

ウェーブレット変換を利用した文字領域推定

的場 亮太[†] 久保田 光一[‡]中央大学大学院 理工学研究科 情報工学専攻^{†‡}

要約: 写真などのフルカラー画像に対する文字認識は、現在発展途上であるが、その原因として画像のどこに文字があるのかを検出するのが難しいということが挙げられる。本研究では、写真画像に対する高精度の文字認識を実現するための、ウェーブレット変換を中心とした文字領域の推定手法の実装および検証を行う。

キーワード: ウェーブレット変換, OCR, 文字認識

1 背景

現在、活字文書等の画像に対する文字認識は高い精度を誇っており、各種 OCR ソフトは既に一般市場にも流通している。また、手書き文字に対する文字認識も、筆順の情報が見られるという前提条件さえ満たせば高精度であり、PDA やニンテンドー DS などの機器に広く使われている。

その一方で、写真画像などに対する文字認識はより困難である。

2 目的

写真画像に対する文字認識の精度が低い原因の一つに、画像内に文字領域とそうでない領域が混在しているために、文字だけを抽出するのが困難であるということが挙げられる。本研究では、文字領域の空間周波数とその周囲に比べて高いという点に着目し、ウェーブレット変換を中心とした文字領域推定の手法を実装する。それとともに、各入力パラメータの変化やフィルタの適用の有無が推定結果に与える影響を調査し、情景画像に対する高精度の文字認識へと繋げることを目的とする。

3 前処理

カラー画像に含まれる文字を際立たせるために、入力画像を白黒画像へと変換する。

3.1 グレースケール変換

カラー画像を二値化するために、まず一旦モノクロ画像に変換する。本研究では、次の式によって表現される NTSC (National Television System Committee) 規格による係数を用いた。

$$Y = (0.298912 \times R + 0.586611 \times G + 0.114478 \times B)$$

3.2 二値化

グレースケールフィルタによって得られたモノクロ画像を、さらに白黒二値画像へと変換する。本研究では、画像全体の輝度の平均値を閾値として二値化を行った。

4 ウェーブレット変換

ウェーブレット変換とは周波数解析の手法の一つであり、基底関数としてウェーブレット関数を用いる。小さい波 (ウェーブレット) を拡大・縮小・平行移動しながら足し合わせることで、入力波形を近似する [1]。もともとは石油探査のために考案されたものであるが、現在では画像の符号化の国際標準である

JPEG 2000 に利用されている [2] ほか、電子透かしの埋め込み [3]、音声信号の特徴抽出 [4]、画像内の文字の判定 [5] など幅広く利用されている。

4.1 連続ウェーブレット変換

連続ウェーブレット変換は次の式で定義される。

$$W_{\psi}(t)(a, b) = \frac{1}{\sqrt{a}} \int \psi \left(\frac{t-b}{a} \right) h(t) dt$$

4.2 離散ウェーブレット変換

連続ウェーブレット変換を $a = 2^j, b = k \times 2^j$ として離散化したものであり、入力を基底関数によって高周波成分と低周波成分に分解することと等価である。

4.3 二次元離散ウェーブレット変換

離散ウェーブレット変換を二次元へと拡張することで、画像に対する離散ウェーブレット変換が可能になる。画像に対してウェーブレット変換を行うことで、図 1 のような結果が得られる。

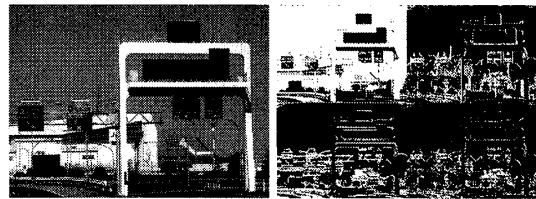


図 1 左: 入力画像 右: 出力結果

ここで、ウェーブレット変換の結果の左上 (LL) 成分は元画像が平滑化された画像である。また、右上 (LH) は水平方向の、左下 (HL) は垂直方向の、右下 (HH) は対角方向の高周波成分をそれぞれ表す。

5 文字領域の推定

ウェーブレット変換の結果、高周波成分が大きい点が多い領域は、元画像の対応する領域が「複雑」だということを示している。この「複雑」さが小さければ、元画像の対応する領域が輝度変換の少ない平坦な領域であり、大きければ輝度変化の大きい領域である。文字が書かれた領域は、基本的に文字色と背景色で構成されていると考えるならば、この「複雑」さがほどほどであればその部分が文字領域である可能性は高いと考える。そこで、以下の手順で文字領域の推定を行う。

1. 元画像を一边が n ピクセルの正方形のブロックに分割する。
2. 上限と下限の閾値を、0 から 1 の範囲でそれぞれ設定する。
3. 各ブロックにおいて、対応する高周波成分の領域における明るいピクセルの数 (輝度値が 128 以上) をカウントする。
4. カウントされたピクセルの数が、(設定した閾値の下限) \times (ブロックの総ピクセル数) から (設定した閾値の上限) \times (ブロックの総ピクセル数) の範囲にあれば、そのブロックを文字領域と推定する。

Character Region Estimation with Wavelet Transform

[†] Ryota MATOBA, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

[‡] Koichi KUBOTA, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

6 数値実験

用意した画像に対して表 1 に示すパラメータを変化させながら、評価値 (後述) が最大となるように、文字領域の推定における閾値を調整し、その結果を比較した。

表 1 パラメータ

パラメータ	取りうる値
マザーウェーブレット	Daubechies, Coiflet
使用する高周波成分	HH 成分, HL 成分, LH 成分
ブロックのサイズ	4 ピクセル, 8 ピクセル, 12 ピクセル
画像フィルタ	オフ, ガウシアン, ラプラシアン

本稿では、用意した画像のうち図 2 の 3 枚の画像についての結果を載せた。

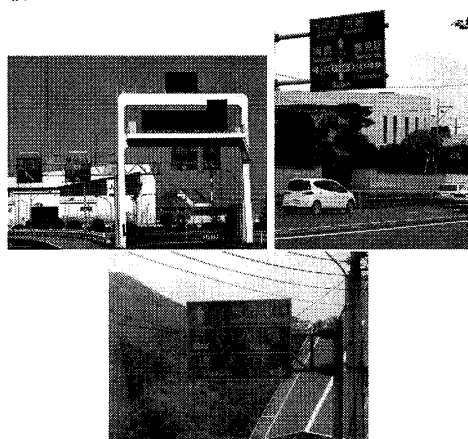


図 2 左上: IMG0 右上: IMG1 下: IMG2

6.1 評価値

推定結果の精度の評価値として、以下の値を用いた。

$$P = \frac{\text{その中で文字領域と推定されたブロックの数}}{\text{文字領域として推定して欲しいブロックの数}}$$

$$Q = \frac{\text{その中で文字領域と推定されなかったブロックの数}}{\text{文字領域として推定して欲しくないブロックの数}}$$

$$\text{評価値} = P \times w_p + Q \times w_q$$

なお、本研究では $w_p = 2/3$, $w_q = 1/3$ として、最大値が 1 になるように評価値を算出した。

6.2 実験結果

6.2.1 マザーウェーブレットによる評価値の変化

ウェーブレット変換における基底 (マザーウェーブレット) として、Daubechies のウェーブレットと Coiflet のウェーブレットを実装し、文字領域推定の精度を比較した。その結果を表 2 に示す。

表 2 マザーウェーブレットによる評価値の変化

ファイル名	Daubechies	Coiflet
IMG0	0.46	0.50
IMG1	0.63	0.77
IMG2	0.40	0.44

結果として、全体的に Coiflet を使用した方が、Daubechies を使用した方に比べて全体的に評価値が若干ではあるが高かった。

6.2.2 ウェーブレットの成分による評価値の変化

文字領域推定においてウェーブレット変換結果の高周波成分を使用する際に、HH 成分・HL 成分・LH 成分をそれぞれ用いて推定を行い、その評価値の最大値を比較した。その結果を表 3 に示す。

表 3 ウェーブレットの成分による評価値の変化

ファイル名	HH	HL	LH
IMG0	0.50	0.47	0.57
IMG1	0.77	0.83	0.80
IMG2	0.44	0.46	0.46

対角方向の高周波成分を使用した時に比べて、水平方向や垂直方向の高周波成分を使用した時の方が評価値が高かった。

6.2.3 ブロックサイズによる評価値の変化

元画像を正方形のブロックに分割する際に、その大きさを変化させて推定結果を比較した。その結果を表 4 に示す。

表 4 ブロックサイズによる評価値の変化

ファイル名	4x4	8x8	12x12
IMG0	0.50	0.58	0.62
IMG1	0.77	0.79	0.83
IMG2	0.44	0.37	0.35

6.2.4 フィルタの有無による評価値の変化

ガウシアンフィルタ (平滑化フィルタ) とラプラシアンフィルタ (鮮鋭化フィルタ) を実装し、それぞれの結果を比較した。その結果を表 5 に示す。

表 5 フィルタの有無による評価値の変化

ファイル名	フィルタ無し	ガウシアン	ラプラシアン
IMG0	0.50	0.33	0.59
IMG1	0.77	0.33	0.87
IMG2	0.44	0.33	0.52

ラプラシアンフィルタを使用すると、評価値が上昇する一方で、ガウシアンフィルタを使用すると評価値が大きく低下する結果となった。

7 まとめと今後の課題

ウェーブレット変換による文字領域推定手法を実装し、作成したシステムに対して準備実験を行うことで、提案手法の有用性のある程度まで確認した。

今後は、その有用性を具体的に検証するためのより大規模な実験を行うとともに、他のマザーウェーブレットを使用した推定を検証するほか、ウェーブレット変換以外での文字領域推定手法との連携を検討していく必要があると考える。

参考文献

- [1] 新井 康平, ウェーブレット解析の基礎理論, 森北出版株式会社, 2000 年 11 月.
- [2] 上野 幾朗, 渥美 栄司, 小野 文孝, 静止画符号化の国際標準方式 (JPEG2000 の概要), 映像情報メディア学会誌: 映像情報メディア 54(2), pp.164-171, 2000 年 2 月.
- [3] 太神 諭, 岡本 仙太郎, 浜辺 隆二, 離散ウェーブレット変換を用いたフラクタル画像符号化に基づく電子透かし法, 電気学会論文誌 C (電子・情報・システム部門誌) 129(7), pp.1381-1382, 2009 年.
- [4] 板井 陽俊, 安川 博, ウェーブレット変換による歩行足音の特徴抽出と SVM を用いた個人識別, 回路とシステム軽井沢ワークショップ論文集 20, pp.191-196, 2007 年 4 月.
- [5] 米津 宏太郎, ウェーブレット変換による文字領域判定, 中央大学理工学部情報工学科卒業論文, 2005 年 3 月.