

## 単語全体の形状に注目した手書き日本語単語の認識

志 久 修† 中 村 彰††  
黒 田 英 夫††† 宮 原 末 治†††

本論文では、手書きされた日本語単語の認識方法について述べる。提案方法では、単語からの個別文字の切出しと認識を行わずに、単語全体を1つのパターンと見なして認識を行う。手書き日本語単語には、文字単体の手書き変形のほかに、文字間隔や文字サイズの変動、単語長さ方向への文字の伸縮が生じており、一般的な個別文字認識方法を単語画像にそのまま適用しても、高い認識性能を得ることは困難であると考えられる。そこで、提案方法では、上記の単語変形に対し、3段階の正規化処理〔図形間隔、図形幅、および単語画像サイズ(非線形正規化)〕と単語長さ方向へのDPマッチングを採用することにより対処している。郵政研究所の手書き漢字画像データベースIPTP CD-ROM2から作成した2種類の単語画像セット(町域部:1655件、都道府県名:174件)を用いて、提案方法と基本的な従来方法(個別文字に注目する方法)との正読率の比較実験を行った。その結果、これらの画像セットに対しては、従来方法において良好な条件(文字切出し正解率100%)での正読率と同等の結果〔正読率:96.50%(町域部),95.98%(都道府県名)〕が得られ、提案方法の有効性を明らかにした。

### A Method for Handwritten Japanese Word Recognition Based on Holistic Strategy

OSAMU SHIKU,<sup>†</sup> AKIRA NAKAMURA,<sup>††</sup> HIDEO KURODA<sup>†††</sup>  
and SUEHARU MIYAHARA<sup>†††</sup>

This paper describes a method for handwritten Japanese words recognition based on holistic strategy which treats a word as a whole unit without segmenting it into individual characters. Our method, to compensated for word shape distortions, adopts three processes of normalization (i.e. figure gap normalization, figure width normalization and word size normalization) and DP matching. We tested the proposed method for 1,655 (town name) and 174 (prefecture name) images in the IPTP CD-ROM2 (the database of handwritten KANJI character images). As a result, 96.50% of 1,655 town names and 95.98% of 174 prefecture names were correctly recognized.

#### 1. はじめに

手書き単語を認識する方法として、①単語から個々の文字の切出し・認識を行い、これらの文字認識結果と単語辞書とを照合(知識処理、後処理)する方法、②単語を個々の文字に分割せずに、単語全体を1つのユニットとして扱う方法(holistic strategy<sup>1)</sup>;ホリスティック法と呼ぶ)がある。

一般的に、手書き日本語単語を認識する場合は、前

者のアプローチがとられ、現在まで多くの方法が研究されている<sup>2)~4)</sup>。しかし、手書き日本語単語では文字の手書き変形のほかに、分離文字、接触文字、入込み文字が多く存在するため、高い文字切出し・認識精度が得られず、知識処理が有効に機能しない場合がある。

ところで、ホリスティック法は個別文字切出しを必要としないため、文字切出しが困難な低品質印刷英単語や筆記体英単語を認識対象として、多くの研究が行われている<sup>1),5)</sup>。しかし、これらと同様に文字切出しが困難な対象である手書き日本語単語については、文献6),7)などがあるだけで、ほとんど研究が行われていない。そのためホリスティック法による手書き日本語単語の認識性能は明らかにされていない。そこで、本論文では、ホリスティック法に基づく日本語手書き単語の認識方法を新たに提案し、郵政研究所の手書き

† 佐世保工業高等専門学校  
Sasebo National College of Technology

†† 熊本工業大学  
Kumamoto Institute of Technology

††† 長崎大学工学部  
Faculty of Engineering, Nagasaki University

漢字画像データベース IPTP CD-ROM2<sup>8)</sup>から作成した手書き宛名画像を例とした実験により提案方法の認識性能を明らかにする。

以下本論文では、2章で提案する認識方法について述べ、3章で提案方法の性能評価実験とその結果について述べる。なお、本論文で対象とする単語画像は、縦書きされた日本語単語の2値画像である。

## 2. 提案する認識方法

### 2.1 基本的な考え方

図1に提案する認識方法の流れを示す。本方法は、従来から研究されている手書き文字認識法を単語認識に拡張したものである。手書き文字認識では、非線形正規化および方向特徴の採用、分類・識別処理の高度化が、認識率向上に有効であることが明らかにされている<sup>9)</sup>。特に、非線形正規化<sup>10)~12)</sup>は、文字単体の手書き変形の吸収に効果的である。しかし、手書き単語では、文字単体の手書き変形に加え、文字間隔や文字サイズの変動、単語長さ方向への文字の伸縮(今回は縦書き単語を対象とするため縦方向への伸縮)が生じており、通常非線形正規化をそのまま適用しても、十分な効果が得られないと考えられる。そこで、本方法ではこれらの変形を吸収するため、入力単語画像(図1(a))に対し、単語を構成する図形の間隔の正規化(図1(b))、図形の幅の正規化(図1(c))、および単語画像のサイズの正規化(図1(d))の3段階の正規化処理を行う。さらに、単語長さ方向の伸縮を吸収するため、手書き文字認識法として提案されているDPマッチング<sup>13),14)</sup>を適用する。

以下、各処理について詳細に説明する。なお、単語画像の左上を原点とし、単語の幅方向を  $x$  方向、長さ方向を  $y$  方向とする。

### 2.2 3段階の正規化

正規化処理では、単語の手書き変形を吸収すること、および単語画像サイズを一定サイズに正規化することを目的とする。

#### (1) 前処理(図1(a))

まず、単語全体に外接する矩形のサイズを、 $x$  サイズが  $L_x$  画素となる倍率で線形正規化(縮小)する。これは、入力単語画像サイズを単語形状が損なわれない程度に縮小することで、以降の画像処理に要する時間を削減するためである。今回は  $L_x = 96$  画素とした。次に、画像に含まれる孤立雑音、および画像端に接触している小図形を除去し、さらに、 $3 \times 3$  平均値フィルタを用いて、輪郭線の凹凸を滑らかにする。

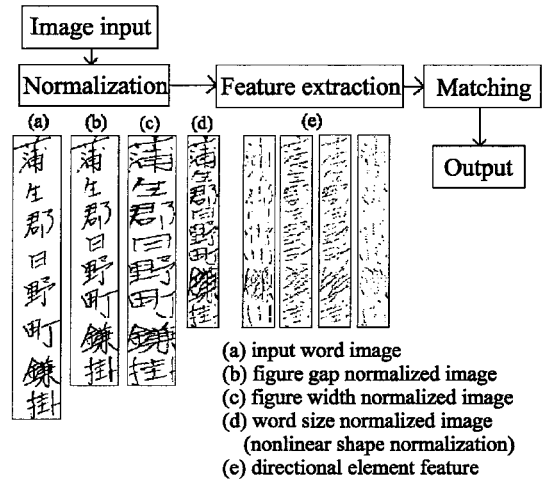


図1 提案する認識方法の流れ

Fig. 1 Flowchart of the proposed method.

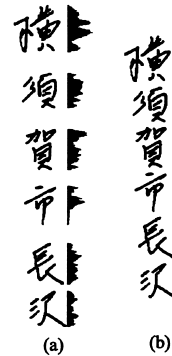


図2 図形間隔の正規化((a) 入力単語画像と射影ヒストグラム、(b) 図形間隔の正規化画像)

Fig. 2 Figure gap normalization ((a) input word image, (b) figure gap normalized image).

#### (2) 図形間隔の正規化 $N1$ (図1(b))

単語を構成する各図形どうしの間隔が広い場合、それらを一定間隔になるように正規化する。これは、図2(a)に示すように、手書き単語画像の  $x$  方向の射影ヒストグラムを求め、射影値0の範囲(図形間の空白部分に対応する範囲)が  $w$  画素以上となる部分を  $w$  画素に縮めることにより行う。これにより、図2(b)に示すように、図形間隔の最大値が  $w$  画素に揃えられ、画像サイズが圧縮できる。

#### (3) 図形幅の正規化 $N2$ (図1(c))

単語を構成する各図形の横幅を画像幅いっぱいになるように広げる。これは、文献11)の非線形正規化を  $x$  方向のみに適用することにより実現する。図3をもとに、具体的な手順を述べる。

①単語画像の任意の白画素に対し、その画素を中心とする  $x$  方向の白ラン長(ストローク間隔)を求め

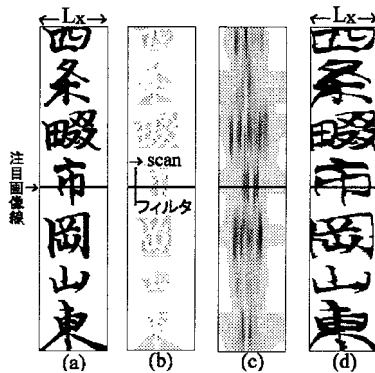


図3 図形幅の正規化 (a) 図形間隔の正規化画像 (入力), (b)  $x$  方向の線密度画像, (c) 平滑化後の線密度画像, (d) 図形幅の正規化画像)

Fig. 3 Figure width normalization ((a) figure gap normalized image (input), (b) line density image, (c) smoothed line density image, (d) figure width normalized image).

る。注目画素の左または右方向の走査線上に黒画素がない場合 (すなわち画像端) は、ストローク間隔は大きな定数とする。

② 求めたストローク間隔の逆数を、注目画素の  $x$  方向の線密度とする。

③ 単語画像のすべての白画素について上記①, ②を行い、 $x$  方向の線密度画像を作成する。注目画素が黒画素のときには、十分小さい線密度を与える。図3(b)は図3(a)から求めた線密度画像の例である。

④ 求めた線密度画像を  $y$  方向に標準偏差  $\sigma$  の1次元ガウシアンフィルタ (フィルタサイズ  $l = 6\sigma$ ) により平滑化する。図3(c)は図3(b)を  $y$  方向に平滑化した例である (フィルタサイズ  $l = L_x$ )。

⑤ 平滑化後の線密度画像から  $x$  方向の走査線上の画素列 (画像線) を1本取り出し (注目画像線と呼ぶ)、その注目画像線上の画素を、平滑化線密度と標準化間隔の積が均等になるように座標変換 (再標準化) を行う。平滑化線密度が大きい部分では標準化間隔が密に、すなわち標本点数が多くなるため、変換後の空間が引き伸ばされ、逆に平滑化線密度が小さい部分では標準化間隔が粗く、すなわち標本点数が少なくなるため、変換後の空間が縮められる。図3では同一の注目画像線における画像の処理過程を示している。

⑥ 注目画像線を  $y$  方向に1画素ずつ移動させつつ、線密度画像のすべての画像線について、⑤を繰り返す。

平滑化線密度が大きくなる図形内部領域が引き伸ばされ、逆に線密度が小さくなる左右の画像端領域が縮められ、結果として図3(d)に示すように各図形幅がほぼ一定に揃えられる。



図4 フィルタサイズ  $l$  を変えた場合の図形幅正規化画像 (a) 図形間隔の正規化画像 (入力), (b) 図形幅の正規化画像  $l = 0.5L_x$ , (c)  $l = 1.0L_x$ , (d)  $l = 3.0L_x$ )

Fig. 4 Effect of filter size  $l$  ((a) figure gap normalized image (input), (b) figure width normalized image  $l = 0.5L_x$ , (c)  $l = 1.0L_x$ , (d)  $l = 3.0L_x$ ).

なお、文献11)では2次元的な線密度が用いられているが、今回は  $x$  方向への正規化効果を高めるため、 $x$  方向の1次元的な線密度<sup>12)</sup>を採用した。

図4にフィルタサイズ  $l$  を変えた場合の図形幅正規化画像の例を示す。フィルタサイズ  $l$  が小さいほど局所性が強くなり、単語中の小さい図形 (たとえば図4(a)の「田」) を引き伸ばす効果 (正規化効果) が大きい。たとえば図4(b)の「郡」に見られる垂直方向のストロークも振動している。逆に  $l$  が大きいほど、正規化効果は小さくなるが、ストロークの振動も小さくなっていることが分かる。

(4) 単語画像サイズの正規化  $N3$  (図1(d))

単語画像サイズを、手書き変形を整える効果があるストローク間隔に注目した非線形正規化<sup>12)</sup>を用いて、 $64 \times 64n$  (倍率  $n = 2, 3, \dots$ ) 画素に正規化する。非線形正規化<sup>12)</sup>は次の手順で行う。

①  $x$  方向の線密度画像を作成し、各点の線密度を、 $x$  軸上に投影する。

②  $x$  軸について、投影値と標準化間隔の積が均等になるように、座標変換を行う。

③  $y$  軸の座標変換については、上記①, ②において  $x$  を  $y$  として同様に行う。

図形幅正規化では  $x$  方向の局所的なバランスを整えたのに対し、ここでは画像全体のバランスを整えることを狙っている。図5(a), (b)に図形幅の正規化後の画像に対し、倍率  $n = 3, 6, 9$  として単語画像サイズの正規化を行った例を示す。ここで、 $n$  は後述する実験により決定する。正規化画像サイズを  $64 \times 64$  画素の倍数とした理由は、ETL9Bから作成する学習パター

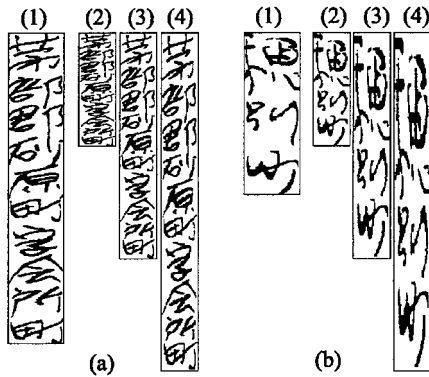


図5 単語画像サイズの正規化( (1) 図形幅の正規化画像(入力;  $L_x = 96$  ), (2) 単語画像サイズの正規化画像  $n = 3$  ( $64 \times 192$  画素), (3)  $n = 6$  ( $64 \times 384$  画素), (4)  $n = 9$  ( $64 \times 576$  画素) )

Fig. 5 Word size normalization ((1) figure width normalized image (input), (2) word size normalized image  $n = 3$ , (3)  $n = 6$ , (4)  $n = 9$ ).

ン(作成法は後述)と同じ大きさにするためである。

### 2.3 特徴抽出

正規化単語画像から, 文字線の輪郭部の方向性に注目した方向線素特徴<sup>15)</sup>を抽出する。正規化単語画像 ( $64 \times 64n$  画素) を, 1つのメッシュサイズが  $8 \times 8$  画素となるように等分割し, 各メッシュに含まれる縦横斜め4方向の線素数をそれぞれカウントする。この場合, メッシュ数は  $x$  方向8個  $\times y$  方向  $8n$  個, 特徴ベクトルの次元数は  $x$  方向8メッシュ  $\times$  線素方向  $4 \times y$  方向  $8n$  メッシュ =  $256n$  次元となる。この特徴ベクトルを,  $y$  方向の時系列信号と見なし, 次のように表す。

$$\mathbf{V} = (v_1, v_2, \dots, v_y, \dots, v_{8n})$$

ただし,  $v_y = (v_{y,1}, v_{y,2}, \dots, v_{y,32})$  である。

### 2.4 DP マッチング

$y$  方向に伸縮する DP マッチングを用いて, 未知特徴ベクトルと標準パターンベクトルの距離を求め, 最短距離をとる単語を認識結果として出力する。ここで, 伸縮方向を  $y$  方向のみとする理由は, 正規化処理で  $x$  方向の変動はある程度吸収していることと, 縦書き単語では  $x$  方向に比べ  $y$  方向の伸縮の変動が大きいと考えられるためである。

いま, 未知特徴ベクトル  $\mathbf{V}$  と単語  $W_k$  ( $k = 1, 2, \dots, K$ ,  $K$ : 単語種類数) の標準パターンベクトル  $\mathbf{U}^k$  をそれぞれ,

$$\mathbf{V} = (v_1, v_2, \dots, v_i, \dots, v_{8n})$$

$$\mathbf{U}^k = (u_1^k, u_2^k, \dots, u_j^k, \dots, u_{8n}^k)$$

とすると, これらのベクトル間の距離  $D(\mathbf{V}, \mathbf{U}^k)$  は, 次の手順で求めることができる<sup>16)</sup>。

(1) 初期条件  $g(1, 1) = d(1, 1)$

ただし,

$$d(i, j) = \|\mathbf{v}_i - \mathbf{u}_j^k\|^2 = \sum_{m=1}^{32} (v_{i,m} - u_{j,m}^k)^2$$

である。

(2)  $i = 2, 3, \dots, 8n$  について (3) ~ (4) を実行

(3)  $j = i - r, i - r + 1, \dots, i + r$  について (4) を実行 (ただし,  $j < 1$  なら  $j = 1$ ,  $j > 8n$  なら  $j = 8n$ ,  $r$  は整合窓  $2r$  の  $1/2$  幅で, 整合窓サイズと定義する)

(4)  $g(i, j)$  を次式で計算する。

$$g(i, j) = \min \begin{pmatrix} g(i-1, j) \\ g(i-1, j-1) \\ g(i, j-1) \end{pmatrix} + d(i, j)$$

(5)  $D(\mathbf{V}, \mathbf{U}^k) = g(8n, 8n)$

上記の手順で, すべての標準パターンベクトルとの距離  $D(\mathbf{V}, \mathbf{U}^k)$ ;  $k = 1, 2, \dots, K$  を求め, 最小距離をとる単語  $W_{k^*}$  を認識結果とする。

## 3. 実験および考察

提案方法の認識性能を調べるため, 次の2つの実験を行う。なお, 実験では, 手書き単語の例として, 郵便宛名の町域部および都道府県名を用いる。

実験1: 提案法で用いるパラメータの決定

実験2: 提案方法と他の方法との性能比較

### 3.1 実験データ

表1に実験データを示す。評価パターンは, 郵政研のIPTP CD-ROM2から作成する。IPTP CD-ROM2には, 11,128件(サンプル名: AI00001~AI11128)の郵便宛名のサンプル画像が含まれている。これらの中から, ①単語の全文字がETL9B内の字種で構成, ②縦書き1段組, の2つ条件を満たすサンプル画像を取り出し, 各サンプルについている付加情報を利用して2値化した後, 町域部または都道府県名のみを抽出した2値画像を評価パターンとする。図6に評価パターン((a)~(c)町域部, (d), (e)都道府県名)の例を示す。筆記具として毛筆, サインペン, ボールペンなどが使われている。

学習パターンは電総研のETL9Bから作成する<sup>6)</sup>。まず, 評価パターンに対応する単語を文字コード列の形式で作成する。次に, 各文字コード列の先頭文字から順に, 対応する文字画像( $64 \times 64$ 画素)をETL9Bの第  $a$  セット ( $a = 1, 2, \dots, 200$ ) から取り出し, 重ねさず縦方向に並べていく。文字数  $L$  の単語画

表 1 実験データ  
Table 1 Experimental data.

実験	実験データ	単語構成文字数											合計 (平均文字数)
		2	3	4	5	6	7	8	9	10	11	12	
実験 1 (町域部)	学習パターン ( AI00*** の単語種 )	15	10	40	198	201	100	85	42	17	9	2	719 (6.2)
	評価パターン ( AI00*** )	16	12	43	226	239	136	93	45	17	9	2	838 (6.2)
実験 2 データセット A (町域部)	学習パターン ( AI00*** ~ AI02*** の単語種 )	21	28	95	587	581	289	249	124	52	16	4	2,046 (6.2)
	評価パターン ( AI01*** ~ AI02*** )	6	18	76	504	487	244	186	88	36	8	2	1,655 (6.2)
実験 2 データセット B (都道府県名)	学習パターン	44	3	0	0	0	0	0	0	0	0	0	47 (2.1)
	評価パターン ( AI01*** ~ AI02*** )	159	15	0	0	0	0	0	0	0	0	0	174 (2.1)

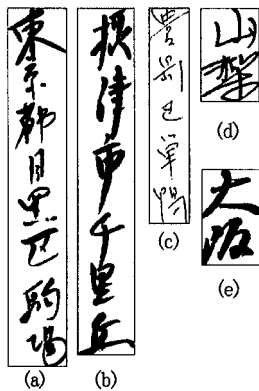


図 6 評価パターンの例  
Fig. 6 Examples of test patterns.

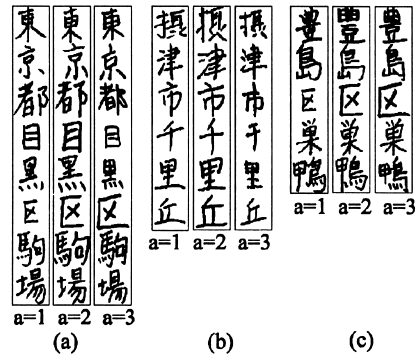


図 7 学習パターンの例 ( ETL9B の第 1, 2, 3 セットの文字画像から作成 )  
Fig. 7 Examples of learn patterns.

像サイズは、横  $64 \times$  縦  $(64 \times L)$  画素となる。以上の手順で、1 種類の単語 ( 文字コード列 ) につき、200 個ずつの単語画像を作成する。これらを学習パターンとする。図 7 に、図 6 (a) ~ (b) ( 東京都目黒区的場、摂津市千里丘、豊島区巣鴨 ) に対応する学習パターンの一部 ( $a = 1, 2, 3$ ) を示す。

また、単語  $W_k$  の標準パターンベクトル  $U^k$  は、単語  $W_k$  について作成した学習パターン 200 個に対し、それぞれ正規化 (2.2 節,  $L_x = 64$  画素)、特徴抽出 (2.3 節) を行い得られた特徴ベクトルの各要素の平均値である。すなわち、単語 1 種につき、1 つの標準パターンベクトルを持つ。

3.2 実験 1

表 1 に示す実験データ ( 学習パターン 719 種、評価パターン 838 ) を用いて、正規化法、正規化画像サイ

ズの倍率  $n$ 、DP マッチングの窓サイズ  $r$  を変えたときの認識性能を調べる。

(1) 正規化処理の効果

表 2 に正規化法を変えた場合の正読率を示す。正規化サイズ  $n = 6$ 、DP マッチングの窓サイズ  $r = 5$  としている ( $n, r$  の決定理由は次の (2) で述べる)。

個別文字認識と同様に単語画像サイズの正規化 ( 非線形正規化 )  $N3$  だけ行う場合に比べ、図形間隔の正規化  $N1$  と図形幅の正規化  $N2$  をそれぞれ併用する場合 ( $N1 + N3, N1 + N2 + N3$ ) が正読率が高くなっている。特に、図形幅の正規化  $N2$  の効果は大きく、フィルタサイズ  $l = 1.0L_x$  のときに最高正読率 97.14% となっている。これより、縦書き手書き日本語単語の認識において、 $N1 + N2 + N3$  の 3 段階の正規化が有効であることが分かる。なお、図形間隔の正規化  $N1$  において  $w$  は 5 画素とした。これは  $w$  を

表2 正規化処理の効果 (N1: 図形間隔正規化, N2: 図形幅正規化, N3: 単語画像サイズの正規化)  
Table 2 Effect of normalization.

正規化処理	N3	N1+N3	N1+N2+N3				
			$l = 0.5L_x$	$1.0L_x$	$1.5L_x$	$2.0L_x$	$3.0L_x$
正読率 [%]	90.93	91.29	95.70	97.14	96.78	96.30	95.82

表3 パラメータ  $n$  と窓サイズ  $r$  を変えた場合の正読率 (下線部は各  $n$  における最高正読率を表す)  
Table 3 Recognition rate for combination of  $n$  and  $r$ .

$n$ (画像サイズ, $y$ 系列長)	窓サイズ $r$								
	0	1	2	3	4	5	6	7	8
2 (64×128,16)	73.15	<u>82.22</u>	81.98	81.62	81.74	81.74	81.74	81.74	81.74
3 (64×192,24)	80.55	92.12	<u>93.68</u>	93.56	93.44	93.32	93.32	93.32	93.32
4 (64×256,32)	82.58	92.00	94.63	<u>95.46</u>	95.35	95.35	95.35	95.35	95.35
5 (64×320,40)	81.86	92.24	94.87	96.18	<u>96.42</u>	96.42	96.42	96.42	96.42
6 (64×384,48)	80.91	89.98	94.39	96.06	96.90	<u>97.14</u>	97.02	97.02	97.02
7 (64×448,56)	80.67	89.50	93.44	95.11	96.66	<u>97.02</u>	96.90	96.90	96.90
8 (64×512,64)	79.48	88.43	92.60	94.75	95.94	96.42	96.54	<u>96.66</u>	96.66
9 (64×576,72)	78.28	87.11	90.93	93.44	95.23	96.30	<u>96.66</u>	96.54	96.54
10 (64×640,80)	77.09	85.44	89.86	92.96	94.03	94.87	95.58	<u>95.70</u>	95.58

0~11 画素まで変化させたときに最も高い正読率が得られた  $w$  の最小値である。  $w$  による正読率の変化は大きなものではなかったが、画像サイズが正規化前の約 90% に縮小され、それともなう処理速度の向上が得られた。

(2) 正規化サイズ (特徴ベクトル次元数) と整合窓サイズ

表 3 に正規化画像の  $y$  サイズ (=  $64n$ ) および特徴ベクトルの  $y$  系列長 (=  $8n$ ) を決めるパラメータ  $n$  と、DP マッチングの窓サイズ  $r$  をそれぞれ変化させたときの正読率を示す。このとき、 $N1+N2+N3$  の 3 段階の正規化 ( $N2$  のフィルタサイズは  $l = L_x$ ) を用いた。

①すべての  $n$  において、窓サイズ  $r = 0$  (線形マッチング, すなわちユークリッド距離) で正読率は最低となり、窓サイズ  $r$  が特徴ベクトル  $y$  系列長  $8n$  の約 6~10% の大きさのとき最高正読率 (表 3 の下線部) となる。  $y$  方向の位置ずれの吸収に DP マッチングが有効であることが分かる。また、  $y$  方向の位置ずれの大きさは、特徴ベクトル  $y$  系列長の約 6~10% 程度であることが推測できる。

②正規化画像サイズの倍率  $n$  が、評価パターンの平均文字数 (= 6.2; 表 1 参照) にほぼ等しい ( $n = 6$ ) とき、最も高い正読率 97.14% ( $r = 5$ ) となる。  $n = 3, 6, 9$  に対する最適 (最高正読率を与える) 窓サイズは表 3 より  $r = 2, 5, 6$  と求められる。図 8 はこの最適窓サイズ条件の 3 状態 ( $n, r$ ) = (3, 2), (6, 5), (9, 6) の下で、種々の単語 (構成文字数 2~12) の認識結果を構成文字数ごとの正読率で示したものである。評価パターン数が十分多い 4~9 文字の範囲に注目すると、

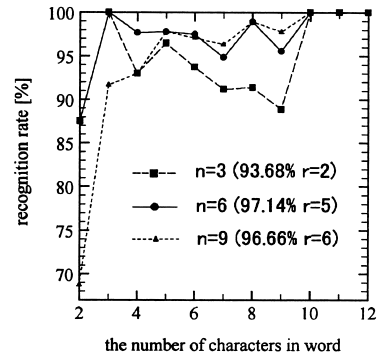


図8 単語構成文字数と正読率の関係

Fig. 8 The relation between the number of characters in word and recognition rate.

$n = 6$  では構成文字数による正読率の変動が ( $n = 3, 9$  の場合に比べ) 比較的少ないのに対し、  $n = 3$  では文字数が多い単語、  $n = 9$  では文字数が少ない単語の正読率が低くなる傾向がある。これは、文字数が多い単語を極端に縮めると大きな形状変形 (たとえば図 5(a) の  $n = 3$ ) や文字線の消失が生じ、逆に文字数が少ない単語を極端に伸ばすと  $y$  方向への文字線の大きな位置ずれ (たとえば図 5(b) の  $n = 9$ ) が生じる場合があるためと考えられる。

なお、  $n$  の値を入力単語の文字数にあわせて変えることも考えられるが、文字数の推定間違いによる認識失敗などの影響を除き、提案方法の性能だけを調べるため、今回は固定とした。

3.3 実験 2

表 1 のデータセット A (町域部) とデータセット B (都道府県名) の 2 種類の単語データを用いて、以下の 5 つの方法の比較を行う。なお、この実験で用いる

表 4 方法 4 で用いた文字認識の性能

Table 4 Performance of character recognition in method 4.

文字データ	学習用文字パターン (ETL9B から作成)	評価用文字パターン	累積文字認識率 [%]			
			1 位	10 位	20 位	30 位
ETL9B 全字種	3,036 字種 × 先頭 180 セット	3,036 字種 × 残り 20 セット	91.33	99.00	—	—
データセット A の評価パターン	830 字種 × 先頭 180 セット	10,272 文字 (1,655 件)	58.01	83.88	88.81	91.18
データセット B の評価パターン	72 字種 × 先頭 180 セット	363 文字 (174 件)	77.41	95.59	98.07	98.90

表 5 各方法での単語正読率の比較

Table 5 Comparison of the proposed method with other methods.

実験データ	提案法	方法 1	方法 2	方法 3	方法 4 ( $m$ : 文字候補数)		
					$m = 10$	$m = 20$	$m = 30$
データセット A: 1,655 件 (類似単語 127 件)	96.50 (94.49)	65.44	80.06	87.92	94.20 (79.53)	96.31 (85.04)	97.16 (86.61)
データセット B: 174 件 (類似単語 99 件)	95.98 (96.97)	89.08	93.68	94.83	89.08 (77.78)	91.38 (89.90)	91.38 (89.90)

評価パターンはすべて実験 1 では使用していないものばかりである。

提案法: 3 段階正規化 + DP マッチング

方法 1<sup>6)</sup>: 正規化  $N_3$  + 線形マッチング

方法 2: 3 段階正規化 + 線形マッチング

方法 3: 正規化  $N_3$  + DP マッチング

方法 4: 個別文字切出し + 個別文字認識 + 知識処理

提案法および方法 1~3 において、データセット A では  $n = 6$ ,  $r = 5$ , データセット B では  $n = 2$ ,  $r = 1$  とした。これは、実験 1 により  $n$  は平均文字数に近い値で最高正読率が得られること、表 3 より  $n = 6$  では  $r = 5$  のとき、 $n = 2$  では  $r = 1$  のとき、それぞれ最高正読率となったためである。また、線形マッチングは DP マッチングの  $r = 0$  (すなわちユークリッド距離) のときである。

方法 4 は個別文字に注目する基本的な方法で以下のとおりである。

文字切出し: 入力単語画像から付加情報の文字数および文字位置座標をもとに文字領域を切り出す。すなわち、切出し正解率 100% である。

文字認識: 入力単語画像の  $i$  番目の文字領域に対し文字認識を行い、上位  $m$  位までの文字候補 (文字種  $C_{i,j}$ , 順位  $j = 1, 2, \dots, m$ ) を出力する。文字認識は、線形正規化 (文字サイズを  $64 \times 64$  画素にする)、非線形正規化<sup>12)</sup> (字形を整える)、方向線素特徴量<sup>15)</sup> ( $8 \times 8$  メッシュ  $\times 4$  方向 = 256 次元) およびユークリッド距離により行う。表 4 にこの文字認識法の性能を示す。表 4 は、ETL9B の先頭 180 セットを学習用文字パターンとした標準文字特徴ベクトルを用いて、ETL9B の残り 20 セット、データセット A の

評価パターン 1,655 件を構成する 10,272 文字 (830 字種)、データセット B の評価パターン 174 件を構成する 363 文字 (72 字種) をそれぞれ認識したときの累積文字認識率を示している。10 位までの累積文字認識率が、ETL9B 文字パターンでは 99.00% であるのに対し、データセット A, B の文字パターンでは字種数が少ないにもかかわらず、それぞれ 83.88%, 95.59% と低くなっている。これより、データセット A, B の文字品質が悪く、文字単位の非線形正規化を用いても十分に変形が吸収されていないことが分かる。

知識処理: 文字候補順位の和を評価値とする簡単な方法<sup>4)</sup>を用いる。単語辞書 (データセット A: 2,046 種類、データセット B: 47 種類の単語がそれぞれ登録) から、入力単語と同一文字数の単語をすべて取り出す。取り出した単語集合の中の 1 つに注目し、注目単語の  $i$  番目の文字  $c_i$  が文字候補  $C_{i,j}$  ( $j = 1, 2, \dots, m$ ) に含まれている場合、 $C_{i,j}$  の順位  $j$  を注目単語の評価値として累積加算する。含まれていない場合は、 $m + 1$  を累積加算する。そして、単語集合の中で、最小の評価値となる単語を認識結果として出力する。なお、最小評価値となる単語が複数存在する場合は、その中に正解単語が含まれていても、出力を 1 つに絞り込めないため、読み取り失敗としている。

表 5 に、各方法での単語正読率の比較を示す。まず、提案法と方法 1~3 とを比較する。方法 1~3 は構成文字数が少ない単語 (データセット B) では有効であるが、文字数が多い単語 (データセット A) では単語変形が大きくなりそれらを吸収できないため正読率が低下している。それに対し、提案法では文字数が多い単語でも、正読率の低下は見られない。これより、手書

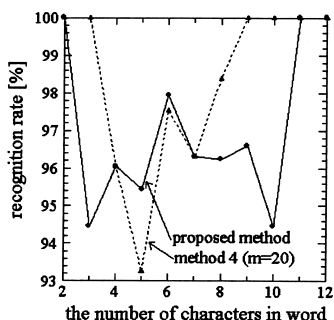


図9 提案方法と方法4 ( $m = 20$ )における単語構成文字数ごとの正読率

Fig. 9 Recognition rate for proposed method and method 4 ( $m = 20$ ).

き単語の変形吸収に3段階の正規化およびDPマッチングの併用が有効であると考えられる。

次に、提案法と方法4(文字候補数  $m = 10, 20, 30$ )とを比較する。表5より、データセットAに対し、提案方法の正読率96.50%は、方法4の  $m = 20$  のときの正読率96.31%とほぼ等しくなっている。図9は、このときの単語構成文字数ごとの正読率である。図9より、構成文字数が少ない単語(4~7文字)では提案法の正読率が高く、構成文字数が多い単語(7~10文字)では方法4が高くなっている(構成文字数2~3, 11~12の単語は、評価パターン数が少ないため無視する)。さらに表5のデータセットB(構成文字数2~3文字)では、提案法の正読率(95.98%)が方法4(91.38%)より高くなっている。これらより、構成文字数が少ない単語に対しては、提案方法が方法4よりも有利であると考えられる。方法4では、単語構成文字のうち正しく認識できた文字数が多いほど知識処理が有効に機能するが、文字数が少ない単語では、文字を正しく読める確率が低くなるため知識処理が機能せずに正読率が低くなると考えられる。なお、文字候補数  $m$  を増やすと文字を読める確率が高くなる(累積文字認識率が向上する)が、同時に文字の組合せの増大により誤った単語が構成される場合がある。

表5の( )内の数値は、データセットA, Bに含まれる類似単語を持つ単語数およびそれらに対する正読率である。ここで、「類似単語を持つ単語」とは、1文字だけが異なる以外は同一文字数・同一文字順を持つ単語(類似単語)が別に存在する単語と定義する。たとえば、「長崎県」は、別に「長野県」が存在するので、類似単語を持つ単語である。データセットA, Bの類似単語における提案法と方法4( $m = 20$ )の正読率の変化を比較すると、提案法(96.50% → 94.49%, 95.98% → 96.97%)に比べ、方法4(96.31% → 85.04%, 91.38%

→ 89.90%)の低下が大きい。今回のサンプルでは類似単語間の相違文字の形状の違いが比較的大きかったため(たとえば長崎県, 長野県の「崎」と「野」), 提案法ではその形状の違いにより距離値が変わり類似単語を比較的良好に見分けることができたと考えられる。方法4では相違文字の認識に失敗した( $m$ 位候補に入らなかった)場合、類似単語に対し同一の評価値  $m + 1$  が累積加算されるため、類似単語の評価値が同じになり正解を1つに絞り込めずに失敗となったものである。

ところで、方法4において文字切出しを自動化する場合、文字切出し失敗による文字認識精度の低下、高い累積文字認識率を維持するための文字候補数の増加などの問題が生じ、単語の正読率は今回の値より大きく低下すると考えられる。第3回文字認識技術コンテスト<sup>17)</sup>でも、文字切出しを行った場合、正読率が低下することが報告されている。特に接触・入込み文字の切出しは困難であるため、方法4ではこのような文字を持つ単語の正読率は非常に悪くなると考えられる。ちなみに提案法では、データセットA内の接触・入込み文字を持つ評価パターン604サンプルに対する正読率は94.37%であった。すなわち、提案法では、約2%(96.50% → 94.37%)の正読率の低下はあるが、接触・入込み文字の影響は大きくはなかった。なお、接触・入込み文字の判定は、各文字領域(サンプルの付加情報)の接触・重畳関係をもとに行った。

一方、提案法では方法4に比べ、次のような問題点がある。①方法4では文字認識辞書と単語辞書を必要とするが、単語は文字コード列として記憶されているので単語種類が増加しても単語辞書の容量はそれほど大きくなる。それに対し提案法では、単語を特徴ベクトルとして記憶しているため、単語種類が増加した場合、記憶容量が大きく増加する。実験2のデータセットAでは、記憶容量は約6MB(=1536次元×2,046単語種;1要素2バイトで表現)となる。②方法4の知識処理は文字コードレベルの照合であるのに対し、提案法では特徴ベクトルレベルでの照合となる。そのため単語種類が増加した場合、照合に要する時間が大きくなる。実験2のデータセットAでは、単語1つあたりの認識時間は約0.9秒(PentiumII 400MHz)であった。③方法4では文字を単位としているため文章認識にも拡張が可能であるのに対し、提案法では単語を単位としているため文章を単語ごとに分割するなどの困難な問題が生じ、拡張は簡単ではない。

このような問題点があるが、認識対象が今回の実験で扱った程度の手書き日本語単語(平均文字数6文



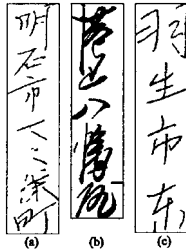


図 10 失敗例

Fig. 10 Examples of mis-recognition.

字程度、単語 2000 種程度)に限定されるような場合(たとえば地名、商品名、氏名などの認識)は、単語の個々の文字に注目する方法(方法 4;特に自動文字切出しを含めた場合)よりも、文字切出しを必要としない単語全体の形状に注目する提案法が有効であると考えられる。提案法において今回採用した DP マッチングではなく擬似ベイズ識別関数など高度な統計的識別関数の適用<sup>7)</sup>も考えられるが、この場合特徴ベクトルの次元数に比べて十分な学習パターン数が必要であるため、今回のように次元数の多い特徴( $n=6$  のとき 1536 次元)を扱うには実用上適していない。

図 1~図 6 の単語画像は提案法で正読できた単語画像の例である。特に、図 6 は方法 4 では正読できなかった例である。図 10 は提案法で正読できなかった単語の例である。失敗原因は、文字のかすれ(「明石市大久保町」)、大きな手書き変形(「港区八幡屋」)、類似単語(正解「羽生市東」→結果「羽生市南」)であった。なお、図 10 (a) は方法 4 で正しく認識できたが、図 10 (b), (c) は上記と同様の理由で失敗した。

#### 4. おわりに

本論文では、手書き日本語単語を、単語全体を 1 つのユニットと見なして認識する方法を提案した。提案方法では、単語の手書き変形を、3 段階の正規化(単語中の図形間隔、図形幅および単語画像サイズ(非線形正規化))により整え、さらに単語長さ方向への図形の伸縮を同方向へ伸縮する DP マッチングにより吸収している。

郵政研の手書き漢字画像データベース IPTP CD-ROM2 から作成した単語画像を用いた実験により、

- ① 3 段階の正規化と DP マッチングが単語の手書き変形の吸収に効果的であること、
- ② 3 段階正規化に用いるパラメータおよび DP マッチングの整合窓サイズが正読率に与える影響、
- ③ 従来方法(個別文字に注目する方法)に対する提案方法の優位性、

を明らかにした。

今後の課題は、IPTP CD-ROM2 の全サンプルを対象として単語の種類が増加したときの認識性能、文字切出しを含むさらに高度な知識処理を行う単語認識法との比較、などについて調べることである。

謝辞 本研究を行うにあたり、手書き文字データベース ETL9B と提供していただいた電子技術総合研究所と、手書き宛名サンプル IPTP CD-ROM2 を提供していただいた郵政研究所に感謝します。

#### 参考文献

- 1) Casey, R.G. and Lecolinet, E.: A Survey of Methods and Strategies in Character Segmentation, *IEEE Trans. PAMI*, Vol.18, No.7, pp.690-706 (1996).
- 2) 西野文人: 文字認識における自然言語処理, 情報処理, Vol.34, No.10, pp.1274-1280 (1993).
- 3) 仲林 清, 北村 正, 河岡 司: あいまい用語検索を用いた高速枠なし手書き文字列読み取り方式, 信学論 (D-II), Vol.74-D-II, No.11, pp.1528-1537 (1991).
- 4) 堤田敏夫, 川又文男: 住所読取りのための漢字認識複合化方式の一検討, 1995 信学情報・システムソサイエティ大会 D-201 (1995).
- 5) Ho, T.K., Hull, J.J. and Srihari, S.N.: A word shape analysis approach to lexicon based word recognition, *Pattern Recognition Letters*, Vol.13, No.11, pp.821-826 (1992).
- 6) 志久 修, 中村 彰, 高比良秀彰, 黒田英夫: パターン整合法による手書き文字列の分類実験, 信学論 (D-II), Vol.80-D-II, No.5, pp.452-454 (1997).
- 7) 堀田悦伸, 直井 聡, 諏訪美佐子, 平井淳一: セグメンテーションの負荷を軽減した手書き住所認識, 信学技報, PRMU98-161 (1998).
- 8) 堤田敏夫, 城戸 賛, 太田一浩: 手書き漢字画像データベース「IPTP CD-ROM2」とその諸性質の分析結果, 信学'97 総大, D-12-97 (1997).
- 9) 梅田三千雄: 手書き漢字認識手法の個性的なサーベイ「手書き文字認識技術の過去・現在・未来」シンポジウム講演論文集, pp.54-63 (1993).
- 10) Lee, S.-W. and Park, J.-S.: Nonlinear shape normalization methods for the recognition of large-set handwritten characters, *Pattern Recognition*, Vol.27, No.7, pp.895-902 (1994).
- 11) 堀内隆彦, 春木亮二, 山田博三, 山本和彦: 線密度を用いた非線形正規化法の 2 次元の拡張, 信学論 (D-II), Vol.80-D-II, No.6, pp.1600-1607 (1997).
- 12) 津雲 淳: 手書き漢字認識における非線形正規化処理, 信学情報・システム全大, No.68 (1987).
- 13) 津雲 淳: 方向パタンマッチング法の改良と手

書き漢字認識への応用, 信学技報, PRMU90-20 (1990).

- 14) 田中直樹, 塩野 充, 真田英彦, 手塚慶一: 動的方向性整合法による手書き漢字認識, 信学論, Vol.68-D, No.1, pp.56-63 (1985).
- 15) 郭 軍, 孫 寧, 根元義章, 佐藤利三郎: 整形変換を用いた手書き漢字データベース ETL9B の高精度認識, 信学論 (D-II), Vol.76-D-II, No.5, pp.1015-1022 (1993).
- 16) 中川聖一: 確率モデルによる音声認識, 電子情報通信学会編, コロナ社 (1988).
- 17) 川又文男, 若原 徹, 松井敏弘, 能見 正, 山下郁生, 堤田敏夫: 手書き漢字宛名認識に関する第3回文字認識技術コンテストの実施, 1994 信学秋季大会, D-321 (1994).

(平成 11 年 5 月 17 日受付)

(平成 12 年 2 月 4 日採録)



志久 修 (正会員)

平成元年長崎大学工学部電子工学科卒業。平成3年同大学大学院工学研究科修士課程修了。同年より国立佐世保工業高等専門学校助手, 平成6年講師, 平成8年助教授。現在に至る。博士(工学)。画像処理, パターン認識の研究に従事。電子情報通信学会会員。



中村 彰 (正会員)

昭和33年九州大学工学部通信工学科卒業。同年電電公社に入社。電気通信研究所にて主として論理設計の自動化, 計算機の開発等に従事。昭和54年~平成8年長崎大学電気情報工学科教授。平成8年~現在熊本工業大学電子工学科教授。画像処理とパターン認識に興味を持つ。工学博士。電子情報通信学会, 画像電子学会, IEEE 各会員。



黒田 英夫 (正会員)

昭和46年3月九州工業大学大学院修士課程修了。同年4月電電公社電気通信研究所入社。平成元年より長崎大学工学部教授。その間平成6年シドニー大学客員教授。画像信号の高能率符号化, 画像処理, ネットワーク, ロボットビジョン等の研究に従事。工学博士。IEEE, 電子情報通信学会, 映像情報メディア学会, 画像電子学会会員。



宮原 未治 (正会員)

昭和44年熊本大学工学部電気工学科卒業。昭和46年同大学大学院工学研究科修士課程修了。同年電電公社研究所入社。平成9年長崎大学工学部教授。工学博士。音声情報処理, 文字認識装置の実用化, ペン入力型文字認識, 文書検索・言語処理等の研究に従事。電子情報通信学会会員。