

情景画像からの判別分析に基づいた文字領域抽出手法に関する研究

Study on Text Detection from Natural Scene Images based on a discriminant analysis method

研究学生 小河祐一 指導教員 江崎修央

1 はじめに

近年、福祉情報工学分野の研究が盛んに行われており、なかでも、視覚障害者の数は全国で30万人を超えるといわれている。我々が生活する空間上には、交通標識・看板など多くの文字情報が存在しているが、視覚障害者はこれらの情報を自身で得ることは出来ない。そこで、我々は、先行研究[1]を参考にしつつ、視覚障害者向けの環境内文字読上げシステムの開発を目的としている。本稿では情景画像からの小さい文字領域抽出手法について報告する。

2 システムの基本設計

本システムの構成要素は、携帯可能な情報端末とパン・チルト・ズームイン機能付のカメラである。システムの流れとして画像のキャプチャ、文字領域の抽出、文字認識、文字情報の読上げを順に行っていくが、初期キャプチャ画像に含まれる文字のサイズは小さいことが予想されるため、まず小さいサイズ文字領域を次項で説明する文字抽出手法で候補文字領域を抽出し、候補文字領域にズームインした後、文字認識用の画像を改めてキャプチャすることになっている。

3 小さい文字領域の抽出手法

3.1 FDR よるしきい値の決定

我々はこれまでに、小さい文字領域を抽出するためにモルフロジー演算のTop-hat処理を応用した手法[2]を提案してきたが、実用にはまだまだ不十分である。そこで、文字認識用領域抽出手法のひとつであるRGB値の各チャネルに対して判別分析法を利用して2値化を行う手法(8color)[3]を応用した手法を新たに開発することとした。8colorでは、濃度値の分布をしきい値によって2クラスに分類し、そのクラス間分散式(1)が最大になるようにしきい値 th_global を決めるので、小さい文字領域は背景と同化してしまい抽出できない。

$$S_B^2 = w_1 w_2 (m_1 - m_2)^2 \quad (1)$$

そこで、画像を小領域(32×32pixel)に分割しその分割された各画像のRGBの各チャネルに対して2値化を適用する場合のしきい値を th_local とする。

ここで、小領域に文字がある場合はしきい値 th_local を利用し、文字が無い場合には th_global を利用して2値化する。どちらのしきい値を利用するかは式(2)に示すFDRの値によって判断する。

$$FDR = S_B^2 / S_W^2 \quad (2)$$

ここで、 S_W^2 はクラス内分散であり式(3)で求められる。

$$S_W^2 = w_1 S_1^2 + w_2 S_2^2 \quad (3)$$

画像のパターンは図1に示すとおり、3種類に大別できる。図1(a)は文字領域の場合で濃度ヒストグラムは双峰性を示す。 FDR は式(2)から解るように、双峰性があるほど値が大きくなる。このとき2値化のしきい値は th_local を利用する。また、図1(b)は背景領域などで、ヒストグラムは単峰性を示す場合。図1(c)はさまざまな要素が画像に含まれている場合で、ヒストグラムは分散する場合である。このような画像は th_global を使って2値化する。

今回の実験では $FDR > 5$ の場合、ヒストグラムは双峰性があると判断した。

3.2 文字列規則の適用

3.1の処理によってRGB各カラーチャネルに2値化された画像に対して、ラベリングを行い、文字列規則の適用[2]を行う。ラベリングされた領域のうち、面積が20ピクセル

以上かつ高さが5以上、アスペクト比が0.1~2の全ての条件を満たした領域が2つ以上並んでいれば、候補文字領域として抽出する。

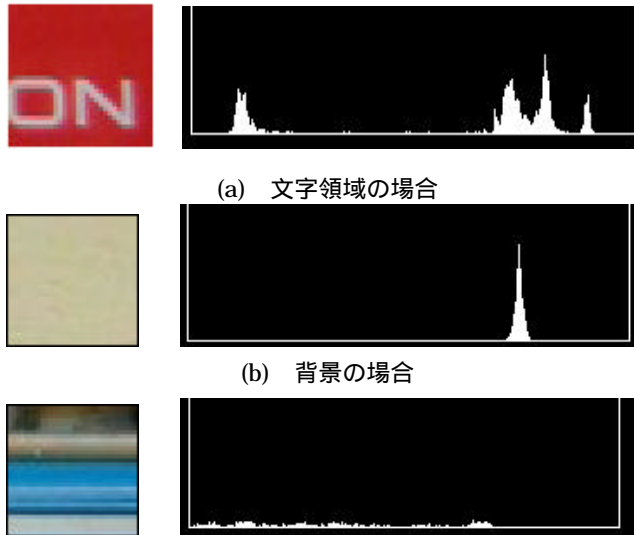


図1 小領域とその濃度値ヒストグラム

4 評価実験結果

評価画像データはICDAR2003の文字列抽出コンテスト[4]において利用されたものから、文字サイズが小さい文字(30pixel以下)を含むデータセットを作成し利用した。ICDAR2003と同様の評価値を利用し $Precision=c/E$, $Recall=c/T$ (E : 候補文字領域, T : 正解文字領域, c : T と E の交差領域)とした。また F はPrecisionとRecallの平均である。

評価実験の結果提案手法のPrecisionが高いことがわかった。また、従来の手法と提案手法のOR領域では $Precision=0.46$, $Recall=0.50$ となった。これはそれぞれの手法が違う文字領域を抽出しているため、互いに補うことによって高い精度で候補文字領域を抽出しているといえる。

表1 各手法の評価実験結果

	Precision	Recall	F
提案手法	0.51	0.31	0.39
従来手法[2]	0.47	0.35	0.40
上記のOR領域	0.46	0.50	0.48

参考文献

- [1]視覚障害者のための環境内の文字情報抽出システム, 鈴木, 大西他, 電気学会論文誌(C), vol.124-c, no.6, pp.1280-1287, 2003-6.
- [2]Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons, Nobuo Ezaki 他, International conference on pattern recognition 2004
- [3]シーン中の文字領域検出における周波数特性の分析と比較, 齋藤 靖二, 後藤 英昭, 小林 広明, 電子情報通信学会技術研究報告, PRMU2004 - 123 ~ 133, p13-36
- [4]S.Mlucas, A.Panaretos, L.Sosa, A.Tang, s.Wong, and R. Young, "ICDAR 2003 Robust Reading Competitions", Proc.of the ICDAR, 2003, pp.682-687.