

字幕パターン認識におけるエッジ方向成分射影法の有効性検討

2L-9

吉村 浩至¹⁾ 近藤 堅司²⁾ 栄藤 稔¹⁾³⁾ 横矢 直和¹⁾

1) 奈良先端科学技術大学院大学 情報科学研究科

2) 松下電器産業株式会社 先端技術研究所

3) 松下電器産業株式会社 東京通信システム研究所

1 はじめに

膨大な映像データから必要なデータを抽出する手段として、映像内の文字をキーとする検索が有効であると考えられる。そのためには映像内の文字認識を行う必要がある。その一手法として栄藤が提案しているエッジ方向成分射影による多値文字パターン認識手法[1]を低品質の字幕認識に適用し、評価実験を行った。

2 エッジ方向成分射影法による特徴抽出

特徴抽出は、Gabor フィルタ [2] によって 4 方向のエッジ成分を抽出し、その出力画像から輝度を累積し、特徴ベクトルを得る。以下に、Gabor フィルタ、特徴抽出手順について詳しく述べる。

2.1 Gabor フィルタ

文字画像からのエッジ抽出には Gabor フィルタを用いた。Gabor フィルタの方向を θ とすると、Gabor フィルタは以下の式 $h(x, y)$ で表される。

$$\begin{aligned} h(x, y) &= g(x, y) \exp(2\pi i \omega x) \\ g(x, y) &= \frac{\exp\{ -[(x/\sigma_x)^2 + (y/\sigma_y)^2]/2 \}}{2\pi\sigma_x\sigma_y} \\ (x', y') &= (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta) \end{aligned}$$

上式の $h(x, y)$ と文字画像上の各画素の輝度をたたみ込み、虚部と実部の絶対値をとることにより、フィルタ出力を得る。今回の実験では、上記の式のパラメータは $\omega = 0.23$, $\sigma_x = 2.0$, $\sigma_y = 2.0$, $\theta = 0, \pi/4, \pi/2, 3\pi/4$ として実験を行った。文字画像に Gabor フィルタをかけたときの出力例を図 1 に示す。

2.2 特徴ベクトル抽出手順

特徴ベクトルを得る手順は以下の通りである。

- (1) 64 画素 × 64 画素の入力画像に対して 4 方向の Gabor フィルタにより $\theta = 0, \pi/4, \pi/2, 3\pi/4$

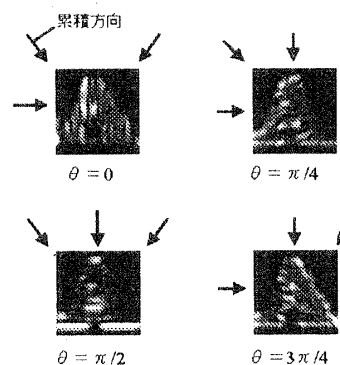


図 1: Gabor フィルタ出力例と各エッジ画像に対する累積方向

のエッジ成分を出力し、そのフィルタ出力の実部と虚部の絶対値を出力画像とする。

- (2) 各方向の出力画像から輝度を累積する。このとき抽出されるエッジに平行な方向の累積は行わない (図 1 参照)。エッジに平行な方向の累積はパターン位置の変動に頑健でないと考えられるからである。累積する画素を 2 ライン間引き一方向につき 16 の特徴を得る。これにより 4 (フィルタ出力方向) × 16 × 3 (累積方向) = 192 次元の特徴ベクトルを得る。
- (3) 特徴ベクトルの大きさを特徴ベクトルの各要素を正規化する。

3 実験

前述の特徴抽出法を用いて、文字認識実験を行った。対象となる文字画像は、5 本の外国映画のエンドロールから名前の部分を手動で切り出し、64 画素 × 64 画素の大きさに正規化した (図 2 参照)。正規化の方法は、文字の縦あるいは横の長さの大きい方を 64 画素に合わせているため、文字自体の縦横比は変わらない。5 本の映画から切り出した総文字数は 758 文字、カテゴリはアルファベットの大文字の 26 カテゴリである。

また同じ文字画像を 2 値化し、同様の方法で 2 値画像からの特徴抽出を行い認識実験を行った。その場合の入力画像の 2 値化には、大津の自動しきい値選定法 [2] を用いた。この方法は、画像全体の濃淡

Efficacy Evaluation of Directional Edge Component Projection Method for Caption Recognition
Hiroshi Yoshimura 1), Kenji Kondo 2), Minoru Etoh 1) 3), Naokazu Yokoya 1)

1) Nara Institute of Science and Technology (NAIST)
8916-5 Takayama, Ikoma, Nara 630-0101, Japan.

2) Advanced Technology Research Laboratories,
Matsushita Electric Industrial Co., Ltd.

3) Tokyo Communication System Research Laboratory,
Matsushita Electric Industrial Co., Ltd.

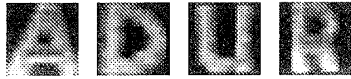


図 2：入力画像の例



図 3：入力画像の2値化処理後の画像

値分布の判別分析により、文字と背景の2つのクラス分離性を最大にする最適なしきい値を発見するものである。入力画像を2値化した例を図3に示す。

3.1 実験方法

5本の映画のうち4本を学習データとし、残り1本を評価データとする。評価データとなる映画を代えて5回実験をし、認識率の平均を実験結果とした。また学習には学習ベクトル量子化法 (LVQ) [4] を用いた。LVQの参照ベクトルの個数は、1カテゴリにつき1個である。

3.2 実験結果

実験結果を表1に示す。表の左に書いてある映画を評価データとし、それ以外の4本の映画を学習データに用いている。

3.3 考察

映画字幕認識で81%の認識率を得た。2値の画像からの認識でもほぼ同じ認識率であり、2値化の必要なく認識できることがわかった。各映画に対する考察を以下に示す。

- 映画1で2値の認識率が低いのは、文字の輪郭付近にのったノイズのために、Gaborフィルタ出力でも本来のストローク方向以外のエッジ成分が出たためと考えられる。多値の認識の場合も同様の傾向があるが、ストローク方向のエッジ成分に比べ、比較的小さい出力となるため、認識に対する影響は減少していると考えられる。

表 1: 実験結果 (認識率)

評価データ	多値画像 (%)	2値画像 (%)
映画1	80.1	62.9
映画2	92.3	91.6
映画3	74.1	80.4
映画4	72.8	89.0
映画5	89.7	84.7
平均	81.8	81.7

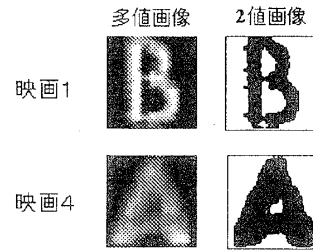


図 4：映画1と映画4の文字例

- 映画4で多値の認識率が低いのは、文字画像の全体的なコントラストが低いため、エッジを抽出するのが困難になっているからと考えられる。Gaborフィルタをかける前に入力画像の輝度を正規化することにより、改善されると思われる。

映画1と映画4の文字例を図4に示す。

4 まとめと今後の課題

エッジ方向成分射影による多値文字パターン認識手法を低品質の字幕認識に適用した。低品質の字幕文字の認識実験において81%の認識率を得た。2値画像に対しても多値画像と同様のアルゴリズムでほぼ同じ認識率が得られたので、入力画像の2値化を必要とすることなく認識ができることがわかった。文字の輪郭付近でノイズがある場合、2値画像では、ノイズの影響のため認識率が低下するのに対し、多値の認識ではノイズの影響が少ないことがわかった。

今後、特徴ベクトルの構成方法を検討し、認識率の向上を図る。そしてこれまで文字認識で有効とされている方向寄与密度特徴による手法 [5] と比較し、多値画像を直接認識する手法の有効性を確認して行く予定である。

参考文献

- [1] 栄藤：“パターン位置決め装置とパターン類別装置”，日本国特許公開 H04-329488 (1991)
- [2] D.Gabor：“Theory of communication”，*J.Inst.Elect. Engr.*, Vol.93, pp.429-457 (1946)
- [3] 大津：“判別および最小2乗規準に基づく自動しきい値選定法”，信学論，Vol.J63-D No.4 (1980)
- [4] T.Kohonen：“Learning Vector Quantization for Pattern Recognition”，Helsinki University of Technology, Report TKK-F-A601, Nov. (1986)
- [5] 萩田，他：“大局的・局所的方向寄与度特徴による手書き漢字認識方式”，信学論，Vol.J-66D, No.6, pp.722-729 (1983)