

5 D-9 カラーイメージからの文字抽出法についての試案

原 正一郎、高須 淳宏、桂 英史

学術情報センター

1. 目的

書誌データと目次イメージの共存するデータベースシステムを構築し、マルチメディア・ドキュメントシステムとしての電子図書館を意図したデータ入力システムの開発を行っている[1, 2]。具体的には目次のイメージデータから書誌情報(論文名、著者名、ページなど)の抽出を目指している。印刷媒体を電子媒体上のコードへ変換する方法としてOCRの研究が行なわれているが、OCRの実用化を目指す上で克服しなければならない問題は数多く指摘されている[3, 4]。これらの問題に加え、本システムが対象としている目次イメージの認識では、1)ページ中に写真・模様・野線などの非テキストイメージが散在している、2)論文名の長さ・著者名の長さや数が不定である、という問題もある。これまでは目次イメージ用OCRのかかえる諸問題のうち、1)ページイメージをテキストとそれ以外の領域にセグメント化する、2)テキスト領域をより詳細にセグメント化する、方法についての検討を行い、一定の成果をあげつつある[5]。

ところで、これまでのOCRでは地が白で文字は黒である(あるいはモノクロ走査において、地の輝度が文字の輝度に対して有意に高い)ことが暗黙の前提条件であった。しかし、デザインなど都合上から目次イメージではこの前提を満たさないことが多い。本稿では、地と文字が多様なカラー特性を示す目次イメージから文字情報を抽出する方法について考案する。ここで考案した方法では、地の輝度が文字の輝度よりも高いといったこれまでのOCRにおける制約を受けることはないが、「文字や地の特性はある程度の広さの領域内でほぼ均一である」という拘束条件が存在する。

2. 方法

本法では画像イメージの特徴として、ある領域内の輝度変化をRGBごとに2次元フーリエ変換した空間周波数分布を用いる。イメージ上の水平方向をX軸、垂直方向をY軸で表わしたとき、座標(X, Y)の濃度分布を $f_i(X, Y)$ 、X方向の画素分割数をN、Y方向の画素分割数をM、X方向の空間周波数を U_i 、Y方向の空間

周波数を V_i とすると、離散2次元フーリエ変換は次のように表わされる。ただし、 i はR、G、Bを表す。

$$F_i(U, V) = \sum_{X=0}^{N-1} \sum_{Y=0}^{M-1} f_i(X, Y) E$$

$$E = \exp[-j2\pi \left(\frac{UX}{N} + \frac{VY}{M} \right)]$$

この方法の基本的な考え方は、テキスト領域とそれ以外の領域では2次元フーリエ変換による空間周波数パターンが異なるということである[5, 6]。一般にテキスト領域では、文字列は一定の行間隔で書かれているので、この間隔がテキスト領域をフーリエ変換して得られた周波数空間上のパワースペクトルで大きなピークを示す傾向がある。これに対して写真や絵などの輝度変化はより複雑であるので、周波数パターンはテキスト領域に比べるとかなり異なったパターンを示す。

さらに、テキスト領域の周波数分布からイメージの傾きを推定できることができる。つまり、傾きのない横書きテキストでは、文字列の周期性はX軸方向よりY軸方向の方が高いので、パワースペクトルのピークは原点上とV軸上に現われる。しかしイメージが傾いている場合、傾きに応じて空間周波数のピークは (U_p, V_p) へ移動する。このとき傾き θ とピークの関係は

$$\theta = \tan^{-1} \left(\frac{U_p}{V_p} \right)$$

となる。

横書きの目次イメージを前提とした文字認識の手順は以下の通りである。

- ① スキャナにより目次イメージをR、G、Bごとに取り込む。
- ② イメージ全体にR、G、Bごとの2次元FFTを実行し、その結果を $(R(f), G(f), B(f))_i$ とする。
- ③ パワースペクトルの最大ピークからイメージの傾きを推定する。もし傾きが一定角度以上であった場合は傾きを補正する。
- ④ 前回*i*番目の処理対象であった領域を $1/4$ に等分割し、各領域内で2次元FFTを実行する。その結果を $(R(f), G(f), B(f))_{i+1}$ とする。

Character Recognition from Color Images

Shoichiro HARA, Atsuhiko TAKASU and Eishi KATSURA
National Center for Science Information System

- ⑥ i 番目の処理結果 $(R(f), G(f), B(f))_i$ と異なった周波数パターンを示す領域に対して④の処理を繰り返す。全ての領域の周波数パターンに有意の差が見られなくなったら⑥に進む。この処理により、目次イメージをほぼ均一な周波数パターンを持った領域にまで分解する(図1)。

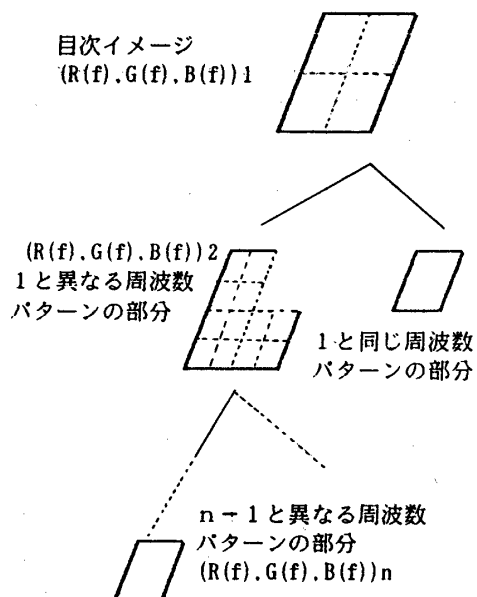


図1. 目次イメージの分解

- ⑥ 「もし文字列が主要な領域であるならば、縦方向の周波数分布上(V軸上)の直流成分以外の位置に強いピークが1つ検出される」というヒューリスティックにより、文字列が主要な領域を1つ選択する。領域の選択ができないときは処理を終了する。
- ⑦ 選択した領域内でR、G、Bごとにエッジ検出を行い、エッジを形成する地と文字部分のR、G、Bの輝度の候補を抽出し $(R, G, B)_{e1}$ と $(R, G, B)_{e2}$ とする。この段階ではどちらが文字

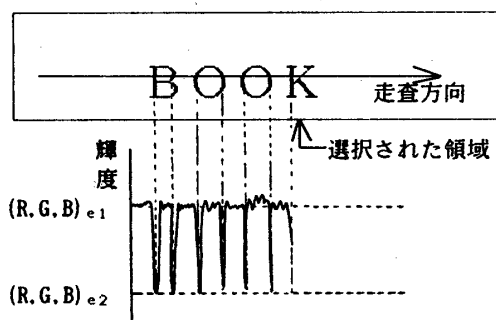


図2. エッジの抽出

を表現しているのか、あるいは両者とも文字でないのかは不明である(図2)。

- ⑧ $(R, G, B)_{e1}$ 、 $(R, G, B)_{e2}$ についての2値画像をそれぞれ作り、これらの2値画像を文字認識システムに適用して文字が抽出できるか検証する。文字列が抽出できなければ⑥にもどる。文字を検出できた輝度を $(R, G, B)_c$ として⑨に進む。
- ⑨ $(R, G, B)_c$ を全領域に適用した2値画像を作り、これまでに開発したOCRに適用する。⑥に戻る。

3. 展望

本法ではカラーの分布特性から文字の抽出を試みているので、これまでのOCRのような白地に黒文字といった制約を受けることなく文字認識を行うことができる。しかし、文字を構成するカラー特性がある程度の広さ領域内ではほぼ一定でないと、FFTによって良好な周波数特性を検出することは困難である。したがってカラー特性が位置によって著しく変化するような場合(例えば、文字ごとに色が変化したり、文字列が系統的に7色に推移しているようなデザインなど)は、本法でも文字を抽出することは困難である。

現在、試験的システムの作成と機能の評価を進めている。

参考文献

- [1] E. Katsura, A. Takasu, S. Hara and A. Nakahara: "An Approach to Electronic Contents Service Based on a Multimedia Document System", Proc. of 14th Int. Online Information Meeting, 1990.
- [2] 高須 淳宏, 桂 英史, 相澤 彰子, 原 正一郎: "マルチメディア文献システムにおけるデータの構造をビュー", 第20回情報基礎研究会, 1990.
- [3] Philippe Vincent: "Progress in automatic reading of complex typeset pages", Imaging Workstations and Document Input Systems, SPIE, Vol. 1074, pp. 146-149, 1989.
- [4] Joachim Kreich, Achim Luhn and Gerd Manderlechner, "Knowledge-Based Interpretation of Scanned Business Letters", SPIE, Vol. 1074, pp. 158-163, 1989.
- [5] 原, 高須, 桂: "目次イメージのセグメント化と文字認識", 1990年度人工知能学会全国大会論文集, pp. 403-406, 1990.
- [6] 長谷 星野: "2次元フーリエ変換を用いた文章画像領域抽出", 信学論(D), Vol. J67, No. 9, pp. 1044-1051, 1989.