞれ、 $O(MN^2)$, $O(M^2N^2)$ である。また、固有ベクトルの計算量は $O(N^3)$ である。従って、 $n = 1$ の場合は、 $M = N$ の場合でも、固有ベクトルの計算量と同程度であり、固有ベクトル計算を行う識別法ならば本手法が適用可能である。 $n = 2$ の場合は、パターンベクトルの次元が特に低い場合を除いて、次元圧縮などによって $M \ll N$ とする必要があると考えられる。

5. 計算機実験

通常の部分空間法と重み付き相関行列による局所部分空間法を比較する計算機実験を行う。元となるデータとして、手書き数字のデータセットである MNIST を用いる [21]。データ数は、学習用 60,000、およびテスト用 10,000 で、それぞれ、(高次) 相関、および認識率を求めるために用いる。今回は、両識別法の比較が目的であるため、エッジ特徴抽出などの特別な特徴抽出法は用いていない。1 つのパターンは 784 次元 (28×28 画素) のベクトルからなる。ここでは計算量削減のため、上下、

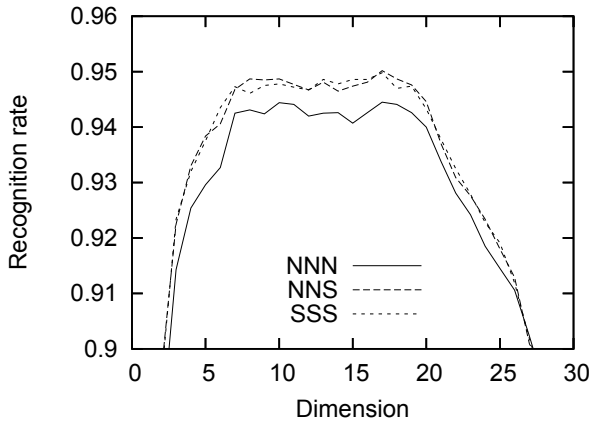


(a) Dimension

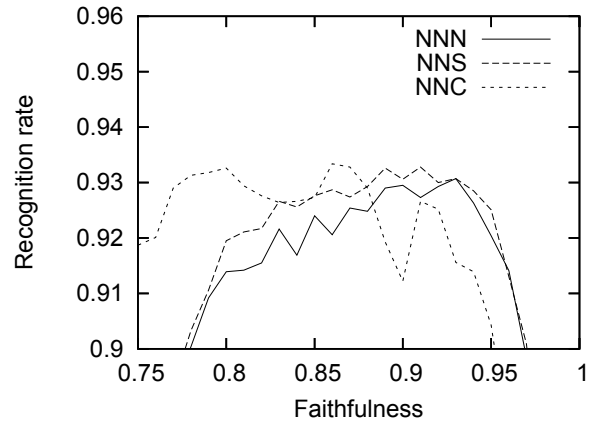


(b) Faithfulness

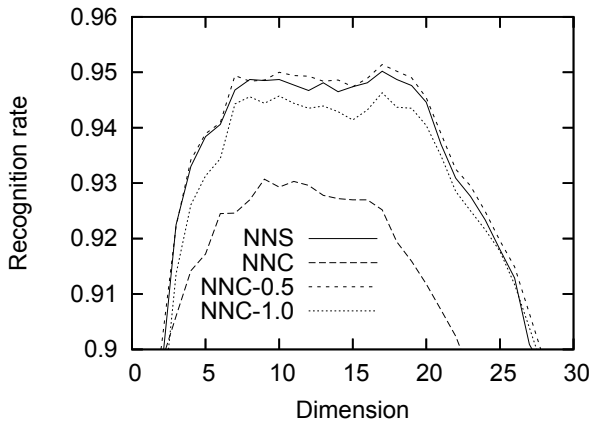
图 1 Subspace method



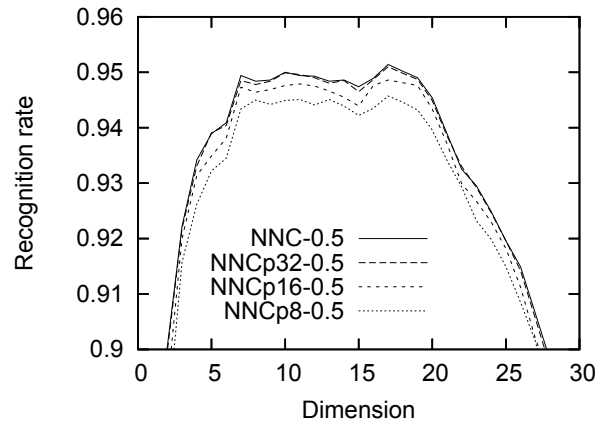
(a) Dimension



(b) Faithfulness

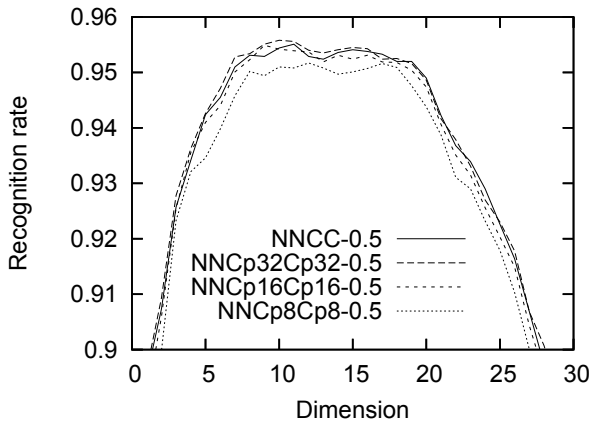


(c) Parameter

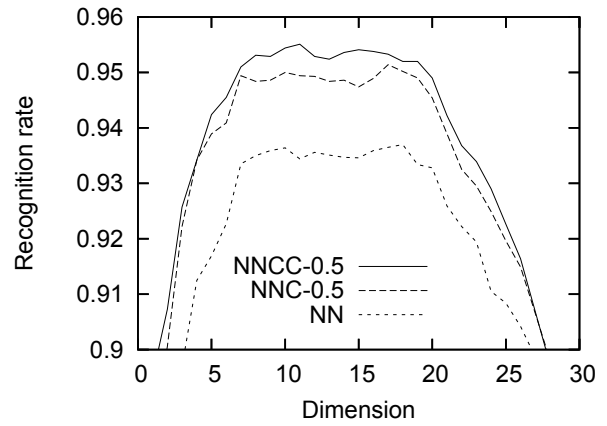


(d) Dimension of $g(\mathbf{f})$

图 2 Local subspace method with third order correlations



(a) With fourth order correlations



(b) Summary

図3 Local subspace method with fourth order correlations and summary

表2 Data type and parameters in experiments

Symbol	$p(x)$	n	$q(x)$	a
'NN'	'N'	0	N.A.	N.A.
'SS'	'S'	0	N.A.	N.A.
'NNN'	'N'	1	'N'	0
'NNS'	'N'	1	'S'	0
'NNC'	'N'	1	'C'	0
'NNC-0.5'	'N'	1	'C'	0.5
'NNC-1.0'	'N'	1	'C'	1.0
'NNC8-0.5'	'N'	1	'C8'	0.5
'NNC16-0.5'	'N'	1	'C16'	0.5
'NNC32-0.5'	'N'	1	'C32'	0.5
'NNCC-0.5'	'N'	2	'C'	0.5
'NNC8C8-0.5'	'N'	2	'C8'	0.5
'NNC8C16-0.5'	'N'	2	'C16'	0.5
'NNC32C32-0.5'	'N'	2	'C32'	0.5

左右の2つのラインを切り取り、 3×3 画素の値の単純平均を取ることによって、64次元のベクトルに変換している。

上記のようにして64次元に変換したパターンベクトルをノルムで正規化したデータの種別を'N'で表す。種別'N'のデータのそれぞれのベクトルの成分の平均を計算し、全ての要素からその平均を引き、ノルムで正規化したデータの種別を'S'で表す。種別'S'のデータに対して、全てのカテゴリの全てのパターンの平均ベクトルを計算し、種別'S'のそれぞれのパターンベクトルからその平均ベクトルを引き、ノルムで正規化したデータの種別を'C'で表す。また、種別'C'の全てのカテゴリの全てのパターンに対するPCAを計算し、ベクトルを k 次元に次元圧縮したデータの種別を'Ck'で表す。

実験結果の種別を示す記号と、使用したデータの種別とパラメータをまとめたものを表2に示す。この表で、例えば、実験結果NNCp16Cp16-0.5は、部分空間を構成するデータとして種別'N'を、式(15)の重みの次数 $n=2$ を、重みを計算するデータとして種別'C16'を、パラメータ a として0.5を使用していることを示している。

図1は、通常の部分空間法の実験結果である。図1(a)およ

び(b)は、それぞれ、全てのカテゴリで部分空間の次元を一定にしたもの、および、全てのカテゴリで寄与率が一定になるように各カテゴリの部分空間の次元を定めたものである。これらの図の結果の種別'NN'は、データの種別'N'に対するものであり、'SS'は、データの種別'S'に対するものである。最良の結果93.64%は、次元を11に固定した結果の種別'SS'から得られている。

図2は、式(15)で $n=1$ のときの提案手法の結果である。2(a)と(b)は、それぞれ、次元一定と寄与率一定の場合の結果である。また、結果の'NNN'および'SSS'は、それぞれ、 $p(x)$ と $q(x)$ の両方にデータの種別'N'および'S'を使ったものである。'NNS'は、 $p(x)$ および $q(x)$ に、それぞれ、データの種別'N'および'S'を使ったものである。最良の結果95.02%は、次元を17に固定した結果の種別'NNS'から得られている。

次元を固定した方が良い結果が得られているため、以下の実験では次元を固定した場合だけについて行うことにする。また、 $p(x)$ を変更してもあまり結果が変わらないため、 $p(x)$ には、データの種別'N'だけを使うことにする。

図2(c)は、 $q(x)$ にデータの種別'C'を使ったものである。式(15)の重みの値が負になることがあるため、正の定数 a を加算する。結果の種別'NNC'、'NNC-0.5'、および、'NNC-1.0'は、それぞれ、 $a=0.0, 0.5, 1.0$ のときの結果を表している。最良の結果95.15%は、次元を17に固定した結果の種別'NNC-0.5'から得られている。以下では、 $a=0.5$ に固定して実験を行う。

図2(d)は、次に計算量を削減するために、重みのためのベクトル $q(x)$ に、PCAによって次元削減を行った場合の結果を示す。結果の種別'NNC-0.5'、'NNC32-0.5'、'NNC16-0.5'、および、'NNC8-0.5'は、それぞれ、次元を圧縮しない場合、32, 16, 8次元に圧縮した場合を表している。実験結果より、次元を圧縮すると共に認識率は低下するが、32次元、すなわち、1/2程度に次元を圧縮しても認識率が大きく変化しないことがわかる。

図3(a)は、式(15)で $n=2$ のときの提案手法の結果である。この場合、4次相関が必要となり計算量が非常に大きくなる。そのため、重みのためのベクトル $q(x)$ に次元圧縮を行った場合の結果も示す。結果の種別'NNCC-0.5'、'NNC32C32-

0.5', 'NNC16C16-0.5', および, 'NNC8C8-0.5' は, それぞれ, 次元を圧縮しない場合, 32, 16, 8 次元に圧縮した場合を表している. 実験結果より, 次元を圧縮すると認識率は低下するが, 16 次元, すなわち, 1/4 程度に次元を圧縮しても認識率が大きく変化しないことがわかる.

図 3(b) は, 実験結果をまとめたものである. この図より, 重み付き相関行列によって, 部分空間法の認識率を向上させることができることがわかる.

6. ま と め

本稿では, 筆者らが提案してきた部分空間法に関するパターン認識法として, RKL 変換法, KWF 法を紹介した. さらに, 重み付き相関行列による局所部分空間法を提案し, その有効性を計算機実験によって確認した. 本手法は, クラスタリングのような複雑な処理を行うことなく, 部分空間法の認識率を向上させることができる. 従って, パターンベクトルの次元があまり高くないときに, 部分空間法の認識率を手軽に向上させたいときに適していると考えられる. 今後の課題として, 重みを計算するための次元を削減法, および, クラスタリングしたサブカテゴリーに対する本手法の有効性に関して研究する必要がある.

文 献

- [1] 飯島: “パターン認識”, 日刊工業新聞社, 東京 (1969).
- [2] 飯島: “視覚情報の基礎理論—パターン認識問題の源流—”, コロナ社, 東京 (1999).
- [3] S. Watanabe and N. Pakvasa: “Subspace method in pattern recognition”, Proc. 1st International Joint Conference on Pattern Recognition, Washington DC, pp. 25–32 (1973).
- [4] J. Laaksonen: “Local subspace classifier”, Proceedings of the 7th International Conference on Artificial Neural Networks, pp. 637–642 (1997).
- [5] 堀田, 喜安, 宮原: “局所部分空間法と変形不変性を用いた画像パターン分類”, 部分空間法研究会 Subspace2006, pp. 111–118 (2006).
- [6] H. Cevikalp, D. Larlus, M. Douze and F. Jurie: “Local subspace classifier: Linear and nonlinear approaches”, 2007 IEEE Workshop on Machine Learning for Signal Processing, pp. 57–62 (2007).
- [7] J. Laaksonen, M. Aksela, E. Oja and J. Kangas: “Adaptive local subspace classifier in on-line recognition of handwritten character”, Proceedings of International Joint Conference on Neural Networks, Vol. 4, pp. 2812–2815 (1999).
- [8] 福井, 山口, 鈴木, 前田: “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”, 電子情報通信学会論文誌 (D-II), **J82-D-II**, 4, pp. 613–620 (1999).
- [9] 坂野, 武川, 中村: “核非線形相互部分空間法による物体認識”, 電子情報通信学会論文誌 (D-II), **J84-D-II**, 8, pp. 1549–1556 (2001).
- [10] 福井, 山口: “カーネル非線形制約相互部分空間法による物体認識”, 電子情報通信学会論文誌 (D-II), **J88-D-II**, 8, pp. 1349–1356 (2005).
- [11] 前田, 村瀬: “カーネル非線形部分空間法によるパターン認識”, 信学論 (D-II), **J82-D-II**, 4, pp. 600–612 (1999).
- [12] 津田: “ヒルベルト空間における部分空間法”, 信学論 (D-II), **J82-D-II**, 4, pp. 592–599 (1999).
- [13] V. N. Vapnik: “Statistical Learning Theory”, Wiley, New-York (1998).
- [14] Y. Yamashita and H. Ogawa: “Relative Karhunen-Loève transform”, IEEE Trans. on Signal Processing, **44**, 2, pp.

371–378 (1996).

- [15] 池野, 山下, 小川: “相対 KL 変換法によるパターン認識”, 信学論 (D-II), **J80-D-II**, 2, pp. 541–547 (1997).
- [16] 鷲沢, 正田, 田中, 山下: “パターン認識のための相対 KL 変換法の高精度化”, 情報科学技術フォーラム 2002 一般講演論文集, I-48, pp. 95–96 (2002).
- [17] Y. Washizawa and Y. Yamashita: “non-linear Wiener filter in reproducing kernel Hilbert space”, Proceedings of 18th International Conference on Pattern Recognition, Vol. 1, pp. 967–970 (2006).
- [18] H. Yoshino and Y. Yamashita: “Pattern recognition by kernel Wiener filter”, Proceedings of Signal Processing, Pattern Recognition, and Applications, pp. 7–12 (2008).
- [19] L. Scharf: “The SVD and reduced rank signal processing”, Signal Processing, **25**, 2, pp. 113–133 (1991).
- [20] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda and B. Schölkopf: “An introduction to kernel-based learning algorithms”, IEEE Transactions on Neural Networks, **12**, 2, pp. 181–201 (2001).
- [21] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, **86**, 11, pp. 2278–2324 (1998).