

骨格ベクトルによるオフライン手書き文字の構造解析

鈴木 道孝†
Michitaka Suzuki

1. はじめに

これまでのオフライン文字認識の研究では、文字画像データから輪郭線を求め、その角度分布を統計的に処理して解析することが主に行われてきた[1]。しかし、この方法では、濁点と半濁点のように明らかに異なる文字の差異を発見できないことが多く、このような類似文字の存在が全体の認識率の向上を妨げている。まず、この原因について考えてみる。

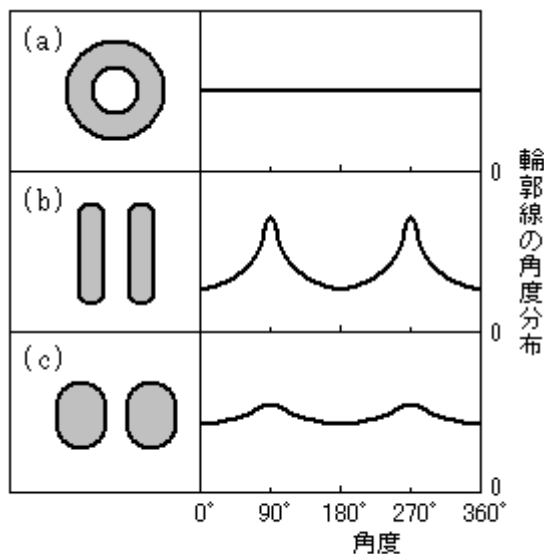


図1 濁点と半濁点の輪郭線の方角分布

図1は半濁点と濁点の輪郭線の方角の角度分布を模式的に表したものである。(a)の半濁点では角度分布は一様になるが、(b)の濁点では縦方向にピークが生じるので、これらの識別は容易である。一方、(c)は人間の目には濁点と認識できるものの、その角度分布は、小さなピークを伴ったほぼ一様な分布になり、これと(a)の角度分布を識別することは難しい。(c)の一様分布は濁点の上下の二つの半円部に起因し、ピーク成分はそれらをつなぐ側線に起因する。すなわち、前者は筆記具の特性、特に太さに起因するが、後者は筆記者の意図の反映である。カスレなどの原因で輪郭線が乱れると、ピーク成分だけが減少する。また、太い筆記具を使うと、一様分布の成分が増加し、相対的にピーク成分が弱くなる。このような場合には、輪郭線を抽出した時点で筆記者の意図を伝える情報はほとんど残っておらず、以後の計算のいかんにかかわらず、これらの文字の識別は困難になる。

輪郭線に頼らない方法の一つに、骨格ベクトルによる方法がある。本報告では、骨格ベクトルの抽出により類似文字を識別し、認識率を算出する。

2. 骨格ベクトルの計算

骨格ベクトルは、幅を持った線分の連なりで入力画像データをフィッティングすることによって求まる[2]。本報告では、最小自乗法を用いてフィッティングを行ない、幅を持って降下する部分の形状を2次式で表すことにする。

\mathbf{r} を2次元の位置ベクトルとし、二値化をしていない生の入力画像値を $g(\mathbf{r})$ 、モデル値を $f(\mathbf{r})$ とすると、最小化すべき量は

$$T = \frac{1}{2} \int [f(\mathbf{r}) - g(\mathbf{r})]^2 d^2r \quad (1)$$

となる。線分が重なった場所では、モデルとしての値は、各線分の寄与のうちの最大値とする。

$$f(\mathbf{r}) = \max_{i,j} Q(\mathbf{r}, \mathbf{r}_{i,j}, \mathbf{r}_{i,j+1}) \quad (2)$$

ここで、 Q は単一の線分の濃淡分布を表し、 i はストローク番号、 j はストローク内の線分番号を表している。

$$Q(\mathbf{r}, \mathbf{r}_{ij}, \mathbf{r}_{i,j+1})$$

$$= \begin{cases} s(|\mathbf{r} - \mathbf{r}_{ij}| / w_{ij}) & (\text{when } (\mathbf{r} - \mathbf{r}_{ij}) \cdot (\mathbf{r}_{i,j+1} - \mathbf{r}_{ij}) \leq 0) \\ s(|\mathbf{r} - \mathbf{r}_{i,j+1}| / w_{ij}) & (\text{when } (\mathbf{r} - \mathbf{r}_{i,j+1}) \cdot (\mathbf{r}_{i,j+1} - \mathbf{r}_{ij}) \geq 0) \\ s(|\mathbf{r} - \mathbf{r}_{ij}| \times (\mathbf{r}_{i,j+1} - \mathbf{r}_{ij}) / (|\mathbf{r}_{i,j+1} - \mathbf{r}_{ij}| w_{ij})) & (\text{otherwise}) \end{cases} \quad (3)$$

ここで、 $s(x)$ は線分の稜線から下降していく関数で、 w_{ij} は線分の幅をあらわしている。 $s(x)$ を次の2次式で表す。

$$s(x) = \begin{cases} 1 - x^2 & (\text{when } -1 \leq x \leq 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (4)$$

$f(\mathbf{r})$ の例を図2に示す。

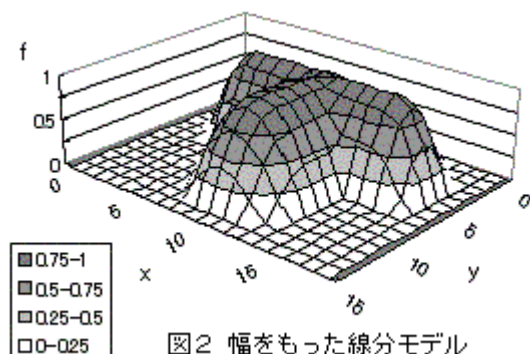


図2 幅をもった線分モデル

ここでは、線分{(5,5),w=3,(15,5)}と線分{(10,5),w=4,(10,10)}が交わり、「T」字型を形成している。

† 〒232-0053 横浜市南区井土ヶ谷下町 215-1-2-210
msuzuki@mxh.mesh.ne.jp

誤差 T の最小値を求める際に、なるべく適切な初期値から計算を始める方が混乱が少なく、計算時間も短縮できる。そのために、輪郭線の曲率を調べ、ストロークの端点候補を探し、その部分の輪郭線の半円の中心を r_{00} とし、半径を w_{00} とする。さらにそこから幅が 0 の線分を伸ばし、単位長さあたりの誤差の減少が最大になるようにもう一方の点 r_{01} を定める。これを第一近似の線分として、誤差を最小化する。次に、さらにその線分の先に同様の方法で点を探し、新しい線分を付け足して、誤差を最小化する。これを繰り返すことによって、入力データのほとんどの領域を幅つきの線分でカバーすることができる。端点を持たない領域については、領域内から任意の点を選び、同様の計算を繰り返す。最後に、後の解析を容易にするために、求めたストローク同士の接続可能性、および、ストローク内の線分の併合可能性を調べ、入力画像との相関を大きく損なわない程度に、点の数を減らしていく。このようにして求めた骨格線の例を図 3 に示す。

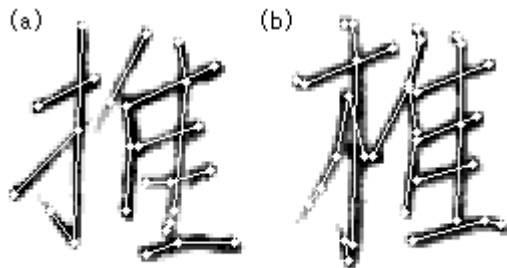


図3 文字画像データと計算された骨格線(白線)

(a) ETL9G-1512-No1 (b) ETL9G-1965-No1

図 3 は、濃淡調整済みの画像データの上に求めた骨格線を白色で重ねて表示したものである。濃淡調整は、フィッティングの計算に先立って行われ、入力濃淡値の度数分布がモデルのそれと同じになるように入力濃淡値が調整される。これによってカスレの問題が緩和される。「推」の下部では、画像データのカスレが影響して、孤立した小さな線分が出現しているように見える。実際には、線分が幅を持ち、その裾野は上方の線分の裾野と共有領域をもつので、二線分は中間的な接続をしているといえる。

3. 類似文字対の識別

得られた骨格ベクトルを使って、類似文字対の二者択一問題の正解率を算出し、それをここでは認識率と呼ぶことにする。対象は、類似文字対の中から「土」と「土」、「末」と「末」、濁点と半濁点(「ば」と「ば」等の 5 組)の 7 組を選んだ。実験データとして、ETL9G の多値の文字画像データ 2800 件を使った。

「土」と「土」については、図 4 のテンプレートを共通のものとして用い、動的計画法により各ストロークを対応付け、上下の二つの横線の長さを比較し、上が長ければ「土」、下が長ければ「土」と判定する。「末」と「末」についても同様である。

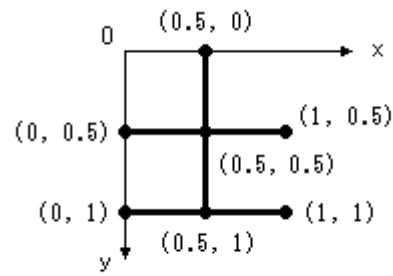


図4 「土」と「土」に共通に用いられたテンプレート

濁点と半濁点については、上右部に環状のストロークを発見できるか否かによって識別した。ただし、前述のように、ストローク端同士の接続/非接続の判定は繊細な問題なので、完全接続していない環であっても、そのストロークが 240° 以上方向を変化させていて、二つの端が十分接近している場合には、半濁点記号とみなした。得られた認識率を表 1 に示す。

表 1 類似文字対の認識率

項番	対象文字対	認識率(%)
(1)	「土」と「土」	97.75
(2)	「末」と「末」	96.50
(3)	「ば」と「ば」	96.25
(4)	「び」と「び」	96.75
(5)	「ぶ」と「ぶ」	96.75
(6)	「べ」と「べ」	98.50
(7)	「ぼ」と「ぼ」	96.00

表 1 の認識率が 100% に至らない原因は、(1) と (2) に関しては、データベース中に不良のデータが含まれていることが主なものである。(3) から (7) については、半濁点の内側の白い領域が著しく小さい、あるいは存在しない文字データが含まれていることが主要な原因になっている。また、閉じていない半濁点や、他のストロークとの接触による悪影響も原因している。

計算は、1.4GHz の Pentium 4 プロセッサ、メモリ 256MB、Java パーチャルマシン上で行われ、一文字あたりの全工程の計算時間は、1 秒程度であった。

4. むすび

骨格ベクトルによる構造解析が類似文字の識別に有効であることが示された。本手法は従来の手法と併用し、文字認識の最終段階の判定の際に利用可能である。

参考文献

- [1] 加藤, 安倍, 根元: "改良型マハラノビス距離を用いた高精度な手書き文字認識", 信学論(D-II), J79-D-II, 1, pp. 45-52 (1996).
 [2] 鈴木 道孝: "手書き文字画像データから骨格ベクトルを抽出する", 情報科学技術フォーラム(FIT) 第 3 分冊 pp.37-38 (2003).