

稜線特徴量による手書き文字認識

Ridge features for handwritten character recognition

鈴木 道孝[†] 石川 幸司[†] 伊藤 彰義[‡]

Michitaka Suzuki Kohji Ishikawa Akiyoshi Itoh

1 まえがき

これまで、統計的文字認識において多くの特徴量が提案されてきた。その中で、位置情報として相対的な位置を使う特徴量としては、PDC, P-LOVE などがあるが、絶対的な位置情報を使うものとしては、方向線素特徴量(DEF) [1] が代表的である。DEF は輪郭線上の点の数を方向ごとに小領域内で合計した数を特徴ベクトルの要素とするものである。DEF は、輪郭線上の点について計算されるので、一つのストロークに対して、その両脇に特徴量が生じるという冗長度がある(図1(b))。また、DEF では、特徴量のどちら側にストロークが存在するかという情報を持たないという問題点がある(図1(c)(d))。

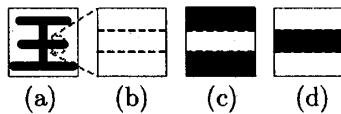


図1 輪郭線情報の問題点

輪郭線を利用しない特徴量としては、Gabor 特徴量 [2] がある。これは、Gabor フィルタによって、ストロークの位置を方向ごとに抽出した特徴量である。Gabor 特徴量は、ストロークの位置をより正確に表しているが、ストロークの太さによって特徴量の変動するため、手書き文字に対しては DEF ほどの高い精度は得られていない。

本報告では、これらの代表的な特徴量の問題点を解消する新しい特徴量 - 稜線特徴量 - を提案し、その優越性を実証する。

2 稜線特徴量

二値の文字画像データが与えられたとき、平滑化処理によって滑らかな多値の文字画像データを作ることができる。多値の文字画像データには、ストロークの中心を走る稜線がある。画素点 r における画素値を g_r (黒1, 白0) とし、縦横斜めの4方向のうち2方向以上で、次の条件(1)あるいは条件(2)を満たす点を稜線上の点とする。

$$g_{r+u} < g_r \quad \wedge \quad g_{r-u} < g_r \quad (1)$$

$$\begin{aligned} &(g_{r+u} = g_r \quad \wedge \quad g_{r-u} < g_r) \quad \vee \\ &(g_{r+u} < g_r \quad \wedge \quad g_{r-u} = g_r) \end{aligned} \quad (2)$$

ここで、 $u \in \{(1,0), (1,1), (0,1), (-1,1)\}$ である。2方向以上という制限は、ノイズや不規則な微細構造によつてできる稜線部分を除外するためである。

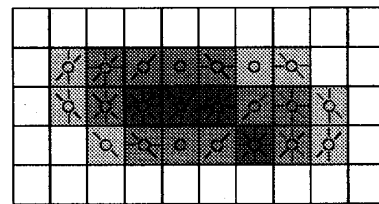


図2 多値画像上の極大の方向

稜線上の点のうち、条件(1)を満たす方向が1つでも存在する点を純粋な稜線上の点とし、その特徴強度 f を1とする。そうではない場合は、さらに $g_r < 1$ の場合と $g_r = 1$ の場合に分けられる。前者は、1つの稜線が2点からなる場合で、その特徴強度を $1/2$ とする。後者は、太いストロークやつぶれなどの平滑化処理の及ばない領域の境界で起こる。太いストロークのケースでは、1つのストロークを両脇の2点で表すので、特徴強度を $1/2$ とし、つぶれ領域では、2つ以上のストロークを2点で表すので、特徴強度は1あるいはそれ以上とするのが妥当である。しかし、二値の入力データからこの二つのケースを区別することはできないので、このような点においては、二つのケースの平均をとり、特徴強度を $3/4$ とする。特徴強度の例を図3に示す。図3(a)は、周辺ノイズ除去、縁取り処理の後の画像である。これに 2×2 の平均フィルタを4回かけて図3(b)の多値画像を得る。それから得られる稜線領域を図3(c)に示し、そのうち強度が $3/4$, $1/2$ の領域をそれぞれ、図3(d),(e)に示す。

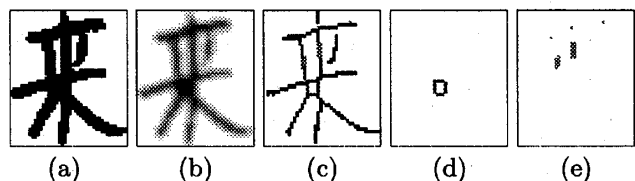


図3 稜線の計算 (a) 縁取り (b) 平滑化 (c) 強度 (d) 強度=3/4 (e) 強度=1/2

[†] 日本大学大学院理工学研究科[‡] 日本大学理工学部

特徴量の方向を求めるために、注目点付近の画像の濃淡値を最小二乗法によって2次曲面で近似する。このさい、注目点と近傍の8点をデータ点として使う。得られた2次曲面の2次微分係数の大きい方の主軸の方向を特徴量の方向 θ ($0 \leq \theta < \pi$) とする。これに縦横比の補正をして (θ')、縦横斜めの4方向のうちの前後の2方向に、特徴量を振り分ける。

$$\theta' = \operatorname{arccot}\left(\frac{Y}{X} \cot \theta\right) \quad (3)$$

$$f_k = \begin{cases} f \cos(2\theta' - \frac{k\pi}{2}) & \text{if } \cos(2\theta' - \frac{k\pi}{2}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

for $k = 0, 1, 2, 3$

ここで、 X, Y は、それぞれ画像の横、縦のドット数である。各点について求められた稜線特徴量の強度と方向成分を図4に示す。

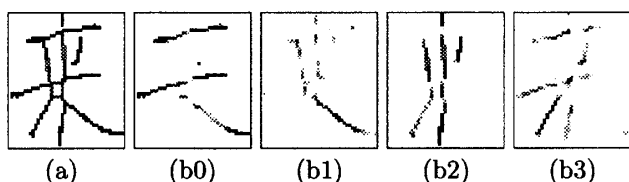


図4 稜線特徴量 (a) 強度 (再掲) (b0)-(b3) 方向成分 (-, \, |, /)

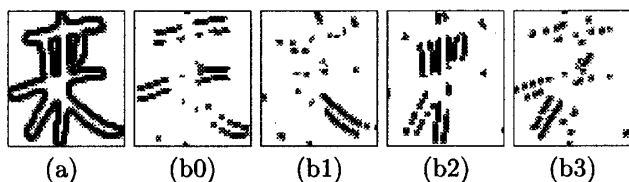


図5 輪郭特徴量 (a) 強度 (b0)-(b3) 方向成分 (-, \, |, /)

3 輪郭特徴量との比較

稜線特徴量と、従来のDEFで使われてきた輪郭線上の特徴量(輪郭特徴量と呼ぶ)を構造と認識率について、比較してみる。

3.1 構造の比較

輪郭特徴量を求めるには、まず、縁取りした二値画像にSobelフィルタをかけてこう配ベクトルを求める。このこう配ベクトルの大きさを特徴強度とし、このこう配ベクトルの方向に90°足したものを特徴量の方向とする。稜線特徴量と同様に、(3)、(4)式により各方向成分を求める。こうして求めた輪郭特徴量の強度と方向成分を図5に示す。図4と図5を比べると、稜線特徴量の方が輪郭特徴量よ

り、冗長度が少なく、構造が単純であることがわかる。特に、ストロークの交差部、先端部、密集している領域で、稜線特徴量の方が的確な情報を提供している。

3.2 認識率の比較

識別計算をするために、各画素点ごとに求めた特徴量の方向成分を小領域内で積分し、次元の小さな特徴ベクトルを作る。小領域は、隣同士で半分ずつ重なり合った7×7個の長方形とし、196次元のベクトルにする。長方形の配置は、山田の非線形正規化に従いストローク密度が可能な限り均一になるように求める。特徴ベクトルの成分が正規分布に近づくように0.7乗して、最終的な特徴ベクトルとする。

文字画像データは、ETL9Bのすべてのサンプル607,200件をそのまま用いる。ETL9Bの各字種200個のサンプルのうち、21~180番のサンプルを学習文字として用い、それらの特徴ベクトルの平均を標準パターンとする。1~20番および181~200番のサンプルを未知文字として用い、標準パターンとのユークリッド距離を計算し、最も近い標準パターンを認識結果とする。結果を表1に掲げる。得られた輪郭特徴量による認識率は、文

表1 輪郭特徴量と稜線特徴量による認識率

特徴量	認識率
輪郭特徴量	95.41
稜線特徴量	95.73

献[1]とほぼ一致する。稜線特徴量による認識率は、それよりも0.32%高いという結果が得られた。

4 結論

二値の文字画像データに平滑化処理を施して得られた滑らかな多値データから稜線を計算することができる。稜線上の点について計算された稜線特徴量は、輪郭特徴量よりの確にストロークの位置と方向を表すことができ、より高い精度の認識を可能にする。

参考文献

- [1] 若林 哲史 他, "非線形正規化と特徴量の圧縮による手書き漢字認識の高精度化", 信学論 (D-II), vol. J78-D-II, no. 5, pp. 765-774, 1996.
- [2] X. Wang, X. Ding, C. Liu, "Gabor filters-based feature extraction for character recognition," Pattern Recognit., vol. 38, pp. 369-379, 2005.