

逐次学習型の線形判別分析アルゴリズムについて — 学習アルゴリズムの提案と局所収束性の証明 —

平岡 和幸 濱平 仁

埼玉大学 工学部 情報システム工学科
埼玉県 浦和市 下大久保 255

電話：048-858-5542 fax：048-858-3716

email：hira@me.ics.saitama-u.ac.jp

あらまし

線形判別分析は、画像認識等で広く用いられている。しかし、次元数 N の大きいデータに線形判別分析を適用しようとすると、 $N \times N$ の大規模行列を扱わなくてはならず、計算が困難になる。

本研究では、オンライン型の線形判別分析アルゴリズムを提案し、その局所収束性を示す。このアルゴリズムは、「新たなデータに応じて、少ない計算量で解の修正を行える」という意味で逐次学習が可能であり、環境変化への適応能力を持つ。その上、高次元のデータに対しても大規模な行列を保持・更新する必要がない、という利点を持つ。

キーワード

線形判別分析, オンライン学習, 確率近似法, 行列力学系, 局所収束性

On Successive Learning Type Algorithm for Linear Discriminant Analysis:

Proposal of Learning Algorithm and Proof of Local Convergence

Kazuyuki HIRAOKA, Masashi HAMAHIRA

Dept. of Information and Computer Sciences Saitama University

255 Shimo-okubo, Urawa, 338, Japan

tel: +81-48-852-2111, fax: +81-48-858-3716

email: hira@me.ics.saitama-u.ac.jp

Abstract

Linear Discriminant Analysis (LDA) is widely used in various areas, e.g. image recognition. However, when we try to apply LDA for large dimensional data, we are confronted with a problem that we have to treat $N \times N$ huge matrix when the dimension of data is N .

In the present study, an online algorithm for LDA is proposed and its local convergence is proved. This algorithm has the ability of successive learning in the sense that correction of the solution according to new additional data can be executed with small computational cost. Thus it has the ability of adaptation to new environment. Moreover, our algorithm has the advantage that there is no need to maintain or calculate a huge matrix.

linear discriminant analysis, online learning, stochastic approximation, matrix dynamics, local convergence

key words

1 はじめに

線形判別分析 (Linear Discriminant Analysis, LDA) は、画像認識等で広く用いられている [8][9]。しかし、次元数 N の大きいデータに線形判別分析を適用しようとする、 $N \times N$ の大規模行列を扱わなくてはならず、計算が困難になる。

一方、主成分分析については、逐次学習が可能なオンラインアルゴリズムが提案されており [1][4]、これを用いれば、 $N \times N$ 行列を生成することなく主成分を求めることができる。これと同様のアルゴリズムを、線形判別分析に関して構成することが、本研究の目的である。

これまでに、オンライン型の線形判別分析は、いくつか提案されている [5][6]。これらは、「新たなデータに応じて、少ない計算量で解の修正を行える」という意味では逐次型である。しかし、これらには、 $N \times N$ 行列を保持・更新しなくてはならないという欠点がある。本研究で提案するアルゴリズムは、上の意味で逐次型であり、環境変化に適応する能力を持つ。その上、 $N \times N$ 行列を生成する必要がないという利点を持つ。

以下、第2節で、基本となる行列力学系を示し、第3節で、提案するアルゴリズムを述べる。アルゴリズムの局所収束性を保証するために、第4節で不動点を、第5節で安定性を調べる。第6節ではシミュレーション結果を示し、最後に、第7節でまとめを行う。

2 線形判別分析のための行列力学系

線形判別分析は、結局、一般化固有値問題

$$Bp = \lambda Wp, \quad (\lambda \in \mathbf{R}) \quad (1)$$

を解くことに帰着される [10]。与えられた B, W に対して、(1) を満たす N 次元ベクトル p と正数 λ を求める問題が、一般化固有値問題である。線

形判別分析の場合、 B はクラス間分散行列 ($N \times N$ 、非負定値)、 W はクラス内分散行列 ($N \times N$ 、通常は正定値) に対応する。そして、できるだけ大きな λ に対応する p を、比較的少数本、求めることが要求される。

本研究では、この一般化固有値問題を解くために、行列力学系

$$\frac{dA}{dt} = BA - \frac{1}{2}BAA^TWA - \frac{1}{2}WAA^TBA \quad (2)$$

を利用する。 $A(t)$ は $N \times K$ の行列である。この力学系の不動点 A_0 が一般化固有値問題の解に対応することが、後の節で示される。特に、安定な不動点は、「大きい方から K 個の一般化固有値に対応する一般化固有ベクトルを並べた行列」となっている¹。これはまさに、線形判別分析で求めたかった行列である。

3 オンライン線形判別分析アルゴリズム

行列力学系 (2) に基づいて、以下のように、オンライン線形判別分析アルゴリズムが得られる。

例題として、 N 次元データベクトルとクラス名の組 $(\mathbf{x}(1), c(1)), (\mathbf{x}(2), c(2)), \dots$ が、時々刻々与えられるとする。新たな例題が与えられるたびに、データの平均ベクトル $\bar{\mathbf{x}}$ と、クラス c のクラス内平均ベクトル $\bar{\mathbf{x}}^c$ を、次のように更新していく。

$$\bar{\mathbf{x}}^c(t) = \bar{\mathbf{x}}^c(t-1) + \eta \delta(c, c(t)) (\mathbf{x}(t) - \bar{\mathbf{x}}^c(t-1)) \quad (3)$$

$$\bar{\mathbf{x}}(t) = \bar{\mathbf{x}}(t-1) + \eta (\mathbf{x}(t) - \bar{\mathbf{x}}(t-1)) \quad (4)$$

ここに、 $\delta(c, c')$ は、 $c = c'$ のとき 1、 $c \neq c'$ のとき 0 とする。また、 η は学習係数 (小さな正の数)

¹ 正確に言うと、それらの一般化固有ベクトルで張られる線形部分空間の W -正規直交基底が得られる。判別分析の場合には、このような基底が得られれば十分である。

である。さらに、クラス間分散行列 B 、および、
データとクラス平均の誤差ベクトル $w(t)$ を、次
のように求める²。

$$w(t) = x(t) - \bar{x}^{c(t)}(t) \quad (5)$$

$$B(t) = \frac{1}{C} \sum_c (\bar{x}^c(t) - \bar{x}(t)) (\bar{x}^c(t) - \bar{x}(t))^T \quad (6)$$

ここに、 C はクラス数で、 $C \ll N$ を想定している。
これらを用いて、

$$A(t) = A(t-1) + \eta \left(BA - \frac{1}{2} BAA^T w w^T A - \frac{1}{2} w w^T AA^T BA \right) \quad (7)$$

という更新則で、 $A(t)$ を修正していく。その際、
 η を適当な速さでゼロに近づけていけば、 $A(t)$ は
(2) の安定不動点に局所収束することが、確率近
似法の一般論からわかる [7]。

そこで、以下の節では、(2) の不動点とその局
所安定性について述べる。

4 不動点について

Lemma 1 行列力学系 (2) は、(ユークリッド計
量での) 直交変換に関して、次の意味で不変であ
る： $A = A(t)$ が、行列力学系 (2) を満たすとす
る。このとき、 $N \times N$ 直交行列 Q および $K \times K$
直交行列 R を用いて

$$\hat{A}(t) = QA(t)R \quad (8)$$

と変数変換すると、 $\hat{A}(t)$ の満たす行列力学系は、

$$\frac{d\hat{A}}{dt} = \hat{B}\hat{A} - \frac{1}{2}\hat{B}\hat{A}\hat{A}^T\hat{W}\hat{A} - \frac{1}{2}\hat{W}\hat{A}\hat{A}^T\hat{B}\hat{A} \quad (9)$$

となる。ここに、

$$\hat{W} = QWQ^T \quad (10)$$

$$\hat{B} = QBQ^T \quad (11)$$

²実際には、 B 自体を陽に計算する必要はない。(7) で
は $BA = \frac{1}{C} \sum_c (\bar{x}^c(t) - \bar{x}(t)) [(\bar{x}^c(t) - \bar{x}(t))^T A]$ しか現
れないので、 $N \times N$ 行列を作る必要はない。

Proof: 略

Lemma 2 $A(t) = A_0$ を力学系 (2) の不動点と
する。このとき、

$$\Gamma = A_0^T B A_0 \quad (12)$$

とおくと、

$$A_0^T W A_0 \Gamma = \Gamma \quad (13)$$

が成り立つ。

Proof: 不動点では、

$$B A_0 A_0^T W A_0 + W A_0 A_0^T B A_0 = 2 B A_0 \quad (14)$$

が成り立っている。すると特に、 $Z = A_0^T W A_0 - I$
(I は単位行列) において、

$$\Gamma Z + Z \Gamma = 0 \quad (15)$$

が成り立つ。ところが一般に、「二つの対称行列
 X, Y の積 XY が反対称なら、 $XY = 0$ 」が示さ
れる。よって、

$$Z \Gamma = (A_0^T W A_0 - I) \Gamma = 0 \quad (16)$$

である。

Theorem 1 (不動点の条件)

$\Gamma = A_0^T B A_0$ が正則であると仮定する。このと
き、 $A(t) = A_0$ が力学系 (2) の不動点となる必要
十分条件は、

$$B A_0 = W A_0 \Gamma \quad (17)$$

$$A_0^T W A_0 = I \quad (18)$$

が成り立つことである。

Proof: 十分性は自明である。必要性について、
(18) は (13) から明らか。また、(17) も、(18) を
(14) にすれば得られる。

Corollary 1 (不動点の条件) $\Gamma = A_0^T B A_0$ が正則であると仮定する。このとき、 $A(t) = A_0$ が力学系 (2) の不動点となる必要十分条件は、行列 P と直交行列 Q により

$$A_0 = PQ \quad (19)$$

と表されることである。ここに、 P は、各列ベクトル $\mathbf{p}_1, \dots, \mathbf{p}_K$ が一般化固有値問題 (1) の解となっており、さらに

$$\mathbf{p}_k^T W \mathbf{p}_l = \begin{cases} 1 & (k=l) \\ 0 & (k \neq l) \end{cases} \quad (20)$$

を満たすような行列である。□

Proof: 十分性は定理 1 より自明。必要性を示す。まず、 $\hat{B} = \sqrt{W}^{-1} B \sqrt{W}^{-1}$, $\hat{A}_0 = \sqrt{W} A_0$ を改めて B, A_0 とおきなおすことにより、はじめから $W = I$ と仮定して一般性を失わない。このとき、定理 1 から、

$$B A_0 = A_0 \Gamma \quad (21)$$

が成り立っている。よって特に、

$$B(\text{Image} A_0) \subset \text{Image} A_0 \quad (22)$$

である。これは、 A_0 の列ベクトル達が、 K 本の固有ベクトル $\mathbf{p}_1, \dots, \mathbf{p}_K$ で張られることを意味している。したがって、正方行列 Q により、

$$A_0 = PQ \quad (23)$$

と表すことができる。さらに、定理 1 から $A_0^T W A_0 = I$ なので、

$$Q^T Q = Q^T P^T W P Q = I \quad (24)$$

も言える。■

5 安定性について

Theorem 2 (安定性の条件)

一般化固有値問題 (1)(20) の解を $(\lambda_1, \mathbf{p}_1), \dots, (\lambda_N, \mathbf{p}_N)$ とし、 $\lambda_1 \geq \dots \geq \lambda_N$ とする。さらに、縦ベクトル $\mathbf{p}_1, \dots, \mathbf{p}_K$ を並べた行列を P とおく。このとき、直交行列 Q により

$$A_0 = PQ \quad (25)$$

と表される行列 A_0 は、行列力学系 (2) の安定不動点である。逆に、(2) の安定不動点 A_0 について、 $\Gamma = A_0^T B A_0$ が正則であれば、 A_0 は上の形で表すことができる。□

Proof: $N \times K$ 行列 X, Y に対し、内積を

$$\langle X, Y \rangle = \text{Tr}(X^T Y) = \sum_{i=1}^N \sum_{j=1}^K X_{ij} Y_{ij} \quad (26)$$

で定義する。 A_0 を力学系 (2) の不動点とし、 Γ が正則と仮定する。 $A(t)$ を、 A_0 の近傍から出発する (2) の解とし、

$$\Delta A(t) = A(t) - A_0 \quad (27)$$

とおく。

$$\begin{aligned} \frac{d\Delta A}{dt} &= \frac{1}{2} B \Delta A (I - A_0^T W A_0) \\ &\quad - \frac{1}{2} B A_0 (\Delta A^T W A_0 + A_0^T W \Delta A) \\ &\quad + \frac{1}{2} (I - W A_0 A_0^T) B \Delta A \\ &\quad - \frac{1}{2} W (\Delta A A_0^T + A_0 \Delta A^T) B A_0 \\ &\quad + o(\Delta A) \end{aligned} \quad (28)$$

$$\begin{aligned} &= -\frac{1}{2} W A_0 \Gamma \Delta A^T W A_0 - \frac{1}{2} W A_0 \Gamma A_0^T W \Delta A \\ &\quad + \frac{1}{2} B \Delta A - \frac{1}{2} W A_0 \Gamma A_0^T W \Delta A \\ &\quad - \frac{1}{2} W \Delta A \Gamma - \frac{1}{2} W A_0^T \Delta A^T W A_0 \Gamma \\ &\quad + o(\Delta A) \end{aligned} \quad (29)$$

$$\equiv D(\Delta A) + o(\Delta A) \quad (30)$$

より、

$$\begin{aligned} \frac{d}{dt} \langle \Delta A, \Delta A \rangle &= 2\text{Tr}(\Delta A^T \frac{d\Delta A}{dt}) \quad (31) \\ &= \text{Tr}(\Delta A^T B \Delta A - \Delta A^T W \Delta A \Gamma) \\ &\quad - 2\text{Tr}((\Delta A^T W A_0)^2 \Gamma) \\ &\quad - 2\text{Tr}((\Delta A^T W A_0)^T (\Delta A^T W A_0) \Gamma) \\ &\quad + o(\|\Delta A\|^2) \quad (32) \end{aligned}$$

である。これが常に負であれば、不動点 A_0 は安定ということになる。逆に、これが正になる ΔA が存在すれば、 A_0 は不安定である。このことは、線形演算子 \mathcal{D} が内積 $\langle \cdot, \cdot \rangle$ に関して対称なことから言える。

さて、 $A_0 = PQ$ と表される A_0 について考えよう。 $\mathbf{p}_1, \dots, \mathbf{p}_N$ の線形独立性から、 ΔA は

$$\Delta A = A_0 G_0 + A_1 G_1 \quad (33)$$

と表すことができる。ここに、 G_0 は $K \times K$ の任意行列、 G_1 は $(N-K) \times K$ の任意行列であり、

$$A_1 = (\mathbf{p}_{K+1}, \dots, \mathbf{p}_N) \quad (34)$$

とおく。 $\mathbf{p}_1, \dots, \mathbf{p}_N$ の W -正規直交性から、

$$A_0^T W A_0 = I \quad (35)$$

$$A_1^T W A_1 = I \quad (36)$$

$$A_0^T W A_1 = O \quad (37)$$

である。また、

$$A_0^T B A_0 = \Gamma \equiv \Gamma_0 \quad (38)$$

$$A_1^T B A_1 = \text{diag}(\lambda_{K+1}, \dots, \lambda_N) \equiv \Gamma_1 \quad (39)$$

$$A_0^T B A_1 = O \quad (40)$$

である。すると、

$$\begin{aligned} \frac{d}{dt} \langle \Delta A, \Delta A \rangle &= \text{Tr} \left(\Delta A^T (W A_0 \Gamma_0 G_0 + W A_1 \Gamma_1 G_1) \right. \\ &\quad \left. - G_0^T G_0 \Gamma_0 - G_1^T G_1 \Gamma_0 \right) \\ &\quad - 2\text{Tr}(G_0^T G_0 \Gamma_0) - 2\text{Tr}(G_0 G_0^T \Gamma_0) \quad (41) \end{aligned}$$

$$\begin{aligned} &= \text{Tr} \left(G_0^T \Gamma_0 G_0 + G_1^T \Gamma_1 G_1 \right. \\ &\quad \left. - G_0^T G_0 \Gamma_0 - G_1^T G_1 \Gamma_0 \right) \\ &\quad - 2\text{Tr}(G_0^2 \Gamma_0) - 2\text{Tr}(G_0 G_0^T \Gamma_0) \quad (42) \end{aligned}$$

$$\begin{aligned} &= -\text{Tr} \left(G_0 G_0^T \Gamma_0 + G_0^T G_0 \Gamma_0 \right. \\ &\quad \left. + G_0 G_0 \Gamma_0 + G_0^T G_0^T \Gamma_0 \right) \\ &\quad - \text{Tr}(G_1^T G_1 \Gamma_0 - G_1 G_1^T \Gamma_1) \quad (43) \end{aligned}$$

$$\begin{aligned} &= -\text{Tr}((G_0 + G_0^T)^2 \Gamma_0) \\ &\quad - \text{Tr}(G_1^T G_1 \Gamma_0 - G_1 G_1^T \Gamma_1) \quad (44) \end{aligned}$$

となる。

ここで、 Γ_0 が正定値であることから、任意の G_0 に対して

$$\text{Tr}((G_0 + G_0^T)^2 \Gamma_0) \geq 0 \quad (45)$$

また、 $\lambda_1 \geq \dots \geq \lambda_N$ なら、

$$\text{Tr}(G_1^T G_1 \Gamma_0) \geq \lambda_K \text{Tr}(G_1^T G_1) \quad (46)$$

$$\geq \lambda_{K+1} \text{Tr}(G_1^T G_1) \quad (47)$$

$$\geq \text{Tr}(G_1 G_1^T \Gamma_1) \quad (48)$$

だから、

$$\frac{d}{dt} \langle \Delta A, \Delta A \rangle \leq 0 \quad (49)$$

が言えて、 A_0 の安定性が示される。なお、等号が成立するのは、 G_0 が反対称で、かつ、 B が重複固有値 $\lambda_K = \dots = \lambda_{K+L}$ を持ち、 $A_1 G_1$ の各列ベクトルが $\mathbf{p}_{K+1}, \dots, \mathbf{p}_{K+L}$ の線形結合となっている場合である。つまり、等号が成立する ΔA は、本定理の条件を満たす不動点 A_0 達のなす曲面 \mathcal{A} の接線方向のみである。

逆については、以下のように示される。今、

$$\min(\lambda_1, \dots, \lambda_K) < \max(\lambda_{K+1}, \dots, \lambda_N) \quad (50)$$

となっていたとしよう。たとえば、 $\lambda_K < \lambda_N$ だったとする。このとき、

$$G_0 = 0, \quad G_1 = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad (51)$$

ととれば、

$$\begin{aligned} \frac{d}{dt} \langle \Delta A, \Delta A \rangle \\ &= -\text{Tr}(G_1^T G_1 \Gamma_0 - G_1 G_1^T \Gamma_1) \quad (52) \\ &= -(\lambda_K - \lambda_N) > 0 \quad (53) \end{aligned}$$

となってしまう。これは、不動点 A_0 が不安定なことを意味する。■

6 シミュレーション

顔画像からの個人識別について、本アルゴリズムを適用した例を示す。

顔画像は、10ピクセル×10ピクセルとした。したがって、データの次元数は $10 \times 10 = 100$ となる。各ピクセルは -1 から $+1$ の256段階のグレースケール値を持つ (Fig. 1)。

6.1 学習能力の確認

提案したアルゴリズムを、人物 A, B の2人の顔画像の識別問題に適用した。各人について50枚の顔画像を撮影し、実験用のサンプルとして使用した。これらの画像が、交互に与えられる：

$$\begin{aligned} \mathbf{x}(1) &= \text{image}_{A,1}, & \mathbf{x}(2) &= \text{image}_{B,1}, \\ \mathbf{x}(3) &= \text{image}_{A,2}, & \mathbf{x}(4) &= \text{image}_{B,2}, \quad \dots, \\ \mathbf{x}(99) &= \text{image}_{A,50}, & \mathbf{x}(100) &= \text{image}_{B,50} \end{aligned}$$

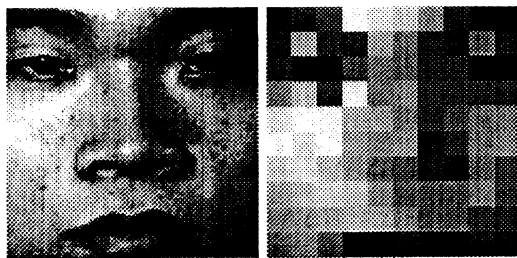


図 1: 実験に用いた画像の例 (左: 元画像、右: 10 × 10 ピクセル画像)

線形判別分析で求める特徴数 (A の列数) は $K = 2$ とし、学習係数 η は 0.1 とした。

学習の結果、これらの画像の97%を正しく判別できるようになった。結果を図2 ($t = 100$ まで) に示す。ステップ t は、提示された画像の総枚数に一致する。判別性能は、10ステップごとに評価した。評価には、全画像100枚を提示し、正しく判別する比率を調べた。なお、評価の間、学習は停止した。

6.2 環境変化

前節で述べたように、 $t = 100$ までは2人の人物 A, B ($c = 1, 2$) を提示した。 $t = 100$ において、突然、第3の人物 C を登場させ、C に対応するクラス $c = 3$ を追加した。

$t = 100$ 以降は、識別性能は、人物 A, B, C の全画像150枚で評価した。識別性能は、環境変化により、一時的に80%近くまで低下した。

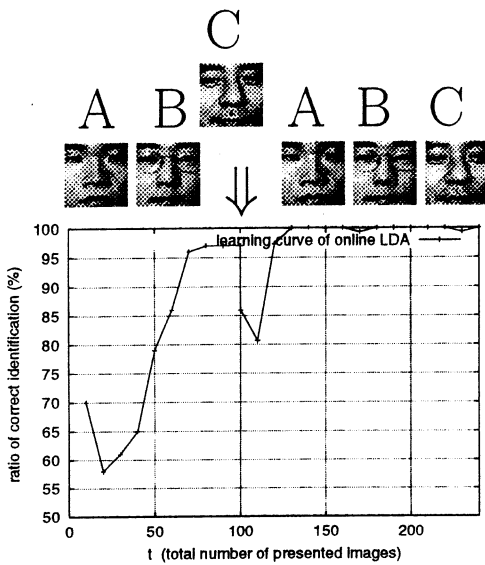


図 2: 実験で得られた学習曲線。(t = 100 までは、2 人の人物 (A, B) の顔画像を交互に提示した。t = 100 において、第 3 の人物 (C) のクラスを追加し、以降は 3 人の顔画像を順に提示した。本研究のアルゴリズムは、環境の変化に適応することができる。)

6.3 適応能力の確認

t = 100 以降は、人物 A, B, C それぞれ 50 枚ずつの画像を、順に提示した。

$$\begin{aligned} \mathbf{x}(101) &= \text{image}_{A,1}, & \mathbf{x}(102) &= \text{image}_{B,1}, \\ \mathbf{x}(103) &= \text{image}_{C,1}, & \mathbf{x}(104) &= \text{image}_{A,2}, \\ \mathbf{x}(105) &= \text{image}_{B,2}, & \mathbf{x}(106) &= \text{image}_{C,2}, \quad \dots \end{aligned}$$

提案したアルゴリズムは、新しい環境に適応し、ほぼ 100% の識別率に回復した。

7 おわりに

行列力学系 (2) に基づき、逐次学習が可能なオンライン線形判別分析アルゴリズムを構成した。

さらに、(2) の不動点とその安定性を調べることで、アルゴリズムの局所収束性を示した。また、本アルゴリズムを顔画像識別に適用し、学習能力と、環境変化への適応能力を確認した。今後は、大域収束性の解析を行うとともに、本手法をより大規模な問題に適用し有効性を検証する予定である。This work has been partly supported by CREST of JST (Japan Science and Technology) 279102.

参考文献

- [1] E. Oja, "A simplified neuron model as a principal component analyzer", J. Math. Biol., Vol. 15, pp. 267-273, 1982.
- [2] Wie-Yong Yan, Uwe Helmke, and John B. Moore, "Global analysis of Oja's flow for neural networks", IEEE Tr. NN, Vol. 5, No. 5, pp. 674-683, 1994.
- [3] K. Hornik and C.-M. Kuan, "Convergence analysis of local feature extraction algorithms", Neural Networks, Vol. 5, No. 2, pp. 229-240, 1992.
- [4] Terence D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network", Neural Networks, Vol. 2, No. 6, pp. 459-473, 1989.
- [5] Chanchal Chatterjee and Vwani P. Roychowdhury, "On self-organizing algorithms and networks for class-separability features", IEEE Tr. NN, Vol. 8, No. 3, pp. 663-678, 1997.
- [6] Jianchang Mao and Anil K. Jain, "Artificial neural networks for feature extraction and multivariate data projection", IEEE Tr. NN, Vol. 6, No. 2, pp. 296-317, 1995.

- [7] M. B. Nevel'son and R. Z. Has'minskiĭ, *Stochastic Approximation and Sequential Estimation*, Nauka, 1968 (in Russian; translated into Japanese by T. Kitagawa and K. Tajima, 1983).
- [8] Takio Kurita and Satoru Hayamizu, "Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer", Proceedings of the third international conference on automatic face and gesture recognition, April, 1998, pp. 422-427.
- [9] W. Zhao and R. Chellappa and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition", Proceedings of the third international conference on automatic face and gesture recognition, April, 1998, pp. 422-427. pp. 336-341.
- [10] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, New York, 1989.