

標本共分散行列の固有ベクトルを用いた真のマハラノビス距離の推定法

岩村 雅一[†] 大町真一郎[†] 阿曾 弘具[†]

Estimation of True Mahalanobis Distance from Eigenvectors of Sample Covariance Matrix

Masakazu IWAMURA[†], Shinichiro OMACHI[†], and Hiroto ASO[†]

あらまし パターン認識で用いられる、ベイズの定理から導かれる統計的識別関数は、パターンの確率分布が正しく与えられたときに誤識別率が最小になる。パターンの確率分布は一般に未知であるため、正規分布を仮定して、その平均ベクトルと共分散行列は学習サンプルからの推定値を用いることが多い。しかし、推定に用いる学習サンプルが不足すると、推定値が誤差を含み、認識性能が低下することが知られている。平均ベクトルや共分散行列の固有値の推定誤差についてはよく調査され、認識性能の低下を回避する方法が提案されている。しかし、共分散行列の固有ベクトルの推定誤差については、これまでほとんど考慮されてこなかった。本論文では、学習サンプルが十分用意できない場合においても高精度な認識を行うことを目的とし、共分散行列の固有ベクトルに誤差がある場合でも、2次識別関数の主要部分であるマハラノビス距離を正しく推定する手法を提案する。提案手法を文字認識に適用した結果、パラメータの推定に用いる学習サンプルが少ない場合でも、マハラノビス距離を正しく推定することができ、認識性能の改善が見られた。

キーワード パターン認識, 識別関数, 固有ベクトル, 推定誤差, マハラノビス距離

1. ま え が き

ベイズの定理から導かれる統計的識別関数は、パターンの確率分布が正しく与えられたときに、誤認識によって生じる損失の期待値を最小にとどめる最適性をもつ。しかし、ほとんどのパターン認識問題ではパターンの確率分布は未知であるため、パターンの確率分布として特定の分布を仮定し、分布パラメータを学習用のサンプルから推定することが多い。このとき、学習サンプル数が不足すると、パラメータの推定誤差のために認識性能が低下することが知られている。これは、特徴量の次元数が増えると、分布の推定に必要な学習サンプル数が指数関数的に増加することが原因で、次元の呪いとして知られている [1]。例えば文字認識では 1536 次元の特徴が用いられるなど [2]、パターン認識では高次元の特徴量が用いられることも多く、パラメータの推定誤差は非常に重要な問題である。

パターンの確率分布を多次元正規分布と仮定した識別関数として、2次識別関数やマハラノビス距離が知られている。多次元正規分布の分布パラメータである平均ベクトルと共分散行列を学習サンプルから推定したとき、推定値である標本平均ベクトルと標本共分散行列はそれぞれ誤差を含む。

竹下ら [3] は、2次識別関数の値を実質的に決定するマハラノビス距離が、平均ベクトルと共分散行列のそれぞれの推定誤差に対してどのような影響を受けるのかを理論的に調査した。その結果、学習サンプル数が 10 を超えるときは、平均ベクトルの推定誤差の影響は共分散行列の推定誤差に比べてほとんど無視できることを導いた。したがって (識別器としての) 2次識別関数やマハラノビス距離では共分散行列の推定誤差が認識精度を低下させているといえる。

共分散行列の推定誤差については、共分散行列が固有値展開により、その固有値と固有ベクトルに分解できることから、固有値の推定誤差と固有ベクトルの推定誤差に分けて考えられてきた。マハラノビス距離の計算の際、共分散行列の固有値と固有ベクトルは“各

[†] 東北大学大学院工学研究科, 仙台市
Graduate School of Engineering, Tohoku University,
Sendai-shi, 980-8579 Japan

固有ベクトル軸方向のユークリッド距離に、対応する固有値の逆数を掛ける” という意味をもち、互いに相補的である。

このうち、固有値の推定誤差については、特に小さな固有値の推定誤差が認識性能に悪影響を及ぼすことが知られ、信頼性の低い推定値を用いない手法が提案されている。このような手法には、木村らの修正 2 次識別関数 (MQDF) [4]、加藤らの改良型マハラノビス距離 (MMD) [5]、孫らの簡素化マハラノビス距離 (SMD) [6]、Friedman の Regularized Discriminant Analysis [7] などがある。

それに対して、信頼性の低い固有値を補正して用いる手法も提案されている。分布のパラメータを補正する手法は、信頼性の低いパラメータを用いない手法に比べて難しい反面、補正が正しく行われればより高い認識性能が得られる可能性がある。補正法を導く理論自体がパターン認識に与える貢献も大きい。James と Stein [8] は、従来の標本共分散行列は非許容的であり、よりよい推定量が存在することを指摘し、標本共分散行列を改良する推定量 (James-Stein 推定量) を導いた。Takeshita ら [9] は実験によって James-Stein 推定量がパターン認識に有効であることを確認した。また、酒井ら [10] は標本共分散行列の固有値の平均的な偏りから標本固有値を補正する手法 (RQDF) を提案し、良好な認識結果を得ている。

たとえ真の固有値を正しく推定して用いたとしても、固有ベクトルが推定誤差を含む場合にはマハラノビス距離が正しく計算されないことが考えられる。しかし、固有ベクトルの推定誤差が認識性能に及ぼす影響についてはこれまであまり考慮されてこなかった。酒井ら [11] は、標本固有ベクトルの揺らぎのために、パターン認識で推定すべき共分散行列は真の共分散行列ではなく、真の共分散行列をぼかした行列 (ぼけ共分散行列) であると主張し、固有値を 2 回補正する手法 (WRQDF) を提案している。しかし、なぜぼけ共分散行列を用いるべきかについての議論は行われていない。また、2 回目の補正では標本固有値も積分変数としているため、ぼけ共分散行列が標本固有値のもつ情報を十分に反映していないと考えられる。

本論文では、共分散行列の固有ベクトルの推定誤差を考慮することで、学習サンプルが十分用意できない場合においても高精度な認識を行うことを目的とし、マハラノビス距離を正しく推定するために真の固有値を修正する手法を提案する。2. では、固有ベクトル

の推定誤差によってマハラノビス距離に推定誤差が生じること、マハラノビス距離の推定誤差は標本固有ベクトルが真の固有値を「ゆがめる」ことによって生じると解釈できることを示す。そして、真の固有値を修正することにより、辞書の作成に用いるサンプルが少ない場合でも、マハラノビス距離を正しく推定する手法を提案する。3. では、提案手法を文字認識に適用し、マハラノビス距離が正しく推定されること、認識性能が改善されることを示す。

なお、本論文では確率変数と定数を区別するため、 \hat{A} や \hat{A} のように確率変数には $\hat{\cdot}$ を付ける。

2. 真のマハラノビス距離の推定法

2.1 誤差を含む固有ベクトルを用いたマハラノビス距離の推定

真の共分散行列の第 i 固有値、固有ベクトルを λ_i 、 ϕ_i とすると、真の固有値、固有ベクトルからなる行列は

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d) \quad (1)$$

$$\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_d] \quad (2)$$

とおける。ここで d は特徴量の次元数で、固有値は降順に並んでいるとする。真の平均ベクトルを μ 、未知入力ベクトルを x とすれば、マハラノビス距離 $d(x)$ は次式で表される。

$$d(x) = (x - \mu)^T \Phi \Lambda^{-1} \Phi^T (x - \mu) \quad (3)$$

一方、標本共分散行列の第 i 固有ベクトルを $\hat{\phi}_i$ とし、標本共分散行列の固有ベクトルからなる行列を

$$\hat{\Phi} = [\hat{\phi}_1 \ \hat{\phi}_2 \ \dots \ \hat{\phi}_d] \quad (4)$$

とおく。固有ベクトルとして、推定誤差を含む標本固有ベクトルを用いたときのマハラノビス距離は次式で表される。

$$\hat{d}(x) = (x - \mu)^T \hat{\Phi} \Lambda^{-1} \hat{\Phi}^T (x - \mu) \quad (5)$$

今、正規直交基底である固有ベクトルが作る座標系を考える。このように考えると、真の固有ベクトルが作る座標系と標本固有ベクトルが作る座標系のずれ (回転) が固有ベクトルの推定誤差であると考えられる。そこで、第 j 標本固有ベクトルが真の第 i 固有ベクトル方向へずれた量を ij 成分にもつ行列 $\hat{\Psi}$ を、固有ベクトルの推定誤差を表す行列として次のように定

義する .

$$\hat{\Psi} \equiv \Phi^T \hat{\Phi} \quad (6)$$

Φ , $\hat{\Phi}$ は正規直交基底 (回転行列) であるので, $\hat{\Psi}$ も正規直交基底である. 式 (6) から

$$\hat{\Phi} = \Phi \hat{\Psi} \quad (7)$$

と表せるので, 式 (7) を式 (5) に代入する .

$$\hat{d}(x) = (x - \mu)^T \Phi \hat{\Psi} \Lambda^{-1} \hat{\Psi}^T \Phi^T (x - \mu) \quad (8)$$

$$= (x - \mu)^T \Phi \left(\hat{\Psi} \Lambda \hat{\Psi}^T \right)^{-1} \Phi^T (x - \mu) \quad (9)$$

式 (9) を式 (3) と比較すれば, 式 (9) の $\left(\hat{\Psi} \Lambda \hat{\Psi}^T \right)^{-1}$ と式 (3) の固有値 Λ^{-1} は明らかに異なり, 標本固有ベクトルを用いて求めたマハラノビス距離 $\hat{d}(x)$ は真のマハラノビス距離 $d(x)$ との誤差を含むことがわかる. そして, その誤差は Λ^{-1} となるべき部分が $\left(\hat{\Psi} \Lambda \hat{\Psi}^T \right)^{-1}$ となっていることから生じており, いわば, $\hat{\Psi}$ が固有値を「ゆがめている」とみなすことができる. 本論文では, 学習サンプル数の不足によって生じる固有値の推定誤差に対して, 固有ベクトルの推定誤差 $\hat{\Psi}$ が原因で生じる固有値の誤差を「固有値のゆがみ」と表現する.

固有値にゆがみが生じることを前提に考えると, ゆがみの傾向からあらかじめ真の固有値を修正しておけば, 固有値がゆがんだ結果, マハラノビス距離が正しく推定できると考えられる. そこで, 式 (5) で Λ の代わりに適当な対角行列 $\tilde{\Lambda}$ を用いる. すると, 式 (9) の固有値に相当する部分は $\left(\hat{\Psi} \tilde{\Lambda} \hat{\Psi}^T \right)^{-1}$ となるので, これを Λ^{-1} に近づける. すなわち,

$$\Lambda^{-1} \approx \left(\hat{\Psi} \tilde{\Lambda} \hat{\Psi}^T \right)^{-1} \quad (10)$$

となる $\tilde{\Lambda}$ を選ぶことを考える. 式 (10) から,

$$\tilde{\Lambda} \approx \hat{\Psi}^T \Lambda \hat{\Psi} \quad (11)$$

が成立するので,

$$\tilde{\Lambda} = D \left(\hat{\Psi}^T \Lambda \hat{\Psi} \right) \quad (12)$$

とおく. ここで $D(\cdot)$ は行列の対角成分を返す関数である. $\hat{\Psi}^T \Lambda \hat{\Psi}$ の非対角成分を無視することの妥当性は 3.1 の実験で確認する. 式 (12) を式 (5) の Λ の代わりに用いることにより, 次式が得られる.

$$\tilde{d}(x) = (x - \mu)^T \hat{\Phi} \tilde{\Lambda}^{-1} \hat{\Phi}^T (x - \mu) \quad (13)$$

式 (13) は, 誤差を含んだ固有ベクトルを用いて真のマハラノビス距離のよりよい推定値を与えると期待される.

2.2 固有ベクトルの推定誤差を補償する固有値

式 (12) では, $\tilde{\Lambda}$ が真の固有値 Λ と固有ベクトルの推定誤差 $\hat{\Psi}$ で定義されている. 固有ベクトルの推定誤差 $\hat{\Psi}$ は, 式 (6) で真の固有ベクトル Φ を用いて定義されている. 今, 真の固有ベクトルが未知である場合を考えているので, $\hat{\Psi}$ は観測できない. しかし, 付録で示すように, 固有ベクトルの推定誤差の確率分布は特徴量の次元数, 学習サンプル数, 真の固有値, 標本固有値のみから決まり, 真の固有ベクトルとは無関係である. したがって, 真の固有ベクトルが未知であっても, 固有ベクトルの推定誤差の期待値であれば観測可能である. そこで, 固有ベクトルの推定誤差を表す確率変数 $\hat{\Psi}$ を考え, 式 (12) の Ψ を $\hat{\Psi}$ で置き換える. すると, $\tilde{\Lambda}$ も確率変数となるので, $\tilde{\Lambda}$ と記すと, 式 (12) は次のようになる.

$$\tilde{\Lambda} = D \left(\hat{\Psi}^T \Lambda \hat{\Psi} \right) \quad (14)$$

式 (14) の両辺の標本固有値 $\hat{\Lambda}$ に関する条件付き期待値をとると, 次のようになる.

$$\begin{aligned} \tilde{\Lambda} &= E \left[\tilde{\Lambda} \mid \hat{\Lambda} \right] \\ &= E \left[D \left(\hat{\Psi}^T \Lambda \hat{\Psi} \right) \mid \hat{\Lambda} \right] \\ &= D \left(E \left[\hat{\Psi}^T \Lambda \hat{\Psi} \mid \hat{\Lambda} \right] \right) \end{aligned} \quad (15)$$

ここで $\tilde{\Lambda}$ は, 固有ベクトルの推定誤差を補償する固有値 $\tilde{\lambda}_i$ からなる行列

$$\tilde{\Lambda} = \text{diag} \left(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_d \right) \quad (16)$$

である. 式 (15) の第 i 対角要素を取り出すと,

$$\tilde{\lambda}_i = E \left[\sum_{j=1}^d \left\{ \hat{\psi}_{ji} \right\}^2 \lambda_j \mid \hat{\Lambda} \right] \quad (17)$$

$$= \sum_{j=1}^d E \left[\left\{ \hat{\psi}_{ji} \right\}^2 \mid \hat{\Lambda} \right] \lambda_j \quad (18)$$

となる. 更に

$$\left\{ \tilde{\psi}_{ji} \right\}^2 = E \left[\left\{ \hat{\psi}_{ji} \right\}^2 \mid \hat{\Lambda} \right] \quad (19)$$

とおくと, $\tilde{\lambda}_i$ は次式で得られる.

$$\tilde{\lambda}_i = \sum_{j=1}^d \left\{ \tilde{\psi}_{ji} \right\}^2 \lambda_j \quad (20)$$

2.3 式 (20) の計算法

本節では式 (20) の計算法を与える．式 (20) は $\left\{ \tilde{\psi}_{ji} \right\}^2$ が求まれば計算できる． $\left\{ \tilde{\psi}_{ji} \right\}^2$ を計算するため，式 (19) の標本固有値 $\hat{\Lambda}$ に関する条件付き期待値を一般化して， $\hat{\Psi}$ の関数 $f(\hat{\Psi})$ に関する条件付き期待値を考える．この積分表現は次のようになる．

$$E \left[f(\hat{\Psi}) \mid \hat{\Lambda} \right] = \int_{\hat{\Psi}} f(\hat{\Psi}) P(\hat{\Psi} \mid \hat{\Lambda}) d\hat{\Psi} \quad (21)$$

ここで $P(\hat{\Psi} \mid \hat{\Lambda})$ は，固有ベクトルの推定誤差 $\hat{\Psi}$ の確率密度関数である．高次元の場合，この期待値を厳密に求めるのは困難であるので，数値的に求めることを考える．モンテカルル口法を用いて計算する場合，正規分布に従うサンプルから標本固有値が $\hat{\Lambda}$ であるときの標本固有ベクトルを作成して，その誤差から期待値を計算する必要がある．しかし，これは実現不可能である．そこで本論文では，式 (21) を条件付きモンテカルル口法 [12] を用いて計算する． $f(\hat{\Psi}) = \left\{ \tilde{\psi}_{ji} \right\}^2$ として式 (21) を計算すれば，式 (19) から $\left\{ \tilde{\psi}_{ji} \right\}^2$ は式

(21) の右辺で与えられ，式 (20) の $\tilde{\lambda}_i$ が計算できる．

条件付きモンテカルル口法を実現するために，式 (21) の右辺を変形する．確率変数 $\hat{\Sigma}$ と $\hat{\Lambda}$ をそれぞれ

$$\hat{\Sigma} = \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T \quad (22)$$

を満たす対称行列と対角行列とする．固有ベクトルの推定誤差の分布は真の固有ベクトルによらないので，真の固有ベクトルが単位行列の場合 ($\Phi = I$) を考えても一般性を失わない．このとき，式 (6) から標本固有ベクトルと固有ベクトルの推定誤差が一致する ($\hat{\Phi} = \hat{\Psi}$) ので， $\hat{\Sigma} = \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T$ の分布は Wishart 分布 (付録参照) で与えられる．式 (22) の関係から，次が成立する．

$$\begin{aligned} P(\hat{\Sigma}) &= P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T) \\ &= P(\hat{\Psi}, \hat{\Lambda}) J(\hat{\Psi}, \hat{\Lambda}) \end{aligned} \quad (23)$$

ただし， $J(\hat{\Psi}, \hat{\Lambda})$ は変換のヤコビアンで，

$$J(\hat{\Psi}, \hat{\Lambda}) = \frac{d\hat{\Psi} d\hat{\Lambda}}{d\hat{\Sigma}} \quad (24)$$

である． $\hat{\Lambda}$ は確率変数 Λ の実現値であるから，

$$P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T) = P(\hat{\Psi}, \hat{\Lambda}) J(\hat{\Psi}, \hat{\Lambda}) \quad (25)$$

が成り立つ．更に，任意の関数 $g(\hat{\Lambda})$ を定義し，

$$G = \int_{\hat{\Lambda}} g(\hat{\Lambda}) d\hat{\Lambda} \quad (26)$$

とおく．以上の準備のもとで，次式が成り立つ．

$$\begin{aligned} & \int_{\hat{\Psi}} f(\hat{\Psi}) P(\hat{\Psi} \mid \hat{\Lambda}) d\hat{\Psi} \\ &= \int_{\hat{\Psi}} f(\hat{\Psi}) \frac{P(\hat{\Psi} \mid \hat{\Lambda})}{G} \left[\int_{\hat{\Lambda}} g(\hat{\Lambda}) d\hat{\Lambda} \right] d\hat{\Psi} \\ &= \int_{\hat{\Psi} \times \hat{\Lambda}} f(\hat{\Psi}) \frac{P(\hat{\Psi}, \hat{\Lambda})}{P(\hat{\Lambda})} \frac{g(\hat{\Lambda})}{G} d\hat{\Psi} d\hat{\Lambda} \\ &= \frac{1}{P(\hat{\Lambda})} \int_{\hat{\Sigma}} f(\hat{\Psi}) \frac{P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)}{J(\hat{\Psi}, \hat{\Lambda})} \frac{g(\hat{\Lambda})}{G} J(\hat{\Psi}, \hat{\Lambda}) d\hat{\Sigma} \\ &= \frac{1}{P(\hat{\Lambda})} \int_{\hat{\Sigma}} f(\hat{\Psi}) w_0(\hat{\Sigma}; \hat{\Lambda}) P(\hat{\Sigma}) d\hat{\Sigma} \end{aligned} \quad (27)$$

ただし，

$$w_0(\hat{\Sigma}; \hat{\Lambda}) = \frac{P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)}{J(\hat{\Psi}, \hat{\Lambda})} \frac{J(\hat{\Psi}, \hat{\Lambda})}{P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)} \frac{g(\hat{\Lambda})}{G} \quad (28)$$

である．式 (27) は，確率密度 $P(\hat{\Psi} \mid \hat{\Lambda})$ をもつ固有ベクトルを作成して $f(\hat{\Psi})$ の期待値をとることが，確率密度 $P(\hat{\Sigma})$ をもつ共分散行列を作成して $f(\hat{\Psi}) w_0(\hat{\Sigma}; \hat{\Lambda}) \frac{1}{P(\hat{\Lambda})}$ の期待値をとることに等しいことを意味する．この式の変形によって，乱数を用いた数値計算が可能になる．

$w_0(\hat{\Sigma}; \hat{\Lambda})$ は付録に示す式 (A-9)，(A-11) を式 (28) に代入して変形することにより，

$$\begin{aligned} w_0(\hat{\Sigma}; \hat{\Lambda}) &= \frac{|\hat{\Lambda}|^{\frac{1}{2}(n-d-2)} \prod_{i < j}^d (\lambda_i - \lambda_j)}{|\hat{\Lambda}|^{\frac{1}{2}(n-d-2)} \prod_{i < j}^d (\lambda_i - \lambda_j)} \\ & \cdot \frac{\exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T\right)}{\exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T\right)} \frac{g(\hat{\Lambda})}{G} \end{aligned} \quad (29)$$

となる．式 (27) の $P(\hat{\Lambda})$ は解析的に得られるが，特に高次元の場合，数値計算を行うのは困難である．しかし，式 (27) で $f(\hat{\Psi}) = 1$ の場合を考えると，確率の公理から $\int_{\hat{\Psi}} P(\hat{\Psi} \mid \hat{\Lambda}) d\hat{\Psi} = 1$ となり，

$$\frac{1}{P(\hat{\Lambda})} \int_{\hat{\Sigma}} w_0(\hat{\Sigma}; \hat{\Lambda}) P(\hat{\Sigma}) d\hat{\Sigma} = 1 \quad (30)$$

Algorithm 1 $\tilde{\lambda}_i$ の推定

- 1: 乱数を用いて正規分布 $N(0, \hat{\Lambda})$ に従う nt 個の人工サンプル X_1, \dots, X_{nt} を作成する .
- 2: for $k = 1$ to t do
- 3: n 個の人工サンプル $X_{n(k-1)+1}, \dots, X_{nk}$ から標本共分散行列 $\hat{\Sigma}_k$ を求め、標本固有値 $\hat{\Lambda}_k$ と固有ベクトルの推定誤差 $\hat{\Psi}_k$ を得る .
- 4: end for
- 5: 式 (33) を用いて $\{\hat{\Psi}_{ji}\}^2$ を求める .
- 6: 式 (20) で $\tilde{\lambda}_i$ を得る .

が成り立つので、 $P(\hat{\Lambda})$ は

$$P(\hat{\Lambda}) = \int_{\hat{\Sigma}} w_0(\hat{\Sigma}; \hat{\Lambda}) P(\hat{\Sigma}) d\hat{\Sigma} \quad (31)$$

で与えられる .

ここで改めて $f(\hat{\Psi}) = \{\hat{\psi}_{ji}\}^2$ の場合を考えると、式 (27) は式 (19) の右辺に等しいので、式 (27) の $P(\hat{\Lambda})$ に式 (31) を代入すると、式 (19) は次式で表される .

$$\{\tilde{\psi}_{ji}\}^2 = \frac{\int_{\hat{\Sigma}} \{\hat{\psi}_{ji}\}^2 w_0(\hat{\Sigma}; \hat{\Lambda}) P(\hat{\Sigma}) d\hat{\Sigma}}{\int_{\hat{\Sigma}} w_0(\hat{\Sigma}; \hat{\Lambda}) P(\hat{\Sigma}) d\hat{\Sigma}} \quad (32)$$

式 (32) の分子、分母の積分を Wishart 分布に従って生成した $\hat{\Sigma}$ の t 個のサンプル $\hat{\Sigma}_k$ による算術平均で表し、分子、分母の係数 $\frac{1}{t}$ を約分すると次式が得られる .

$$\{\tilde{\psi}_{ji}\}^2 = \frac{\sum_{k=1}^t \{\hat{\psi}_{k,ji}\}^2 w_0(\hat{\Psi}_k \hat{\Lambda}_k \hat{\Psi}_k^T; \hat{\Lambda})}{\sum_{k=1}^t w_0(\hat{\Psi}_k \hat{\Lambda}_k \hat{\Psi}_k^T; \hat{\Lambda})} \quad (33)$$

ここで、 $\hat{\Lambda}_k$ 、 $\hat{\Psi}_k$ は $\hat{\Sigma}_k$ の固有値、固有ベクトルからなる行列、 $\hat{\psi}_{k,ji}$ は $\hat{\Psi}_k$ の ji 要素である .

以上の議論から、条件付きモンテカルロ法を用いて式 (20) の $\tilde{\lambda}_i$ を推定する手順を Algorithm 1 に示す . n は実際のパターン認識において学習に使用できるサンプル数、 $\hat{\Lambda}$ は真の固有値、若しくは既存の固有値補正法で補正された標本固有値である . 本論文では、 $g(\hat{\Lambda}_k) = \frac{1}{t}$ とした .

2.4 標本固有ベクトルを用いて真のマハラノビス距離を推定する手順

実際のパターン認識において学習サンプルが n 個あるとき、以下の手順により、標本固有ベクトルを用いて真のマハラノビス距離を推定することができる .

(1) 辞書の作成に使用可能な n 個のサンプルから標本平均ベクトルと標本共分散行列を求め、標本固

有値 $\hat{\Lambda}$ と標本固有ベクトル $\hat{\Phi}$ を得る .

(2) 標本固有値を固有値の推定誤差を補正する固有値補正法 (例えば酒井らの手法 [10], James-Stein 推定量 [8] など) を適用して、標本固有値を補正した $\tilde{\Lambda}$ を得る .

(3) Algorithm 1 を用いて $\tilde{\lambda}_i$ を推定する .

(4) $\tilde{\lambda}_i$ を固有値として、標本固有ベクトル $\hat{\Phi}$ とともに認識に用いる .

3. 提案手法の性能評価

3.1 マハラノビス距離の推定値

標本固有ベクトルと提案手法で修正した固有値を用いたときに、マハラノビス距離が正しく推定できることを実験により確認する . 特徴量の次元数を固定して学習サンプル数を変化させる実験と、サンプル数を固定して特徴量の次元数を変化させる実験を行う .

3.1.1 次元数固定

特徴量の次元数を固定して、辞書の作成に用いる学習サンプル数を変化させる .

最初に、4 種類のマハラノビス距離 t_{ij} 、 s_{ij} 、 e_{ij} 、 p_{ij} を定義する . t_{ij} をクラス i の平均ベクトルとクラス i に属する j 番目の未知入力サンプルとの真のマハラノビス距離 $d(x)$ (式 (3)) とする . s_{ij} 、 e_{ij} 、 p_{ij} を、それぞれ異なる固有値と、共通の真の平均ベクトル、標本固有ベクトルを用いて求めたマハラノビス距離とする . s_{ij} は固有値として標本固有値を用いて求めたマハラノビス距離、 e_{ij} は真の固有値を用いて求めたマハラノビス距離 $\hat{d}(x)$ (式 (5))、 p_{ij} は提案手法で求めた固有値を用いて求めたマハラノビス距離 $\tilde{d}(x)$ (式 (13)) である . s_{ij} は共分散行列の推定誤差を一切補正しない場合、 e_{ij} は固有値の推定誤差のみ補正した場合、 p_{ij} は固有値の推定誤差を補正し、更に固有ベクトルの推定誤差を補償した場合に相当する . t_{ij} に対する s_{ij} の比の平均値 r_s 、 t_{ij} に対する e_{ij} の比の平均値 r_e 、 t_{ij} に対する p_{ij} の比の平均値 r_p を次のように定義する .

$$r_s = \frac{1}{cs} \sum_{i=1}^c \sum_{j=1}^s \frac{s_{ij}}{t_{ij}} \quad (34)$$

$$r_e = \frac{1}{cs} \sum_{i=1}^c \sum_{j=1}^s \frac{e_{ij}}{t_{ij}} \quad (35)$$

$$r_p = \frac{1}{cs} \sum_{i=1}^c \sum_{j=1}^s \frac{p_{ij}}{t_{ij}} \quad (36)$$

ここで c はクラス数, s はクラスごとのテストサンプル数である. r_s, r_e, r_p が 1 に近いほど, マハラノビス距離が正しく推定されているといえる. 本実験では r_e と r_p を比較する. 同時に r_s の結果も示す.

実験では, 正規分布に従う人工特徴量を用いる. 最初に文字画像の特徴量を作成し, 文字画像の特徴量から推定した分布パラメータを用いて人工特徴量を作成する. まず, NIST Special Database 19 [13] の数字サンプルを 64×64 の大きさに非線形正規化 [14] した後, 196 次元の方向線素特徴量 [15] を抽出し, 文字画像の特徴量とする. 次に, クラス i の平均ベクトル μ_i と共分散行列 Σ_i をクラスごとに文字画像の特徴量 36,000 文字分から推定する. 人工サンプルはこの推定値を真の値として, 正規分布 $N(\mu_i, \Sigma_i)$ に従うように乱数を用いて作成する. このように作成すると, 同じクラスの文字画像の特徴量と人工特徴量は平均がほぼ等しく, 分布の形が異なる特徴量になる.

辞書の作成に用いる学習サンプル数を各クラス 200 個から 100,000 個まで変化させ, 学習サンプルには含まれないテストサンプルを 1,000 個ずつ用いる. また, Algorithm 1 のパラメータ t を 10,000 とした.

r_s, r_e, r_p を図 1 に示す. r_e では辞書の推定に用いる学習サンプル数が少ないほど, マハラノビス距離の推定値が大きく偏っているが, r_p では学習サンプル数によらずほぼ正しく推定できている. r_s は r_e よりも更に大きく偏っている.

3.1.2 サンプル数固定

辞書の作成に用いる学習サンプル数を固定し, 特徴量の次元数を変化させる. 実験は 3.1.1 と同様, $r_s,$

r_e, r_p を比較する.

実験では, 3.1.1 とは別の 4 種類の正規分布に従う人工特徴量を用いる. 4 種類の特徴量の次元数はそれぞれ 4, 16, 64, 256 である. 3.1.1 と同様, 最初に文字画像の特徴量を作成し, 文字画像の特徴量から人工特徴量を作成する. 次元数 4 の文字画像の特徴量は, NIST Special Database 19 [13] の数字サンプルを 64×64 の大きさに非線形正規化 [14] した後, 重複しない 32×32 の大きさの黒画素数を 1 次元とする. 同様に, 次元数 16, 64, 256 の特徴量は非線形正規化後, 重複しない $16 \times 16, 8 \times 8, 4 \times 4$ の大きさの黒画素数を 1 次元とする. 文字画像の特徴量から人工サンプルを作成する手順は 3.1.1 と同じである.

辞書の作成に用いる学習サンプル数を 300 個, 学習サンプルには含まれないテストサンプルを 1,000 個ずつ用いる. また, Algorithm 1 のパラメータ t を 10,000 とした.

マハラノビス距離の比の平均値を図 2 に示す. r_e は特徴量の次元数が大きくなるほど, マハラノビス距離の推定値が大きく偏っているが, r_p は特徴量の次元数によらずほぼ正しく推定できている.

図 1, 図 2 から, 学習サンプル数が特徴量の次元数に比べて少ない場合でも, 提案手法を用いることでマハラノビス距離が正しく推定できることが確認できた. また, r_s は r_e よりも更に大きく偏っている.

2. では, $\hat{\Psi}^T \Lambda \hat{\Psi}$ の非対角成分を無視した. 非対角成分を無視しないとき, 式 (11) の導出過程から, マハラノビス距離の比の平均値の理論値はほぼ 1 になると考えられる. 図 1, 図 2 では r_p が学習サンプル数,

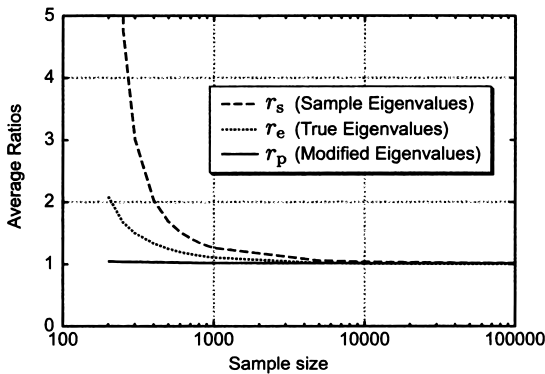


図 1 マハラノビス距離の比の平均値 (次元数固定)
Fig.1 The average ratio of Mahalanobis distances (fixed dimensionality).

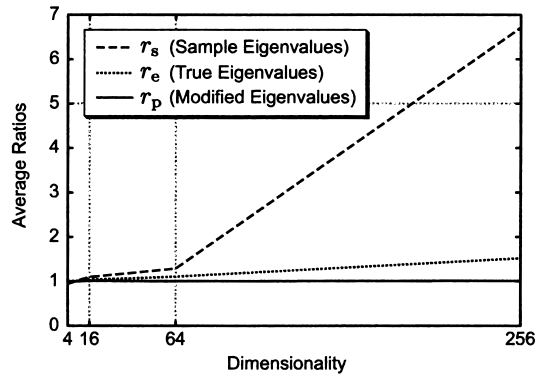


図 2 マハラノビス距離の比の平均値 (サンプル数固定)
Fig.2 The average ratio of Mahalanobis distances (fixed sample size).

特徴量の次元数に関係なく、ほぼ 1 であることから、 $\hat{\Psi}^T \Lambda \hat{\Psi}$ の非対角成分を無視することは妥当であると考えられる。

3.2 認識性能の確認

提案手法を統計的識別手法に適用したときの認識性能を確認するために、識別関数としてマハラノビス距離を用いた認識実験を行う。実験は、固有値だけが異なる 3 種類の辞書を作成し、それぞれの辞書を用いたときの認識率を比較する。3 種類の辞書は、固有値として標本固有値を用いる“Control”、真の固有値を用いる“True eigenvalue”、提案手法で修正した固有値を用いる“Proposed method”である。すべての辞書で共通の標本平均ベクトル、標本固有ベクトルを用いる。

サンプルは 3.1.1 で作成した文字画像の特徴量と正規分布に従う人工特徴量を用いる。辞書の作成に用いる学習サンプル数を文字画像の特徴量では各クラス 200 個から 35,000 個まで、人工サンプルでは 200 個から 100,000 個まで変化させる。テストサンプルは学習サンプルに含まれないサンプルを 1,000 個ずつ用いる。実際の文字画像では真の固有値が得られないため、酒井らの手法 [10] で補正した固有値を真の固有値の代わりに用いる。また、Algorithm 1 のパラメータ t を 10,000 とした。

実際の文字画像を用いた認識実験の結果を図 3 に、人工サンプルを用いたときの結果を図 4 に示す。実際の文字画像、人工サンプルともに、“Proposed method” の認識率はほぼすべての学習サンプル数で“True eigenvalue”より高く、提案手法を適用した識別関数は認識に有効であることが確認できた。“True eigenvalue”

と“Proposed method”の認識率の差を学習サンプル数との関係を考慮して見てみると、学習サンプル数が多いときには認識率にほとんど差はないが、少なくなるほど認識率の差は大きくなる。これは学習サンプル数が多いときには、補正されるべき推定誤差がもともと小さく、少ないときには認識性能に影響する固有ベクトルの推定誤差が大きいためであると考えられる。また、真の固有値を用いた“True eigenvalue”（文字画像の特徴量では補正した標本固有値）は、標本固有値をそのまま用いた“Control”よりも高い認識率が得られた。

3.3 Algorithm 1 のパラメータ t

提案手法を文字認識に適用する場合、Algorithm 1 で人工サンプルから共分散行列を推定する回数 t をどれくらいにすればよいのかを明らかにするため、提案手法が特に有効である学習サンプル数が少ない場合について認識実験を行う。実験はパラメータ t と学習サンプル数 n を変えることによって、3.2 の“Proposed method”の認識率がどう変化するかを調べる。

辞書の推定に用いる学習サンプル数を各クラス 200, 300, 500, 1,000, 5,000 ずつ、テストサンプルは学習サンプルに含まれないサンプルを 1,000 個ずつ用いる。 t は 5 から 1,500 まで 5 ずつ変化させる。

認識結果を図 5 に示す。まずパラメータ t と認識率の関係については、 t が小さい場合には認識率が変動し、 t が大きくなるほど認識率の変動が小さくなり、認識率の段階的な上昇が見られる。学習サンプル数 n とパラメータ t の関係については、 t が小さい場合の認識率の変動は、 n が大きいほど早く収まっている。こ

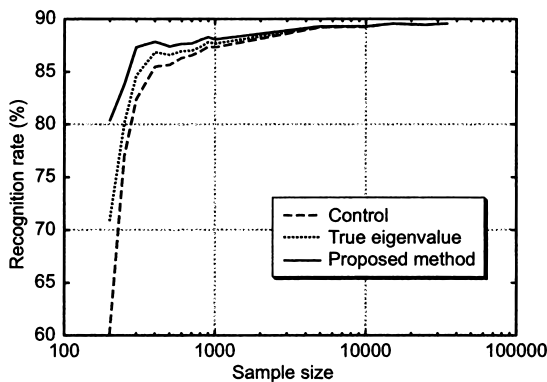


図 3 認識実験の結果 (文字画像)

Fig. 3 The recognition rates of the experiments on feature vectors of actual character samples.

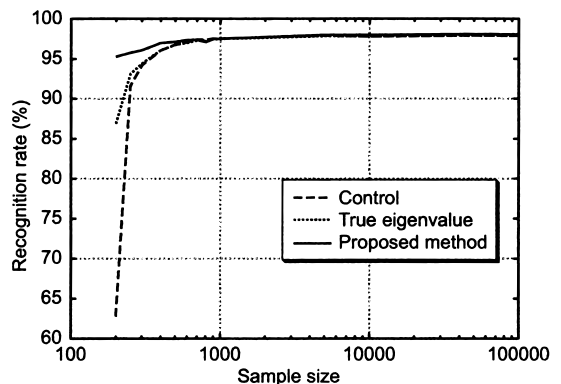


図 4 認識実験の結果 (人工サンプル)

Fig. 4 The recognition rates of the experiments on artificial feature vectors.

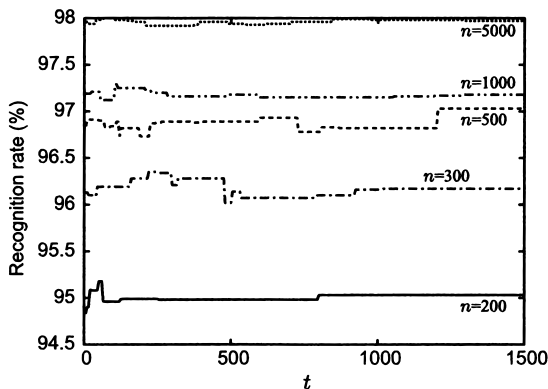


図5 学習サンプル数 n とパラメータ t と認識率の関係
Fig.5 The relation between sample size n , parameter t and recognition rates.

れは、 n が小さいほど、固有ベクトルの推定誤差の確率分布が大きく広がっているためであると考えられる。

4. む す び

本論文では、共分散行列の固有ベクトルの推定誤差によってマハラノビス距離に推定誤差が生じ、マハラノビス距離の推定誤差は標本固有ベクトルが真の固有値を「ゆがめる」ことによって生じると解釈できることを示した。そして、辞書の作成に用いるサンプルが少ない場合でも、固有ベクトルの推定誤差を補償するように真の固有値を修正することで、マハラノビス距離を正しく推定する手法を提案した。提案手法を文字認識に適用し、マハラノビス距離が正しく推定されること、認識性能が改善されることを確認した。マハラノビス距離の推定値を改善する効果はパラメータの推定に用いる学習サンプルが少ないほど大きく、本手法は学習サンプルを十分用意できない場合において特に有効である。

文 献

- [1] 坂野 鋭, 山田敬嗣, “怪奇!! 次元の呪い—識別問題, パターン認識, データマイニングの初心者のために (前編);” 情報処理, vol.43, no.5, pp.562-567, March 2002.
- [2] 萩田紀博, 内藤誠一郎, 増田 功, “外郭方向寄与度特徴による手書き漢字の識別” 信学論 (D), vol.J66-D, no.10, pp.1185-1192, Oct. 1983.
- [3] 竹下鉄夫, 鳥脇純一郎, “マハラノビス距離の推定値の確率分布に関する一性質” 信学論 (D-II), vol.J76-D-II, no.1, pp.59-64, Jan. 1993.
- [4] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, “Modified quadratic discriminant functions and the application to Chinese character recogni-

- tion,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.9, no.1, pp.149-153, Jan. 1987.
- [5] 加藤 寧, 安倍正人, 根元義章, “改良型マハラノビス距離を用いた高精度な手書き文字認識” 信学論 (D-II), vol.J79-D-II, no.1, pp.45-52, Jan. 1996.
- [6] F. Sun, S. Omachi, and H. Aso, “Precise selection of candidates for handwritten character recognition using feature regions,” IEICE Trans. Inf. & Syst., vol.E79-D, no.5, pp.510-515, May 1996.
- [7] J.H. Friedman, “Regularized discriminant analysis,” J. American Statistical Association, vol.84, no.405, pp.165-175, 1989.
- [8] W. James and C. Stein, “Estimation with quadratic loss,” Proc. 4th Berkeley Symp. on Math. Statist. and Prob., pp.361-379, 1961.
- [9] T. Takeshita and J. Toriwaki, “Experimental study of performance of pattern classifiers and the size of design samples,” Pattern Recognit. Lett., vol.16, no.3, pp.307-312, March 1995.
- [10] 酒井 充, 米田政明, 長谷博行, 丸山 博, 直江美知子, “固有値の偏り補正に基づく2次識別関数” 信学論 (D-II), vol.J82-D-II, no.4, pp.631-640, April 1999.
- [11] 酒井 充, 米田政明, 長谷博行, 丸山 博, “固有値を二重に補正した共分散行列を用いた2次識別関数” 信学技報, PRMU99-30, June 1999.
- [12] 津田孝夫, モンテカルロ法とシミュレーション—電子計算機の確率論的応用, 三訂版, 培風館, 東京, 1995.
- [13] P.J. Grother, “NIST special database 19 — Hand-printed forms and characters database,” Technical Report, National Institute of Standards and Technology, March 1995.
- [14] 山田博三, 斉藤泰一, 山本和彦, “線密度イコライゼーション—相関法のための非線形正規化法” 信学論 (D), vol.J67-D, no.11, pp.1379-1383, Nov. 1984.
- [15] 孫 寧, 田原 透, 阿曾弘具, 木村正行, “方向線素特徴量を用いた高精度文字認識” 信学論 (D-II), vol.J74-D-II, no.3, pp.330-339, March 1991.
- [16] 竹村彰通, 多変量推測統計の基礎, 共立出版, 東京, 1991.
- [17] R.J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons, New York, 1982.

付 録

固有ベクトルの推定誤差の確率密度関数

(i) Wishart 分布

正規分布 $N(\mathbf{0}, \Sigma)$ に従う各々独立な n 個の d 次元確率ベクトル $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ の平方和積和行列 \mathbf{W} を

$$\mathbf{W} = \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t^T \quad (\text{A}\cdot 1)$$

とおくと、 Σ が与えられたときの \mathbf{W} の分布 $W_d(n, \Sigma)$ は Wishart 分布として知られている [16], [17]。その確

率密度関数は次式で与えられる .

$$P(\mathbf{W}|\Sigma) = \frac{v(d, n) |\mathbf{W}|^{\frac{1}{2}(n-d-1)}}{|\Sigma|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{W}\right) \quad (\text{A}\cdot 2)$$

ここで $v(d, n)$ は

$$v(d, n) = \frac{1}{2^{\frac{1}{2}nd} \pi^{\frac{1}{4}d(d-1)} \prod_{i=1}^d \Gamma\left[\frac{1}{2}(n+1-i)\right]} \quad (\text{A}\cdot 3)$$

である .

(ii) 標本共分散行列の確率密度関数
標本共分散行列 $\hat{\Sigma}$, 標本平均ベクトル $\hat{\mu}$ を

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{X}_t - \hat{\mu})(\mathbf{X}_t - \hat{\mu})^T \quad (\text{A}\cdot 4)$$

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t \quad (\text{A}\cdot 5)$$

とおくと, $\hat{\Sigma}$ の分布は $W_d(n-1, \frac{1}{n-1}\Sigma)$ で与えられ, $\hat{\Sigma}$ の確率密度関数は

$$P(\hat{\Sigma}|\Sigma) = (n-1)^{\frac{1}{2}(n-1)d} v(d, n-1) \frac{|\hat{\Sigma}|^{\frac{1}{2}(n-d-2)} \exp\left(-\frac{n-1}{2} \text{tr} \Sigma^{-1} \hat{\Sigma}\right)}{|\Sigma|^{\frac{1}{2}(n-1)}} \quad (\text{A}\cdot 6)$$

となる [16], [17] .

(iii) 固有値, 固有ベクトルの推定誤差の同時確率密度関数

共分散行列 Σ , 標本共分散行列 $\hat{\Sigma}$ は固有値展開により,

$$\Sigma = \Phi \Lambda \Phi^T \quad (\text{A}\cdot 7)$$

$$\begin{aligned} \hat{\Sigma} &= \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T \\ &= \Phi \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T \Phi^T \end{aligned} \quad (\text{A}\cdot 8)$$

となる . 式 (A.6) は (A.7), (A.8) を用いて次式のよ
うに書き直すことができる .

$$P(\Phi \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T \Phi^T | \Phi \Lambda \Phi^T) = (n-1)^{\frac{1}{2}(n-1)d} v(d, n-1) \frac{|\hat{\Lambda}|^{\frac{1}{2}(n-d-2)} \exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \hat{\Psi} \hat{\Lambda} \hat{\Psi}^T\right)}{|\Lambda|^{\frac{1}{2}(n-1)}} \quad (\text{A}\cdot 9)$$

式 (A.9) は Φ に依存しないので, $P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T | \Lambda)$ となり, $P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)$ と書く . この $P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)$ は

$$P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T) = P(\hat{\Psi}, \hat{\Lambda}) J(\hat{\Psi}, \hat{\Lambda}) \quad (\text{A}\cdot 10)$$

と変形できる . ヤコビアン $J(\hat{\Psi}, \hat{\Lambda})$ は式 (A.9) より,

$$J(\hat{\Psi}, \hat{\Lambda}) = \frac{\prod_{i=1}^d \Gamma\left[\frac{1}{2}(d+1-i)\right]}{2^d \pi^{\frac{1}{4}d(d+1)}} \cdot \frac{1}{\prod_{i<j}^d (\hat{\lambda}_i - \hat{\lambda}_j)} \quad (\text{A}\cdot 11)$$

である [16], [17] .

(iv) 固有ベクトルの推定誤差の確率密度関数

固有ベクトルの推定誤差の確率密度関数 $P(\hat{\Psi} | \hat{\Lambda})$ は式 (A.10) を用いて,

$$P(\hat{\Psi} | \hat{\Lambda}) = \frac{P(\hat{\Psi}, \hat{\Lambda})}{P(\hat{\Lambda})} \quad (\text{A}\cdot 12)$$

$$= \frac{P(\hat{\Psi} \hat{\Lambda} \hat{\Psi}^T)}{P(\hat{\Lambda}) J(\hat{\Psi}, \hat{\Lambda})} \quad (\text{A}\cdot 13)$$

となる . 式 (A.10), (A.11) から, $P(\hat{\Psi} | \hat{\Lambda})$ は標本固有値 $\hat{\Lambda}$ が与えられたとき, 次元数 d , サンプル数 n , 真の固有値 Λ のみから決まり, 真の固有ベクトル Φ に依存していないことがわかる .

(平成 14 年 2 月 21 日受付, 8 月 27 日再受付)

岩村 雅一 (学生員)



平 10 東北大・工・通信卒 . 平 12 同大大学院博士課程前期了 . 現在, 同大学院博士課程後期在学中 . パターン認識に関する研究に従事 .



大町真一郎 (正員)

昭 63 東北大・工・情報卒。平 5 同大大学院博士課程了。同年同大情報処理教育センター助手。平 8 同大工学部助手。平 11 同大大学院工学研究科助教授，現在に至る。博士(工学)。その間，平 12~13 米国ブラウン大学客員助教授。パターン認識，コンピュータビジョン，並列処理，文字認識システムの開発などの研究に従事。IEEE，情報処理学会，人工知能学会，Pattern Recognition Society 各会員。



阿曾 弘具 (正員)

昭 43 東北大・工・電気卒。昭 49 同大大学院博士課程了。昭 48 東北大・工・助手，昭 54 名大・工・講師。昭 57 同助教授，昭 61 東北大・工・助教授を経て，平 3 同教授。現在，同大学院工学研究科教授。工博。その間，学習オートマトン，セル構造オートマトン，並行処理理論，シストリックアルゴリズム設計論，文字認識，音声認識，ニューラルネットワークなどの研究に従事。平 3 年度本学会業績賞受賞。IEEE，ACM，EATCS，情報処理学会，人工知能学会，日本認知科学会，LA 各会員。