

サンプル数が少ない状況下における 正規直交判別ベクトル法の検討

宮本 貴宣[†] 浜本 義彦[†]

特徴抽出系の設計法の1つである正規直交判別ベクトル法では、平均ベクトルと共分散行列を有限個の訓練サンプルから推定する必要がある。訓練サンプル数が少ない状況下では、標本平均ベクトルと標本共分散行列に推定誤差が生じ、これらの推定誤差が特徴空間を劣化させる原因となる。ここでは、標本共分散行列の推定誤差の低減についてのみ議論する。本論文では、共分散行列の推定法として、テプリッツ法、正則化法、Hoffbeckらの手法、摂動法を比較評価する。さらに従来手法との比較を行い、標本共分散行列の推定誤差を低減することの効果を検討する。

Study on the Orthonormal Discriminant Vector Method in Small Sample Size Situations

TAKANOBU MIYAMOTO[†] and YOSHIHIKO HAMAMOTO[†]

The orthonormal discriminant vector method requires the estimation of a mean vector and covariance matrix. If possible, one should estimate the mean vector and the covariance matrix by using a large number of training samples. However, in practice, these must be estimated from a finite number of training samples. The estimation error due to the finite training samples leads to the degradation of the performance of the resulting orthonormal discriminant vector method. In this paper, we discuss the estimators of the covariance matrix for the orthonormal discriminant vector method. We compare the performance of the orthonormal discriminant vector method based on several estimators such as the Toeplitz approximation, regularized method, Hoffbeck's method and perturbation method, particularly in the small training sample size situations. The effect of reducing the estimation error of the covariance matrix is discussed in terms of the error rate.

1. はじめに

パターン認識における特徴抽出法の1つとして、正規直交判別ベクトル法が提案されている^{1),2)}。これは、特徴軸の直交性のもとでFisher評価関数を最大にする特徴軸を求める手法である。この手法では、平均ベクトルと共分散行列とを有限個の訓練サンプルから推定する必要がある。訓練サンプル数が少ない状況下では、標本平均ベクトルと標本共分散行列に推定誤差が生じ、これらの推定誤差が特徴空間を劣化させる原因となる。ここでは、標本共分散行列の推定誤差の低減についてのみ議論する。

本論文では、共分散行列の推定法として、テプリッツ法、正則化法、Hoffbeckらの手法、摂動法を比較評価し、訓練サンプル数が少ない状況下における正規直

交判別ベクトル法を検討する。

以下、2章で、訓練サンプル数が少ない状況下において、標本平均ベクトルと標本共分散行列の推定誤差が特徴空間に与える影響について検討する。3章で、4つの共分散行列の推定法を述べ、それらの正規直交判別ベクトル法への適用を議論するとともに、それらの諸性質を明らかにする。4章で、実データと人工データを用いて、誤識別率とCPU時間の両面から共分散行列の推定法の比較実験を行う。5章で、結論を述べる。

2. 準備

本論文では、 n 次元ユークリッド空間 R^n をパターン空間とし、その各点をパターンと見なす。 \mathbf{x}^T は \mathbf{x} の転置を表す。以下では、次の記号を用いる。

m : クラス数

N_i : クラス i の訓練サンプル数

N_T : 全訓練サンプル数

[†] 山口大学工学部

Faculty of Engineering, Yamaguchi University

p_i : クラス i の事前確率
 \mathbf{x}_{ij} : クラス i の j 番目の訓練サンプル
 $\boldsymbol{\mu}_i$: クラス i の標本平均ベクトル
 Σ_i : クラス i の標本共分散行列
 $\boldsymbol{\tau}_r$: 正規直交判別ベクトル法により得られる r 番目の特徴軸
 I_n : n 次の単位行列

2.1 正規直交判別ベクトル法

正規直交判別ベクトル法は、特徴軸の直交性の下で Fisher 評価関数を最大にする特徴軸を求める手法である。Fisher 評価関数は式 (1) で定義される。

$$J(\boldsymbol{\tau}) = \frac{\boldsymbol{\tau}^T B \boldsymbol{\tau}}{\boldsymbol{\tau}^T W \boldsymbol{\tau}} \quad (1)$$

ここで、 W はクラス内共分散行列、 B はクラス間共分散行列であり、それぞれ以下に定義される。

$$W = \sum_{i=1}^m p_i \Sigma_i \quad (2)$$

$$B = \sum_{i=1}^{m-1} \sum_{j=i+1}^m p_i p_j (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \quad (3)$$

ここで、 $\boldsymbol{\mu}_i$, Σ_i は、それぞれ訓練サンプルを用いて推定されるクラス i の標本平均ベクトル、標本共分散行列である。式 (1) を最大にする特徴軸は、式 (4) の固有値問題を解いて得られる最大固有値に対応する固有ベクトルである。得られた特徴軸は第 1 特徴軸 $\boldsymbol{\tau}_1$ となる。

$$W^{-1} B \boldsymbol{\tau} = \lambda \boldsymbol{\tau} \quad (4)$$

一般に、第 r 特徴軸は次の手順で求めることができる。

(1) 直交補空間 S^{n-r+1} を生成する $n-r+1$ 個の n 次元正規直交基底ベクトル $\boldsymbol{\nu}_s$ ($r \leq s \leq n$) を式 (5) により求める。

$$\boldsymbol{\nu}_s = c_s \left(I_n - \sum_{t=1}^{s-1} \boldsymbol{\nu}_t \boldsymbol{\nu}_t^T \right) \boldsymbol{n}_s \quad (5)$$

ここで、 $\boldsymbol{\nu}_t = \boldsymbol{\tau}_t$ ($1 \leq t \leq r-1$)、 \boldsymbol{n}_s は $\boldsymbol{\nu}_t$ ($1 \leq t \leq s-1$) と一次独立な任意の n 次元ベクトル、 c_s は $\|\boldsymbol{\nu}_s\| = 1$ とするための正規化定数である。

(2) 正規直交基底ベクトル $\boldsymbol{\nu}_s$ を用いて

$$P_{r-1} = [\boldsymbol{\nu}_r \boldsymbol{\nu}_{r+1} \cdots \boldsymbol{\nu}_n] \quad (6)$$

なる $n \times (n-r+1)$ 行列 P_{r-1} を構成する。

(3) 直交補空間 S^{n-r+1} における W と B を求める。

$$W_r = P_{r-1}^T W P_{r-1} \quad (7)$$

$$B_r = P_{r-1}^T B P_{r-1} \quad (8)$$

(4) $W_r^{-1} B_r$ の最大固有値に対応する $n-r+1$ 次

元固有ベクトル $\boldsymbol{\tau}^{(r)}$ を式 (9) により n 次元ベクトルに変換することで第 r 特徴軸 $\boldsymbol{\tau}_r$ を得る。

$$\boldsymbol{\tau}_r = P_{r-1} \boldsymbol{\tau}^{(r)} \quad (9)$$

2.2 推定誤差の影響

パラメトリックな特徴抽出系や識別系では、有限個の訓練サンプルを用いて、標本平均ベクトル $\boldsymbol{\mu}_i$ 、標本共分散行列 Σ_i を求め、それらを母平均ベクトル、母共分散行列の代わりに用いる。実際には訓練サンプル数が有限であるがゆえに、標本平均ベクトルと標本共分散行列には推定誤差が生じる。パラメトリックな識別系においては、標本平均ベクトルの推定誤差よりも標本共分散行列の推定誤差の方が識別器の識別能力に大きく影響することが知られているため、標本共分散行列の推定誤差を低減するべく、多くの手法が提案されている^{3)~5)}。一方、パラメトリックな特徴抽出系においては、いずれの推定誤差がより強い影響を与えるかは十分に明らかにされていない。そこで、まず、標本平均ベクトルと標本共分散行列のいずれの推定誤差が特徴抽出系を設計のうえで注意すべきかを明らかにする。

平均ベクトルと共分散行列の 4 通りの組合せに対して、それぞれ特徴抽出系を設計する。すなわち、標本平均ベクトルと標本共分散行列を直接用いるのが従来の正規直交判別ベクトル法であり、母平均ベクトルと母共分散行列を用いるのが推定誤差の存在しない場合を表す。標本平均ベクトルと母共分散行列を用いる場合は、標本平均ベクトルの推定誤差の影響を調べることになる。また、母平均ベクトルと標本共分散行列を用いる場合は、標本共分散行列の推定誤差の影響を調べることになる。ここでは、特徴抽出系の影響のみを調べるために、識別系には母平均ベクトル、母共分散行列を用いる。得られる特徴空間の性能を調べるために、特徴空間上でテストサンプルに対する誤識別率を推定する。このテストサンプルは訓練サンプルとは独立とする。訓練サンプル数については、Jain らはパターン空間の次元数に対する訓練サンプル数の比が少なくとも 5 以上あるべきであると指摘している⁶⁾。本論文では、訓練サンプル数が少ない状況下を想定するため、上述の比が 1 未満から十分多数と考えられる 25 までの範囲で実験した。一方、テストサンプル数については、その影響を受けないように、各クラス 1000 個のテストサンプルを用いた。まず、実験で用いた 2 つの人工データ ($I-\Lambda$ データ, Ness データ) を説明する。

$I-\Lambda$ データは、以下の母平均ベクトル $\boldsymbol{\mu}_i^*$ 、母共

分散行列 Σ_i^* ($i = 1, 2$) の正規分布に従う 8 次元, 2 クラスのデータで, Fukunaga らにより多くの論文で標準データとして用いられている. このデータの Bayes 誤識別率は 1.9% である.

$$\begin{aligned} \mu_1^* &= \mathbf{o}, \Sigma_1^* = I_8 \\ \mu_2^* &= [3.86 \ 3.10 \ 0.84 \ 0.84 \ 1.64 \ 1.08 \ 0.26 \\ &\qquad\qquad\qquad 0.01]^T \\ \Sigma_2^* &= \text{diag}[8.41, 12.06, 0.12, 0.22, 1.49, 1.77, \\ &\qquad\qquad\qquad 0.35, 2.73] \end{aligned}$$

Ness データは以下の母平均ベクトル $\hat{\mu}_i$, 母共分散行列 $\hat{\Sigma}_i$ ($i = 1, 2$) の正規分布に従う n 次元, 2 クラスのデータである⁷⁾. この Ness データは, これまで多くの識別実験で用いられてきた^{8)~10)}. ここで, Δ は Mahalanobis 距離を表すパラメータである. $n = 8$ の場合, $\Delta = 2, 6$ の Bayes 誤識別率はそれぞれ 18.64%, 0.95% となる.

$$\begin{aligned} \hat{\mu}_1 &= \mathbf{o}, \hat{\Sigma}_1 = I_n \\ \hat{\mu}_2 &= [\Delta/2, 0, \dots, 0, \Delta/2]^T \\ \hat{\Sigma}_2 &= \begin{bmatrix} I_{n/2} & 0 \\ 0 & \frac{1}{2}I_{n/2} \end{bmatrix} \end{aligned}$$

実験を以下の条件下で行った.

データ : I- Λ データ,
Ness データ

パターン空間の次元数 : $n = 8$
特徴空間の次元数 : 2 次元
訓練サンプル数 : 各クラス 7, 8, 16, 24,
32, 48, 100, 200 個
テストサンプル数 : 各クラス 1000 個
識別器 : Fisher の線形識別器
独立試行回数 : 100 回
 Δ : 2, 6

I- Λ データ, Ness データ上の実験結果をそれぞれ図 1, 図 2 に示す. 横軸は訓練サンプル数, 縦軸は Fisher の線形識別器による真の誤識別率を表す. ここで真の誤識別率とは, 訓練サンプルとは独立に生成した大量のテストサンプルに対する誤識別率であり, 将来入力されるパターンに対する識別性能を表すものになる. 図 2 (a) から, クラス間の重なりが大きいつきは, クラスが互いに接近しているため, 平均ベクトルの情報が相対的に重要となり, 標本平均ベクトルの推定誤差の低減がより効果的であることが分かる. 一方, 図 1, 図 2 (b) から, クラス間の重なりが小さいときは標本共分散行列の推定誤差の低減が相対的に効果的となることが分かる. さて, クラス間の重なりが大きい場合, 線形の特徴抽出系により次元圧縮することには問題があると思われる. 一般に, 次元圧縮によりク

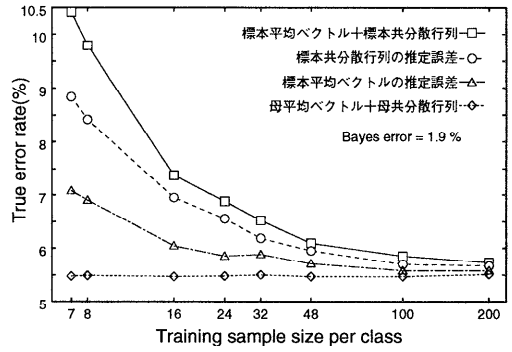
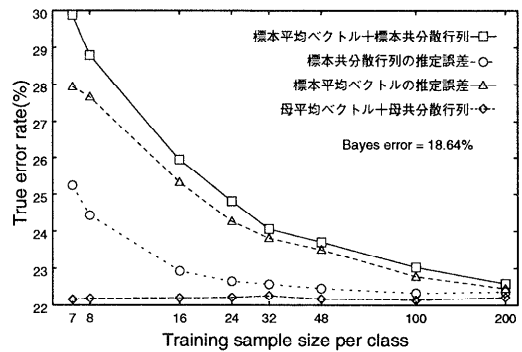
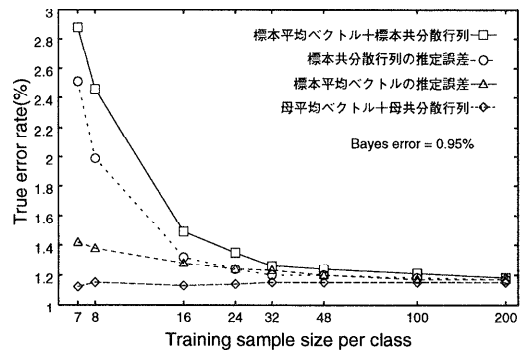


図 1 推定誤差の影響

Fig. 1 Influence of estimation errors on the true error rate on the I- Λ data set.



(a) $\Delta = 2$



(b) $\Delta = 6$

図 2 推定誤差の影響

Fig. 2 Influence of estimation errors on the true error rate on the Ness data set.

ラスが互いにより接近することになるため, クラス間の重なりが大きい場合には, 次元圧縮をせずパターン空間上で識別を行うか, または原特徴を見直すことでパターン空間を再構成することが考えられる. そこで本論文では, 標本共分散行列の推定誤差の低減についてのみ議論することにする.

3. 共分散行列の推定法

以下で述べられる正則化法, Hoffbeck らの手法は二次識別器の設計に用いられているが, それらの特徴抽出器設計への適用はこれまでになされていない. テプリッツ法は, これまで二次識別器の設計³⁾に用いられているほか, 音声処理の分野で線形予測法にも用いられている¹¹⁾. しかし, テプリッツ法はこれまで正規直交判別ベクトル法に適用されていない. 一方, 摂動法は, Hong らにより判別分析の性能改善を試みる手法として用いられた¹²⁾. Hong らは, 摂動法において逆行列を安定に求めることができるように, 後述する摂動パラメータ δ の値を大きく, かつ, Frobenius ノルムの観点から最良近似となるように摂動パラメータ δ の値を小さく, という相反する 2 つの条件を同時に満たすことを考えた. しかし, 具体的に δ のパラメータ値の与え方については何も論じていない. それゆえ, 実際的とはいえない. 本論文では, 後述の誤識別率をもとにパラメータ δ の最適化を行うことで, 摂動法を用いることにする.

3.1 テプリッツ法

ここでは記述を簡単にするためクラス名を表すラベル i を省略することにする. 標本共分散行列 Σ を

$$\Sigma = \begin{bmatrix} \sigma_1^2 & c_{12} & \cdots & c_{1n} \\ c_{21} & \sigma_2^2 & & c_{2n} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \quad (10)$$

と表すと, テプリッツ法では, 標本共分散行列は

$$\Sigma_T = \Gamma R \Gamma \quad (11)$$

となる³⁾. ここで,

$$\Gamma = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_n] \quad (12)$$

$$R = \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho^{n-1} & \cdots & \rho & 1 \end{bmatrix} \quad (13)$$

$$\rho = \frac{1}{n-1} \sum_{j=1}^{n-1} \frac{c_{jj+1}}{\sigma_j \sigma_{j+1}} \quad (14)$$

である. このとき, Σ_T^{-1} は

$$\Sigma_T^{-1} = \Gamma^{-1} R^{-1} \Gamma^{-1} \quad (15)$$

と表すことができる. ここで,

$$\Gamma^{-1} = \text{diag}[1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_n] \quad (16)$$

$$R^{-1} = \frac{1}{1-\rho^2} \times \begin{bmatrix} 1 & -\rho & & & 0 \\ -\rho & 1+\rho^2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{bmatrix} \quad (17)$$

である. テプリッツ法では, ρ と n 個の σ_i のみを推定すればよく, そのため訓練サンプル数が少ない場合でも標本共分散行列の推定精度が向上することになる.

各クラスの標本共分散行列 Σ_i にテプリッツ法を適用し, それらを式 (2) に代入することで, クラス内共分散行列を求めることができる. これをテプリッツ 1 法と呼ぶ. また, 式 (2) のクラス内共分散行列 W にテプリッツ法を適用することも可能である. これをテプリッツ 2 法と呼ぶ.

3.2 正則化法

正則化法による共分散行列の推定式は

$$\Sigma_i(\alpha, \beta) = (1-\beta)\Sigma_i(\alpha) + \frac{\beta}{n} \text{tr}[\Sigma_i(\alpha)] I_n \quad (18)$$

で与えられる⁴⁾. ここで,

$$\Sigma_i(\alpha) = S_i(\alpha)/N_i(\alpha) \quad (19)$$

$$S_i(\alpha) = (1-\alpha)S_i + \alpha S \quad (20)$$

$$S_i = \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \boldsymbol{\mu}_i)(\mathbf{x}_{ij} - \boldsymbol{\mu}_i)^T \quad (21)$$

$$S = \sum_{i=1}^m S_i \quad (22)$$

$$N_i(\alpha) = (1-\alpha)N_i + \alpha N_T \quad (23)$$

である.

正則化法では, クラス内共分散行列を 2 通りで与えることができる. 1 つはクラスごとにパラメータ α , β を与える手法, もう 1 つは, 各クラスのパラメータ α , β を共通とする手法である. 前者を正則化 1 法, 後者を正則化 2 法と呼ぶ. いずれにせよ, 各クラスの共分散行列を推定した後, それらを式 (2) に代入することでクラス内共分散行列 W を得ることができる.

正則化 2 法では, 特に各クラスの訓練サンプル数が等しいとき, クラス内共分散行列を

$$W_R(\beta) = \frac{1}{m} \left\{ (1-\beta) \frac{S}{N} + \frac{\beta}{n} \text{tr} \left[\frac{S}{N} \right] I_n \right\} \quad (24)$$

と表すことができる (付録 A.1 参照). ここで, $N =$

表1 パラメータ $\gamma_i = [\gamma_{i1}, \gamma_{i2}, \gamma_{i3}, \gamma_{i4}]$ の組合せ
Table 1 Values of γ_i .

γ_{i1}	1.0	.75	.50	.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
γ_{i2}	0.0	.25	.50	.75	1.0	.75	.50	.25	0.0	0.0	0.0	0.0	0.0
γ_{i3}	0.0	0.0	0.0	0.0	0.0	.25	.50	.75	1.0	.75	.50	.25	.00
γ_{i4}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.25	.50	.75	1.0	0.0

$N_1 = N_2 = \dots = N_m$ である。この場合、クラス内共分散行列がパラメータ β のみの関数となる。

3.3 Hoffbeck らの手法

Hoffbeck らにより提案された共分散行列の推定式は

$$\Sigma_H(\gamma_i) = \gamma_{i1} \text{diag}(\Sigma_i) + \gamma_{i2} \Sigma_i + \gamma_{i3} W + \gamma_{i4} \text{diag}(W) \quad (25)$$

である⁵⁾。この推定に必要とされる4つのパラメータ $[\gamma_{i1}, \gamma_{i2}, \gamma_{i3}, \gamma_{i4}]$ の組合せは13通りある(表1参照)。クラスごとにパラメータを与える手法を Hoffbeck1 法、各クラスのパラメータを共通とする手法を Hoffbeck2 法と呼ぶことにする。

特に Hoffbeck2 法では、共分散行列は

$$\Sigma_H(\gamma_{i1}) = \gamma_{i1} \text{diag}(\Sigma_i) + (1 - \gamma_{i1}) \Sigma_i \quad (26)$$

と表される(付録A.2参照)。この場合、共分散行列はパラメータ γ_{i1} のみの関数となる。得られた各クラスの共分散行列を式(2)に代入してクラス内共分散行列 W を得ることになる。

3.4 摂動法

W は式(27)のように特異値分解することができる。

$$W = U \Lambda U^T \quad (27)$$

ここで、

$$\Lambda = \text{diag}[\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_n^{1/2}] \quad (28)$$

$$U = [u_1 \ u_2 \ \dots \ u_n] \quad (29)$$

である。 λ_j ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$) は WW^T の固有値、 u_j は λ_j に対応する固有ベクトルである。

摂動法ではクラス内共分散行列の推定式は

$$W_P = U \hat{\Lambda} U^T \quad (30)$$

で与えられる¹²⁾。ここで、

$$\hat{\Lambda} = \text{diag}[\lambda_1^{1/2}, \dots, \lambda_r^{1/2}, \delta^{1/2}, \dots, \delta^{1/2}] \quad (31)$$

$$\lambda_{r+1}^{1/2} < \delta^{1/2} < \lambda_r^{1/2} \quad (32)$$

であり、 $\delta^{1/2}$ は設計者により与えられるパラメータである。

クラス内共分散行列 W に摂動法を適用する手法を摂動1法と呼ぶ。各クラスの共分散行列に摂動法を適用することも考えられ、この場合は2通りの推定法がある。1つはクラスごとに異なるパラメータ $\delta^{1/2}$ を

用いる手法で、これを摂動2法と呼ぶ。もう1つは、各クラスの共分散行列に共通のパラメータ $\delta^{1/2}$ を用いる手法で、これを摂動3法と呼ぶ。

3.5 パラメータの最適化法

パラメータ最適化には様々なアプローチが考えられる。ここでは、Fisher の線形識別器による誤識別率を評価関数とし、その最小化を考える。つまり、識別志向のアプローチである。具体的には、あらかじめパラメータの候補値を複数用意し、各候補値に対し訓練サンプルのみを用いて誤識別率を推定し、その誤識別率が最小となるパラメータ値を解とする。誤識別率が最小となるパラメータ値が複数存在する場合は、それら誤識別率を最小とするパラメータ値の平均値を求め、誤識別率を最小とするパラメータ値の中でこの平均値に最も近いパラメータ値を、解とする。訓練サンプル上での誤識別率の推定法として、再代入法と leave-one-out 法 (LOO) が考えられる。以降、たとえば、共分散行列の推定法として正規化1法を、訓練サンプル上での誤識別率として再代入法を用いた場合は、それを正規化1法(再代入)と表すことにする。

3.6 各推定法の諸性質

$I - \Lambda$ データを用いて、各推定法の諸性質を調べる。実験を次の条件下で行った。この実験では、次元数に対する訓練サンプル数の比は、1未満から最大4までの範囲を考えている。

データ:	$I - \Lambda$ データ
パターン空間の次元数:	$n = 8$
特徴空間の次元数:	2次元
訓練サンプル数:	$N = 7, 8, 10, 12, 14, 16, 24, 32$
テストサンプル数:	各クラス1000個
識別器:	Fisher の線形識別器
パラメータ α :	0.0, .125, .354, .650, 1.0
パラメータ β :	0.0, .25, .50, .75, 1.0
パラメータ $\delta^{1/2}$:	0.0~4.5 (0.5 刻み)
独立試行回数:	100回

正規化族, Hoffbeck 族, 摂動族の実験結果には、各手法の性能の限界も示す。ここで、性能の限界とは、テストパターンに対する誤識別率の最小値を意味する。つまり、その手法において、いかにパラメータの値を変えてもそれ以上の性能を与えることができないという意味から限界と称することにした。

まず、誤識別率の観点から検討する。テプリッツ族

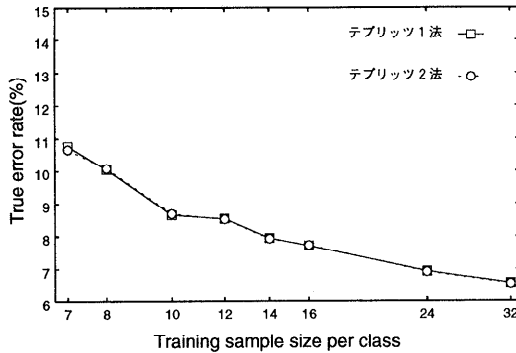


図3 テプリッツ族における比較
Fig. 3 Comparison of Toeplitz-based methods.

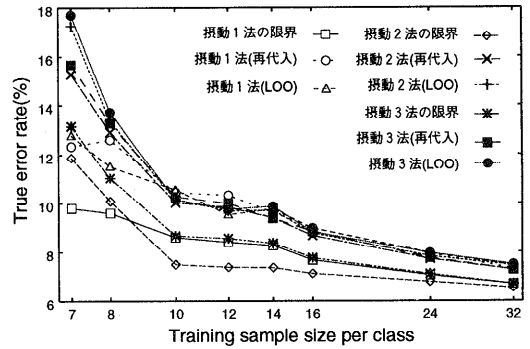


図6 摂動族における比較
Fig. 6 Comparison of perturbation methods.

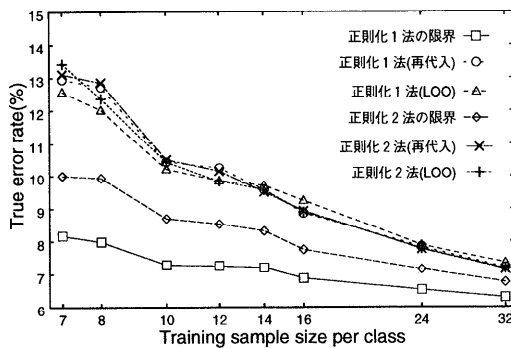


図4 正則化族における比較
Fig. 4 Comparison of regularized methods.

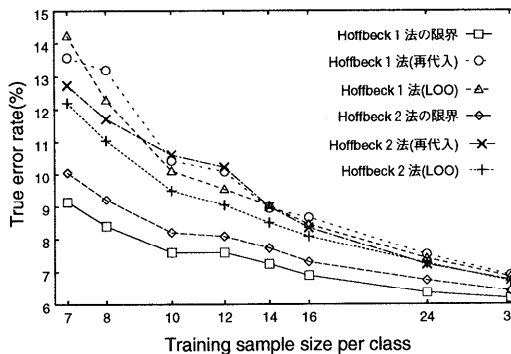


図5 Hoffbeck族における比較
Fig. 5 Comparison of Hoffbeck's methods.

の実験結果を図3に示す。テプリッツ1法とテプリッツ2法とは真の誤識率に差がないことが分かる。正則化族の実験結果を図4に示す。識別性能から、どの手法が最良であるかを特定することはできない。Hoffbeck族の実験結果を図5に示す。実験結果から、Hoffbeck2法(LOO)が最良であることが分かる。最後に摂動族の実験結果を図6に示す。摂動法も識別性能からは、どの手法が最良の手法であるかを特定する

表2 特徴空間の生成に要するCPU時間の比

Table 2 The ratio of the CPU time for generating a 2-dimensional feature space from a 4-dimensional pattern space.

共分散行列の推定法	CPU時間の比
テプリッツ1法	1.7
テプリッツ2法*	1.5
正則化1法(再代入)	5.9×10^4
正則化1法(LOO)	8.8×10^5
正則化2法(再代入)*	2.1×10
正則化2法(LOO)	3.2×10^2
摂動1法(再代入)*	2.6×10
摂動1法(LOO)	2.5×10^2
摂動2法(再代入)	2.4×10^3
摂動2法(LOO)	2.5×10^4
摂動3法(再代入)	7.1×10
摂動3法(LOO)	7.3×10^2

*は各族においてCPU時間が最小であることを表す。

ことはできない。そこで誤識別率では優劣がつかない、テプリッツ族、正則化族、摂動族に対して、CPU時間による検討を行う。そのため、2次元特徴空間を10回生成し、特徴空間を1回生成するのに必要なCPU時間の平均値を求める。以上の処理を各共分散行列の推定法に対して行い、従来手法のCPU時間に対する各推定法を取り入れた場合のCPU時間の比を求める。

実験条件は次のとおりである。

- データ: Irisデータ
- パターン空間の次元数: $n = 4$
- 特徴空間の次元数: 2次元
- 訓練サンプル数: $N = 10$
- 識別器: Fisherの線形識別器
- パラメータ α : 0.0, .125, .354, .650, 1.0
- パラメータ β : 0.0, .25, .50, .75, 1.0
- パラメータ $\delta^{1/2}$: 0~0.09 (0.01刻み)
- 独立試行回数: 10回

Irisデータは、3種類のアイリスの花、150本から4つの特徴を抽出したデータである¹³⁾。実験結果を表2に示す。実験結果から、テプリッツ2法、正則

表3 Zernike データに対する比較実験の条件
Table 3 Conditions of experiments for the Zernike data set.

	比較実験 1	比較実験 2	比較実験 3	比較実験 4
特徴空間の次元数	10		5, 10, ..., 45, 47	
訓練サンプル数	$N = 100, 200, 300, 400$		$N = 100$	
テストサンプル数	各クラス 1000 個			
識別器	Fisher の線形識別器	二次線形識別器	Fisher の線形識別器	二次線形識別器
独立試行回数	5 回			
パラメータ β	0.0, .25, .50, .75, 1.0			
パラメータ $\delta^{1/2}$	0~900 (100 刻み)			

化2法(再代入), 摂動1法(再代入)がCPU時間の観点から最良であることが分かる。以上の実験結果から, 4章の比較実験では, テプリッツ2法, 正則化2法(再代入), Hoffbeck2法(LOO), 摂動1法(再代入)を用い, それぞれを単にテプリッツ法, 正則化法, Hoffbeck法, 摂動法と呼ぶことにする。

4. 比較実験

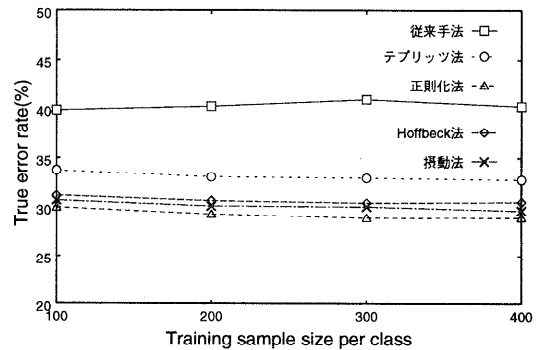
本章では, 標本共分散行列をそのままクラス内共分散行列に用いる正規直交判別ベクトル法(以後, 従来手法と呼ぶ), テプリッツ法, 正則化法, Hoffbeck法, 摂動法の優劣を検討する。

4.1 実在データを用いた認識実験

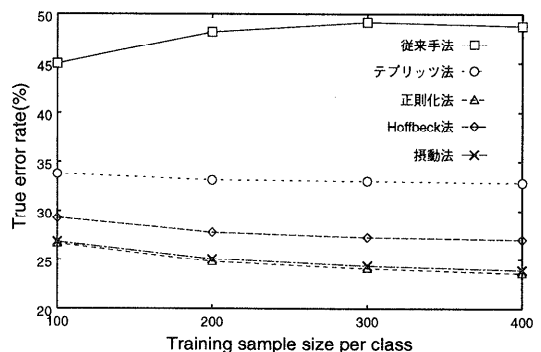
まず, 47次元, 10クラスのZernikeデータを用いて比較実験を行う。Zernikeデータは電子技術総合研究所の手書き文字データベースETL-1¹⁴⁾の数字10字種をZernikeモーメント¹⁵⁾を用いて抽出したデータである。

比較実験1では, 識別系をFisherの線形識別器とし, 訓練サンプル数を増加させて比較を行う。比較実験2では二次識別器を用いて訓練サンプル数増加による比較実験を行う。比較実験3ではFisherの線形識別器を用いて特徴軸数を増加させ, 比較を行う。比較実験4では二次識別器を用いて特徴数増加の実験を行う。以上の実験条件を表3にまとめて示す。

比較実験1, 2の結果を図7に示す。従来手法に比べ, テプリッツ法, 正則化法, Hoffbeck法, 摂動法は, いずれも低い誤識別率を与え, 標本共分散行列の推定誤差を低減することの効果がある。中でも正則化法が比較的有望と考えられる。比較実験3の結果を図8(a)に示す。この実験結果からも, 正則化法が有効であることが分かる。比較実験4の結果を図8(b)に示す。実験結果から, 摂動法が有効であることが分かる。比較実験3, 4から, 特徴空間の次元数が小さいほど, 標本共分散行列の推定誤差を低減することの効果を確認することができた。比較実験3のFisherの線形識別器では生じないピーキング現象¹⁶⁾が, 二次



(a) Fisher の線形識別器の場合



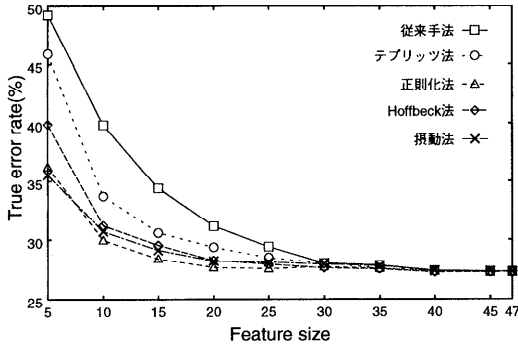
(b) 二次識別器の場合

図7 訓練サンプル数の影響

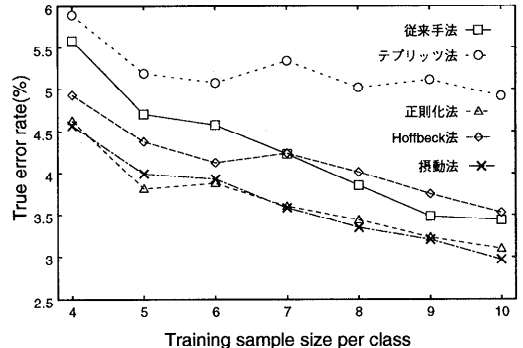
Fig. 7 Influence of the training sample size on the true error rate on the Zernike data set.

識別器を用いる比較実験4で生じている。また, 47次元パターン空間よりも小さな誤識別率を与える特徴空間が生成されたことから, 適当な共分散行列の推定法を用いることにより, 訓練サンプル数が少ない状況下であってもパターン空間よりも識別能力の高い特徴空間を生成することができるという見通しを得た。

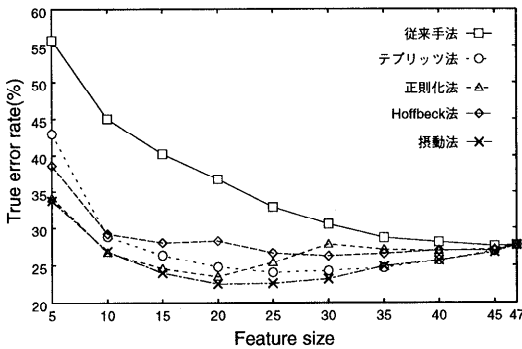
次に, Irisデータを用いて比較実験を行う。本実験では, 訓練サンプルとテストサンプルとの独立性を保つため, 各クラス50個のサンプルを10個の訓練サンプルと40個のテストサンプルとにランダムに分割し



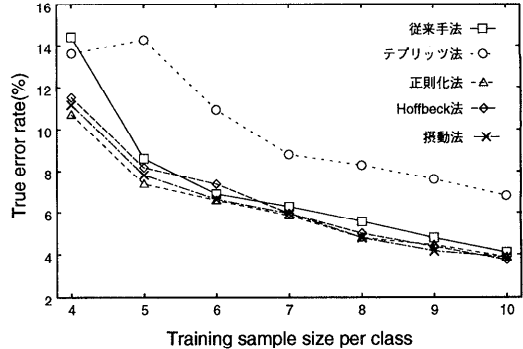
(a) Fisher の線形識別器の場合



(a) Fisher の線形識別器を用いた場合



(b) 二次識別器の場合



(b) 二次識別器を用いた場合

図 8 特徴軸数の影響

Fig. 8 Influence of the feature size on the true error rate on the Zernike data set.

図 9 訓練サンプル数の影響

Fig. 9 Influence of the training sample size on the true error rate on the Iris data set.

た。実験を次の条件下で行った。

- データ: Iris データ
- パターン空間の次元数: $n = 4$
- 特徴空間の次元数: 2次元
- 訓練サンプル数: $N = 4, 5, 6, 7, 8, 9, 10$
- テストサンプル数: 各クラス 40 個
- 識別器: Fisher の線形識別器
二次識別器
- パラメータ β : 0.0, .25, .50, .75, 1.0
- パラメータ $\delta^{1/2}$: 0~0.09 (0.01 刻み)
- 独立試行回数: 100 回

実験結果を図 9 に示す。図 9 から、正則化法と摂動法は従来手法に対してつねに小さい誤識別率を与えていることが分かる。Zernike データと Iris データを用いた実験結果から、実在データに対しては、正則化法、摂動法のいずれかが有効であると考えられる。

4.2 人工データを用いた認識実験

人工データである $I-\Lambda$ データ、Ness データを用いて各手法の比較を行う。実験条件は次のとおりである。

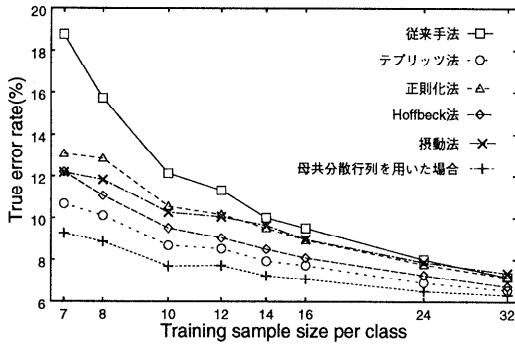
- パターン空間の次元数: $n = 8$
- 特徴空間の次元数: 2次元
- 訓練サンプル数: $N = 7, 8, 10, 12, 14, 16, 24, 32$
- テストサンプル数: 各クラス 1000 個
- 識別器: Fisher の線形識別器,
二次識別器
- パラメータ β : 0.0, .25, .50, .75, 1.0
- パラメータ $\delta^{1/2}$: 0.0~4.5 (0.5 刻み)
- パラメータ Δ : 6
- 独立試行回数: 100 回

Fisher の線形識別器を用いた場合の実験結果を図 10 に示す。実験結果から、テプリッツ法が有効な手法であることが分かる。二次識別器を用いた実験結果を図 11 に示す。実験結果から、テプリッツ法と摂動法が有効な手法であることが分かる。

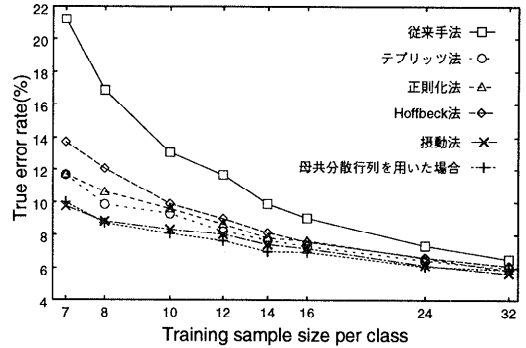
4.3 CPU 時間による比較実験

本実験では、特徴抽出系の設計、すなわち特徴空間の生成に要する CPU 時間から各手法を比較する。

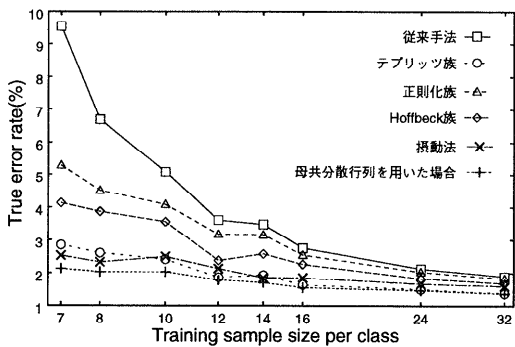
実験条件は次のとおりである。



(a) I-H データ上での比較



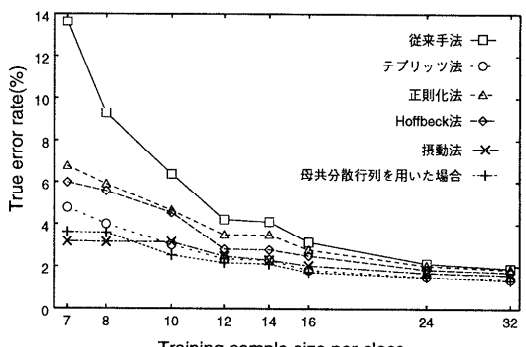
(a) I-H データ上での比較



(b) Ness データ上での比較

図 10 Fisher の線形識別器の場合

Fig. 10 Results for the Fisher's linear classifier.



(b) Ness データ上での比較

図 11 二次識別器の場合

Fig. 11 Results for the quadratic classifier.

データ: Ness データ
 パターン空間の次元数: $n = 4, 8, 12, 16, 20, 24, 28$
 特徴空間の次元数: 2 次元
 訓練サンプル数: $N = 48$
 識別器: Fisher の線形識別器
 パラメータ β : 0.0, .25, .50, .75, 1.0
 パラメータ $\delta^{1/2}$: 0~4.5 (0.5 刻み)
 パラメータ Δ : 6

実験結果を図 12 に示す。図 12 の横軸は次元数を、縦軸は特徴抽出系の設計に要する CPU 時間の比を表す。縦軸は対数表示である。実験結果から、特徴抽出系の設計に要する CPU 時間に関しては、テブリッツ法が最小であり、次に正則化法、摂動法の順となる。正則化法と摂動法とは、次元数の増加とともに CPU 時間の相違がさらに著しくなる。たとえば、次元数が 28 のとき、正則化法は摂動法の 3 分の 1 の CPU 時間で実行できる。それゆえ、次元数の大きいパターン認識問題では、摂動法よりも正則化法が实际的と考えられる。

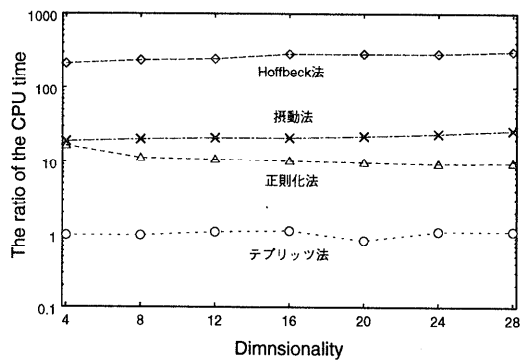


図 12 特徴空間の生成に要する CPU 時間の比

Fig. 12 The ratio of the CPU time for generating a feature space.

5. ま と め

本論文では訓練サンプル数が少ない状況下における正規直交判別ベクトル法を検討した。まず、標本平均ベクトルと標本共分散行列の推定誤差が特徴空間の識別能力に与える影響について検討した。共分散行列の

推定法として、テプリッツ法、正則化法、Hoffbeck法、摂動法を用い、それらの手法の諸性質を調べ、さらに4つの推定法を誤識別率とCPU時間の両面から比較した。

実験結果から、以下の見通しを得た。まず正規分布に準じるデータに対しては、誤識別率とCPU時間の観点から、テプリッツ法が有効と考えられる。一方、実在データに対しては、誤識別率の観点から正則化法と摂動法とはほとんどの場合同等であり、CPU時間をも考慮すると、正則化法が推奨される。

今後の課題として、特徴空間の次元数の最適化、大規模データでの有効性の検討、一般化Fisher評価基準への適用などが考えられる。

謝辞 手書き文字データベースETL-1を作成、提供していただいた電子技術総合研究所の関係諸氏に深く感謝いたします。

参 考 文 献

- 1) 岡田敏彦, 富田真吾: 正規直交判別ベクトル法による特徴抽出論, 信学論 (A), Vol.J65-A, No.8, pp.767-771 (1982).
- 2) 浜本義彦, 金岡泰保, 富田真吾: パターン認識のための正規直交判別ベクトル法に関する理論的考察, 信学論 (A), Vol.J77-A, No.1, pp.76-83 (1994).
- 3) Fukunaga, K.: *Introduction to statistical pattern recognition, Second edition*, Academic Press (1990).
- 4) Friedman, J.H.: Regularized discriminant analysis, *Journal of the American Statistical Association*, Vol.84, No.405, pp.165-175 (1989).
- 5) Hoffbeck, J.P. and Landgrebe, D.A.: Covariance matrix estimation and classification with limited training data, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.18, No.7, pp.763-767 (1996).
- 6) Jain, A.K. and Chandrasekaran, B.: Dimensionality and sample size considerations in pattern recognition practice, *Handbook of Statistics*, Krishnaiah, P.R. and Kanal, L.N. (Eds.), Vol.2, pp.835-855, North Holland (1982).
- 7) Van Ness, J. On the dominance of non-parametric Bayes rule discriminant algorithms in high dimensions, *Pattern Recognition*, Vol.12, No.6, pp.355-368 (1980).
- 8) Hamamoto, Y., Fujimoto, Y. and Tomita, S.: On the estimation of a covariance matrix in designing Parzen classifiers, *Pattern Recognition*, Vol.29, No.10, pp.1751-1759 (1996).
- 9) Hamamoto, Y., Uchimura, S. and Tomita, S.:

A bootstrap technique for nearest neighbor classifier design, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.1, pp.1-6 (1997).

- 10) 浜本義彦, 常田祐司, 金山知余, 富田真吾: 遺伝的アルゴリズムを用いた最近傍識別器のための代表サンプル選択, 信学論 (A), Vol.J80-A, No.2, pp.371-378 (1997).
- 11) 古井貞熙: 音声情報処理, 森北出版 (1998).
- 12) Hong, Z.-Q. and Yang, J.-Y.: Optimal discriminant plane for a small number of samples and design method of classifier on the plane, *Pattern Recognition*, Vol.24, No.4, pp.317-324 (1991).
- 13) Fisher, R.A.: The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, Vol.7, pp.280-322 (1936).
- 14) 斎藤泰一, 山田博三, 森 俊二: 手書き文字データベースの解析 (III), 電子技術総合研究所, Vol.42, No.5 (1978).
- 15) Khotanzad, A. and Hong, Y.H.: Invariant image recognition by Zernike moments, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.12, No.5, pp.489-497 (1990).
- 16) 若林哲史, 鶴岡信治, 木村文隆, 三宅康二: 少クラス分類問題における特徴選択に関する考察, 信学論 (D-II), Vol. J80-DII, No.1, pp. 73-80 (1997).

付 録

A.1 正則化法における式 (24) の導出

各クラスの訓練サンプル数が s N のとき、正則化法による共分散行列 W_R を次のように変形することができる。

$$\begin{aligned}
 W_R(\alpha, \beta) &= \sum_{i=1}^m \left\{ \frac{1}{m} \Sigma_i(\alpha, \beta) \right\} \\
 &= \frac{1}{m} \sum_{i=1}^m \left\{ (1 - \beta) [\Sigma_i(\alpha)] + \frac{\beta}{n} \text{tr} [\Sigma_i(\alpha)] I_n \right\} \\
 &= \frac{1}{m} \left\{ (1 - \beta) \left[\sum_{i=1}^m \Sigma_i(\alpha) \right] \right. \\
 &\quad \left. + \frac{\beta}{n} \text{tr} \left[\sum_{i=1}^m \Sigma_i(\alpha) \right] I_n \right\} \quad (33)
 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^m \Sigma_i(\alpha) &= \frac{\sum_{i=1}^m \{(1-\alpha)S_i + \alpha S\}}{(1-\alpha)N + \alpha N_T} \\ &= \frac{\{(1-\alpha) + \alpha m\}S}{\{(1-\alpha) + \alpha m\}N} \\ &= \frac{S}{N} \end{aligned} \tag{34}$$

式 (34) を式 (33) に代入すれば、式 (25) を得る。

A.2 Hoffbeck らの手法における式 (26) の導出

$$\begin{aligned} \sum_{i=1}^m p_i \Sigma(\gamma_{i1}, \gamma_{i2}) &= \sum_{i=1}^m p_i \{\gamma_{i1} \text{diag}(\Sigma_i) + \gamma_{i2} \Sigma_i\} \\ &= \sum_{i=1}^m p_i \{\gamma_{i1} \text{diag}(\Sigma_i) + (1-\gamma_{i1}) \Sigma_i\} \\ &= \gamma_{i1} \text{diag} \left(\sum_{i=1}^m p_i \Sigma_i \right) + (1-\gamma_{i1}) \sum_{i=1}^m p_i \Sigma_i \\ &= \gamma_{i1} \text{diag}(W) + (1-\gamma_{i1})W \end{aligned} \tag{35}$$

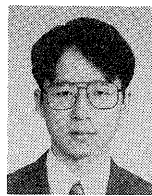
$$\begin{aligned} \sum_{i=1}^m p_i \Sigma(\gamma_{i2}, \gamma_{i3}) &= \sum_{i=1}^m p_i (\gamma_{i2} \Sigma_i + \gamma_{i3} W) \\ &= \sum_{i=1}^m p_i \{\gamma_{i2} \Sigma_i + (1-\gamma_{i2})W\} \\ &= \gamma_{i2} \sum_{i=1}^m p_i \Sigma_i + (1-\gamma_{i2})W \sum_{i=1}^m p_i \\ &= \gamma_{i2} W + (1-\gamma_{i2})W \\ &= W \end{aligned} \tag{36}$$

$$\begin{aligned} \sum_{i=1}^m p_i \Sigma(\gamma_{i3}, \gamma_{i4}) &= \sum_{i=1}^m p_i \{\gamma_{i3} W + \gamma_{i4} \text{diag}(W)\} \\ &= \sum_{i=1}^m p_i \{(1-\gamma_{i4})W + \gamma_{i4} \text{diag}(W)\} \\ &= (1-\gamma_{i4})W \sum_{i=1}^m p_i + \gamma_{i4} \text{diag}(W) \sum_{i=1}^m p_i \\ &= (1-\gamma_{i4})W + \gamma_{i4} \text{diag}(W) \end{aligned} \tag{37}$$

式 (35) と式 (37) とは等しく、式 (35) に式 (36) は含まれるので、Hoffbeck2 法では、共分散行列は式 (26) のみでよい。

(平成 10 年 12 月 24 日受付)

(平成 11 年 6 月 3 日採録)



宮本 貴宣 (学生会員)

昭和 50 年生。平成 9 年山口大学工学部知能情報システム工学科卒業。平成 11 年同大学大学院理工学研究科知能情報システム工学専攻博士前期課程修了。現在、同大学大学院博士後期課程。パターン認識に関する研究に従事。電子情報通信学会学生会員。



浜本 義彦

昭和 32 年生。昭和 56 年山口大学工学部電子工学科卒業。昭和 58 年同大学大学院修士課程修了。同年日本電気 (株) 入社。昭和 62 年山口大学工学部助手。現在同大学教授、工学博士。パターン認識に関する研究に従事。電子情報通信学会会員。