

## 文字の構造情報を活用した印刷文字認識方式

木村正行† 馬場伊美子‡

† 北陸先端科学技術大学院大学  
〒 923-12 石川県能美郡辰口町旭台 15  
‡ 七尾短期大学  
〒 926 石川県七尾市藤橋町西部 1

あらし 従来のパターンマッチングをベースにした文字認識の方式は、印刷文字と手書き文字とを問わず適用し得る優れた特色がある。しかし、未知入力に対する候補字種を絞る大分類の段階で、基本的には、認識対象となる字種の数だけマッチングのための距離計算が必要であって、このことが認識の高速化の大きな障害になっている。本論文では、文字の構造情報の活用によって飛躍的に高速な文字認識を可能とする新しい文字認識方式を提案する。ここでは特に、文字の縦と横の2方向の各ヒストグラムに反映される構造情報を系統的に順次に活用することによって、効率よく候補字種を絞り込む大分類のアルゴリズムを中心に本方式の基本的考え方について述べる。

和文キーワード 文字認識、高速化、パターンマッチング、文字の構造情報、大分類アルゴリズム

## Printed Character Recognition Based on Structural Information of Characters

Masayuki KIMURA† and Imiko BABA‡

† Japan Advanced Institute of Science and Technology, Hokuriku  
15, Asahidai, Tatsunokuchi, Ishikawa, 923-12, JAPAN  
‡ Nanao Junior College  
1, Seibu, Fujihashi, Nanao, Ishikawa, 926, JAPAN

**Abstract** The method of pattern matching heretofore in use requires the calculation of distance between an unknown input pattern and every standard one stored in the dictionary. Such a requirement causes the greatest difficulty in making character recognition faster especially in the case where characters in the dictionary reach into thousands. This paper proposes a new method of character recognition that makes it possible to overcome that difficulty by using the structural information of a character. A basic idea to extract a structural information from the image of a character and to devise an algorithm to identify an unknown input based on structural information, is discussed and made clear.

英文 keywords character recognition, pattern matching, structural information

## 1 はしがき

文字認識の原点は、文字の構造に関する情報を活用することにあるといえる。しかし、どのように活用するかについては多くの問題が残されている。特に、手書き文字は、書く人によって字体が様々であって、かつ変化の度も多様である。一方、印刷文字にも様々な字体はあるが、変化の度が小さく、その複雑さは手書き文字とは比較にならないほど単純である。したがって、手書き文字と印刷文字の認識において、文字の構造情報の活用の仕方が異なってくるのは、しごく当然のことであろう。

本稿では印刷文字に焦点を絞って、文字の構造情報をどのようにに認識に活用するかについて考察し、印刷文字を対象にした新たな文字認識方式を提案する。まず最初に、従来の文字認識について概観しよう。

### パターンマッチング法

従来の方式では、文字画像上を画素(ビット)単位でスキャンし、スキャンされている当該画素とその近傍の画素との関係に着目して文字の特徴、言い換えれば構造情報を抽出している。しかし、画素単位で探索して個々の特徴を抽出しているのがノイズの影響を受け易く、この種の特徴をそのまま活用して、順次候補字種を効率よく絞って行くことは困難である。

そこで、次のような大分類の方法が考案された。即ち、数量化された個々の特徴(特徴量)で作られる特徴空間において、未知入力文字(未知パターン)と辞書のすべての字種の標準パターンとの間で、両者の総合的な類似度を計るパターンマッチング(距離計算)が行われ、距離の小さい順に候補字種が選択される。ただし、標準パターンは、同一字種の多くのサンプルの特徴量を用いて構成された統計量(平均値など)である。

したがってこの方式では、1個の未知入力文字に対して、パターンマッチング(数百次元の距離計算)を基本的には、標準パターンの数、即ち認識の対象とする字種数と同じ回数だけ繰り返すことになり、日本語のように3000乃至5000字種を対象とする場合には、文字認識の高速化にとって大きな障壁となっている。

### 構造解析法

文字の構造情報を直接的に活用する方法として、構造解析法がある。この手法は文字を構成する線分などの要素間の関係の解析に基づいて未知文字を同定しようとするものであるが、文字の構造が複雑になると計算量が急激に増加するばかりでなくノイズ対策が極めて困難となるという大きな欠点がある。

## ニューラルネットワーク

ニューラルネットワークの学習能力を活用して文字認識を行う方法が研究されているが、この種の学習に関しても多くの問題が残されている。特に、この方法については、完成した文字認識用のネットワークに対して、新たな字種の追加、相性の悪い字体に対する対策が非常に困難になるように思われる。今後の研究の進展が期待される。

### 提案する文字認識方式

以上に考察したように、文字認識はノイズと上手に付き合って構造情報を効率よく活用する点にあるといえよう。そこで、クローズアップされのが文字画像のヒストグラムである。ヒストグラムは、バースト的ノイズ(潰れなど)以外のランダムなノイズによる影響が少ないこと、縦、横、斜めの線分やスペースなど文字の構造情報を比較的によく反映すること、画素単位でスキャンして特徴を抽出するという操作が不要となること、などの利点がある。

文字の構造に関する情報源として、文字画像の縦と横及び $\pm 45$ 度の4方向の4つのヒストグラムを用いる。

以下では、これらのヒストグラムに反映される文字の構造情報を活用する新しい文字認識方式の基本的考え方について述べる。

## 2 文字の構造情報の抽出

### 2.1 ヒストグラムの量子化-0-1パターン表現

文字画像の縦と横の2方向のヒストグラム(図1)は第一段階の分類(大分類)で、 $\pm 45$ 度の2方向の2つのヒストグラムは第二段階の分類(細分類)で用いる。ここでは、縦と横方向のヒストグラムの量子化-0-1パターン表現(図2)を得る方法について述べる。

縦と横の各ヒストグラムの横軸を、それぞれ、 $l_1$ 個及び $l_2$ 個の区間に分割する。字体の変動やノイズを考慮して隣接する区間を適当な $\delta$ だけ重複させる。ついで、縦及び横方向のヒストグラムの高さを、それぞれ、文字画像の高さ及び巾に対する百分率(%)で表す。縦軸の区間 $[0, 100\%]$ を $m$ 個の区間に分割し、 $m * (l_1 + l_2)$ 個の各領域において( $*$ は乗算記号)、その領域を下から上に通過するヒストグラムが存在するときその領域に1を、存在しないとき0を割り当てる。このようにして、縦と横方向の両ヒストグラムを1セットとした場合の量子化-0-1パターン表現を得

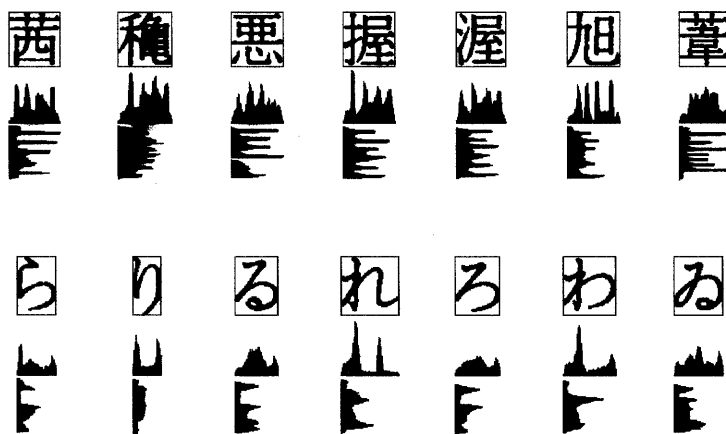


図 1: 文字画像のヒストグラム

る。但し、高さが100%に等しいヒストグラムは、縦軸100%の境界を通過したと見なす(図2)。

横軸の各区間の番号を  $i$  で表す、 $i = 1, 2, \dots, l_1 + l_2$ 。以下では、この番号  $i$  を座標と呼ぶ。縦軸をしきい値  $\theta$  で表し、両端と分割点を  $\theta_t (t = 1, 2, \dots, m + 1)$  で表す。但し、 $\theta_1 = 100\%$ ,  $\theta_{m+1} = 0\%$  とおく。すると、ヒストグラムの量子化-0-1パターン表現の各領域は、 $(i, \theta_t)$ ,  $i = 1, 2, \dots, l_1 + l_2$ ,  $t = 1, \dots, m$  で表される。

例えば座標を左から  $i = 1, 2, \dots$  とし、しきい値  $\theta_t$  を  $\theta_1 = 100\%$ ,  $\dots$ ,  $\theta_{m+1} = 0\%$  とすると、第1列の各領域は、上から  $(1, \theta_1)$ ,  $(1, \theta_2)$ ,  $\dots$ ,  $(1, \theta_m)$  で表される。

## 2.2 量子化-0-1パターン表現の相補性

量子化-0-1パターン表現の1(文字の構成要素)に着目して構造情報を抽出する場合には、次のような性質を利用する。

- (a) しきい値  $\theta_t$  を、 $\theta_1 = 100\%$  から下に向けて順次  $t = 1, 2, 3, \dots, m$  とスキャンして行く。この過程では、領域  $(i, \theta_t)$  が1ならば、 $t' \geq t$  となるすべての領域  $(i, \theta_{t'})$  で1であり、領域  $(i, \theta_t)$  が0ならば、 $t' \leq t$  となるすべての領域  $(i, \theta_{t'})$  で0である。

一方、0(スペース)に着目して、構造情報を抽出する場合には、次のような性質を利用する。

- (b) しきい値  $\theta_t$  の目盛りは、(a) で100%のところを0%に、0%のところを100%とし、 $\theta_1 = 100\%$

から上に向けて、順次  $t = 1, 2, 3, \dots, m$  とスキャンして行く。この過程では、領域  $(i, \theta_t)$  が0ならば、 $t' \geq t$  となるすべての領域  $(i, \theta_{t'})$  で0であり、領域  $(i, \theta_t)$  が1ならば、 $t' \leq t$  となるすべての領域  $(i, \theta_{t'})$  で1である。

同一量子化-0-1パターン表現を対象としているので、上からスキャンを開始すると(a)が、下からスキャンを開始すると(b)が成り立つ。但し、開始点の  $\theta$  を  $\theta_1 = 100\%$  になるように目盛りを書き換える。このような関係を、ここでは、量子化-0-1パターン表現の相補性と呼ぶ。

## 2.3 構造情報の抽出

量子化-0-1パターン表現を基にして、未知入力文字の構造情報を抽出する。ノイズによる変動を吸収するために、注目する領域をしきい値  $\theta$  のある領域を定める。例えば、

$$(i, \theta_t), (i, \theta_{t+1}), (i, \theta_{t+2}) \text{ 及び } (i, \theta_{t+3}),$$

$$i = 1, 2, \dots, l_1 + l_2$$

の領域に着目し、その各々の0,1の配列(横の配列)を、それぞれ、

$$u, v, w, x$$

で表す(図2)。そして、かく  $i$  について、この4つの配列がすべて同じ値を持つ部分のみをその領域における文字の顕著な構造情報(特徴)として、その構造情

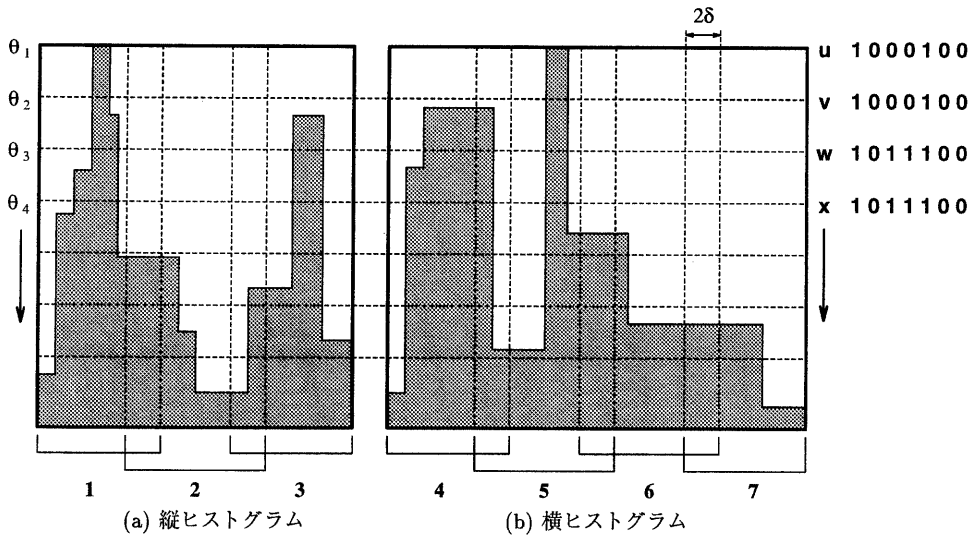


図 2: ヒストグラムの量子化 -0-1 パターン表現

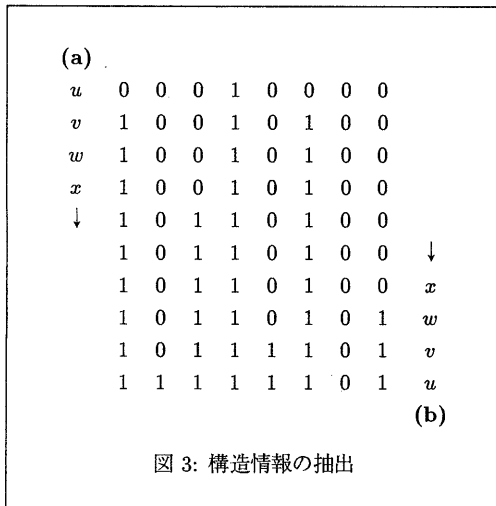


図 3: 構造情報の抽出

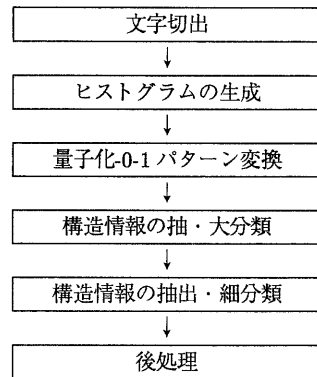
報を、例えば  $w$  の対応する部分の配列で代表させる (図 3)。

注目する領域を、例えば、 $\theta_i$  の 1 区間づつシフトして以上の操作を繰返す。このような抽出方法では、1 の部分は文字の構成要素を、0 の部分はスペースを表すので、文字の構成要素とスペースの両面から構造情報を抽出することになる。

### 3 印刷文字認識方式

#### 3.1 文字認識の手順

以上の考察から印刷文字を対象とした場合には、次のような手順で文字認識を行うことが出来る。即ち、



以下では、本方式の中核となる大分類について述べる。細分類については今回は省略する。

#### 3.2 大分類

##### 3.2.1 分類アルゴリズム A

###### ① 定義と初期値の設定

区間の重複中  $\delta$  及び候補字種数を絞り込む目安となる定数  $\kappa$  を定める。初期値として、分類の対象となる字種の集合  $\Omega$  を

$$\Omega = \{ \text{対象とする全字種の集合} \}$$

とおき、縦軸を  $m$  個の区間に分割し  $t$  の初期値を

$$t = 1, \theta_1 = 100\%,$$

$$(\text{例: } \theta_{t+1} = \theta_t - 5\%, t = 1, 2, \dots, m = 20)$$

とおく。また、

座標  $i$  は最初は何れも固定されていない

とする。但し、座標  $i$  を固定するとは、座標  $i$  を候補字種を絞るための考察の対象から除外することを意味する。

また、 $(i, \theta_i)$ ,  $i = 1, 2, \dots, l_1 + l_2$  の各領域がとる 0, 1 の配列 (横の配列) を  $u$  で、同様に  $(i, \theta_{i+1})$ ,  $(i, \theta_{i+2})$  及び  $(i, \theta_{i+3})$  に対する 0, 1 の配列を、それぞれ、 $v$ ,  $w$  及び  $x$  で表す。以下では、 $u$ ,  $v$ ,  $w$  及び  $x$  を、それぞれ、単に配列と呼ぶ。

②  $\Omega$  に属する未知入力文字  $X$  のヒストグラムについて、しきい値  $\theta_i, \theta_{i+1}, \theta_{i+2}$  及び  $\theta_{i+3}$  について、それぞれ、配列  $u, v, w$  及び  $x$  を求める。

③ a. 固定座標以外の座標で、 $u$  の値が 1 となる座標が存在せず、かつ  $x$  の値が 0 となる座標も存在しない場合  $\theta_i = \theta_{i+1}$ ,  $\theta_{i+1} = \theta_{i+2}$ ,  $\theta_{i+2} = \theta_{i+3}$ ,  $\theta_{i+3} = \theta_{i+4}$  において手順②へ。

b. その他の場合、即ち、固定座標以外の座標で、 $u$  の値が 1 となる座標を  $i'$ 、 $x$  の値が 0 となる座標を  $i''$  とするとき、少なくともその何れかが存在する場合

2.2 の (a) から、 $i'$  と  $i''$  は異なる、さらに、 $u$  の値が 1 となる座標  $i'$  では  $w$  の値は 1、 $x$  の値が 0 となる座標  $i''$  では  $w$  の値は 0 となる。そこで、 $\Omega$  の中でこのようなすべての座標  $i' (1 \leq i' \leq l_1 + l_2)$  及び  $i'' (1 \leq i'' \leq l_1 + l_2)$  について、 $\theta_{i+2}$  における配列  $w$  の値が、それぞれ、1 及び 0 となるすべての字種からなる集合に候補字種を絞り、この集合を  $S_w(i' = 1, i'' = 0)$  で表す (註 1.)。

$|S_w(i' = 1, i'' = 0)| \leq \kappa$  ならば終了。候補字種の集合を  $S_w(i' = 1, i'' = 0)$  とする。

$|S_w(i' = 1, i'' = 0)| > \kappa$  ならば、ここで用いたすべての  $i'$  を固定座標として、手順④へ。

- ④ a. 座標  $i (1 \leq i \leq l_1 + l_2)$  がすべて固定されているならば終了。候補字種の集合を  $S_w(i' = 1, i'' = 0)$  とする。
- b. 固定座標でない座標  $i$  が存在するとき、 $\Omega = S_w(i' = 1, i'' = 0)$  とおき、 $\theta_t = \theta_{t+1}$ ,  $\theta_{t+1} = \theta_{t+2}$ ,  $\theta_{t+2} = \theta_{t+3}$ ,  $\theta_{t+3} = \theta_{t+4}$  において、手順②へ。

註 1.: ③の b を次のような手順で  $b'$  と  $b''$  置き換えると、文字の構成要素である黒画素を表す 1 をスペースを表す 0 より優先させた手順となる。

③ b'. 固定座標以外の座標で、 $u$  の値が 1 となる座標が存在する場合

そのような座標を  $i'$  で表す。2.2 の (a) から、 $u$  の値が 1 となる座標  $i'$  では  $w$  の値は 1 となる。そこで、 $\Omega$  の中でこのようなすべての座標  $i' (1 \leq i' \leq l_1 + l_2)$  について、 $\theta_{i+2}$  における配列  $w$  の値が 1 となるすべての字種からなる集合に候補字種を絞り、この集合を  $S_w(i' = 1)$  で表す。

$|S_w(i' = 1)| \leq \kappa$  ならば終了。候補字種の集合を  $S_w(i' = 1)$  とする。

$|S_w(i' = 1)| > \kappa$  ならば、ここで用いた  $i'$  をすべて固定座標として、手順④へ。ただし、④で  $\Omega = S_w(i' = 1)$  とする。

b''. 固定座標以外の座標で、 $u$  の値が 1 となる座標  $i'$  が存在せず、 $x$  の値が 0 となる座標が存在する場合

そのような座標を  $i''$  で表す。2.2 の (a) から、 $x$  の値が 0 となる座標  $i''$  では  $w$  の値は 0 となる。そこで、 $\Omega$  の中でこのようなすべての座標  $i'' (1 \leq i'' \leq l_1 + l_2)$  について、 $\theta_{i+2}$  における配列  $w$  の値が 0 となるすべての字種からなる集合に候補字種を絞り、この集合を  $S_w(i'' = 0)$  で表す。

$|S_w(i'' = 0)| \leq \kappa$  ならば終了。候補字種の集合を  $S_w(i'' = 0)$  とする。

$|S_w(i'' = 0)| > \kappa$  ならば、 $\Omega = S_w(i'' = 0)$  とおき、 $\theta_t = \theta_{t+1}$ ,  $\theta_{t+1} = \theta_{t+2}$ ,  $\theta_{t+2} = \theta_{t+3}$ ,  $\theta_{t+3} = \theta_{t+4}$  において、手順②へ。

### 3.2.2 分類アルゴリズム B

$\Omega$  に属する未知入力文字  $X$  の量子化-0-1 パターンについて、下から上にスキャンして行く場合を考えると、2.2 の (b) にしたがう分類アルゴリズム B を得る。このアルゴリズムは分類アルゴリズム A から次のようにして簡単に求められる。

- (1) (b) の目盛りをそのままとし、量子化-0-1 パターンの 1 を 0 に、0 を 1 に置き換えたものに対して、分類アルゴリズム A を適用すれば、即、分類アルゴリズム B となる。
- (2) 分類アルゴリズム A において、1 を 0 に、0 を 1 に、それぞれ置き換えて、(b) に適用すれば、即、分類アルゴリズム B となる。

### 3.3 大分類と細分類の手順

分類アルゴリズム A と B は相補的であるので、前者で得られた候補に新しい候補が含まれない場合でも、後者で得られた候補に新しい候補が含まれる場合があり、その逆の場合もある。したがって、大分類の手順としては、この 2 つのアルゴリズムを未知入力文字  $X$  に適用して、それぞれ得られた候補字種の和集合をとる。

一方、細分類の方法及び手順は、一般的には、大分類の結果得られた候補字種の集合の大きさに依存して選択されることになる。日本語の印刷文字の場合には、斜めの線分を持つ字種が比較的に多いので、 $\pm 45$  度の 2 方向のヒストグラムを生成し、分類アルゴリズム A 及び B を適用して候補字種数をさらに少数に絞り込むことを考えているが、これだけでは不十分であろう。細分類の具体的な方法については、現在検討中である。

## 4 むすび

構造情報に基づく文字認識方式の基本的な考え方について述べた。本方式は、少なくとも字種数が数千に及ぶ印刷文字を対象とする場合には、文字認識の飛躍的高速化が期待される。一方、ノイズに比較的に強い文字画像のヒストグラムに着目して、ヒストグラムに反映される文字の構造情報を効率よく活用するという点で、印刷文字認識方式の新たな展開の手がかりが得られればと期待している。次の機会には、日本語印刷文字の認識実験についてご報告したいと考えている。

## 参考文献

- [1] 木村正行: “知識を活用する漢字認識システム” 電子通信学会誌, J 68, 9, pp.986-989 (1985).
- [2] 孫 寧、阿曾弘具、木村正行: “連想整合法に基づく高速文字認識アルゴリズム” 情報処理学会論文誌, vol.32 No.3, pp.29-38 (1991).
- [3] 山本和彦: “弛緩整合法による手書き教育漢字認識” 信学論 (D), J85-D, 9, pp.1169-1174 (1982).