

# 高次元特徴ベクトルの次元圧縮と重みつき K-最近傍法によるパターン認識

長谷川 修<sup>†,††,†††</sup> 栗田 多喜夫<sup>††</sup>

本稿では、(1) 識別のために有効と思われる特徴を幅広く盛り込んだ高次元の特徴ベクトルの構成、(2) 高次元の特徴ベクトルを識別に有効な次元を選択しつつ圧縮、(3) 圧縮後のベクトルに重みつき K-最近傍法 (以下 K-NN) を適用して識別、という枠組みに基づく多クラスパターンの認識法を提案し、実験によりその有効性を示す。高次元の入力特徴ベクトルは、認識率の向上に有効と考えられる、複数の異なる性質の特徴を組み合わせる。K-NN は、多クラスパターンを識別する non-parametric な手法の 1 つであり、その識別誤差は “Bayes Error” の 2 倍を超えないとされる。しかし入力特徴ベクトルの次元数が高い場合その性能は保証されず、また総演算量が膨大になる。そこで本研究では、一般化線形モデルの一種である多項ロジットモデル (Multinomial Logit Model) を用いて入力特徴ベクトルの次元を圧縮し、圧縮後のベクトルに K-NN を適用する。これにより K-NN の本来の性能を引き出し、同時に識別処理時の演算量を大幅に削減することを狙いとする。手書き文字データベース ETL6 中の、(1) 36 クラス (数字 + 英大文字)、(2) 82 クラス (数字 + 英大文字 + 片仮名) のデータを用いて評価実験を行ったところ、(1) 36 クラス (7,200 個の未知サンプル) に対して 100.0%、(2) 82 クラス (16,400 個の未知サンプル) に対して 99.93% の識別率を得た。

## Pattern Recognition by Compression of High Dimension Vectors and Weighted K-Nearest Neighbor Rule

OSAMU HASEGAWA<sup>†,††,†††</sup> and TAKIO KURITA<sup>††</sup>

This paper proposes a simple method for multi-class pattern classification by combined use of Multinomial Logit Model (MLM) and weighted K-nearest neighbor rule (K-NN). MLM is one of the generalized linear models and is one of the simplest neural networks for multiple-class pattern recognition. K-NN is a simple but powerful non-parametric classification tool whose error probability does not exceed double of “Bayes error”. However, it is also known that such high performance of K-NN reduces if the number of dimensions of input feature vector space is large. Therefore, first we train MLM using the training vectors, and then apply weighted K-NN to the output vector of the MLM. By this, since K-NN is applied to the compressed low dimension vectors, it is expected not only to bring out natural performance of K-NN but also to shorten computation time. Evaluation experiments were conducted by using non-artificial samples extracted from the handwritten character image database “ETL6”. Those are (1) 36-classes (number + English capital letter), and (2) 82-classes (number + English capital letter + “katakana” letter). Consequently, we obtained the following recognition rates: (1) 36-classes: 7,200 unknown samples  $\Rightarrow$  100.0%, and (2) 82-classes: 16,400 unknown samples  $\Rightarrow$  99.93%.

### 1. はじめに

本稿では、代表的な non-parametric パターン認識

手法の 1 つである K-最近傍法 (以下 K-NN) に着目した多クラスパターンの認識手法を提案する。

K-NN は、十分な訓練サンプルが与えられれば、未学習データに対する識別誤差が “Bayes Error” の 2 倍を超えないという顕著な特長を持つことが知られている<sup>1)</sup>。しかし入力特徴ベクトルの次元が高い場合には K-NN のこのような性能は保証されず<sup>2)</sup>、また識別のための演算量も膨大となる。したがって、K-NN の入力としてカーネル特徴などの高次元の特徴ベクトルを直接用いることは必ずしも得策ではない。

† 東京工業大学大学院理工学研究科像情報工学研究施設  
Imaging Science and Engineering Laboratory, Tokyo  
Institute of Technology

†† 産業技術総合研究所脳神経情報研究部門  
Neuroscience Research Institute, Advanced Industrial  
Science and Technology (AIST)

††† 科学技術振興事業団さきがけ研究 21  
PRESTO, Japan Science and Technology Corp. (JST)

一方、一般にパターン認識に適した特徴を事前に知ることはできないため、パターンの特徴ベクトルを構成する際にはクラス間の識別に有効と思われる特徴を幅広く盛り込むことが望ましいが、その結果としてパターンの特徴ベクトルは高次元となる<sup>3)</sup>。

そこで提案手法では、まずパターン認識に有効と考えられる特徴を広く盛り込んだ特徴ベクトルからクラス間の識別に有効なものを選択・合成することにより、元の空間より高い識別力を持つ低次元の特徴ベクトル空間を構成する。こうしたアプローチの有効性は確認されており、線形判別分析 (Linear Discriminant Analysis: 以下 LDA) や boosting<sup>4)</sup> と組み合わせられて利用されている<sup>5),6)</sup>。

次に、構成した低次元の特徴空間内で K-NN を適用する。これにより K-NN 本来の高い識別性能を引き出すとともに、識別処理時の演算量を大幅に削減することを狙いとす。

また本研究では、初期の特徴ベクトルを構成する際に盛り込む特徴として、異なる性質の特徴を複数組み合わせることを試みる。これにより初期の特徴ベクトル内に互いに相関の少ない特徴が効率的に増え、結果として特徴選択・合成後の空間の各軸の独立性 (識別力) が向上すると考える。具体的には、本研究では文字認識分野で提案され有効性が確認されている方向性特徴<sup>7)</sup> と、近年パターン認識の分野で有効性が確認されているカーネル特徴<sup>8)</sup> を組み合わせて利用した。

入力特徴ベクトルに含まれる特徴のうち、クラス間の識別に有効なものを選択・合成しつつ全体を圧縮する (低次元の新たな特徴ベクトル空間を構成する) 手法にはいくつか考えられるが、本研究では一般化線形モデルの一種である多項ロジットモデル (Multinomial Logit Model: 以下 MLM<sup>9)</sup>) を用いた。MLM は、多クラスパターン認識のための簡素なニューラルネットワークモデルの 1 つであり、実装が容易でありながら、その識別能力は線形識別器<sup>10)</sup> と同等か若干優れるとされる。

MLM の学習時には汎化性能の向上に寄与するとされる 3 つの工夫を導入した。また K-NN は近傍ベクトルとの距離でサンプルに重みをつける重みつき K-NN とし、識別境界付近のサンプル数の多寡による識別率への影響の抑制を図った。

提案手法の有効性を評価するため、手書き文字データベース ETL6<sup>11)</sup> 中の (1) 36 クラス (数字 + 英大文字)、および (2) 82 クラス (数字 + 英大文字 + 片仮名) のデータを用いて評価実験を行い、(1) 36 クラス (7,200 個の未知サンプル) に対して 100.0%、(2) 82 ク

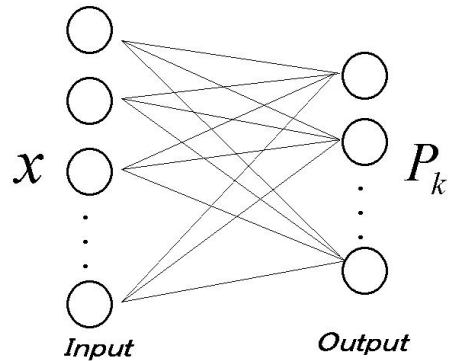


図 1 Multinomial Logit Model の基本構成  
Fig.1 Structure of Multinomial Logit Model.

ラス (16,400 個の未知サンプル) に対して 99.93% の識別率を得た。

## 2. Multinomial Logit Model: MLM

本章では、MLM の概要について述べる。図 1 に MLM の基本構成を示す。

いま入力特徴ベクトルを  $x = (x_1, \dots, x_N)^T \in R^N$  とし、これを  $K$  個のクラス  $\{C_1, C_2, \dots, C_K\}$  に識別するとする。また各入力特徴ベクトルに対応する教師ベクトルは、正解のクラス  $C_j$  に対応する要素  $t_j$  のみが 1 で、それ以外の要素がすべて 0 の 2 値ベクトル  $t = (t_1, \dots, t_K)^T$  とする。

このとき MLM では、識別器の  $k$  番目の出力素子の出力を、入力ベクトル  $x$  とパラメータベクトル  $a$  との線形結合  $\eta_k = a_k^T x$  の “softmax” として以下のように算出する。

$$p_k = \frac{\exp(\eta_k)}{1 + \sum_{m=1}^{K-1} \exp(\eta_m)}, \quad (k=1, \dots, K-1) \quad (1)$$

$$p_K = \frac{1}{1 + \sum_{m=1}^{K-1} \exp(\eta_m)}, \quad (k=K) \quad (2)$$

この式 (1) を上位 (クラス数 -1) 個の各クラスに対応する MLM の出力素子からの出力値とし、式 (2) の  $p_K (= 1 - \sum_{m=1}^{K-1} p_m)$  を最終クラスに対応する出力値とする。これにより、MLM の出力素子の数、および圧縮後のベクトルの次元数は (クラス数 -1) となる。これは式 (1)、(2) の出力の総和をつねに 1 にする工夫であり、この結果 MLM の出力を事後確率と見なすことができる。

パラメータ  $A = (a_1, \dots, a_K)^T$  を入力層から出力層への結合加重と見なすと、MLM の確率モデルは以下となる。

$$P(t|\mathbf{x}; A) = \prod_{k=1}^K p_k^{t_k} \quad (3)$$

ここで式 (3) の両辺の対数をとると、対数尤度  $l(t|\mathbf{x}; A) = \log P(t|\mathbf{x}; A)$  が求まり、

$$l(t|\mathbf{x}; A) = \sum_{k=1}^{K-1} t_k \eta_k - \log \left( 1 + \sum_{m=1}^{K-1} \exp(\eta_m) \right) \quad (4)$$

となる。MLM の学習アルゴリズムは、この対数尤度に関する最急降下法として導ける。具体的には、対数尤度の勾配は以下となる。

$$\frac{\partial l}{\partial \eta_k} = t_k - p_k \quad (5)$$

$$\frac{\partial \eta_k}{\partial \mathbf{a}_k} = \mathbf{x} \quad (6)$$

$$\frac{\partial l}{\partial \mathbf{a}_k} = \frac{\partial l}{\partial \eta_k} \frac{\partial \eta_k}{\partial \mathbf{a}_k} = (t_k - p_k) \mathbf{x} \quad (7)$$

これらより、結合加重ベクトル  $\{\mathbf{a}_k\}$  の更新式は、簡素に

$$\mathbf{a}_k \leftarrow \mathbf{a}_k + \alpha (t_k - p_k) \mathbf{x} \quad (8)$$

となる。ここで  $\alpha$  は学習係数である。

### 3. 学習・評価用特徴ベクトルの算出

#### 3.1 学習・評価用データ

本研究では、提案手法の有効性を手書き文字データベース ETL6<sup>11)</sup> に含まれる、

(1) 36 クラス：数字 + 英大文字

(2) 82 クラス：数字 + 英大文字 + 片仮名

を用いて評価した。ここで数字は 0~9 の 10 字種、英大文字は 26 字種、片仮名は 46 字種である。具体的には、ETL6 の 36 クラス (数字, 英大文字) および 82 クラス (数字, 英大文字, 片仮名) のデータのうち、各字種の偶数番, 奇数番, 各上位 200 個のデータをそれぞれ学習用, 評価用に用いた。またカーネル特徴複合ベクトルを用いた 36 クラスの実験 (後述) では、各字種の偶数番, 奇数番, 上位 100 個をそれぞれ学習用, 評価用に用いた。

図 2 に ETL6 中の手書き文字データ “A” の例を示す。図に示されるように、実験に用いたデータは比較的粒揃いであり、極端に文字が小さいものや形が歪んでいるもの、また隣の文字が大きく含まれるといったものはみられなかった。

これらの画像データから、次に述べる特徴ベクトルを構成した。



図 2 ETL6 中の手書き文字データ例「A」

Fig. 2 Examples of “A” extracted from handwritten English capital letter images in ETL6.

#### 3.2 基本特徴ベクトル

まず ETL6 中の手書き文字画像から以下のようにして「基本特徴ベクトル」を構成した。

Zero-cross 法で文字画像からエッジを抽出し、ノイズ除去を行う。この画像から 4 枚の 4 方向特徴画像<sup>7)</sup>を求め、15 × 15 ピクセルに縮小する。縮小した 4 枚の画像を 30 × 30 ピクセルの 1 枚の画像にまとめ、これを 2 次元 Gauss 関数でぼかす。図 3 にこのようにして求めた “A” の 4 方向特徴画像を示す。図 3 で、左列の上下 2 個では横方向成分が、右列の上下 2 個では縦方向成分が抽出されている。

次にこの 30 × 30 ピクセルの画像データを 900 × 1 サイズに変換し、900 次元 (= 4 × 15 × 15) のベクトルとする。こうしたベクトルを学習用/評価用独立に、各クラス 200 ずつ用意した。これにより、学習用/評価用のベクトルを、(1) 36 クラスでは 7,200 個、(2) 82 クラスでは 16,400 個構成した。

図 4 に 900 次元の 4 方向特徴ベクトル 7,200 個を 7,200 × 900 の行列として可視化した例を示す。横軸方向の各行が 1 文字分のベクトルデータである。図 4 にみられる 4 周期は 4 方向特徴に対応している。縦軸方向のパターンの変化は字種の違いに対応している。

#### 3.3 カーネル特徴複合ベクトル

上記の「基本特徴ベクトル」は、各入力画像に含まれる幾何学的特徴を要素とするベクトルである。そこでこれをカーネル特徴ベクトルと結合したベクトルを構成すると、互いに相関の少ない特徴がベクトル内に増え、結果的に入力特徴空間における各クラスの分離が促進されて、認識率の向上に寄与すると期待される。そこで本研究では、36 クラスのデータに対して下記

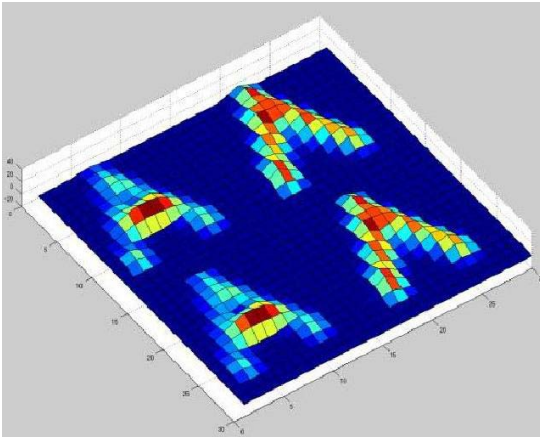


図 3 4 方向特徴画像 (画像のサイズは全体で  $30 \times 30$  ピクセル。左列の上下 2 個は横方向成分を、右列の上下 2 個は縦方向成分を抽出している)

Fig. 3 4-direction edge image. ( $30 \times 30$  pixel size. Two images of the left side column are extracting the horizontal direction features, and two images of the right side column are extracting the vertical features.)

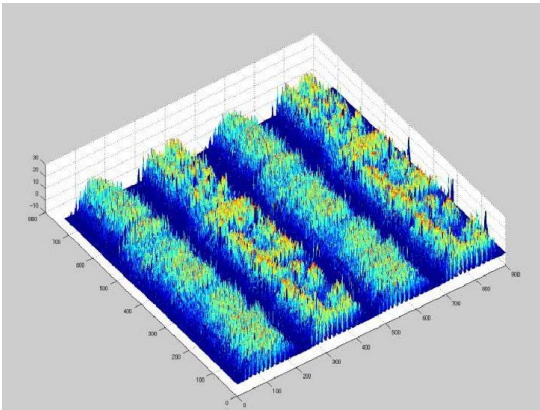


図 4 900 次元特徴ベクトル (横軸方向の 4 周期は 4 方向特徴に対応し、縦軸方向のパターンの変化は字種の違いに対応している)

Fig. 4 Visualized  $7,200 \times 900$  size matrix constituted from 7,200 basic feature vectors.

の「カーネル特徴「複合」ベクトル」を構成し、「基本特徴ベクトル」「カーネル特徴ベクトル」を用いた場合と認識率を比較した。

まず学習用の基本特徴ベクトルの各クラスから任意に 100 個 (総数 3,600 個) を抽出して「基準ベクトル」とし、次いでこれらの基準ベクトルと各クラスのベクトルから 3,600 次元のカーネル特徴ベクトル<sup>8)</sup> を構成する。カーネル関数には Gauss 関数を用いた。

カーネル特徴複合ベクトルは、カーネル特徴ベクトルと基本特徴ベクトル (900 次元) を結合したベクトル

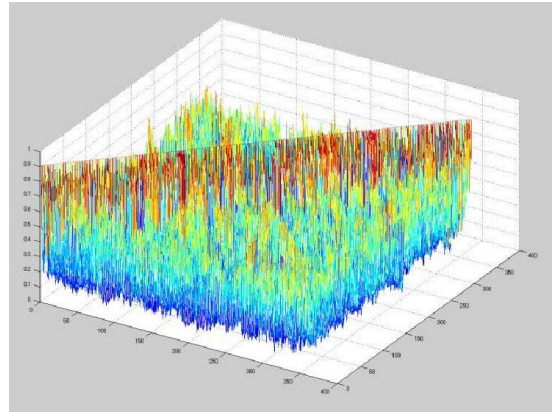


図 5 学習用のカーネル特徴ベクトルの例

Fig. 5 Examples of Kernel feature vectors for training.

ルであり、4,500 ( $3,600 + 900$ ) 次元となる。

具体的には、各クラスの基本特徴ベクトルを

$$\mathbf{x}_k = (x_{k1}, \dots, x_{kN})^T, \quad (N = 900) \quad (9)$$

とし、「基準ベクトル」を

$$\mathbf{x}_i = (x_{i1}, \dots, x_{iN})^T, \quad (N = 900) \quad (10)$$

とすると、

$$y_{ik} = \exp\left(\frac{-\|\mathbf{x}_k - \mathbf{x}_i\|^2}{2 \times \sigma}\right) \quad (11)$$

$$\mathbf{y}_i = (y_{i1}, \dots, y_{ik})^T \quad (12)$$

(ただし式 (11), (12) の  $\{i, k\} = 1, \dots, 3600$ )

で求まる  $\mathbf{y}_i$  をカーネル特徴ベクトルとする。このカーネル特徴ベクトルと基本特徴ベクトルと結合したベクトルが「カーネル特徴「複合」ベクトル」である。

なお、式 (9) の基本特徴ベクトルには、学習用のベクトルの構成時には上記の基準ベクトルをそのまま使い、評価用のベクトルの構成時には評価用の基本特徴ベクトルから抽出した各クラス 100 個 (総数 3,600 個) のベクトルを用いた。

図 5 と図 6 に、学習用および評価用の 3,600 個/各 3,600 次元のカーネル特徴ベクトルを、それぞれ  $3,600 \times 3,600$  の行列として可視化した例を示す。図の各行が各文字のカーネル特徴ベクトルに対応している。これに図 4 の基本特徴ベクトルを結合したものがカーネル特徴複合ベクトルである。

図 5 の対角線上は同一のベクトルから算出したカーネル特徴値であるため、式 (11) では  $k = i$  となることから値は 1 となる。図 5 の対角線上以外の値は低く抑えられているが、これは学習用ベクトルの各クラスの分離が良好であることを意味している。

図 6 は評価用のカーネル特徴ベクトルである。ここにおいても対角線上の値が相対的に大きく、各クラスの分離が進んでいることが理解される。



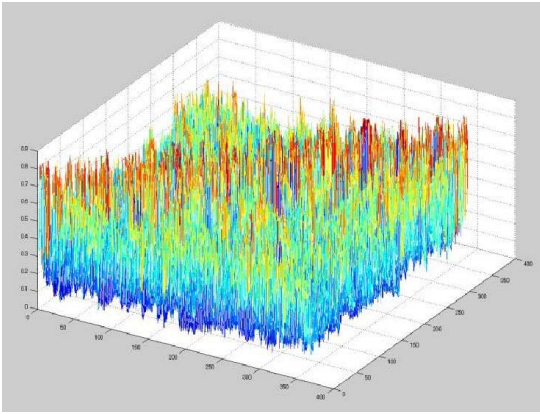


図 6 評価用のカーネル特徴ベクトルの例

Fig. 6 Examples of Kernel feature vectors for test.

本研究では、こうしたカーネル特徴ベクトルと前出の基本特徴ベクトルを組み合わせ「カーネル特徴複合ベクトル」とし、ここから識別に有効な特徴を選択・合成させることによって、より識別に適した低次元の特徴空間の構成を試みた。

#### 4. MLM の汎化性向上のための工夫

##### 4.1 3つの手法

先に述べたように、本研究ではニューラルネットワークの汎化性の向上に寄与するとされる手法のうち、異なる観点から独立に提案された3つの手法をMLMの学習時に導入した。以下にそれらの手法を示す。

##### 4.1.1 ノイズの付加

栗田らは、多層パーセプトロンの学習中、中間層にノイズを付加するとネットワークが構造化され、汎化性が向上することを示した<sup>12)</sup>。これは擬似的に学習用のサンプル数を増やすことに相当する。

そこで本研究では、入力ベクトル $x$ とパラメータベクトル $a$ との線形結合出力 $\eta_k = a_k^T x$ に一樣乱数ノイズを加えた。

##### 4.1.2 エントロピーに基づく学習係数の算定

良好な認識率を得るためには、識別境界付近のサンプルを正しく識別する境界面をいかに形成するかが重要なポイントとなる。そこで学習の際、識別クラスが不明瞭なベクトル(識別境界付近のサンプル)の重みを大きくして学習させると、それらをより明瞭に識別するように境界面が形成(修正)され、好ましいと考えられる。また、実質的に識別境界に重点を置いた学習過程となるため、すべてのサンプルを等しく扱う場合に比べて学習に要する時間が短縮されると期待される。

そこで本研究では、式(1)の出力から式(13)によ

り各データのエントロピーを求め、これを「識別クラスの明瞭さ」を示す量と定義し、式(8)の学習係数 $\alpha$ に加えて用いることとした。これによれば、識別クラスが不明瞭であるほど学習係数 $\alpha'$ の値が大きくなる。

$$\alpha' = \beta \times \sum_{i=1}^{K-1} (-p_i \times \log(p_i)) \quad (13)$$

ここで、 $\beta$ は学習係数である。

##### 4.1.3 Weight Decay

Hansonら<sup>13)</sup>は、ネットワークの結線のうち、学習時の評価基準に照らして寄与の少ないものを取り除く(結合加重を強制的に0にする)のではなく、徐々に0に近づけるような項を更新式に加える手法を提案しており、*WeightDecay*と呼んでいる。

本研究では、この項を以下のようにまとめ、更新式(8)に加えた。

$$-\lambda x \quad (14)$$

ここで、 $\lambda$ は学習係数である。

#### 4.2 3つの手法の複合的導入の意味

上記の3つの手法はそれぞれ工夫を加えるポイントが異なっている。本研究では、上記の3つの手法を複合的に(組み合わせる)導入する意味を以下のように解釈している。

「まず限られた数のサンプルを擬似的に増やし、各クラスの分布のより密な(より精度の高い)近似を図る。次にサンプルの中で、識別境界付近に位置するものに優先的に着目して識別面を形成する。具体的には、更新式に式(13)、(14)を導入し、それらの学習係数を調整することによって、各クラスの分布境界をより高い精度で近似する識別境界面の形成と、過学習の抑制を図る」。

3つの手法を複合的に導入した後の更新式は以下となる。

$$a_k \leftarrow a_k + (\alpha + \alpha')(t_k - p_k)x - \lambda x \quad (15)$$

#### 4.3 MLMの学習の停止条件

MLMはニューラルネットワークの一種であるが構造が簡素なため自由度は高くなく、学習時には大きな過学習はみられなかった。特に更新式に式(15)を用いた場合は式(13)、(14)の項が機能し、さらに過学習が抑えられた。

そこでMLMの学習の停止条件は、学習用サンプルに対して最良の識別率が得られた時点から500万回学習を進めた後とした。すなわち、学習データで最良の識別率が得られた時点からしばらく学習を進め、十分に収束したと考えられる時点で学習を停止させた。

図7に、36クラスの学習用基本特徴ベクトルと式

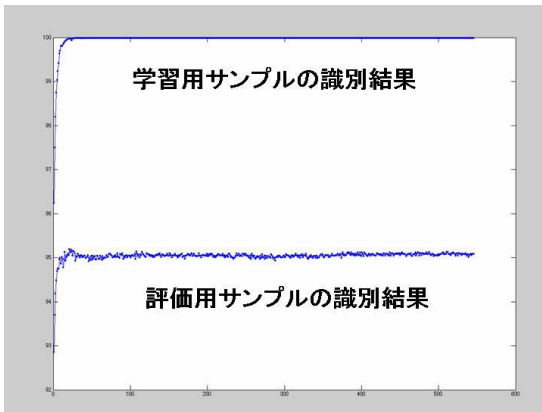


図7 学習時の汎化性能の推移(36クラスの学習用基本特徴ベクトルを用いてMLMを学習させた例)

Fig. 7 Transition of generalization performance in training process. (An example of the case that a MLM is trained using 36-class basic feature vectors.)

(15)を用いてMLMを学習させた際の、学習用サンプルと評価用サンプルに対する識別結果例を示す。学習の初期の段階で学習用サンプルに対し100%の識別率(縦軸)が得られた後、さらに約2,500万回学習を続けたが(横軸)、評価用サンプルに対する識別率は95%前後で推移し、過学習による汎化性の低下はみられていない。

## 5. 実験と結果

### 5.1 標準的なMLMによる認識結果

以下の各実験では、MLMの学習時に4章に述べた手法は導入していない。

まず36クラスのサンプルから構成した学習用の「基本特徴ベクトル」、「カーネル特徴ベクトル」、「カーネル特徴複合ベクトル」を用いて3つのMLMの学習を行った。次に学習後のMLMにそれぞれ対応する評価用の「基本特徴ベクトル」、「カーネル特徴ベクトル」、「カーネル特徴複合ベクトル」を適用して式(1)、(2)の出力 $p_k$ , ( $k=1, \dots, K$ )を求め、そのうち最大値をとるものを認識結果のクラスとして正答と比較した。その結果を、82クラスの「基本特徴ベクトル」を標準的なMLMで学習・認識した場合の結果とあわせて示す。

<基本特徴ベクトル>

36クラス：94.86%， 82クラス：92.97%

<カーネル特徴ベクトル>

36クラス：93.08%

<カーネル特徴複合ベクトル>

36クラス：97.89%

上に示すように、標準的なMLMで学習・認識した場合、「カーネル特徴複合ベクトル」を用いた認識率は「基本特徴ベクトル」もしくは「カーネル特徴ベクトル」を用いた認識率を上回った。

また「基本特徴ベクトル」による認識率と「カーネル特徴ベクトル」による認識率を比較すると、前者が後者を約1.8ポイント上回った。

### 5.2 汎化性向上の工夫を加えたMLMによる認識結果

MLMの学習時に4章に述べた3つの手法を導入し、5.1節と同様の実験を行った。ただし5.1節の結果を受け、「カーネル特徴ベクトル」を用いた実験は省略した。

4章に述べた各手法のパラメータ( $\alpha, \beta, \lambda$ )のうち1つを固定し( $WeightDecay$ の $\lambda=0$ とした)、他の2つを変動させた場合の認識率の変化例を図8、図9に示す。図8、図9はそれぞれ学習用データ、評価用データに対する認識率であり、鉛直軸に認識率をとっている。双方の図より、良好な認識率を与えるパラメータは各水平軸上ではなく面上にあることが分かる。これは4章に述べた各手法をそれぞれ単独に適用するよりも、組み合わせて適用するほうがより良好な認識率が得られることを意味している。そこで上記の3つのパラメータを変化させ、最適なパラメータ値を探索したうえで認識率を比較した。その結果を、82クラスの「基本特徴ベクトル」を用いて得られた結果とあわせて示す。

<基本特徴ベクトル>

36クラス：94.89%， 82クラス：93.13%

<カーネル特徴複合ベクトル>

36クラス：98.19%

ここで学習用、評価用のサンプル数はそれぞれ

36クラス ⇒ 7,200個

82クラス ⇒ 16,400個

である。

これらはいずれも5.1節に示した認識率を超えており、本アプローチ(最適パラメータを探索したうえでの、3手法の複合的・組合せ的適用)の有効性を示すと考える。

なお今回の実験では、最適パラメータの探索時、パラメータを網羅的かつ密に更新したが、図8、図9の認識率の変化は比較的よく連動していることなどから、より効率的な探索法が導入可能と考える。

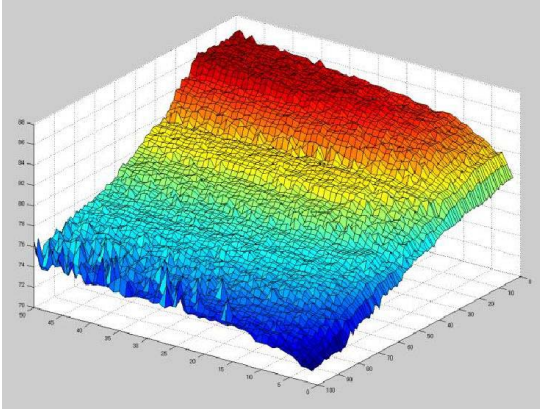


図 8 パラメータを変化させたときの学習用データに対する認識率の変化例

Fig. 8 Change of recognition rate to the data for training by changing parameters.

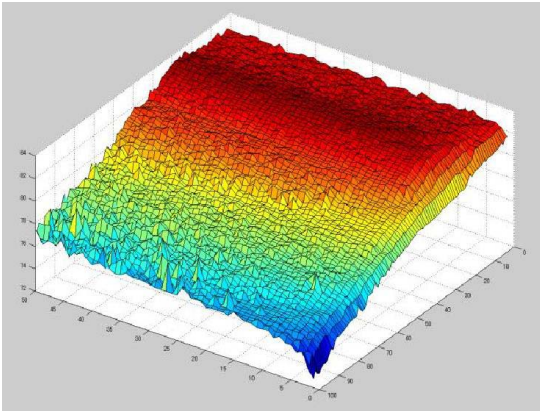


図 9 パラメータを変化させたときの評価用データに対する認識率の変化例

Fig. 9 Change of recognition rate to the data for test by changing parameters.

### 5.3 汎化性向上の工夫を加えた MLM と重みつき K-NN の組合せによる認識結果

#### 5.3.1 基本特徴ベクトルを用いた場合

まず 36 クラス, 82 クラスの学習用の「基本特徴ベクトル」を用い, 4 章の 3 つの工夫を加えつつ 2 つの MLM を学習させた. 次に学習後の MLM にそれぞれ対応する 36 クラス, 82 クラスの評価用の「基本特徴ベクトル」を適用し, 得られる式 (1) の出力をベクトル  $\mathbf{p} = (p_1, \dots, p_{(K-1)})^T$  と見なしてこれに重みつき K-NN を適用した. ここで各ベクトル間の距離  $d_{i,j}$  を,

$$d_{i,k} = \frac{1}{\|\mathbf{p}_k - \mathbf{p}_i\|} \quad (16)$$

で定義し, これを重みとした. 以下に結果を示す.

<基本特徴ベクトル>

36 クラス: 99.99%, 82 クラス: 99.93%

学習用, 評価用のサンプル数はそれぞれ,

36 クラス  $\Rightarrow$  7,200 個

82 クラス  $\Rightarrow$  16,400 個

である.

上記の認識率は 5.1 節, 5.2 節の結果を超えており, 重みつき K-NN の併用が認識率向上の観点から有効に機能していることを示すと考える. なお, 重みつき K-NN は MLM の出力ベクトルに対して適用するため, 処理対象となるベクトルの次元数は,

36 クラス: 900 次元  $\Rightarrow$  36 次元

82 クラス: 900 次元  $\Rightarrow$  82 次元

と, それぞれ約 91%, および約 96% 圧縮されていることを強調する.

#### 5.3.2 カーネル特徴複合ベクトルを用いた場合

36 クラスの学習用/評価用の「カーネル特徴複合ベクトル」を用い, 5.3.1 項と同様の処理を行った. その結果, 以下の認識率を得た.

<カーネル特徴複合ベクトル>

36 クラス: 100.0%

本実験における学習用, 評価用のサンプルの総数は, それぞれ

36 クラス  $\Rightarrow$  3,600 個

である. ここで, 重みつき K-NN を適用する MLM の出力ベクトルの次元数は,

36 クラス: 4,500 次元  $\Rightarrow$  36 次元

と 99.2% 圧縮されていることを強調する.

## 5.4 考 察

### 5.4.1 実験結果に関する考察

表 1, 表 2 および図 10, 図 11 に実験結果をまとめる. 表 1 および図 10 は 36 クラスのサンプルを用いた場合の実験結果であり, 表 2 および図 11 は 82 クラスのサンプルを用いた場合の実験結果である. また各表と図において, (1), (2), (3) はそれぞれ, (1) 標準的な MLM による認識結果, (2) 学習時に 3 つの工夫を加えた MLM による認識結果, (3) 学習時に 3 つの工夫を加えた MLM と重みつき K-NN を併用した認識結果, である.

まず表 1 および図 10 に示すように, 本研究で導入した「カーネル特徴複合ベクトル」を用いた場合, 「基



表 1 基本特徴ベクトルを用いた場合とカーネル特徴複合ベクトルを用いた場合の認識率の比較 (36 クラスのデータ)

Table 1 Comparison of recognition rate by using basic feature vector and Kernel feature compound vector (36 classes).

	(1)	(2)	(3)
36 class basic vector	94.86%	94.89%	99.99%
36 class comp. vector	97.89%	98.19%	100.0%

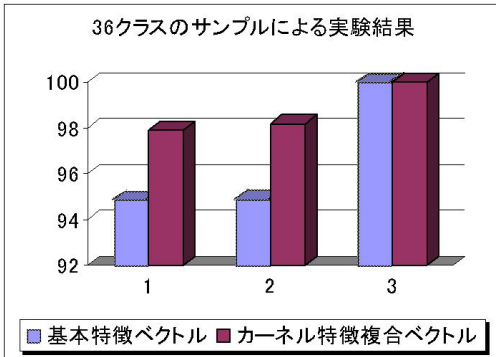


図 10 基本特徴ベクトルを用いた場合とカーネル特徴複合ベクトルを用いた場合の認識率の比較 (36 クラスのデータ)

Fig. 10 Comparison of recognition rate by using basic feature vector and Kernel feature compound vector (36 classes).

本特徴ベクトル」もしくは「カーネル特徴ベクトル」を用いた場合に比べて全般的に認識率が向上することを確認した。

表 1 および図 10 の (1) と (2) において認識率の差が少ないのは、4 章に述べた手法が有効に機能するのは少ないサンプルが非線形に分布するといった場合であり、今回実験に用いたサンプルではその効果を発揮する余地があまりなかったためと考える。

一方、表 1 および図 10 の (1) と (2) の結果の差と (2) と (3) の結果の差を比較すると、「基本特徴ベクトル」、「カーネル特徴複合ベクトル」のいずれを用いた場合も後者が前者を大きく上回っており、このことは、重みつき K-NN は識別面の近似精度で MLM に優れ、MLM の後段に重みつき K-NN を適用することの効果が大いことを示すと考える。

なお、36 クラス、82 クラスの「基本特徴ベクトル」に対して直接重み付きの K-NN を適用したところ、36 クラスに対して 96.25%、82 クラスに対して 95.59% の認識率となった。このことは、重み付きの K-NN を適用する前段の MLM による次元圧縮は、総演算量の削減/処理時間の短縮のみならず、認識率向上の観点からも重要な役割を果たしていることを示すと考える。

表 2 82 クラスの基本特徴ベクトルを用いた場合の認識率比較

Table 2 Comparison of recognition rate by using basic feature vector (82 classes).

	(1)	(2)	(3)
82 class basic vector	92.97%	93.13%	99.93%

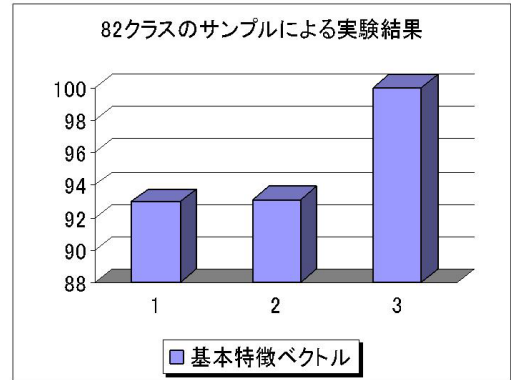


図 11 82 クラスの基本特徴ベクトルを用いた場合の認識率比較

Fig. 11 Comparison of recognition rate by using basic feature vector (82 classes).

「基本特徴ベクトル」を用いて「重みつき K-NN の適用あり」で認識する場合と「カーネル特徴複合ベクトル」を用いて「重みつき K-NN の適用なし」で認識する場合とを比較すると、表 1 の 3 行 1 列と 2 行 2 列に示されるように前者が後者を上回った。

そこでこの結果を受け、82 クラスの「基本特徴ベクトル」を用いて「重みつき K-NN の適用あり」の認識実験を行ったところ、表 2 および図 11 に示すように表 1 および図 10 と同様の傾向がみられ、特に (3) の実験においては顕著な認識率が得られた。この結果は本稿の提案手法の一部である「基本特徴ベクトル + 工夫あり MLM + 重みつき K-NN の適用」でも有効性が期待できることを示すと考える。

36 クラスのサンプルを用いた実験全体を通じて比較すると、最も高い認識率が得られたのは提案手法の枠組みを全般的に適用した「カーネル特徴複合ベクトルを用い、MLM の学習時には 3 つの工夫を加え、重みつき K-NN を併用した場合」であり、このことは提案手法全体の有効性を示すと考える。

#### 5.4.2 カーネル特徴複合ベクトルが良好な結果を与える理由についての考察

1 章にも述べたように、一般にパターンの表現空間を構成する際、あらかじめ様々な観点からなるべく幅広く特徴を盛り込み、そこから識別に有効なものを選ぶようにすれば、より識別に適した低次元の特徴空間が得られると考えられる<sup>5),6)</sup>。5.1~5.3 節の実験で、「カーネル特徴複合ベクトル」を用いた場合の認識率



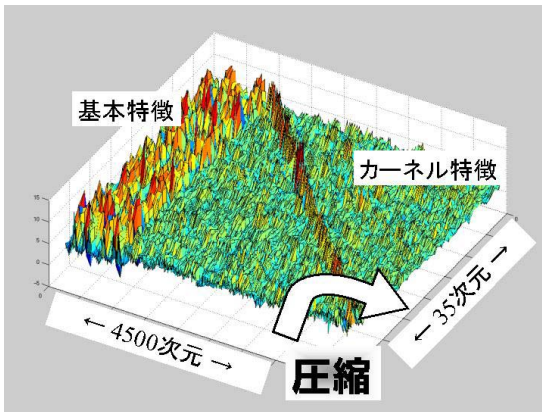


図 12 カーネル特徴複合ベクトル (36 クラス) を学習後の MLM の入出力間結合係数: 鉛直軸に結合係数値を提示

Fig.12 Weight of coefficient obtained by MLM learned with Kernel feature compound vector (36 class): vertical axis corresponds to weight.

がそれ以外のベクトルを用いた場合の認識率を上回ったのも、この原理に基づくと考えられる。

以下では、以上を本研究で用いたデータを用いて検証する。まず学習用の基本特徴ベクトル、カーネル特徴ベクトル、カーネル特徴複合ベクトルを用い、式 (15) の学習係数と学習回数 (500 万回) を同じにしてそれぞれから低次元の特徴空間を構成した。次いでその空間中に評価用のベクトルを写像し、写像後の特徴ベクトルと、それぞれ対応する教師ベクトルとの差分ベクトルのノルムを算出した。以下に、そのようにして求めた各特徴ベクトルのノルムの平均値を、基本特徴ベクトルによる平均値を 1 としたときの比の値で示す。

基本特徴ベクトル	1.000
カーネル特徴ベクトル	1.035
カーネル特徴複合ベクトル	0.931

このことから、カーネル特徴複合ベクトルを用いて構成した特徴空間中の評価ベクトルが最も教師ベクトルとの差が少ないこと、すなわちカーネル特徴複合ベクトルを用いて構成した特徴空間が、最も良好にサンプルを分離していることを確認した。

図 12 には、カーネル特徴複合ベクトルを用いて学習した後の、MLM の入出力間結合パラメータ “A” の結合係数を可視化した画像を示す。鉛直軸が結合係数値に対応している。

この図から、カーネル特徴複合ベクトルを入力とした場合、大きな係数値を持つ結合が基本特徴・カーネル特徴の双方に分布していること、すなわち識別に有用な特徴が広く基本特徴・カーネル特徴の双方から選

択・合成され、圧縮後の特徴空間が構成されていることを確認した。

これらのことから「カーネル特徴複合ベクトルを用いた場合、基本特徴、カーネル特徴の双方からクラスの判別に有用な特徴が選択・合成されることにより、基本特徴、カーネル特徴を単独で用いた場合よりも同じ次元数でより識別力の高い空間が得られる」と結論づけられると考える。

#### 5.4.3 提案手法が機能する理由についての考察

一般に MLM による識別では、式 (1), (2) の条件下で、各クラスに対応する素子出力 (図 1 の Output) のうち最大値を持ったノードに対応するクラスとして行っている。これは幾何学的には、圧縮後の空間における各クラスの識別境界面を直線 (平面) で近似することを意味している。

このため、学習用サンプルに対する識別率が 100% であっても、学習後 (圧縮後) の空間における学習用サンプルの分布と評価用サンプルの分布に差があれば、識別境界面を超えて分布するサンプルも出るものと考えられる。このたびの実験で MLM で誤識別されたのは、そうしたサンプルと考える。K-NN はそうした境界面の凹凸を非線形に (曲面で) 近似し、MLM で生じる誤識別を補っていると考えられる。

なお学習用のサンプル数が各クラス 100~200 個では十分でない可能性もあるが、それについては、本研究では 4.1.1 項に述べたノイズの付加による擬似的なサンプル数の増加の工夫や、K-NN を「重みつき K-NN」とする工夫を加えて対処した。

#### 5.4.4 圧縮器に MLM を用いることの妥当性についての考察

圧縮器に MLM を用いることの妥当性については、まず MLM による識別結果と、線形判別分析 (LDA: 線形の枠内で「識別」の観点から最適な軸を探索しつつ次元圧縮を行う) による識別結果を比較した。

具体的には、36 クラスの学習用「カーネル特徴複合ベクトル」に式 (15) を用いて MLM をかけ、評価用の「カーネル特徴複合ベクトル」を識別させた際の識別率 (98.19% : 5.2 節参照) と、同じデータに LDA をかけて識別を試みた結果 (97.81%) とを比較した。その結果、これらはほぼ同等であることを確認した。

この結果は 1 章の記述を裏づけると同時に、MLM により事後確率の観点から次元圧縮して構成した空間と、LDA により構成した空間がほぼ同等な識別力を持つことを意味すると考える。

他の代表的な線形の次元圧縮法に主成分分析法 (PCA) があるが、「識別」の観点からは LDA によ

り次元圧縮を行うほうが好ましいことが指摘されている<sup>3)</sup>。そこで本研究では、MLM と LDA の識別力がほぼ等しいことを確認すれば十分と判断した。

以上より、線形の範囲では MLM の利用とそれに付随する圧縮次元数は妥当と考える。

## 6. おわりに

本稿では、

- (1) 識別のために有効と思われる特徴を幅広く盛り込んだ高次元の特徴ベクトルの構成、
- (2) 高次元の特徴ベクトルを識別に有効な次元を選択しつつ圧縮、
- (3) 圧縮後のベクトルに重みつき K-NN を適用して識別、

という枠組みに基づく多クラスパターンの認識法を提案し、実験によりその有効性を示した。提案手法は、アルゴリズムが簡素で実装が容易でありながら、良好な認識率が期待できる点にも特長があると考えられる。

最近、ETL6 の 36 クラスから 12 成分の特徴を抽出した相補的特徴場に摂動（平行移動、大きさの変化、傾き）を加え、そこに相関法を施すことによって、「99%代半ばの正読率を得た」とする報告がなされている<sup>14)</sup>。文献 14) でなされた実験と本実験とでは、学習・評価用のサンプルの数が異なるなど若干の違いがあるが、本実験結果は、今回の提案手法により「4 方向成分のみ、摂動なし」の原データを用いて「12 成分の相補的特徴場 + 摂動」に相当する結果が得られたことを示すと考える。

本研究では、入力特徴ベクトルの圧縮器（特徴選択器）として一般化線形モデルの一種である MLM を用いたが、MLM は入力特徴ベクトルを事後確率空間に写像する写像器と見なすことができる。今後同様の機能を持つ非線形の次元圧縮手法（たとえば多層パーセプトロン）も視野に入れ、本提案手法と前出の非線形判別分析の理論<sup>15)</sup> などとの関連を議論することを検討している。

3.3 節に述べたカーネル特徴算出のための基準ベクトルは、今回の実験では各クラスから任意に 100 個を抽出して用いたが、対象の認識・識別に寄与するものを算定して用いるといった改良が考えられる。

本研究では、画像から得る特徴として 4 方向性特徴、カーネル特徴、およびそれらの複合特徴を用いたが、他の特徴の利用/組合せも可能である。他の特徴の利用による性能評価は今後の課題である。

本手法を ETL6 以外の文字データ（たとえば ETL9）や文字以外のパターンに適用しての性能評価なども今

後の課題である。

## 参考文献

- 1) Cover, T. and Hart, H.: Nearest Neighbour Pattern Classification, *IEEE Trans. Inf. Theory*, Vol.IT-13, No.1, pp.1-27 (1967).
- 2) Fukunaga, K.: Bias of Nearest Neighbour Error Estimation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.9, No.1, pp.103-112 (1987).
- 3) 石井健一郎, 上田修功, 前田英作, 村瀬 洋: わかりやすいパターン認識, オーム社 (1998).
- 4) Meir, R. and Ratsch, G.: An introduction to boosting and leveraging, *Advanced Lectures on Machine Learning*, Mendelson, S. and Smola, A.(Eds.), LNCS, pp.119-184, Springer (2003).
- 5) Goudail, F., Lange, E., Iwamoto, T., Kyuma, K. and Otsu, N.: Face Recognition System Using Local Autocorrelations and Multiscale Integration, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.18, No.10, pp.1024-1028 (1996).
- 6) Li, S.Z., Zhang, Z.Q., Shum, H.-Ye. and Zhang, H.J.: FloatBoost Learning for Classification, *Online Proceedings of Neural Information Processing (NIPS2002)*, pp.AA-65 (2002).
- 7) 安田道夫, 山本和彦, 山田博三, 斎藤泰一: 文字認識のための相関法の一改良, 信学論, Vol.J67-D, No.12, pp.1442-1449 (1984).
- 8) たとえば, Scholkopf, B., Burges, C. and Mika, S.(Eds.): *Advances in Kernel Methods*, MIT Press (1998).
- 9) McCullagh, P. and Nelder, J.A. FRS: *Generalized Linear Models*, Chapman and Hall (1983).
- 10) Duda, O., et al.: *Pattern Classification* (2nd edition), John Wiley & Sons, Inc. (2001).
- 11) 斎藤泰一, 山田博三, 森 俊二: 手書文字データベースの解析 (III), 電総研彙報, Vol.42, No.5, pp.385-434 (1978).
- 12) 栗田多喜夫, 麻生英樹, 梅山伸二, 赤穂昭太郎, 細美章隆: 多層パーセプトロンの学習における中間層に付加したノイズの影響とネットワークの構造化, 信学論 D-II, Vol.J79-D-II, No.2, pp.257-266 (1996).
- 13) Hanson, S.J. and Pratt, L.Y.: Comparing Biases for Minimal Network Construction with Back-Propagation, *Advances in Neural Information Processing Systems 1*, Touretzky, D.S.(Ed.), pp.177-185, Morgan Kaufmann (1989).
- 14) 安田道夫: 相関法による摂動法の効果について, 認識型入力方式標準化委員会資料, R01-4-4 (2001).
- 15) 大津展之, 栗田多喜夫, 関田 巖: パターン認

識—理論と応用, 朝倉書店 (1996).

(平成 14 年 9 月 10 日受付)

(平成 15 年 3 月 28 日採録)

(担当編集委員 長尾 健司)



長谷川 修 (正会員)

1993 年東京大学大学院博士課程修了。博士(工学)。同年電子技術総合研究所入所。1999 年 6 月より 1 年間カーネギーメロン大学ロボティクス研究所滞在研究員。2001 年産業技術総合研究所主任研究員。2002 年 5 月東京工業大学大学院理工学研究科附属画像情報工学研究施設助教授。産業技術総合研究所脳神経情報研究部門に併任。2002 年 11 月科技団さきがけ研究 21 に兼任。パターン認識, マルチモーダルシステム等の研究に従事。電子情報通信学会, 人工知能学会, 日本認知科学会, 日本顔学会, IEEE CS 等各会員。



栗田多喜夫

1981 年名古屋工業大学工学部電子工学科卒業。同年電子技術総合研究所入所。1990 年~1991 年カナダ国立科学研究協議会(NRC)招聘研究員。現在, 産業技術総合研究所脳神経情報研究部門副部門長。工学博士。統計的パターン認識および生体模倣型ビジョンの研究に従事。日本神経回路学会, 行動計量学会, 日本顔学会, IEEE CS 各会員。