

手書き文字認識における投影距離法†

池田正幸†† 田中英彦†† 元岡達††

本研究の目的は、「手書き文字認識における効果的な距離評価法」の検討にある。従来の研究の多くは文字認識に有効な特徴を抽出することに主眼をおき、入力サンプルとカテゴリの近さの評価としては、入力サンプルの特徴ベクトルとカテゴリの平均ベクトルとの間のユークリッド距離を用いているものがほとんどであった。それに対して本研究では距離評価法の改良を試みた。従来の距離を「単純距離」とよぶことにし、その拡張として、カテゴリを超平面で近似して、入力サンプルを表す点とその超平面との距離によって評価する「投影距離」を提案した。これに、各要素に推定されるばらつきを考慮して重み付けする(Maharanobis 距離の一つの場合に相当する)「重み付け距離」、印刷漢字認識で用いられていた「複合類似度」を加えた4種類の評価法の検討を、3種類の特徴に対して、認識率を通して、行った。その結果、投影距離を用いることにより、従来の単純距離に比べて認識率がかなり改善され、また重み付け距離、複合類似度よりも高い認識率が得られることがわかった。

1. はじめに

本研究の目的は“手書き文字認識における効果的な距離評価法”を検討することにある。これまでの手書き文字認識の一般的手法として、図1に示すように①原パターンから特徴を抽出し、②辞書を用いてカテゴリとの距離を算出する、といった手順が定着しつつある¹⁾。しかしながら多くの研究は①のみに工夫をこらし、②では入力パターンの特徴ベクトルと各カテゴリの平均特徴ベクトルとのユークリッド距離というごく原理的な距離評価法しか用いていない例が多い。そこで本研究では②に着目して、従来の距離評価方式の問題点を検討した上で、距離評価に関する自然な拡張法を提案し、認識性能の向上を試みた。また、これまでに提案されてきた種々の特徴に対して、正読率・累積認識率等を通じて各種の距離法の評価・検討を行った。

本手法を用いた場合の本来の対象としては、手書き漢字を想定しており、検討を行う特徴はいずれも手書き漢字認識を目的として提案されたものであるが、認識実験の際には対象カテゴリとして手書きひらがなを扱った。その理由は、①手書きひらがなにも多様な変形が存在し、手書き漢字と比較しても必ずしも認識が容易とは限らず、本手法の有効性を十分に検討できるためと②実験規模を縮小するためである。

2. 距離概念の拡張

本章では、まず、従来よく用いられてきた単純距離について述べたあと、それを自然な形で拡張した重み付け距離・投影距離を提案する。

2.1 単純距離²⁾

2.1.1 単純距離の定義

入力サンプルとカテゴリとの間の距離を評価するためには、まずカテゴリを代表するものを何らかの形で表現する必要がある。最も単純かつ従来から用いられてきた方法として、カテゴリ i を学習サンプルの平均特徴ベクトル M_i によって代表させることが考えられる。

すなわち第 i カテゴリ ($i=1, 2, \dots, I$; I はカテゴリ数) の第 j サンプル ($j=1, 2, \dots, J$; J は1カテゴリ当りのサンプル数) を特徴抽出して得られた特徴ベクトルを

$$F_i^j = (f_{i1}^j, f_{i2}^j, \dots, f_{ki}^j)^t$$

(K は特徴ベクトルの次元数、 t は転置を表す) として、平均特徴ベクトル $M_i = (m_{i1}, m_{i2}, \dots, m_{ki})^t$ を次式によって求める。

$$m_{ki} = \frac{1}{J} \sum_{j=1}^J f_{ki}^j \quad (2.1)$$

($k=1, 2, \dots, K$; K は特徴ベクトルの次元数、 $j=1, 2, \dots, J$; J はサンプル数)。入力サンプル X とカテゴリ i との間の単純距離 $D_i(X, i)$ は、入力サンプル X の特徴ベクトル $X = (x_1, x_2, \dots, x_K)^t$ とカテゴリ i の特徴平均ベクトルとの間のユークリッド距離によっ

† Projection Distance Method for Recognition of Hand-written Characters by MASAYUKI IKEDA, HIDEHIKO TANAKA and TOHRU MOTO-OKA (Faculty of Engineering, Tokyo University).

†† 東京大学工学部電気工学科



図 1 一般的な認識手順
Fig. 1 General flow of character recognition.

とおけるから(2.3)式と(2.6)式の条件が一致することがわかる。

この理由から、従来最も単純で直観的にも把握しやすい単純距離が、サンプルとカテゴリとの間の近さの尺度として用いられてきた。しかし、実際の文字認識の場では、この仮定が必ずしも満たされるわけではない。すなわち、「文字パターンに加わる雑音が無ランダムなもののみであって、ばらつきの大きさ・分布がカテゴリによらず一定である」という前節の仮定は保証されず、手書きされた文字の雑音としては、位置ずれ・文字線幅の変動・回転・接触・切れ・ループのつおれ等のように、入力文字のパターン、すなわち入力文字の属するカテゴリに対して強い相関をもっているものが多いと考えられる。

これらの問題に対処するために、各カテゴリ・各 element ごとにその平均からの偏差に違いがあることを考慮して評価を行おうとしたものが、2.2 節で述べる重み付け距離であり、さらに各カテゴリ内での element 間の相関関係までも考慮しようとしたのが 2.3 節で述べる投影距離である。

で定義される。すなわち、

$$D_i^2(X, i) = \sum_{k=1}^K (x_k - m_{ki})^2 \quad (2.2)$$

によって単純距離を定義する。類別の際には、 D_i の小さいカテゴリから順に候補として選択してゆく。また、

$$\min_{i \in I} D_i(X, i) = D_i(X, i_0) \quad (2.3)$$

(ただし、 $i_0 \in I$)

のとき、入力サンプル X の属するカテゴリの第一候補として、カテゴリ i_0 を選択する。

2.1.2 単純距離の意味付けと問題点

カテゴリ i に属するサンプルを特徴抽出した結果、特徴ベクトルが $X = (x_1, x_2, \dots, x_K)$ (K は特徴ベクトルの次元数) として出現する確率密度関数を $P_i(X)$ とする。 $x_k (k=1, 2, \dots, K)$ がそれぞれ統計的に独立であると仮定することにより、

$$P_i(X) = P_i(x_1, x_2, \dots, x_K) = \prod_{k=1}^K p_{ik}(x_k) \quad (2.4)$$

(ただし、 $p_{ik}(x_k)$ は第 k element が x_k となる確率) とおくことができ、さらに各 x_k が (異なる i についても) 分散の等しい正規分布をなすと仮定すると、

$$P_i(X) = \frac{1}{(2\pi\sigma)^{K/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^K (x_k - m_{ki})^2\right\} \quad (2.5)$$

とおける。ここで $M_i = (m_{1i}, m_{2i}, \dots, m_{Ki})^t$ はカテゴリ i の平均特徴ベクトルで(2.1)式をもって推定するものとする。このとき最尤度法により、入力サンプル X に対して

$$\max_{i \in I} P_i(X) = P_{i_0}(X) \quad (i_0 \in I) \quad (2.6)$$

なる i_0 をその X の属するカテゴリとして選択する判定法が損失を最小にする意味で最適となることが知られている。すなわち(2.2)式により(2.5)式は

$$P_i(X) = \frac{1}{(2\pi\sigma)^{K/2}} \exp\left(-\frac{1}{2\sigma^2} D_i^2(X, i)\right) \quad (2.7)$$

2.2 重み付け距離

2.2.1 重み付け距離の定義

前節で述べたように特徴ベクトルの element の値の変動のしかたには、各カテゴリごとにまた各 element ごとにそれぞれ違いがある場合が多いと考えられる。このような状況に対処するためには、カテゴリを代表するものとして平均ベクトル M_i のみではなく、element のばらつきの度合いを示す指標 V_i をも導入してばらつきの大きさによって各 element の平均からの差異に重み付けを行えばよい。すなわち 2.2.1 項で述べた特徴ベクトル $F_i^j = (f_{1i}^j, f_{2i}^j, \dots, f_{Ki}^j)^t$ に対して、(2.1)式で得られるカテゴリ i の平均ベクトルを $M_i = (m_{1i}, m_{2i}, \dots, m_{Ki})^t$ とする。ばらつきの指標 V_i としては、学習サンプルから各カテゴリ・各 element ごとに分散を算出して用いることにする。分散ベクトルを $V_i = (v_{1i}, v_{2i}, \dots, v_{Ki})^t$ とすると、

$$v_{ki} = \frac{1}{J-1} \sum_{j=1}^J (f_{ki}^j - m_{ki})^2 \quad (2.8)$$

によって求めることができる。

これらを用いて、未知サンプル $X = (x_1, x_2, \dots, x_K)^t$ とカテゴリ i との距離の 2 乗を「未知サンプル X と平均ベクトル M_i との各 element の差異の加重 (二乗) 和」によって定義する。これを式で示すと、

$$D_v^2(X, i) = \sum_{k=1}^K \frac{(x_k - m_{ki})^2}{v_{ki}} \quad (2.9)$$

これは, Maharanobis 距離³⁾において分散共分散行列の対角成分のみを用いた場合に相当する. (2.9)式においてすべての k, i について $v_{ki}=1$ とおくと,

$$D_v^2(X, i) \Big|_{v_{ki}=1} = \sum_{k=1}^K \frac{(x_k - m_{ki})^2}{1} = D_i^2(X, i) \quad (2.10)$$

となって重み付け距離は単純距離の自然な拡張の一つであることがわかる.

2.2.2 重み付け距離の問題点

各カテゴリ・各 element ごとの値のばらつき具合を考慮した距離評価法として, 重み付け距離を提案したが, 各 element が独立ではなくその間での相関が強い場合には相関についても考慮する必要がある.

また重み付けのパラメータすなわち 'ばらつき' の度合い v_{ki} の推定の際の限界について考える必要がある. 本方式では(2.8)式のように各 element の標本分散をもって推定値とした, (2.9)式からわかるように, v_{ki} の非常に小さい項 (k_0) については, その分子の項 $(x_{k_0} - m_{k_0i})^2$ の大きさが非常に大きな重みをもっている. 本来の重み付け距離の目的はそこにあったわけだが, v_{k_0i}, m_{k_0i} のわずかな推定誤差によって重み付け距離に大きな誤差をもたらすおそれがある. また未知サンプル X の第 k_0 項の値が必要以上に大きい意味をもつことも考えられ, 他の element の情報が十分に生かされないという危険性もある.

Maharanobis 距離は分散共分散行列を用いて element 間の相関をも考慮にいたした距離評価法であるが, パラメータの推定における誤差を含めた上述の問題が存在する. また, 特徴ベクトルの次元数を K とすると, 1 カテゴリに対する重み付け距離の計算量と辞書に必要な記憶容量は $O(K)$ となるのに対し, Maharanobis 距離では計算量, 辞書の容量はいずれも $O(K^2)$ となる. したがって K が大きくなった際には, Maharanobis 距離は計算量・容量の点でかなり不利と思われる.

これらの問題に対処するため, すなわち,

- (1) element 間の相関が存在する場合も考慮し
- (2) 分散の非常に小さいものの誤差の影響が小さく
- (3) 分散の大きい変動成分を抑圧し
- (4) 計算量・記憶容量が Maharanobis 距離ほど大きくならない

ことを目的として提案したのが次節で述べる投影距離である.

2.3 投影距離

2.3.1 投影距離の定義

入力サンプルとカテゴリとの距離を求める際に, カテゴリを表現するものとして, 単純距離においては特徴ベクトルの平均, 重み付け距離においては平均ベクトルと各 element のばらつき具合を示す分散ベクトルを用いてきた. これに対して投影距離では

$$\begin{cases} M_i: \text{平均ベクトル} \\ \phi_{il}: \text{超平面構成ベクトル}(l=1, 2, \dots, L) \end{cases}$$

を用いてカテゴリを超平面によって近似表現する. そのためには, 重回帰分析の手法として知られる³⁾ ようにカテゴリ i に属する J 個の学習サンプルから分散共分散行列を求めて固有値問題を解き, 適切な次元 L を定めて L 次元超平面を構成すればよい.

投影距離は, 入力サンプルの特徴ベクトルとカテゴリを表現する近似超平面との距離によって定義される. 以上の過程を式で示すと次のようになる.

単純距離の場合と同様に, 特徴ベクトルを F_i^j とし, (2.1)式で求められるカテゴリ i の平均ベクトルを M_i とすると, 分布を示す分散共分散行列は

$$C_i = \frac{1}{J-1} \sum_{j=1}^J (F_i^j - M_i)(F_i^j - M_i)^t \quad (2.11)$$

によって求めることができる³⁾. 近似超平面を構成するためには, (2.11)式で得られた C_i を用いて

$$C_i \phi_{il} = \lambda_{il} \phi_{il} \quad (2.12)$$

ただし,

$$(\phi_{il}, \phi_{il'}) = \begin{cases} 1 (l=l' \text{ のとき}) \\ 0 (l \neq l' \text{ のとき}) \end{cases}$$

なる固有値問題を解き, λ_{il} の大きい方から L 個選びそれに対応する ϕ_{il} によって (M_i を通る) 超平面を張ればよい.

図2 において投影距離は,

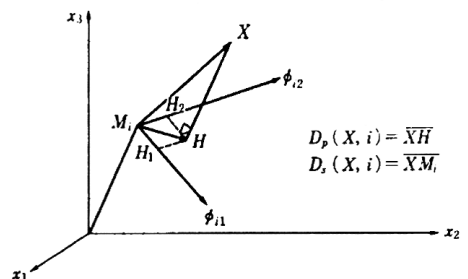


図2 投影距離 ($K=3, L=2$ の場合)
Fig. 2 Projection distance. ($K=3, L=2$)

$$D_p(X, i) = \overline{XH} \tag{2.13}$$

によって定義される。図において M_i はカテゴリ i の平均ベクトル、 ϕ_{i1}, ϕ_{i2} は超平面構成ベクトルであり、 $K=3, L=2$ の場合を示している。すなわち3次元特徴空間でカテゴリ i を2次元平面で近似している。三平方の定理等を用いると投影距離 $D_p(X, i)$ は

$$\begin{aligned} D_p^2(X, i) &= \overline{XM_i}^2 - \overline{M_iH}^2 \\ &= D_p^2(X, i) - \sum_{l=1}^L \{\phi_{il}'(X - M_i)\}^2 \\ &= \sum_{k=1}^K (x_k - m_{ki})^2 - \sum_{l=1}^L \{\phi_{il}'(X - M_i)\}^2 \end{aligned} \tag{2.14}$$

ただし

$$K=3, L=2$$

によって得られる。また一般の K, L (ただし $L \leq K$) についても同様の式によって投影距離を求めることができる。投影距離を用いて識別を行う方式を投影距離法とよぶ。

投影距離の定義式(2.14)式において $L=0$ とおくと

$$D_p^2(X, i) \Big|_{L=0} = \|X - M_i\|^2 = D_s^2(X, i) \tag{2.15}$$

となって単純距離と一致する。0次とは点のことであり、単純距離ではカテゴリの代表を平均ベクトル M_i のみ、すなわち1点で表していたことに対応している。このように、「1点→超平面」の拡張に対応して「単純距離→投影距離」が自然な拡張の一つであることがわかる。

2.3.2 投影距離と複合類似度

カテゴリを1点のみでなく超平面によって近似する考え方のひとつとして、投影距離法の以前に複合類似度法⁶⁾が提案されている。

複合類似度法と投影距離法の違いは、前者が、

- i) カテゴリを原点を通る超平面によって近似し、
- ii) 特徴ベクトルと超平面とのなす角の余弦で相違性を評価する

のに対して、後者では、

- i)' 各カテゴリごとに、その平均ベクトルを通る超平面によって近似し
- ii)' 幾何学的距離(長さ)で相違性を評価する

点にある。どちらがより優れた評価法であるかは、特徴として何を用いるかにも依存し一概に論ずることはむずかしい。

3. 認識実験

距離の評価法の良否は特徴抽出法として何を用いるかに大きく依存する。特徴抽出法の例として、3種の例をとりあげ、それらに対する種々の距離および複合類似度の評価を行った。

3.1 使用データ

電総研手書教育漢字データベース ETL-8⁵⁾ よりひらがな71文字(濁点・半濁点のあるものを含み、小文字「っ」「ゃ」「ゅ」「ょ」を除く)を用いた。160サンプル/カテゴリのうち、はじめの100サンプルを学習サンプルとして辞書を作成し、奇数番目のサンプルを用いて認識実験を行った。すなわち、サンプル1, 2, ..., 99, 100を辞書作成に使い、1, 3, 5, ..., 99の50個を学習サンプル、101, 103, ..., 159の30個を未知サンプルとして認識実験を行った。

3.2 線密度特徴に対する各距離の評価

線密度特徴^{6), 7)}とは、文字図形の構成が単純なものから複雑なものまで種々存在するという事実に着目して、複雑さの度合いを線密度で表現しようとしたものである。大局的線密度特徴は、図3のように水平・垂直・斜め $\pm 45^\circ$ の方向に文字図形を走査し、横切る文字線数(線密度)を計数する。次に各方向について K 区間に等分割して各区間内の線密度の平均を求め、これを特徴ベクトルの1 element とする。実装の際には $K=16$ で行ったため、各方向の特徴ベクトルは16次元、全体で $16 \times 4 = 64$ 次元のベクトルとした。

局所的線密度特徴は、図4に示すように文字図形をいくつかのブロックに分割し、各ブロック内に含まれる線密度を水平・垂直両軸に投影したのち、ブロック内で両軸それぞれ平均をとって特徴ベクトルの element とする(両軸について求めるから1ブロック当り2

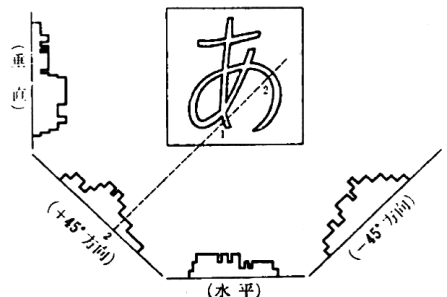


図3 大局的線密度特徴

Fig.3 Global stroke density feature.

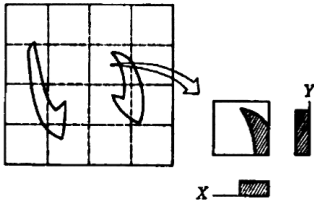


図4 局所的線密度特徴

Fig. 4 Local stroke density feature.

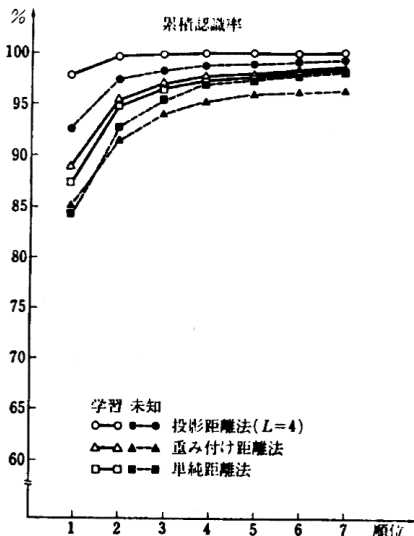


図5 『線密度特徴』を用いた場合の累積認識率
Fig. 5 Cumulative recognition ratio using stroke density feature.

element が得られる). 実装の際には 4×4 のブロックに分割したため, $2 \times 4 \times 4 = 32$ 次元のベクトルとなった.

したがって線密度特徴の特徴ベクトルは, $16 \times 4 + 16 \times 2 = 16 \times 6$ 次元となった*.

異なる3種の距離法を用いた場合の累積認識率を図5に示す. また近似超平面の次元数 L に対する投影距離法および複合類似度法の正読率を図6に示す. 投影距離法で $L=0$ の場合が単純距離に対応する.

図5からわかるように投影距離法の累積認識率は他の二つの距離法に比べてかなり高く, 投影距離法の線密度特徴に対する有効性を示している. また正読率のみに着目すれば, 距離概念の拡張によって重み付け距離・投影距離のいずれも従来の方式よりも優れている

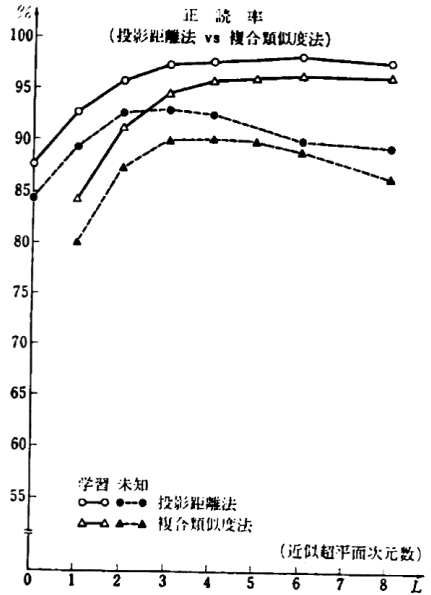


図6 『線密度特徴』を用いた場合の正読率
Fig. 6 Recognition result using stroke density feature.

といえる. 未知サンプルでの第二位以下の累積認識率では重み付け距離法が単純距離法よりも劣っているが, これは2.2.2節でも述べたように sensitive な要素にわずかな雑音のあった場合に, カテゴリとの距離が非常に大きくなって順位がかなり後退してしまうためと思われる.

図6からわかるように一般に投影距離法は複合類似度法に比べて優れている. それぞれの手法ごとに着目した場合, 学習サンプルと未知サンプルとでピークの位置に若干差がみられるのは, L を大にするにつれて学習サンプルをより精度よく近似しようとするために学習サンプル自身の偏りが辞書に反映されてしまうからであろうと考えられる. また, カテゴリに重畳しやすい雑音の成分は, 線密度特徴においては $K=16$ 次元内の $L=2 \sim 4$ 次元程度であると考えられる.

3.3 線方向特徴に対する各距離の評価

線方向特徴は, 文字を構成するおのおのの線分の方方向性に着目して黒点にラベル付けを行い, その方向に垂直な軸に投影してヒストグラム状にしたものである^{6), 8)}. 線分の方方向としては, 右上り・垂直・左上り・水平の4種類を考え図7のように各 element を投影して特徴ベクトルとする. 特徴ベクトルの次元数は 16×4 次元とした.

* 投影距離法において分散共分散行列を扱う際に 96 次元 \times 96 次元の行列では大きすぎて取扱いが不便なため, 特徴ベクトルを 16 次元 \times 6 次元になるように各方向ごとに分け, 16 次元 \times 16 次元の行列を6個用いて近似した.

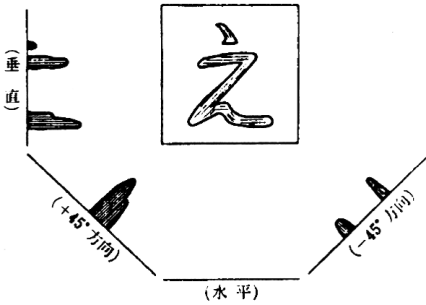


図7 線方向特徴
Fig.7 Stroke direction feature.

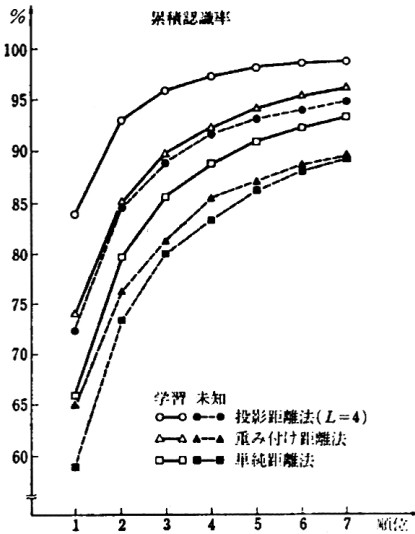


図8 「線方向特徴」を用いた場合の累積認識率
Fig.8 Cumulative recognition ratio using stroke direction feature.

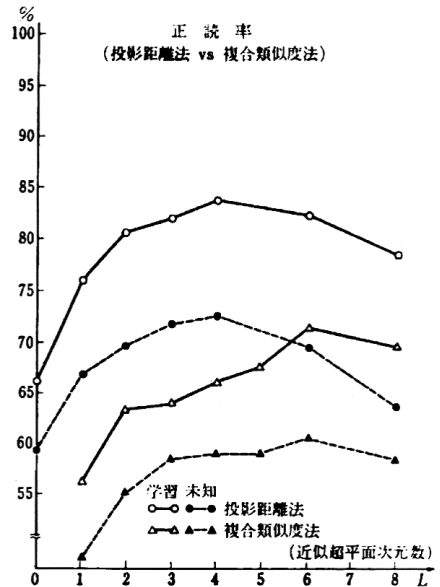


図9 「線方向特徴」を用いた場合の正読率
Fig.9 Recognition result using stroke direction feature.

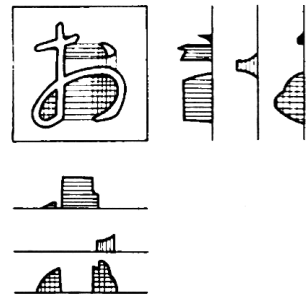


図10 線間領域特徴
Fig.10 Stroke domain feature.

3種の距離法を用いた場合の累積認識率を図8に、また近似超平面の次元数Lに対する投影距離法および複合類似度法の正読率を図9に示す。

図8からわかるように、学習サンプル・未知サンプルいずれも単純・重み付け・投影の順に認識率が改善されている。図5と比較して全体的に認識率が低いのは、特徴そのものの優劣の差であり、線方向特徴が曲線部の比較的多い「ひらがな」には不向きであることを示している。

図8から、投影距離法ではL=4程度が、複合類似度法ではL=6程度が適切であろうと思われる。

3.4 線間領域特徴に対する各距離の評価

線間領域特徴は、文字パターンの背景部に着目した特徴であり、各白点から上下左右方向に触手をのぼし

黒点との交わり具合によって白点にラベル付けを行い、それを各ラベルごとに水平・垂直両軸に投影して特徴ベクトルとしたものである(図10参照)⁶⁾。特徴ベクトルの次元数は1方向・1軸当たり16次元とし、全体で16×6次元のベクトルとなった。

3種の距離法を用いた場合の累積認識率を図11に、また近似超平面次元数Lに対する投影距離法および複合類似度法の正読率を図12に示す。

図11からわかるように、単純・重み付け・投影の順に認識率が改善されており、投影距離法の有効性が確認できる。

図12から、投影距離法ではL=2~4程度が、複合類似度法ではL=2程度が適切であろうと思われる。

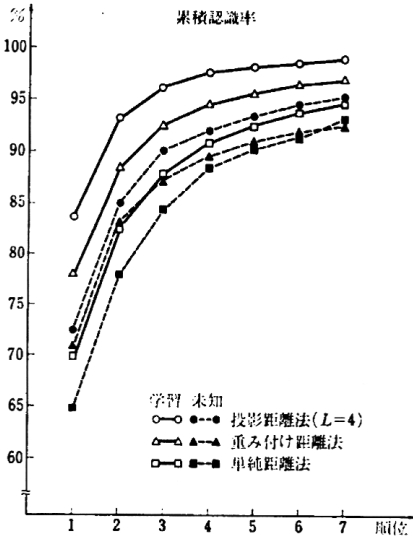


図 11 「線間領域特徴」を用いた場合の累積認識率
Fig.11 Cumulative recognition ratio using stroke domain feature.

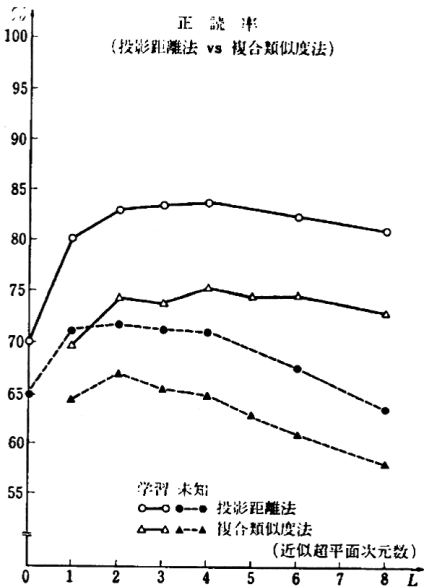


図 12 「線間領域特徴」を用いた場合の正読率
Fig.12 Recognition result using stroke domain feature.

4. おわりに

本研究の目的は、「手書き文字認識における効果的な距離評価法」の検討にある。そのために従来の一般的な評価法である単純距離法・重み付け距離法を拡張

した投影距離法を提案し、実験を通してその有効性を検討した。

本研究の成果として、3章で述べた3種の特徴については、

- 1) 投影距離法を用いることにより、単純距離法・重み付け距離法・複合類似度法のいずれよりも高い認識性能が得られる。
- 2) 投影距離法での近似超平面の次元数としては、特徴にも依存するが、 $K=16$ 次元内の $L=2\sim 4$ 程度が適切である。

等が明らかになった。

今後の課題としては、カテゴリを手書き漢字へ拡張した際の有効性の検討、特徴の組合せの際の評価方式等があげられる。

謝辞 本認識実験に使用した ETL-8 の利用を許可して下さったことに対し、電総研パターン情報部図形処理研究室各位に感謝の意を表する。

参 考 文 献

- 1) 桑原：大分類の段階がほぼ完成した手書き漢字認識，日経エレクトロニクス，No.279，pp.148-167(1981. 12. 7)。
- 2) 森他：パターン認識の現状と将来，日経エレクトロニクス特集 (1975. 1. 27)。
- 3) 奥野他：多変量解析，p.430，日科技連出版，東京(1971)。
- 4) 飯島：文字読取装置 ASPET/71，テレビジョン，Vol.27，No.3，pp.157-164(1973)。
- 5) 森，山本，山田，斉藤：手書教育漢字のデータベースについて，電総研集報，Vol.43，No.11，12，pp.752-773(1979)。
- 6) 藤井他：多面的な特徴による手書漢字認識の検討，信学技報，PRL 81-32(1981)。
- 7) 萩田，増田：大局的・局所的線密度の併用による手書き漢字の大分類，信学技報，PRL 80-23(1980)。
- 8) 塩野，馬場口，真田，手塚：方向性マッチングによる常用手書き文字の認識，電子通信学会論文誌，J 63-D(1980)。
- 9) 池田，田中，元岡：手書き文字認識における識別辞書の構成法に関する一検討，第23回情報処理学会全国大会，pp.675-676(1981)。
- 10) 池田，田中，元岡：手書き文字認識における距離概念の拡張とその評価，第24回情報処理学会全国大会，pp.671-672(1982)。

(昭和 57 年 4 月 7 日受付)
(昭和 57 年 9 月 6 日採録)