

看板画像からの文字抽出における 処理領域の限定と特徴量の補強に関する研究

大原宏太[†] 小嶋 和徳[†] 伊藤 慶明[†] 石亀 昌明[†]
岩手県立大学ソフトウェア情報学部[†]

1. はじめに

OCR 技術による文書画像からの文字認識は現在では実用段階にある。しかし、情景画像からの文字認識は OCR 技術を用いても困難である。これは、情景画像が複雑な背景パターンを有しており、文字位置の特定が難しい為である。もし情景画像における文字認識精度が向上すれば様々な分野での活用が期待できる。本研究では、先行研究¹⁾を基に、文字領域の作成やクラスタリング処理における問題点を改善することで文字抽出の精度向上を図る。

2. 先行研究の手法と問題点

先行研究は、入力画像に対して Sobel フィルタによりエッジを抽出し、位置情報によりエッジ領域を限定する。次に、クラスタリング処理により、2 値画像を作成する。得られた 2 値画像から文字または背景の特徴量を抽出し、SVM により学習することで与えられた入力画像内に文字が存在するか識別を行う。この研究では以下の様な問題点がある。

- (1) エッジ抽出時に弱いエッジが残りやすい為、背景領域が文字候補領域として多く作成残されるほか、文字領域の作成が弱いエッジに阻害される。
- (2) クラスタ数が多い為に、文字領域の画素が 1 つのクラスタに集中せず、うまく特徴量が抽出出来ない。

3. 改善案

本研究では、問題点の(1)に対しては強いエッジのみを用いた 2 値化、(2)に対しては輝度ヒストグラムによるクラスタ数の設定をそれぞれ行うことで問題点の改善を図った。

3.1. 強いエッジのみを用いた 2 値化

Sobel フィルタによるエッジ画像を 2 値化する

際、閾値を画像全体の濃度の平均を用いて決定すると、閾値が低くなり、弱いエッジでも残ってしまう。文字が人に見えやすい様に背景との色差がはっきりしている場合、エッジも強くなるため、本研究では強いエッジだけを残すように処理を行う。具体的には一定値以上の濃度をもつ画素だけを用いて 2 値化の閾値を決定する。今回は、濃度 20 以上の画素を利用して閾値を決定した。

3.2. 輝度ヒストグラムの頂点数を利用した初期クラスタ数設定

先行研究では、クラスタリングの際に、初期クラスタ数を 4 とし、クラスタリングの実行中に代表点同士が近くなった場合にクラスタを結合している。しかし、ほぼ 2 つの色しか見られない領域についても代表点が近くないために結合されず、文字領域の画素が 1 つのクラスタに集中しないことがある。この場合、文字としての特徴量を良好に抽出出来ず、SVM による学習及び認識に悪影響を及ぼす。そこで、本研究では適切なクラスタ数を設定する為に、輝度ヒストグラムを用いる。文字領域であるならば背景と文字とで輝度差が生まれると考えられるため、輝度ヒストグラムの山の数をクラスタ数とすることができる。ここでは、領域ごとに輝度ヒストグラムを求め、閾値以上となった場所を山として数え、山の数を領域ごとの初期クラスタ数として設定する。この時、閾値はヒストグラムの平均とする。そしてクラスタリングの実行中に代表点が近くなった場合、各クラスタを結合することとした。図 1 のような輝度ヒストグラムが得られた場合、閾値以上となる山は 2 つなので、初期クラスタ数は 2 としてクラスタリングを行う。

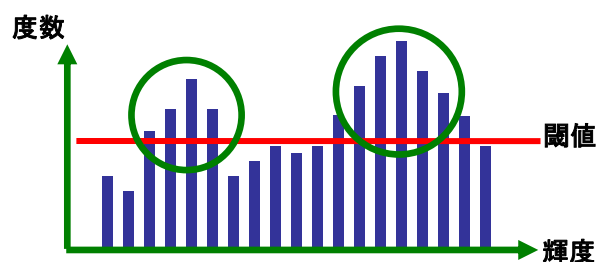


図 1 輝度ヒストグラムからのクラスタ数決定

A study of Pre process for Feature Extraction in Letter Extraction from Scene Image

[†] Kouta O'Hara, Kazunori Kojima, Yoshiaki Ito, Masaaki Ishigame. Faculty of Software and Information Science, Iwate Prefectural University

4. 実験及び評価

4.1. 評価方法

識別には LIBSVM⁴⁾を用いる. カーネルは RBF を使用し, $\gamma = 0.6$ とした. 本研究の評価方法として, 先行研究で用いられた以下の式を利用する.

$$p = \frac{n_m}{n_e} \quad (1)$$

$$r = \frac{n_m}{n_g} \quad (2)$$

$$f = \frac{2pr}{p+r} \quad (3)$$

評価式での f は F 尺度で, p は適合率, r は再現率をそれぞれ表している. n_m は, 評価画像と比較したときに, とともに文字列パターンにある画素数, n_e は抽出された文字列パターンの画素数, n_g は評価画像の文字列パターンの画素数をそれぞれ表している.

4.2. 実験データ

学習および評価データとして ICRDA2003 で使用された画像セット 258 枚を使用した. 学習データとしては画像 200 枚, 評価画像として残りの 58 枚を使用し評価実験を行った.

4.3. 結果・考察

表 1 実験結果

	適合率[%]	再現率[%]	F 尺度[%]
本研究	59	50	53
岩渕	60	44	50

表 1 が評価結果となる. 再現率が 6%向上していることから, より文字領域を取り出せていることが分かる. その一方で適合率が低下しているのは, 図 2 のような文字領域の背景が文字と似た特徴量を得るため, 図 3 のようにどちらも文字と認識し, 塗りつぶしたような出力になったのが原因であると考えられる.



図 2 文字領域における背景



図 3 文字・背景が取得されてしまった結果

図 4, 図 5 に先行研究による識別結果との比較を示す. これらの図から, 適切なクラス数でクラスタリングしたことで, 元の文字の形状を維持したまま, 出力されていることが分かる.



図 4 先行研究



図 5 本手法

また, 本手法を用いた場合, ガラスへの移り込みが弱ければ, 抽出は可能であることが確認できた. しかし, 図 6 のように背景情報がガラスに強く写り込み, 背景のエッジが強くなる場合は文字領域の抽出がやはり困難となった.



図 6 写り込みの激しい画像での識別

5. おわりに

本研究では, 弱いエッジの除去と輝度ヒストグラムを用いたクラス数の設定によって文字領域を取り出しやすくなる方法を提案した. その結果, 精度向上を図ることが出来た. 今後は, 図 2 のような文字領域の背景部分と文字部分を良好に分けるための特徴量を考える必要がある.

参考文献

- 1) 大原宏太, 岩渕慎也, 小嶋和徳, 伊藤慶明, 石亀昌明, "看板画像からの文字抽出に関する特徴量の検討", 平成 22 年度電気関係学会東北支部連合大会 TSJC2010, p.114, 2010.
- 2) 宮本定明, "クラスター分析入門", 森北出版株式会社, 1999.
- 3) 小野田崇, "サポートベクターマシーン (知の科学)", オーム社, 2007.
- 4) "LIBSVM", <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, (閲覧 1 月 11 日).