

複数の領域評価手法の統合による文字認識

Character recognition by integration of multiple region evaluation methods

佐藤沢彦 和泉勇治 田中和之

Sawahiko Satou Yuji Waizumi Kazuyuki Tanaka

東北大学 大学院情報科学研究科

Graduate School of Information Sciences, Tohoku University

1. はじめに

情景画像中の文字認識は視覚障害者向けのナビゲーションシステムや自動的な映像の目録の作成など様々な応用例が考えられている。現在の情景画像中の文字認識の研究においては情景画像から文字領域を抽出した後に認識を行う2段階の手法がある。しかし、これらの手法は文字領域抽出によって文字があると判定された領域に対してのみ認識を行うために、抽出漏れとなった文字に対しては認識処理が行われない。そのため、文字領域の未抽出と文字の誤認識が積み重なってしまうという欠点がある。そこで本稿では画像から抽出されたすべての輪郭線を認識対象とし、文字認識の前段階で認識対象の絞込みを行わない1段階の認識手法を検討する。

画像からの輪郭抽出法としては類似色、色相、輝度、またはそれらを複合したものなどがある。複数の抽出法により抽出されたすべての輪郭線を認識対象とすれば文字領域の抽出漏れへの対処は可能とあるが、文字ではない輪郭線も抽出されてしまう問題も生じてしまう。特に、異なる輪郭抽出方式を利用した場合、それらの抽出特性の違いから、文字以外の不要な輪郭線をより多く認識対象としてしまう可能性が考えられる。そこで、本稿では、1つの輪郭抽出方式を利用し、それを入力する画像に変化を与えることで、不要な輪郭が抽出されることを防ぎつつ文字領域の抽出漏れを低減する認識方式を提案する。

2. 輪郭線の認識

本稿では、画像から抽出された輪郭線から、次で定義される2次元ヒストグラムを特徴量として用いる。文字画像の輪郭の黒画素の重心を原点とし、角度方向と動径方向に分割された領域ごとの黒画素数を全黒画素数で正規化した値をヒストグラムで表したものを特徴量として抽出する。本研究では動径方向10角度方向10の合計100の領域に分割し特徴量となる2次元ヒストグラムを作成した。

抽出された特徴量 \mathbf{x} とし、 C_k を各文字のクラスとした時、式(1)であらわされる事後確率が最

大となる C_k を \mathbf{x} に対する認識結果とする。

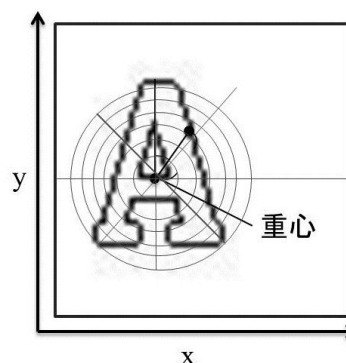


図1 分割される領域の例

$$P(C_k | \mathbf{x}) = \frac{p(\mathbf{x} | C_k)p(C_k)}{p(\mathbf{x})} \propto p(\mathbf{x} | C_k) \quad \dots(1)$$

$$P(\mathbf{x} | C_k) = \prod_{l=1}^L \pi_l \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_l, \Sigma_l) \quad \dots(2)$$

ただし、 $\mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_l, \Sigma_l)$ はガウス分布、 L はコンポーネント数であり、本研究では $L=2$ とした。

3. 情景画像中の文字認識

1で述べたように、輪郭抽出方式に入力する画像に対し変化を与え、1枚の未知画像から複数の画像を生成し、それらから抽出された輪郭に対する認識結果を組み合わせる文字認識を提案する。

3.1 輪郭抽出法

本稿では輪郭抽出方式としてCanny edge detectionを利用する。

本稿では、未知画像をそのままCanny edge detectionに入力し輪郭を抽出した画像(以下、Canny)と、類似色のクラスタリングにより減色処理を施した画像にCanny edge detectionに入力し輪郭抽出した画像(以下、Color)の2種類を1枚の未知入力から生成された認識対象の画像として扱う。ここで、減色処理は、画素の(R, G, B)のデータをK-meansクラスタリングによってクラスタリングし、各クラスタ内のピクセルをそのセントロイドで置き換えることによって4色に減色した画像を得るものである。

3.2 曲線の組み合わせ画像作成

図2のように、輪郭画像から曲線を1つから3つ取出し、それらの曲線を組み合わせて1文字として認識される候補となる画像を作成する。この画像から特徴抽出を行い、2の認識処理を適用することで個別の文字認識を実現する。

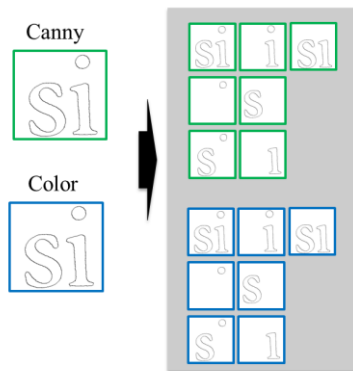


図2 作成される組み合わせ画像の例

3.3 輪郭線の組み合わせ画像の識別

3.2で得られた画像を使って情景画像の文字の認識を行う。CannyとColorの2つの輪郭抽出法で共通する領域を持つ曲線の組み合わせ画像の特徴量 $\mathbf{x}_{\text{Canny}}$ と $\mathbf{x}_{\text{Color}}$ を式(1)にそれぞれ入力し得られた尤度に対し、閾値 h_k と l_k ($h_k > l_k$) を設定し、図3のような決定木によって認識結果を出力していく。ここで k_{max} は事後確率を最大とする文字のクラスである。この過程では同一の曲線に対して複数の認識結果を与えることがあるが、その場合は事後確率の高いほうの結果を優先していく。

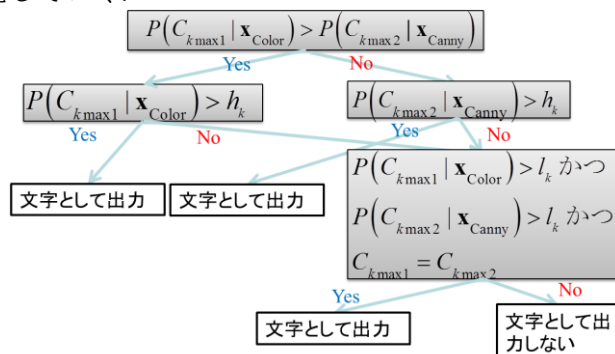


図3 提案手法の出力を決める決定木

3.4 閾値設定方法

ある字種 a に対する閾値 h_k と l_k 設定には、学習データ内の字種 a の文字画像と a 以外の文字画像、文字ではないランダムな曲線の画像を利用する。これらの画像の字種 a のクラスに対す

る事後確率を式(1)から求める。正解データと不正解データの識別誤差を最小する閾値 h_k と正解データとランダムな曲線の識別誤差を最小とする閾値 l_k を設定する。

4. 認識実験

提案手法と重み付き識別器による結果の統合、2つの輪郭抽出法の和という3つの手法によってICDAR 2003 test dataset[1]の10枚の情景画像の認識を行った。重み付き識別機は共通する領域を持つ画像 $\mathbf{x}_{\text{Canny}}$ と $\mathbf{x}_{\text{Color}}$ に対すそれぞれの事後確率に重みを付け新たな識別機を作成し認識を行うものである。2つの輪郭抽出法の和は提案手法の h_k 以下 l_k 以上の事後確率を持つ画像に対する操作がなくなったものである。図4に10枚の画像に対する Fmeasure の値を示す。提案手法が平均して最も高い Fmeasure を得ていることが分かる。特に提案手法を2つの輪郭抽出法の和の結果と比較するとほぼすべての画像で提案手法の認識精度が優れており、 h_k 以下 l_k 以上の事後確率を持つ画像に対する提案手法の操作がほぼ成功していることを示している。

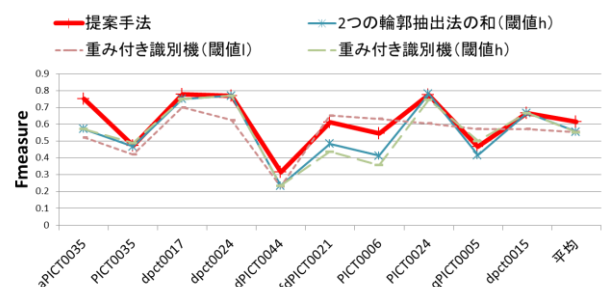


図4 10枚の画像に対する Fmeasure

5. まとめ

本稿では、減色処理を施した画像から抽出した輪郭線と原画像から抽出した輪郭線を組み合わせることにより、より高精度な情景画像中の文字認識を実現する方法について検討した。提案手法の特徴は、一つの輪郭抽出方式利用し、それに入力する画像に変化を与えることで、文字以外の不必要な輪郭抽出を防ぎ、同時に文字の抽出漏れを低減する点にある。10枚の画像を対象した認識実験では、従来手法と比較し、高い認識性能を獲得することが出来た。

参考文献

[1]S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, ICDAR 2003 robust reading competitions, 2003.