

### 共起行列を用いた文書画像からの写真領域の抽出

笹川, 賢太郎 / OGAWA, Koichi / Sasagawa, Kentaro / 尾川, 浩一

---

(出版者 / Publisher)

電子情報通信学会

(雑誌名 / Journal or Publication Title)

電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理 / 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理

(号 / Number)

3

(開始ページ / Start Page)

548

(終了ページ / End Page)

556

(発行年 / Year)

1994-03-25

## 論文

## 共起行列を用いた文書画像からの写真領域の抽出

准員 笹川賢太郎<sup>†</sup> 正員 尾川 浩一<sup>†</sup>Extraction of Picture Areas from a Document Image  
Using a Cooccurrence MatrixKentarō SASAGAWA<sup>†</sup>, Associate Member and Koichi OGAWA<sup>†</sup>, Member

あらまし 文書画像には、2値の階調特性を有する文字領域と連続的な濃淡部分を有する写真領域が含まれるので、同一の処理系による印刷では、これらの画質を同時に満足させることは困難となる。このような異なる特徴の領域に対してそれぞれ最適な階調表現を行うためには、これらの領域を正しく認識し、異なる処理を施す必要がある。本研究では、コピー機や DTP 装置等における文書画像の高画質出力を目的として、写真と文字とが混在する文書画像から写真領域を正しく抽出するアルゴリズムを開発した。提案する手法は、双峰性ヒストグラムの最大分散による2分割アルゴリズムを利用した2値化、共起行列を用いた特徴抽出、およびハフ変換を導入した写真領域拡張という三つのアルゴリズムから構成される。この手法を用いることにより、背景領域との濃度差が小さい写真領域を含む画像に対しても、正確に写真領域を抽出できた。

キーワード DTP (デスクトップパブリッシング)、文書画像処理、共起行列、ハフ変換

## 1. まえがき

現在のコピー機や DTP 装置で任意の文書を出力する場合、文字のような2値の階調特性がある部分と写真のような連続的な濃淡部分とが同一の処理系で印刷されるため、これらの画質を同時に満足させることは困難である。このような特徴の異なる領域に対してそれぞれ最適な階調表現を行うためには、各々の領域を正しく認識し、異なる処理を施す必要がある。本研究では、文字と写真とが混在した文書画像から、濃淡階調特性を有する写真領域を正確に抽出するアルゴリズムを開発した。本論文における文書画像とは、白い紙の上に写真や文字、線図形などが印刷された文書をイメージスキャナによって8ビットの階調に量子化したものである。また、本論文では白い紙の何も書かれていない部分を背景領域、その他の写真や文字の領域を非背景領域と呼ぶ。

文書画像に対する画像処理として、構成要素の分離、画像の蓄積・伝送、編集・変換等<sup>(1)</sup>がある。このうち、本研究と目的を同じくする構成要素の分離とは、文書

画像から写真領域と文字領域とを分離する手法である。この手法には網点領域検出とエッジ領域検出とを組み合わせた像域分離アルゴリズム<sup>(2)</sup>、画像を4×4画素のブロックに分割し非背景領域候補のブロック内の濃度変化およびブロック単位のランレングス等を利用して写真領域を抽出する手法<sup>(3)</sup>、3×3画素の小領域を設定し、それを処理単位とする濃淡パターンの出現の仕方から写真領域を抽出する手法<sup>(4)</sup>、ディザ法および単一しきい値で2値化を行い、各々の2値画像上でランレングスを利用して領域分割を行う手法<sup>(5)</sup>等、さまざまな手法がある。しかし、これらの手法では、非背景領域と背景領域との区別に単一のしきい値を用いたり、しきい値を固定化したディザマトリクスやメッシュパターンを用いるため、しきい値の条件に合う画像や明確に非背景画素と背景画素との濃度差が異なる画像でなければ、適切に写真領域を抽出できないという問題があった。そのため、背景領域と写真領域との濃度差が小さい場合や背景領域が完全な白色でなく淡い色をもつ場合においても正確に写真領域を抽出できるように研究を進めた。

本研究で対象とした画像は、白または淡い背景色の紙の上に、黒または濃い色の文字および写真が印刷され、各構成要素が互いに重ならず、有意な空白で囲ま

<sup>†</sup> 法政大学工学部電気工学科, 小金井市  
College of Engineering, Hosei University, Koganei-shi, 184 Japan

## 論文／共起行列を用いた文書画像からの写真領域の抽出

れたグレイスケール 256 階調 (黒:0, 白:255) の画像である。ここでは、面積 (画素数) が大きく、濃度が多値である非背景領域を写真領域とし、共起行列から計算される特徴量を用いて写真領域を抽出した。また、写真領域拡張法を用いることにより、写真領域と背景領域との濃度差が小さい画像に対しても正確に写真領域を抽出できた。

## 2. 本研究に用いた画像

本研究では、文書画像認識アルゴリズムおよびパラメータ評価のために UNIX マガジンとアサヒグラフ (網点線数 100~150) から得られた  $6.4 \times 6$  cm の多数の画像を用いた。これらの画像は、イメージスキャナ GT6000 (EPSON 製) を介して、横 256 画素、縦 240 画素、1 画素当り 8 ビット (白黒 256 階調)、解像度 100 dpi でデジタル化したものである。顕著に傾いて入力された画像や、中の文字が読みとれないほど小さく縮小された画像等は、本研究には使用しなかった。また、後述する B5 サイズの 3 種類の文書画像を 60 枚用いて本手法の実用性を評価した。

## 3. 方 法

デジタル化された文書画像から写真領域を認識するための方法を、以下に説明する。

まず文書画像に対して、3.1 で述べる双峰性ヒストグラムの分散を最大とする 2 分割アルゴリズム<sup>(6)</sup>を用いた 2 値化のアルゴリズムを適用する。次に、これより求めたしきい値にある一定のオフセットを加えた値 (ここでは 20) より低い濃度値をもつ画素を非背景画素として抽出し、この連続した非背景画素の領域に対してラベル付け<sup>(6)</sup>を行う。このようにして抽出された各々の領域  $1 \sim P$  について、領域拡張による孤立点の除去 (穴埋め) や欠損部の除去 (欠けの埋め) を行う。そして、各々の領域について、その領域に含まれる画素数  $z(l)$  [領域番号  $l: 1 \sim L$ ] を求める。ここで、非背景領域の穴埋め、欠けの埋めは、図 1 のような方法で行う。図 1 (A), (B) は、それぞれ横方向、縦方向に対して等しいラベル (label a) をもった非背景画素に囲まれている背景画素 (background) を非背景画素 (label a) にする処理である。この処理により、写真領域中の欠損部を埋めることができ、また文字領域では背景画素を取り込むことになるので、文字領域の 2 値的な特徴を際立たせることもできる。図 1 (C) は、label a の非背景画素に囲まれている label b の非背景画素を、label a の非

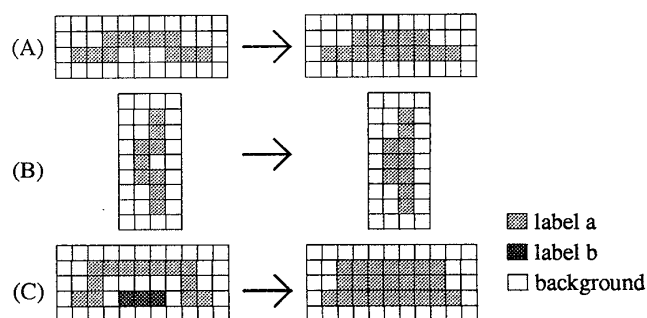


図 1 非背景領域の穴埋め、欠けの埋め  
Fig. 1 Expansion of areas.

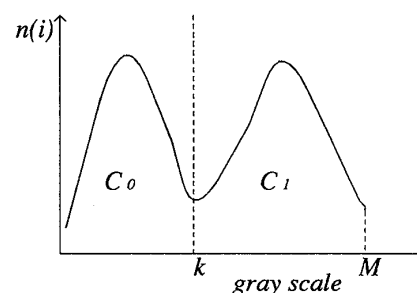


図 2 濃度ヒストグラム  
Fig. 2 Gray scale histogram.

背景画素にする処理である。この処理により、領域認識の妨げとなる孤立点を除去できる。

次に、3.2 に述べる共起行列から得られる特徴量を用いて写真と文字を認識し、面積 (画素数) が大きく、濃度が多値である非背景領域を写真領域として抽出する。この段階では、背景領域と写真領域との接する部分の濃度差が小さいときに写真領域の一部が写真領域として抽出されないことがある。そこで、この欠点を補うために 3.3 で述べる写真領域拡張という処理を行い、背景領域中に埋もれた写真領域部分を抽出する。

次に、これらの手法について詳しく説明する。

## 3.1 2 値化のアルゴリズム

まずはじめに、双峰性ヒストグラムの最大分散による 2 分割アルゴリズム<sup>(6)</sup>を利用して、非背景画素の抽出のための 2 値化を行う。この方法は、濃度ヒストグラムをあるしきい値で二つに分ける際、二つに分けられたグループ間の分散が最大となるようにしきい値を定め、2 分割する方法である。図 2 に示すヒストグラムで、gray level  $i$  ( $0 \leq i \leq M$ ) の頻度を  $n(i)$  とするとき、level  $i$  の濃度が出現する確率  $p(i)$  は  $p(i) = n(i) / \sum_{j=0}^M n(j)$  となる。ここで、level  $k$  で  $\{C_0 | i \in 0, 1, \dots, k\}$ ,  $\{C_1 | i \in k+1, k+2, \dots, M\}$  と二つの群にクラス分けしたとすると、 $C_0$  と  $C_1$  の生起確率はそれぞれ

$$\omega_0(k) = \sum_{i=0}^k p(i) \quad (1)$$

$$\omega_1(k) = \sum_{i=k+1}^M p(i) = 1 - \omega_0(k) \quad (2)$$

と表される。また、 $C_0$ 、 $C_1$ の期待値  $\mu_0(k)$ 、 $\mu_1(k)$ は

$$\mu_0(k) = \sum_{i=0}^k \frac{ip(i)}{\omega_0(k)} \quad (3)$$

$$\mu_1(k) = \sum_{i=k+1}^M \frac{ip(i)}{\omega_1(k)} \quad (4)$$

と表され、二つのクラス間の分散は

$$\begin{aligned} \delta^2(k) &= \omega_0(k)(\mu_0(k) - \mu)^2 + \omega_1(k)(\mu_1(k) - \mu)^2 \\ &= \omega_0(k)\omega_1(k)(\mu_1(k) - \mu_0(k))^2 \end{aligned} \quad (5)$$

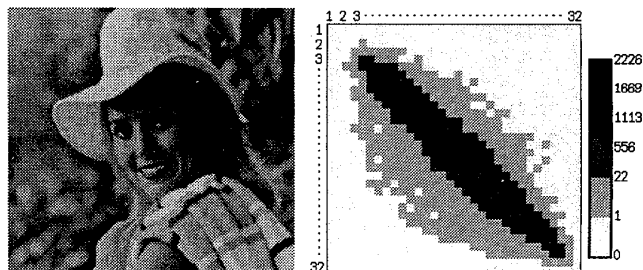
となる。ここで、 $\mu$ は全画素の平均値である。この  $\delta^2(k)$  について、 $\max \delta^2(k)$ を与える  $k$ を求めて、これをしきい値とするのが双峰性ヒストグラムの最大分散による2分割アルゴリズムである。本研究では、双峰的なヒストグラムを得るために、画像  $f(x, y)$  について2次微分  $\nabla^2 f(x, y)$ の絶対値があるしきい値よりも大きな値をもつ画素についてのみ濃度ヒストグラムを作成し、2値化のアルゴリズムを適用した。ここで  $\nabla^2 f(x, y)$ は次式で定義される。

$$\begin{aligned} \nabla^2 f(x, y) &= f(x+1, y) + f(x-1, y) + f(x, y+1) \\ &\quad + f(x, y-1) - 4 \times f(x, y) \end{aligned} \quad (6)$$

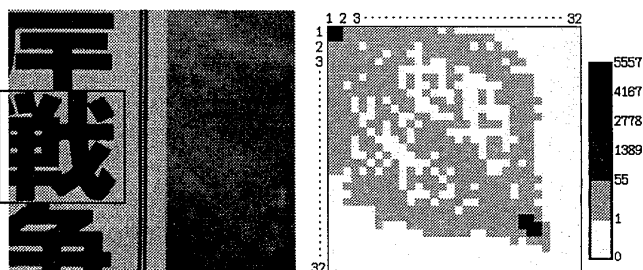
### 3.2 共起行列を用いた特徴量抽出

任意の文書画像に対して、非背景領域の抽出および各々の非背景領域の穴埋め、欠けの埋めを行った後に、各非背景領域が写真か文字かを共起行列を用いた特徴量から判断する。本研究では、各非背景領域ごとに濃度値を256階調から32階調に変換し、任意の画素  $a$ とその一つ右隣の画素  $b$ に対して  $32 \times 32$ の共起行列を作成した。共起行列とは、画像の濃度変化の様子を統計的に表したものであり、具体的には、画素  $a$ の濃度値  $i$ とその特定方向の近隣の画素  $b$ の濃度値  $j$ が出現する頻度が  $i$ 行  $j$ 列の行列として表される。ここで共起行列を求める際に、隣接する二つの画素のうち一つの画素が非背景領域であればよいことにした。本研究で共起行列を用いた理由は、濃度値が連続的に変化する写真領域では隣接する画素間の濃度差が小さく、2値的である文字領域においてはこの濃度差が大きいという特徴を利用するためである。

写真領域は、多値で画素数の大きな非背景領域と定義できるので、主に画素数が1,000以上の領域について写真か文字かを判断した。ここで、画素数1,000以上の領域について区別しているのは、画素数1,000(正方領



(a) 写真領域(全体)における共起行列の例



(b) 枠内の文字領域における共起行列の例

図3 共起行列

Fig. 3 Cooccurrence matrices.

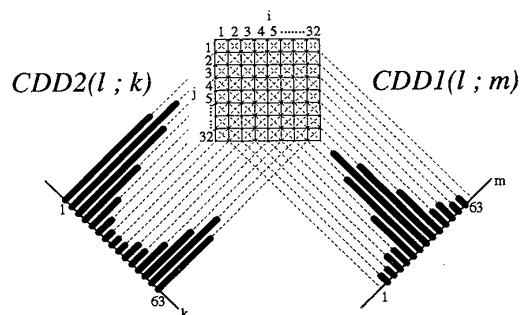


図4 共起行列とその累積分布  $CDD1(l; m)$ ,  $CDD2(l; k)$   
Fig. 4 A cooccurrence matrix and cumulative density distributions  $CDD1(l; m)$  and  $CDD2(l; k)$ .

域とすれば  $8 \text{ mm} \times 8 \text{ mm}$ 程度)より小さい写真領域は実在するとは考えられないからであり、写真領域認識の一つの目安となる画素数として用いた。更に、見出し文字のような画素数が1,000以上の大きな文字領域を文字領域(大)、画素数が1,000未満の文字領域を文字領域(小)と表記し、写真領域と区別した。実際に、写真領域と文字領域(大)における共起行列を作成すると、写真領域では対角線に沿って大きな値が分布するのに対して、文字領域(大)では左上と右下の両端に多くの値が集まることが図3からわかる。これらのパターンを考慮して、共起行列から図4に示すような  $CDD1(l; m)$  [ $m=1 \sim 63$ ],  $CDD2(l; k)$  [ $k=1 \sim 63$ ]という二つの累積分布を考えた。各領域における  $CDD1(l; m)$ ,  $CDD2(l; k)$ を図5に示す。  $CDD1(l; m)$ は、隣接す

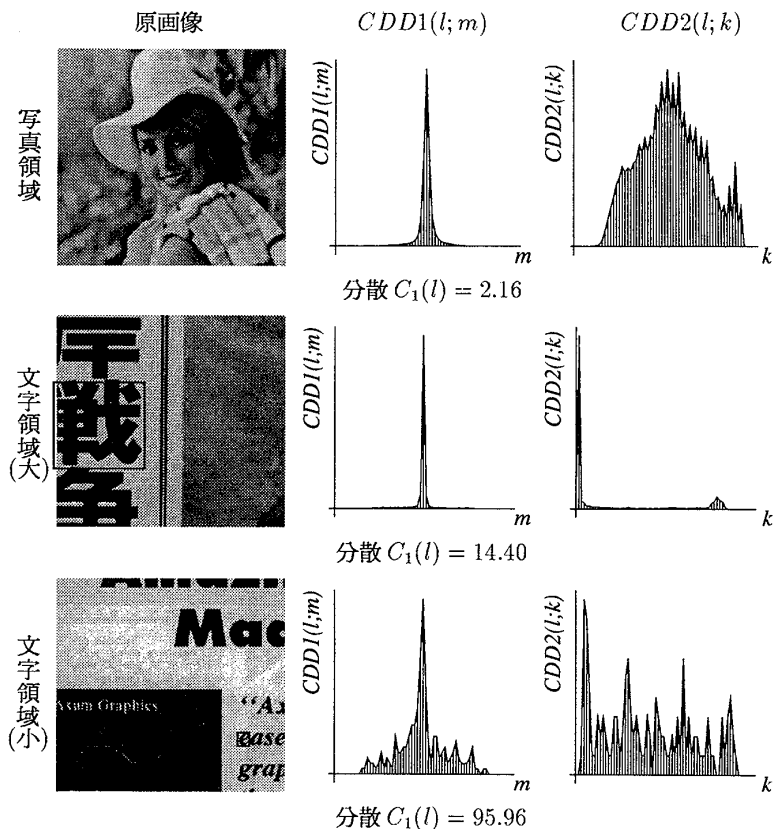


図5  $CDD1(l; m)$ ,  $CDD2(l; k)$ の例  
Fig. 5 Examples of  $CDD1(l; m)$  and  $CDD2(l; k)$ .

る二つの画素間の濃度差を表して、累積分布は中央にピークをもつ形になる。写真領域のように隣接する二つの画素間の濃度差が小さい場合には累積分布の分散が小さくなり、文字領域のように二つの画素間の濃度差が大きい場合には分散が大きくなる。 $CDD2(l; k)$ の場合、2値的な特性をもっている文字領域(大)では累積分布の両端が高くなり、濃度が連続的に変化する写真領域では中央にピークをもつなだらかな山のような形になる。一方、文字領域(小)では、 $CDD1(l; m)$ の分散は大きな値をもつが、スキャナの分解能のため $CDD2(l; k)$ では2値的特性を得ることはできない。そのため、画素数の少ない領域では、 $CDD2(l; k)$ を考えない。

次に、これらの分布関数から写真、文字認識用のパラメータ  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  を定義した。特徴量  $C_1(l)$  は、 $CDD1(l; m)$  の分散であり、文字領域(大)、(小)の方が写真領域より大きな値をもつ。特徴量  $C_2(l)$  は、図6のように  $CDD2(l; k)$  を40%クリッピングした後での  $k_2$  を含むピークの幅であり、このクリッピングにより、文字領域(大)を2値化することが可能となる。ここで  $k_2$  は、 $CDD2(l; k)$  の最大値をとる  $k$  の値

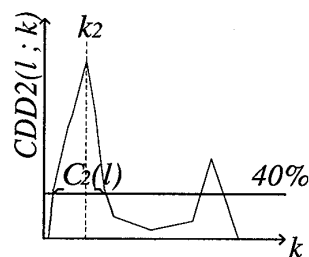


図6  $C_2(l)$  の定義  
Fig. 6 Definition of  $C_2(l)$ .

である。 $C_2(l)$  については、写真領域の方が文字領域より大きな値をもつ。特徴量  $C_3(l)$  は、 $CDD2(l; k)$  を用いて式(7)より定義され、文字領域が2値的な特性をもつことを考慮し、平均濃度から大きくかけ離れた濃度をもつ画素の数を強調したパラメータである。従って、 $C_3(l)$  は文字領域(大)の方が写真領域よりも大きな値となる。

$$C_3(l) = \sum_{k=1}^{63} \left| \left( \frac{k-32}{32} \right)^3 \right| \frac{CDD2(l; k)}{z(l)} \quad (7)$$

本手法では、以下に示す手順で写真領域と文字領域を認識した。画素数が1,000以上の領域に対しては、特徴量  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  を用いて写真か文字かを判断

し、画素数が1,000未満の領域に対しては、特徴量  $C_1(l)$  を用いて文字領域かそれ以外の領域かを判断(後に述べる写真領域拡張の処理のために必要)した。これは、画素数が小さくなると、スキャナの分解能の低さにより文字領域が2値的でなくなり、 $C_2(l)$ 、 $C_3(l)$  で正確に判断することが難しくなるからである。また逆に、スキャナの分解能から、画素数が小さくなるほど  $C_1(l)$  が大きい値をもつ可能性が高くなるという特性も利用した。以下に、領域判別のためのロジックを示す。ここで  $T_1 \sim T_8$  は、あらかじめ設定したしきい値である ( $T_1 > T_3$ ,  $T_2 > T_4$ ,  $T_7 < T_8$ )。

```

if ( $z(l) \geq 1000$ ) {
  if ( $C_1(l) > T_1$ ) then 領域  $l$  は文字領域である
  if ( $C_2(l) > T_2$ ) then 領域  $l$  は写真領域である
  if ( $C_1(l) > T_3$  or  $C_2(l) < T_4$  or  $C_3(l) > T_5$ 
    のうち二つ以上の条件を満たす)
    then 領域  $l$  は文字領域である
  else 領域  $l$  は写真領域である
}

```

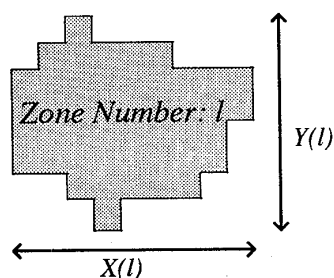


図7 領域  $l$  の幅  $X(l)$  と長さ  $Y(l)$   
Fig. 7 Zone width  $X(l)$  and length  $Y(l)$ .

```

}
else if ( $C_1(l) > T_6$ ) then 領域  $l$  は文字領域である
但し、画素数1,000未満の孤立した文字領域はノイズとした。また図7に示した番号  $l$  の領域の縦、横方向の長さ  $Y(l)$ ,  $X(l)$  を用い、以下の条件を満たす非背景領域を罫線として抽出した。

```

```

if (( $X(l) > T_7$  or  $Y(l) > T_7$ )
  and ( $|X(l) - Y(l)| > T_8$ ))
  then 領域  $l$  は罫線である

```

### 3.3 写真領域拡張法

写真領域と背景領域との境界部分の濃度差が小さいときに写真領域の一部が写真領域として抽出されない場合があるので、写真領域拡張の処理を行った。

まず、図8(a)に示された原画像に対して、3.1、3.2の処理により写真領域と文字領域とが図8(b)、(c)のように抽出される。写真領域拡張法では、原画像を  $8 \times 8$  画素の正方形の小領域に分割し、濃度分散がしきい値  $A$  (本手法では100に設定) 以上の領域に対して小領域内で2次微分値のヒストグラムを作成し、これに2値化のアルゴリズムを適用する。そして、求めたしきい値より高い2次微分値をもつ画素について今度は濃度ヒストグラムを作成し、再び2値化のアルゴリズムを用いてしきい値より低い濃度値をもつ画素を写真候補画素として抽出する(図8(d)参照)。ここで、しきい値  $A$  はノイズを拾いにくくするためのものであり、2値化のアルゴリズムを導入することで細かな濃淡部分に反応させることができた。そして、写真領域と文字領域との間の背景領域に写真候補画素が存在すれば、

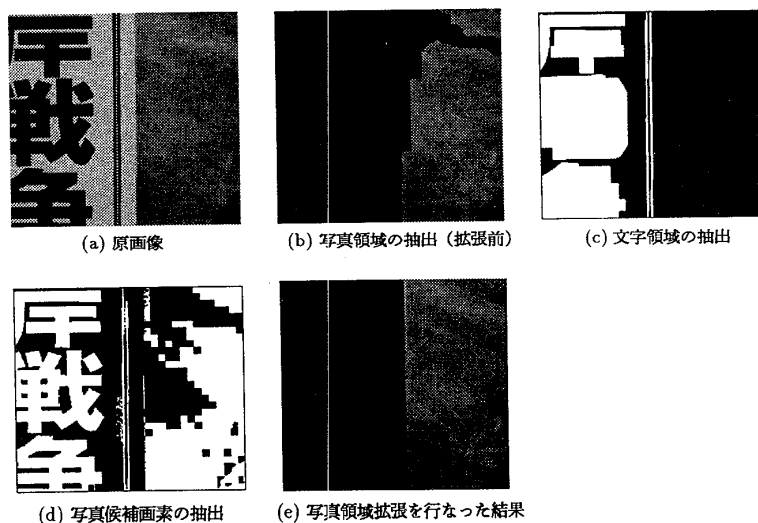


図8 写真領域の拡張(1)  
Fig. 8 Expansion of picture areas (1).

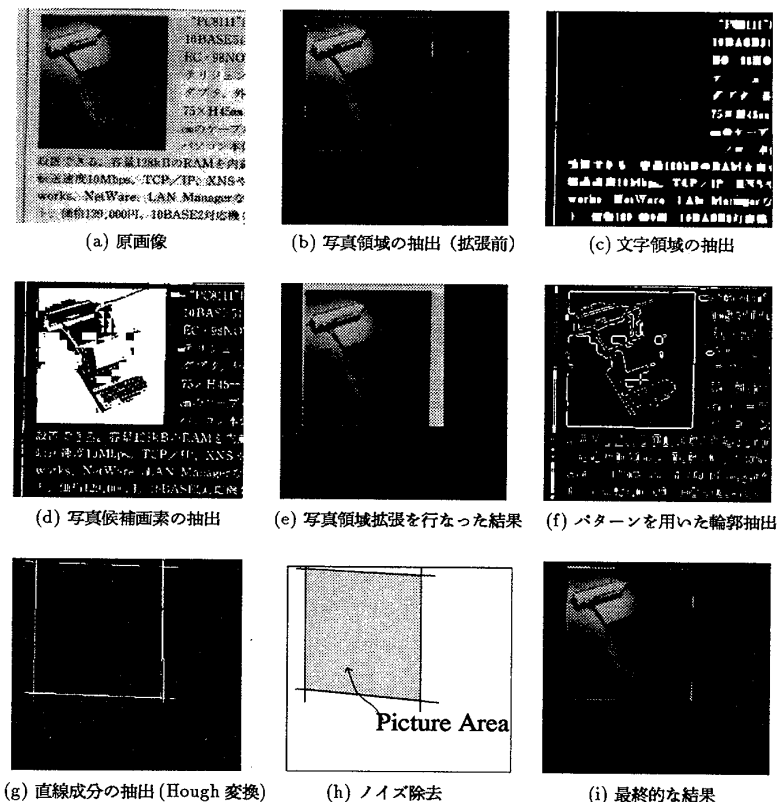


図9 写真領域の拡張(2)  
Fig. 9 Expansion of picture areas (2).

その地点まで写真領域を拡大する(図8(e)参照)。

また、図9(a)の画像では、写真候補画素抽出(図9(d)参照)の際に写真領域と文字領域との間の背景領域のノイズを拾うために、図9(e)のように必要以上に写真領域が拡大されてしまう。そのため、写真候補画素の2値画像上で図10のようなパターンと一致する箇所があれば、 $3 \times 3$ 領域内の中心画素をエッジとして抽出した(図9(f))後に、拡張された写真領域内でハフ変換<sup>(7)</sup>を行い、主走査方向に対し平行な直線、垂直な直線を抽出する(図9(g))。ハフ変換を用いることにより、とぎれている直線を連結することができ、また若干の傾きをもった直線も抽出することができる。そして、写真領域中にハフ変換によって抽出された水平方向、垂直方向の直線があり、その直線の外側にしきい値より少ない数の写真候補画素が存在するならば、直線の外側の領域を写真領域から削除する(図9(h)参照)。但し、写真領域から削除される領域に非背景画素が存在するならば、その画素は削除しない。ノイズ除去を行った最終的な結果を図9(i)に示す。

以上の処理を行うことにより、背景領域中に埋もれた写真領域の一部を抽出することができ、またノイズ

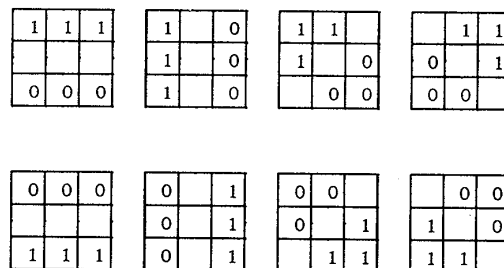


図10 エッジ抽出のために用いたパターン  
Fig. 10 Patterns used for edges detection.

を拾って写真領域が必要以上に拡張されるのを防ぐことができる。

#### 4. 結 果

本研究では、 $T_1=30$ ,  $T_2=20$ ,  $T_3=10$ ,  $T_4=10$ ,  $T_5=0.2$ ,  $T_6=30$ ,  $T_7=5$ ,  $T_8=30$  に設定して、写真領域を抽出した。これらの値は、多数の画像から特徴量  $C_1(I)$ ,  $C_2(I)$ ,  $C_3(I)$  を算出し、チューニングした値である。図11は異なる3種類のタイプの原画像に対して処理した結果であり、(a)写真領域の抽出、(b)文字領域の抽出、(c)写真領域拡張のための写真候補画

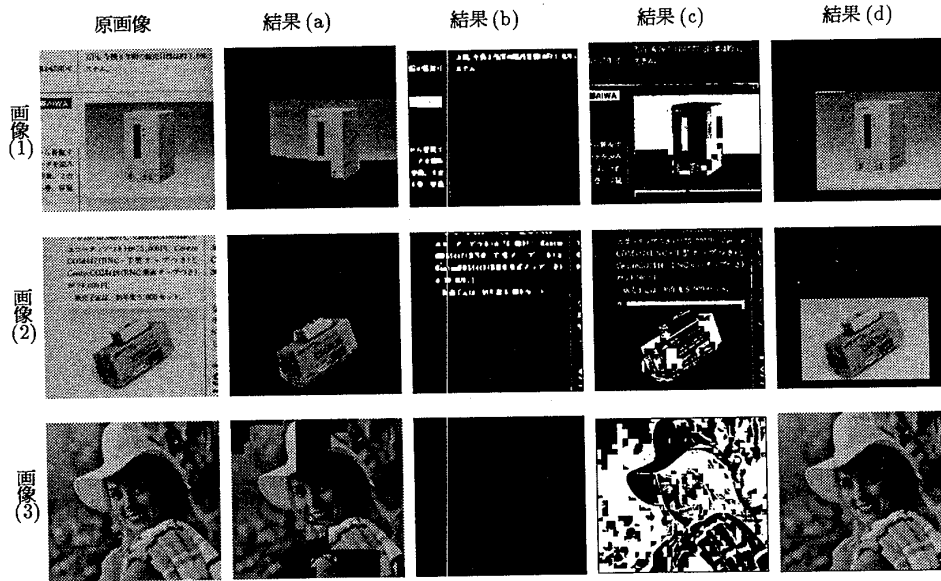


図 11 結果 (A)  
Fig. 11 Results (A).

素の抽出, (d)写真領域拡張法を適用し, 最終的に抽出された写真領域を示している. 図 11 の画像(1), (2) は背景領域との濃度差が小さい写真領域を含む画像, 図 11 の画像(3)は画像全体が写真領域となっている画像である. また, 前述した図 8 は画素数の大きな文字と背景領域との濃度差が小さい写真領域によって構成される画像, 図 9 は標準的な文書画像の例である. これらの結果から, 写真領域を正確に抽出できていることがわかる.

また, 本手法の実用性を確かめるために処理対象の領域を B5 サイズに拡大し, 電子情報通信学会論文誌 (網点線数 100), 法政大学工学部入学案内 (網点線数 150), 法政大学工学部新聞 (網点線数 100) からサンプル画像 20 枚ずつ合計 60 枚をデジタル化 (118 dpi) して実験を行った. これらの画像は, 性質の異なる三つのタイプの文書画像であり, 電子情報通信学会論文誌は学術雑誌の代表として, 法政大学工学部入学案内はカタログ誌の代表として, 法政大学工学部新聞は新聞の代表として用いた. この結果, 電子情報通信学会論文誌と法政大学工学部入学案内はすべて正確に写真領域を抽出でき, 法政大学工学部新聞は 20 枚中 18 枚について正確に抽出できた. 図 12 の画像(1)は罫線が存在する画像, 図 12 の画像(2)は線図形が存在する画像から写真領域を抽出した例である.



図 12 結果 (B)  
Fig. 12 Results (B).

## 5. 考 察

### 5.1 写真領域抽出アルゴリズム

ここでは, 写真領域抽出アルゴリズム, しきい値, 抽出結果の順に考察する.

写真領域抽出アルゴリズムにおける 2 値化のアルゴリズム (3.1) では, 固定しきい値を用いないため, さまざまな文書画像に対して背景と各構成要素を分離できた. また, 共起行列は写真領域では二つの隣接する画素間の濃度差が小さく, 文字領域ではそれが大きいと



いう特徴を非常によく表すため、共起行列から得られた三つの特徴量(3.2)によって非常に高い精度で写真と文字とを判別することができた。写真領域拡張法(3.3)は、背景との濃度差が小さい写真領域を含む画像に対して写真領域の一部が抽出されないという欠点を補うために用いた。また、この手法では円形や多角形の写真領域を含む画像に対しても、方形で領域抽出を行うために若干の背景領域を含むが正確に写真領域を抽出することができる。

次に、写真領域抽出のための判断基準となるしきい値に対して考察する。 $T_1 \sim T_6$ は、多数の画像から特徴量  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  を算出し設定した値であるが、中には写真領域に近い特徴量をもつ文字領域や、文字領域に近い特徴量をもつ写真領域があるので、単一しきい値のみで写真と文字とを完全に分離することができない。このしきい値のうち、 $T_1$ ,  $T_2$ ,  $T_6$  は明らかに文字あるいは写真である領域を抽出するために設定し、 $T_3$ ,  $T_4$ ,  $T_5$  はそれ以外の領域を7~8割ぐらいの確率で写真と文字とを分離できるように設定している。例えば、文字領域(大)が濃度値ヒストグラム上で近接した二つのピークをもつ場合は、 $C_1(l)$ の値が一般的な文字領域(大)のように典型的な大きな値をもたない。しかし、この場合、 $C_2(l)$ ,  $C_3(l)$  が明らかに文字であるという特徴量をもっていることが多い。これにより、本手法では特徴量  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  を組み合わせることにより認識率を高めることができた。

次に、文書画像から写真領域を抽出した結果について考察する。図9は、一般的な文書画像であるが、写真領域拡張における写真候補画素の抽出の際にノイズを拾うため、単純に写真領域の拡張を行うと、背景領域まで写真領域として抽出してしまう。しかし、ハフ変換を用いた画像の直線的なエッジ部分の抽出により、写真領域と背景領域の境界線を抽出することができ、写真領域が必要以上に拡張されるのを防ぐことができた。図11の画像(1), (2)と図8は、背景領域との濃度差が小さい写真領域を含む画像であり、結果(a)では写真領域の一部が欠けているが、写真領域拡張法を用いることで正確に写真領域が抽出できている。図11の画像(1)では、画像のほぼ中央にある横方向の罫線の付近にノイズがあるため、写真領域が罫線のある場所まで拡張されるおそれがあるが、ハフ変換を用いたエッジ抽出によってそれを防ぐことができた。図8では文字領域(大)と写真領域をうまく判別することができた。また、この画像中の写真は砂漠の写真で、一見すると

写真とわかりにくい、このような画像でも正確に写真領域を抽出することができた。2値化のアルゴリズムでは、画素を非背景画素と背景画素に分けるため、図11の画像(3)のような画像全体が写真領域である場合には、結果(a)では正確に写真領域を抽出することができない。しかし、写真領域拡張法を用いることによって、画像全体を写真領域として抽出することができた。

## 5.2 本手法の実用性および限界

まず最初に、B5サイズのサンプル画像を用いた実験について考察する。この実験では、3種類の異なった種類の画像を各20枚ずつ計60枚用いて写真領域を抽出した。そのうち、正確に写真領域が抽出されたのは58枚である。失敗した画像のうち1枚は、写真領域の一つの特徴量  $C_1(l)$ ,  $C_3(l)$  がしきい値より大きな値をもつために、その写真領域を抽出できなかったからである。他の1枚は、 $CDD2(l; k)$  のクリッピングで的確に2値に分けることができなかった文字領域が写真領域として抽出されたからである。現在対応できていない部分(背景に濃い色が着いている場合や黒色でない文字を含む場合、写真と罫線が隣接する場合や各構成要素が罫線によって囲まれる場合等)やしきい値の最適化等の改良は必要であるが、本手法が画像のサイズや文書の種類に依存せず実用的であると考えられる。

本研究では、SUN SPARC station2を用いて実験したが、処理時間は256×240の画像で約25秒かかっている(そのうち90%以上は写真領域拡張法)。これは、処理の各段階で画像表示やパラメータ値を出力するようにできているためである。

次に、本手法の限界および残された課題について述べる。現段階では、特徴量  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  に対し、しきい値  $T_1 \sim T_6$  を用いて写真か文字かを判断しているため、例えば、ある領域で特徴量  $C_1(l)$ ,  $C_2(l)$ ,  $C_3(l)$  の値がすべてしきい値付近にあり、しきい値のとり方によってはどちらとも考えられるような場合や、写真領域だけが文字のように特徴量  $C_2(l)$  が小さな値をもつ場合(カラーの文書をそのまま白黒で読み取ったときにまれに起こることがある)などの悪い条件では、写真と文字を誤認する場合がある。また、 $CDD2(l; k)$  のクリッピングで的確に2値に分けることができなかった文字領域が写真領域として抽出される場合もある。領域判断の基準となる領域の大きさについては、非背景画素を1,000以上の領域と1,000未満の領域に分けて考えているが、実際は万単位の画素数をもつ領域と100前後の画素数の領域を扱う場合がほとんどである。しか

し、認識の精度を向上させるために、今後は画素数が1,000前後の領域に対しても考えていく必要があると考える。

また、本手法では簡単な処理(特に共起行列を用いた特徴量による写真と文字との判別)を適用しているが、今後はしきい値の最適化や背景に濃い色が着いているような場合での処理等について考える必要がある。

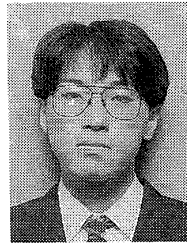
## 6. む す び

本研究では、写真と文字とが混在する画像から写真領域を抽出するために、2値化のアルゴリズム、共起行列を用いた特徴量、写真領域拡張法の三つの手法を組み合わせた認識アルゴリズムを開発した。この手法は、簡便かつ、高精度な認識が可能であるという特徴をもっており、通常の写真画像に対して高い精度で写真領域が抽出できた。また、B5サイズの文書画像を用いたシミュレーションにより、本手法が画像のサイズや文書の種類にほとんど依存せず、正確に写真領域を抽出できることが明らかとなった。但し、現時点では処理時間の短縮を十分に行っていないので処理時間がやや遅いが、今後改善する予定である。

## 文 献

- (1) 美濃導彦, 加藤俊一: “文書画像処理”, 別冊 Oplus E, pp. 158-171 (1986-11).
- (2) 大内 敏, 今尾 薫, 山田和作: “文字/絵柄(網点, 写真)混在画像の像域分離方式”, 信学論(D-II), J75-D-II, 1, pp. 39-47 (1992-01).
- (3) 安居院猛, 田村正司, 長橋 宏, 岡賢一郎, 大西 勝: “ハードコピー装置に適用する GBTC 型符号を用いた文字/自然画像の領域分割”, 信学'92 春大, D-449.
- (4) 王 大成, 美濃導彦, 坂井利之: “3×3 単位メッシュ内の濃淡パターンに基づく写真領域識別装置”, 情報処理学会第 27 回(昭和 58 年後期)全国大会, 3Q-7.
- (5) 村尾充洋, 坂井利之: “文書画像における構造情報の抽出”, 昭和 55 年度情報処理学会第 21 回全国大会 7H-1.
- (6) 長尾 真: “画像認識論”, pp. 38-45, コロナ社(1983).
- (7) 安居院猛, 中嶋正之, 木見尻秀子: “C 言語による画像処理”, pp. 136-144, 昭晃堂(1990).

(平成 5 年 7 月 2 日受付, 10 月 14 日再受付)



菅川賢太郎

平 4 法政大・工・電気卒。現在、同大学院修士課程在学中。画像処理などの研究に従事。



尾川 浩一

昭 55 慶大・工・電気卒。昭 57 同大学院修士課程了。同年同大医学部助手。平 2 同大・医・講師。平 3 法政大・工・助教授。医用画像処理、放射総計測の研究に従事。工博。平 1 日本医用画像工学会論文賞、平 3 日本核医学会賞各受賞。日本医学放射線学会、日本核医学会、IEEE、SNM 等各会員。