

## 解 説

## 人間の視覚機構に着想を得た手書き文字と活字文字の判別

小山 純平<sup>\*1</sup>, 加藤 雅弘<sup>\*2</sup>, 廣瀬 明<sup>\*1</sup>東京大学大学院工学系研究科電子工学専攻<sup>\*1</sup>, 富士ゼロックス株式会社<sup>\*2</sup>Method to Distinguish between Handwritten and Machine-printed Characters  
Inspired by Human Vision SystemJumpei Koyama,<sup>\*1</sup> Masahiro Kato<sup>\*2</sup> and Akira Hirose<sup>\*1</sup>Department of Electronic Engineering, The University of Tokyo<sup>\*1</sup>  
Fuji Xerox Co., Ltd.<sup>\*2</sup>

## 概要

手書き文字と活字文字の自動判別の実現は、画像処理分野の課題として興味深いだけでなく、光学文字認識装置における認識エンジンの使い分けに役立つことが期待されるなど、実用的側面でも大きく期待されている。著者らが近年提案しているスペクトル領域局所ゆらぎ検出 (Spectrum-domain local fluctuation detection, SDLFD) 法は、それを高い性能で実現する方法である。この SDLFD 法は従来法と異なり、文字や文字列の切り出しやテンプレートを必要としない方法である。本稿では、SDLFD 法の考え方と処理の概要を解説するとともに、その特長を紹介する。

## 1. はじめに

手書き文字と活字文字を判別することは、人間にとってはたやすい問題であるが、計算機でこれを実現することは依然として難しい。それは、手書き文字に含まれる手書きに起因する不確定なゆらぎが、文字のアスペクト比や文字列のベースラインのばらつきを生むためである。もし計算機による手書き文字と活字文字の判別が可能になれば、社会に大きく貢献できると考えられる。ひとつの例として、光学文字認識 (Optical character recognition, OCR) 装置の自動選択が挙げられる。手書き文字は文字の歪みや線分の傾きを含むため、活字文字用の OCR アルゴリズムとは異なるアルゴリズムを使用しなければ高い認識率を得ることができない。そのため現状では人間が画像中の文字を手書き文字か活字文字か判別し、適切な OCR エンジンを選択する必要がある。計算機で手書き文字と活字文字を判別する技術は判別から認識までの一連の作業を自動化し、労働コストの削減を実現する。また一般に、手書き文字と活字文字は文書中では異なる役割を担う。納税申告用紙を例に挙げると、活字文字は属性を、手

書き文字はそれに対する具体的な値である属性値を表す。手書き文字と活字文字の判別は、事前知識を必要とせずに属性値のみを抽出することを可能とするため、紙文書のデータベース化など様々な用途が考えられる。

このように手書き文字と活字文字の判別には様々な工学的応用が考えられるため、多くの研究が行われてきた。従来の手書き文字と活字文字の判別法は文字画像をそのまま実空間で解析する手法が主体であった。例えば、Fan らは文字の水平・鉛直方向のヒストグラムを取ることで文字列を抽出し、その基線の直線度を評価する手法を提案した<sup>1)</sup>。Kuhnke らは、文字線分の直線性と文字の対称性を評価し判別を行った<sup>2)</sup>。Pal と Chaudhuri はベンガル文字とデーヴァナーガリー文字が持つ特徴的な線分に注目した判別手法を提案した<sup>3)</sup>。また、手書き文字と活字文字の判別に類似した研究として、文書中に使用されている文字のフォントを判別する研究がある。フォント判別には主に文字装飾の特徴を比較する手法<sup>4,5)</sup>が提案されている。これらの手法では、まず前処理として文字または文字列の位置を特定しておく必要がある。そのため、文字または文字列の抽出精度がその後の処理に大きな影響を与える。しかし手書き文字を含む実際の文書画像の場合、文字の大きさのばらつきや文字列の基線のゆらぎによ

\*1 〒 113-8656 東京都文京区本郷 7-3-1

\*2 〒 259-0157 神奈川県足柄上郡中井町境 430

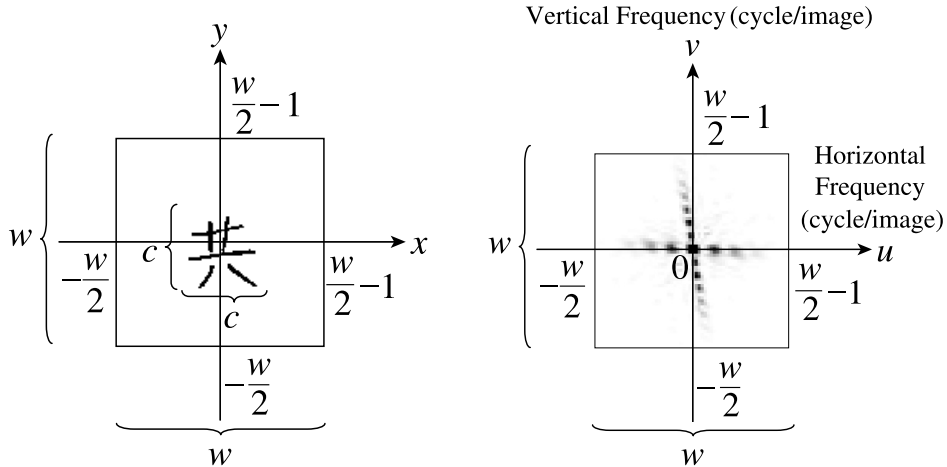


Fig. 1 Example of a local region  $I(x, y)$  in a document image and its power spectrum  $P(u, v)$ .

り、そもそも文字や文字列を抽出すること自体が非常に困難になる。

われわれは文字や文字列の位置特定を必要としない手書き文字と活字文字の判別手法を提案するため、人間の素晴らしい視覚機構に注目してきた。人間の視覚は、文字や文字列を切り出すことなく、瞬間的に、文字種に関係なく、雑音に強い手書き文字と活字文字の判別を行うことができる。この高度な判別は物体のテクスチャ情報を基に行われているという報告がある<sup>6-8)</sup>。これらの研究によると、まず人間の視覚は光刺激により形成される空間周波数特性を一次視野で知覚する。そしてその情報をより高次の視野で処理し、テクスチャの判別を行う。さらに Julesz は、人間の視覚が二次統計量が等しいテクスチャを判別できないことを示した<sup>9)</sup>。これは人間の視覚が空間周波数特性という基礎的な情報を基にテクスチャ判別を行うことを示唆している。

多くの工学的テクスチャ判別手法は上記の知見を基にしている<sup>10-12)</sup>。これらの手法は、画像を周波数領域に変換した後、フィルタリングを行うことで特徴量を取得。この際、離散コサイン変換やウェーブレット変換、ガボールフィルタリング<sup>13,14)</sup>などの周波数解析手段が用いられる。そして得られた特徴量を基にフィッシャー判別を行いテクスチャを判別する。これらのテクスチャ判別手法は、手書き文字と活字文字の判別<sup>15)</sup>や、フォント判別<sup>16-18)</sup>、文字の書き手判別<sup>19)</sup>など、様々な文書画像解析に応用されており、高い判別率を達成した。しかし、文書画像中の広い領域を判別単位とする必要があったため、一文字程度の局所領域の判別が

できないという欠点が存在した。

そこでわれわれは、局所領域も扱うことができ、さらに文字や文字列の位置を特定する前処理を必要としないスペクトル領域局所ゆらぎ検出 (Spectrum-domain local fluctuation detection, SDLFD) 法を提案した<sup>20-22)</sup>。SDLFD 法では文書画像中の局所領域を周波数領域に変換した後に手書きのゆらぎを含む特徴量を抽出する。その特徴量を、事前の学習によって最適化された多層パーセプトロン (Multilayer perceptron, MLP) に与えることで、われわれは手書き文字らしさを得る。その手書き文字らしさを基に手書き文字と活字文字の判別を行う。本稿の以下の章ではわれわれの研究成果を紹介する。まず第 2 章に SDLFD 法の処理を解説する。次に第 3 章で具体的な判別の様子とその性能を紹介する。そして第 4 章に、SDLFD 法の今後の展開も含め、まとめを行う。

## 2. スペクトル領域局所ゆらぎ検出法と多層パーセプトロン

本章ではわれわれの提案する SDLFD 法による文書画像からの特徴量抽出と、手書き文字と活字文字を判別する多層パーセプトロンについて解説をする。

### 2.1 特徴抽出

スペクトル領域ゆらぎ検出法では、はじめに実空間文書画像の局所領域  $I(x, y)$  を周波数領域に変換する。Fig. 1 左に  $w \times w$  サイズの実空間画像の一例を示す。われわれは  $I(x, y)$  に式 (1) に示す二次元離散フーリエ変換を施し、フーリエスペクトル  $F(u, v)$  を得る。

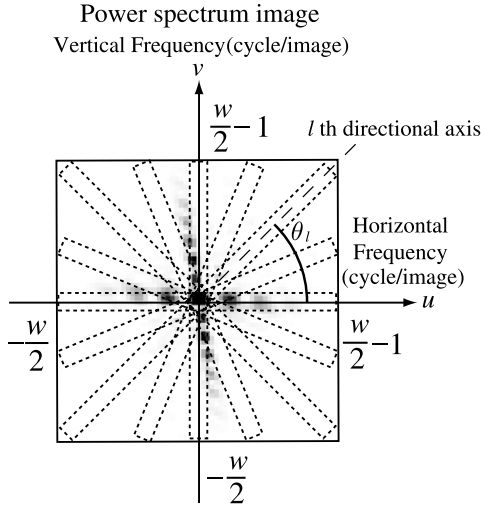


Fig.2 Spectrum-domain local fluctuation detection method. Every dotted rectangle is a power summation region in  $\theta_l$  direction.

$$F(u, v) = \sum_{x=-\frac{w}{2}}^{\frac{w}{2}-1} \sum_{y=-\frac{w}{2}}^{\frac{w}{2}-1} I(x, y) H(x, y) \cdot \exp \left\{ -2\pi j \left( \frac{x}{w} u + \frac{y}{w} v \right) \right\} \quad (1)$$

ここで、 $j = \sqrt{-1}$ である。また  $H(x, y)$  は窓関数を表す。本稿ではハン窓を用いた。実空間画像を周波数領域に変換する手法にウェーブレット変換や離散コサイン変換などを使用することもできる。さらにフーリエスペクトル  $F(u, v)$  から Fig. 1 右に示すようなパワースペクトルを得る。

$$P(u, v) = |F(u, v)|^2 \quad (2)$$

2次元周波数座標の原点  $(u, v)$  は直流成分を表す。周波数領域では、文字線分の方位と文字線分の幅や間隔はそれぞれ極座標系でいうところの偏角と動径で表現される。そのため、Fig. 1 右に見られるように、文字線分は周波数領域で大きいパワーを持つ点列を誘起する。われわれは便宜上それを基軸と呼ぶことにする。

次にパワースペクトル  $P(u, v)$  から手書きに起因するゆらぎを含む特徴量を定量化する。Fig. 2 の点線で描かれた領域に示すように、われわれは原点を中心とし放射状に一定間隔の軸をとり、その軸上のパワーの和を計算する。このパワーの和は文字線分の方位を良く表す。われわれはその軸の方位  $\theta_l$  上のパワーを特徴量  $f_l$  として抽出する。

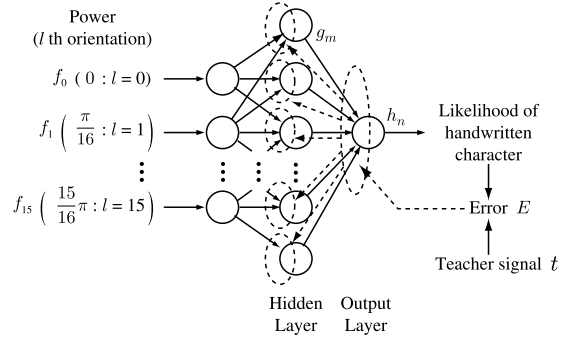


Fig.3 2-layer perceptron.

$$f_l = \frac{\sum_{|v - \tan(\theta_l)u| \leq 1} P(u, v)}{\sum_u \sum_v P(u, v)} \quad (3)$$

本稿では、特徴量を検出する軸の方位の数を  $l = 16$  とした。

人間のテキスト判別は空間周波数特性という基礎的な情報で実現されている。テキストを良くあらわす周波数領域で実空間画像を扱うことで、われわれは人間の視覚システムのような判別が可能になると考えた。そして文書画像のテキストは文字線分により構成される。手書き文字の場合、文字線分が持つ手書きによるゆらぎのため、その基軸は傾きやぼやけを持つ。パワーを計算する軸の間隔を細かくすることで、それらを検出することが可能になる。

前述のように、SDLFD 法は文書画像の局所領域のパワースペクトルから手書き文字のゆらぎを含む特徴量を抽出する。これには2つの利点がある。まず、われわれは文字や文字列の位置を特定する必要がなくなる。なぜならばパワースペクトルでは位相が無視されるためである。2つめに、局所領域を評価するため、従来手法では不可能であった一文字程度の領域の手書き文字と活字文字の判別が可能になる。

### 2.2 多層パーセプトロン

われわれは 2.1 章で得られた特徴量を基に、ニューラルネットワークを用い手書き文字と活字文字の判別を行う。Fig. 3 に実際に使用する2層パーセプトロンを示す。この多層パーセプトロンは特徴量  $f_l$  を入力として受け取り、手書きらしさを表す値  $h_n$  を出力する。本稿ではニューロンの数を、中間層では 64、出力層では 1 とする。つまり、われわれは 1 つの手書きらしさ  $h_1$  を出力に得る。学習法には誤差逆伝搬法を利

Table 1 Mean square error (MSE) and correct-distinction rate of the optimized multilayer perceptron in each letter type class.

Letter type class	Correct-distinction rate (%)	MSE ( $\times 10^{-2}$ )
Alphanumeric	99.70	0.387
Hiragana	98.75	0.550
Katakana	99.75	0.366
Chinese character	97.45	1.31
All letter type	97.00	1.26

用する。誤差逆伝搬法は多層パーセプトロンのための教師あり学習法である。われわれはそれぞれの入力信号に対して、その理想的な出力を表す教師信号  $t$  を用意する。特徴量  $f_i$  を入力し手書きらしさを得る順方向処理を行った後に、われわれは出力信号  $h_1$  と教師信号  $t$  の誤差  $E$  を計算する。その誤差  $E$  が小さくなるように、出力側に存在するニューロンの結合荷重から順にその値を更新していく。教師信号  $t$  は手書き文字の場合 0.9, 活字文字の場合 0.1 と設定した。学習は平均二乗誤差 (Mean square error, MSE) があらかじめ決めておいたしきい値  $\epsilon = 1.0 \times 10^{-6}$  よりも小さくなるか、試行回数がその最大値  $T = 1000$  に到達した場合に打ち切られる。

### 3. 具体的な判別処理の様子とその性能

ここでは SDLFD 法による判別実験の様子を 2 つご紹介する。まず、文字が無作為に配置され、レイアウト解析が困難な文書画像の判別実験を行った。この実験により、提案手法が文字や文字列の切り出し処理によらないことを示すことができる。また同時に、十分な判別が可能な解析窓のインターバルの大きさを実験的に示す。次に、われわれは提案手法の実用状況を考え、様々な解像度でスキャンされた文書画像の判別実験を行った例も紹介する。

#### 3.1 文字が無作為に配置された文書画像の判別

提案手法が文字の切り出し処理によらないことを示すため、われわれは Fig. 4 に示すような文字を無作為に配置した文書画像中の文字の判別を行った。Fig. 4 のサイズは  $640 \times 480$  ピクセルであり、画像は二値化されている。画像中に存在する文字は、文字データベースから無作為に選択され、さらに文字が列を成さないよう、また文字間隔が一定にならないように配置されている。そのため、この文書画像からヒストグラムを計算し文字や文字列の切り出しを行うことは難しい。

##### 3.1.1 多層パーセプトロンの最適化

まずわれわれは判別に使用する多層パーセプトロン

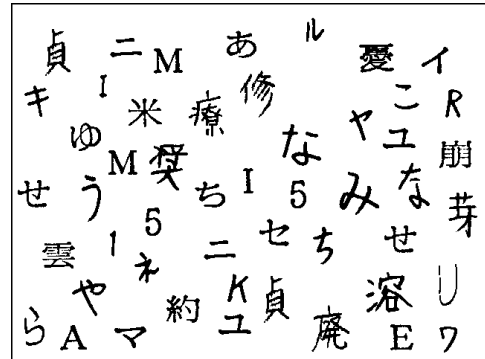


Fig. 4 A document image which has randomly arranged characters.

を最適化した。最適化のために、産業技術総合研究所で公開されている ETL 文字データベースの文字画像を利用した<sup>23)</sup>。このデータベースは手書きと活字あわせて約 120 万の文字画像を持ち、文字認識実験などで広く使用されている。これらの画像サイズ  $w$  と文字サイズ  $c$  はそれぞれ  $(w, c) = (128, 64)$  である。また、文字画像はすべて二値化されている。

ETL 文字データベースのなかの ETL1, ETL2, ETL9 を使用し、‘英数字’、‘ひらがな’、‘カタカナ’、‘漢字’、‘すべて’ という 5 つクラスの学習用データセットを作成した。‘英数字’、‘ひらがな’、‘カタカナ’、‘漢字’は、ETL 文字データベースからそのクラスに属する手書き文字を 1000 サンプル、活字文字を 1000 サンプルずつランダムに選択した。‘すべて’クラスでは、ETL 文字データベースから‘英数字’、‘ひらがな’、‘カタカナ’、‘漢字’に属する文字を、手書き文字は 500 サンプル、活字文字は 500 サンプルだけランダムに選出したため、そのサンプル数は合計 4000 サンプルである。また、同様の作成方法で、それぞれのクラスの判別用データセットも作成した。データセット作成の際にはランダムに文字を選択しているため、学習と判別、それぞれのデータセットに含まれる文字は全く異

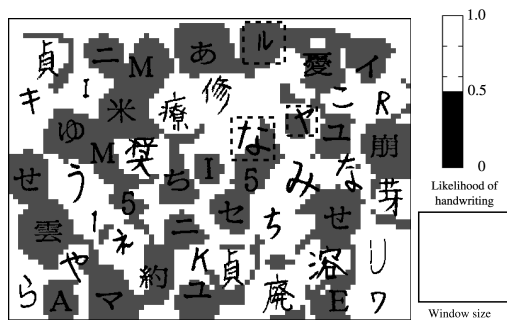


Fig. 5 Binarized handwriting likelihood map with window interval  $d = 8$ . The threshold value for binarization is 0.5. Characters in dashed-line frames are misclassified.

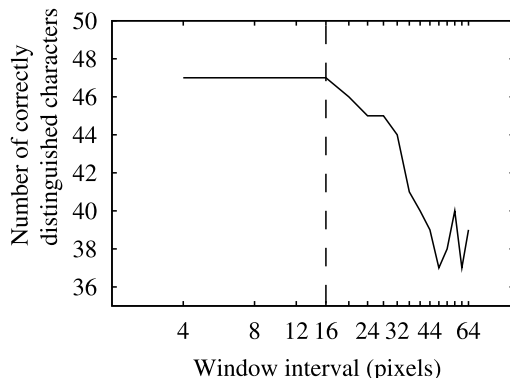


Fig. 6 Relationship between the window interval  $d$  and number of correctly distinguished characters.

なる。

まず、われわれは5つの学習データセットを用い、学習率  $\eta$ 、シグモイド関数の急峻さを決定する  $\beta$  について最適化し、次いで最適化された多層パーセプトロンを用い、判別データセットの文字画像判別実験を行った。文字画像を局所領域と見なし、SDLFD法を用い得られた手書きらしさについて、しきい値を0.5として、しきい値以上ならば手書き文字、しきい値未満ならば活字文字であると判別した。

Table 1 にパラメタを最適化したときの判別率と学習終了時のMSEをそれぞれの文字種クラスについて記す。‘カタカナ’で最も大きい判別率99.75%を、‘漢字’で最も小さい判別率97.45%を得た。カタカナや英数字で判別率が高いのは、文字構造が比較的単純であり、さらにひらがなに多く含まれるような曲線を持つ文字が少ないためだと考えられる。逆に漢字の判別率が最も低いのは、漢字は複雑な構造を持つものが多数存在し、手書きに起因するゆらぎがその複雑さの中に紛れてしまうためだと考えられる。切り出し処理を必要とする従来の手法では、文書画像中の英数字の手書き文字と活字文字の判別で98.1%の判別率が得られており<sup>15)</sup>、SDLFD法はそれと比較しても遜色のない判別率を得ることができた。また、曲線を多く持つ‘ひらがな’や、様々な文字種から構成される‘すべて’で97%の判別率を得た。従来研究は英数字のみ、漢字のみというように文字種を限定して判別を行うものが主であるため、多種多様な文字を含めこの判別率が得られたことは特筆すべきことである。

### 3.1.2 判別手法

文字画像の学習は第3.1.1節で既に行われているため、その学習結果のうち、‘すべて’で最適化された多

層パーセプトロンを使用した。われわれは文書画像上で、解析窓をインターバル  $d$  でずらしながら、特徴量を抽出し、事前に最適化された多層パーセプトロンによる判別を行った。そして本実験ではインターバル  $d$  を様々に変更して、文字の正解判別数の変化を観察した。ETL文字データベースの文字を使用しているため、文書画像中の文字サイズ  $c_p$  は64ピクセルである。処理対象となる解析窓のサイズ  $w_p$  は128ピクセルとする。これは学習時のETL文字データベースの文字サイズ  $c_l$  と窓サイズ  $w_l$  と同じである。ここで添え字の  $p, l$  はそれぞれ判別時 (processing), 学習時 (learning) であることを示す。なお、処理対象とする文書中の局所領域に文字線分が全く存在しない場合は、二次元離散フーリエ変換から多層パーセプトロンによる判別までの一連の処理を行わない。

### 3.1.3 判別結果

Fig. 5に、解析窓をずらす大きさを  $d = 8$  にしたときの判別結果を示す。この判別結果は、文書画像サンプルに多層パーセプトロンから得られる手書き文字らしさを、しきい値を0.5とし二値化したマップを重ね合わせることで作成された。手書き文字らしさは、0.0に近いとき活字文字、1.0に近いとき手書き文字と判別されたことを表す。文字ごとの簡単な定量的評価を行うために、それぞれの文字矩形の70%以上が正しく判別されていたら、その文字は正確に判別されているものと考えられる。するとFig. 5の破線のフレームで囲まれた文字のみが誤って判別された。この文書は50文字含むため、その中の47文字、つまり94%の文字が正しく判別されたことになる。文字や文字列を切り出すことなくこれほどの判別率を得られた判別手法は従来

Table 2 Mean square error (MSE) and correct-distinction rate of the optimized multilayer perceptrons with datasets which have machine-printed character images in MS Gothic font.

Letter type class	Correct-distinction rate (%)	MSE ( $\times 10^{-2}$ )
Alphanumeric	99.9	1.59
Hiragana	98.6	1.69
Katakana	99.8	1.87
Chinese	96.5	1.35
All letter type	97	2.23

## 和谐社会 和谐社会

精华:民主法治、公平正义、诚信友爱、充满活力、安定有序、人与自然和谐相处的社会。  
 解读:民主法治、公平正义、诚信友爱、充满活力、安定有序、人与自然和谐相处的社会。  
 就是社会主义民主得到充分发扬, 依法治国基本方略得到切实落实, 各方面积极因素得到广泛调动。  
 公平正义, 就是社会各方面的利益关系得到妥善协调, 人民内部  
 背景: 十六届四中全会  
 画出了一幅清晰的和谐社会“蓝图”

Fig. 7 A sample document image.

研究に存在せず, この結果は非常に有用なものである。

Fig. 6 にインターバル  $d$  を変更したときの正しく判別された文字数を示す。インターバル  $d$  が 16 ピクセル以下のときに 47 文字を正しく判別することができた。この結果をふまえると, 十分な正解判別数が得られる時のインターバル  $d$  と画像中の文字サイズ  $c_p$  の関係は次のようであると推察される。

$$d \lesssim \frac{c_p}{4}$$

### 3.2 様々な解像度でスキャンされた文書画像の判別

文書画像の解像度は重要な要素である。なぜならば, 企業活動においてはひとつの紙文書が異なる解像度でスキャンされる場合が多々存在し, 判別手法は様々な解像度に対応する必要があるためである。われわれは Fig. 7 に示すような紙文書を用意し, 単一の解像度の文字画像で最適化された多層パーセプトロンを用い, 異なる解像度の文書画像の判別を行った。そして高い判別率を得ることができる解像度  $r$  と解析窓サイズ  $w$  の関係を実験的に示唆した。

#### 3.2.1 多層パーセプトロンの最適化

Fig. 7 に示される文書中の活字文字は MS Gothic 体に形状が似ている中国語のフォントからなる。そこでわれわれは MS Gothic フォントの文字画像を作成し, ETL データベースの活字文字と入れ替えることで,

新たな学習データセットを作成した。このとき, 手書き文字画像のサンプルは ETL 文字データベースの画像サンプルを利用し, サンプルの作成法やサンプル数も第 3.1.1 項で説明したものと同様にした。従ってサンプル数は, 文字種のクラス‘英数字’, ‘ひらがな’, ‘カタカナ’, ‘漢字’では 2000 サンプル, ‘すべて’は 4000 サンプルの手書き文字と活字文字画像を持つ。文字画像のサイズ  $w_l$  は 128 ピクセル, 画像中の文字サイズ  $c_l$  は 64 ピクセルである。

Table 2 に, MS Gothic を活字文字とする新しい学習データセットで多層パーセプトロンのパラメータを最適化したときの平均二乗誤差 (MSE) と判別率 (Correct-distinction rate) を示す。

#### 3.2.2 判別手法

文書画像サンプルの判別実験にはクラス‘漢字’の学習結果を用いた。文書画像の判別手法は, 第 3.1.2 項で説明した手法と同様である。Table 3 に示される各解像度  $r$  とそれに伴う文字サイズ  $c_p$ , 解析窓のサイズ  $w_p$ , そして解析窓のインターバル  $d$  の条件で判別を行った。

#### 3.2.3 判別結果

Fig. 8 に Table 3 のそれぞれの条件下で得られた手書き文字と活字文字の判別結果を示す。判別結果は手書き文字らしさを閾値 0.5 とし二値化されている。暗く表示された領域は活字文字, 明るく表示された領域が手書き文字と判別された領域である。提案手法により, 解像度 150 dpi (Set A) と 300 dpi (Set C), 600 dpi (Set D) の場合, 手書き文字と活字文字が上手く判別されている。いくつかの手書き文字は誤って活字文字と判別されているが, それらは斜線または曲線のみを持つ漢字である (「人」など)。このことは, 多層パーセプトロンが文字線分の水平・鉛直性を判別のための重要な情報として扱っていることを示唆している。

解像度の異なる 4 種類の文書画像を比較すると, 解像度 150 dpi (Set A), 300 dpi (Set C), 600 dpi (Set

Table 3 Parameters of datasets A, B, C and D.

Label	Resolution $r$	Character size $c_p$	Analysis window size $w_p$	Interval $d$
Set A	150	32	64	8
Set B	200	43	64	8
Set C	300	64	128	16
Set D	600	128	256	32

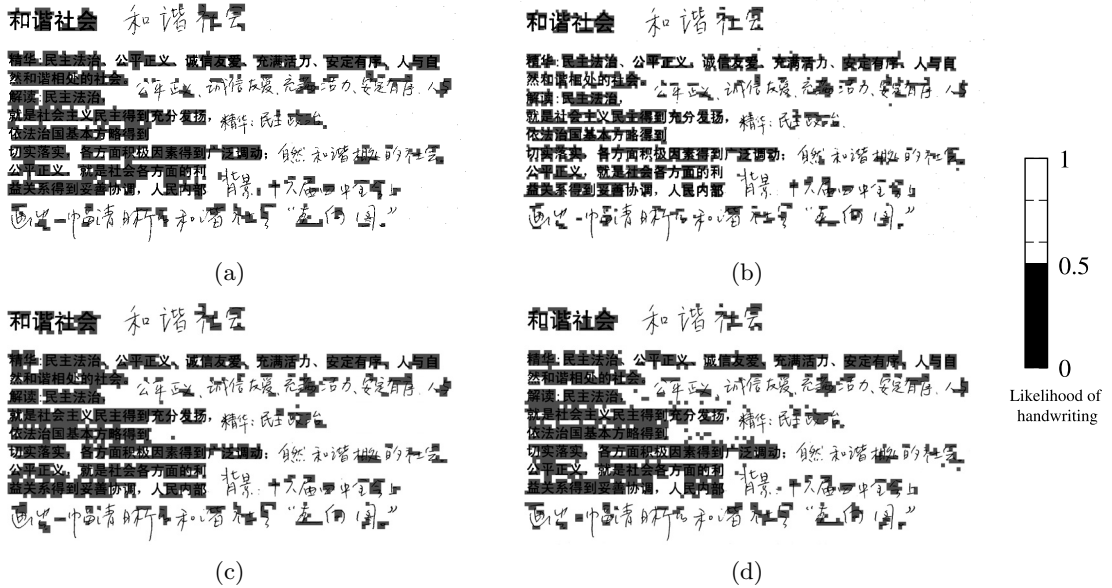


Fig. 8 Binarized handwritten likelihood maps of the document images scanned in various resolutions. Threshold of binarization is 0.5. (a) Set A, (b) Set B, (c) Set C, (d) Set D.

D) の画像は判別結果が似通っている。それらとは異なり、解像度 200 dpi (Set B) の画像のみ若干判別数が低下しているように見られる。この結果は、解析時の文書画像中の解析窓のサイズ  $w_p$  と文字サイズ  $c_p$  の比が、文字画像サンプル学習時の比とやや異なるために起こると考えられる。なぜなら、解像度が 150 dpi (Set A), 300 dpi (Set C), 600 dpi (Set D) の文書画像の場合、解析時の文書画像中の解析窓のサイズ  $w_p$  と文字サイズ  $c_p$  の比  $w_p/c_p$  が学習時の解析窓サイズ  $w_l$  と文字画像サイズ  $c_l$  の比  $w_l/c_l$  と同じ 2 であるのに対し、解像度が 200 dpi (Set B) の場合 1.5 であり、やや異なるからである。これらの結果より、大きい正解判別数を得るための解析窓サイズ  $w_p$  と文書解像度  $r$  によって決まる文字画像サイズ  $c_p$  には次の条件が存在することが示唆される。

$$\frac{w_l}{c_l} \approx \frac{w_p}{c_p} \quad (4)$$

その後、詳しい実験により、(4) で示される緩やかな条

件が存在することが明らかになっている。その実験内容と解析結果の詳細は、現在、論文誌に投稿中である。

#### 4. まとめ

本稿では、文字や文字列の位置を特定する必要のない文書画像中の手書き文字と活字文字の判別を目的としてわれわれが提案してきた、スペクトル領域局所ゆらぎ検出法 (Spectrum-domain local fluctuation detection, SDLFD) 法を紹介した。この手法は人間の視覚機構に習い、文書画像中の局所領域のパワースペクトルから手書きに起因するゆらぎを含む特徴量を検出し、それを基に多層パーセプトロンで判別を行う。適用例として、まずわれわれはレイアウト解析が困難な文書画像に対し判別実験を行い、文字や文字列の切り出し処理を必要としない判別が可能であることを示した。そして解析窓サイズについて実験的に解析を行い、十分な判別率を得ることができる条件を述べた。またわれわれは SDLFD 法の解像度に対する耐性を解析し

た結果を紹介した。ここでは、同一の紙文書を様々な解像度でスキャンし得られた文書画像を、単一解像度の文字画像を学習することで得られた多層パーセプトロンで判別した。実験結果より、良好な判別結果を得るための解像度と窓サイズの条件を述べた。今後、提案手法を OCR システムの中に組み込んで、実用化を図る予定である。

SDLFD 法を提案するにあたり、われわれは人間の手書き文字と活字文字の判別に注目した。例えば、人間はたった一本の線分であってもその形状からそれが手書き文字であるか活字文字であるかを判断できるだろう。また、全く知らない言語の文字種であっても、それが手書き文字であるか活字文字であるか瞬時的に判別できる。人間の視覚機構では、光刺激は網膜から第一次視覚野で処理され、より高次の視覚野へと伝達されてゆく。文書画像を処理する場合、第一次視覚野は文字線分の傾きに反応し、パルス信号を発する。人間はその信号を高次視覚野で統合することで、線分の直線性や曲線の曲率を感じ取り、文字や文字列を認識し、また文書の空間構造を理解している。

一方、現在の工学分野における文書画像処理は、そのほとんどが空間テンプレートを利用し、細かいルールを適用するものである。なぜならば、現在の画像処理は、他の論理処理と同様に、記号処理をベースにしているからである。特に近年、コンピュータが格段に速くなりメモリ容量が増大したため、そのような処理方法の有用性が増している。しかし、それは人間の処理方法のうち、意識上のもののみに対応する。瞬時に手書きか活字か区別できる能力は、意識下のものである。それはパターン処理に基づいており、上に述べた脳の視覚刺激の処理に対応する。今後は、これらの意識上と意識下の各処理を融合した工学システムも検討されるべきだと、われわれは考えている。すなわち、記号処理とパターン処理を融合した手法の実現である。その優れたお手本が、われわれの脳である。

### 参 考 文 献

- 1) Fan, K.C., Wang, L.S., and Tu, Y.T. (1998): Classification of machine-printed and hand-written texts using character block layout variance, *Pattern Recognition*, Vol.31, No.9, pp.1275-1284
- 2) Kuhnke, K., Simoncini, L., and Kovács-V, Zs. M. (1995): A system for machine-written and hand-written character distinction, *Proc. IEEE Int. Conf. on Document Analysis and Recognition*, Vol.2, pp.811-814
- 3) Pal, U. and Chaudhuri, B.B. (2001): Machine-printed and hand-written text lines identification, *Pattern Recognition Letters*, Vol.22, pp.431-441
- 4) Khoubyari, S. and Hull, J.J. (1996): Font and function word identification in document recognition, *Computer Vision and Image Understanding*, Vol.63, No.1, pp.66-74
- 5) Zramdini, A. and Ingold, R. (1998): Optical font recognition using typographical features, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.8, pp.877-882
- 6) Nothdurft, H.C. (1985): Sensitivity for structure gradient in texture discrimination tasks, *Vision Research*, Vol.25, No.12, pp.1957-1968
- 7) Kingdom, F.A.A., Keeble, D., and Moulden, B. (1995): Sensitivity to orientation modulation in micropattern-based textures, *Vision Research*, Vol.35, pp.79-91
- 8) Motoyoshi, I. and Nishida, S. (2002): Spatiotemporal interactions in detection of texture orientation modulations, *Vision Research*, Vol.42, pp.2829-2841
- 9) Julesz, B. (1965): Texture and visual perception, *Scientific American*, Vol.212, pp.38-48
- 10) Papathomas, T.V., Kashi, R.S., and Gorea, A. (1997): A human vision based computational model for chromatic texture segregation, *IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics*, Vol.27, No.3, pp.428-440
- 11) Wouwer, G.V., Scheunders, P., and Dyck, D.V. (1999): Statistical texture characterization from discrete wavelet representations, *IEEE Trans. on Image Processing*, Vol.8, No.4, pp.592-598
- 12) Han, J. and Ma, K.K. (2007): Rotation-invariant and scale-invariant Gabor features for texture image retrieval, *Image and Vision Computing*, Vol.25, pp.1474-1481
- 13) Jones, J.P. and Palmer, L. (1987): An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, *J. Neurophysiology*, Vol.58, No.6, pp.1233-1258
- 14) Daugman, J.G. (1985): Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Opt. Soc. Am. A*, Vol.2, No.7, pp.1160-1169
- 15) Zheng, Y., Li, H., and Doermann, D. (2004): Machine printed text and handwriting identification in noisy document images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.26, No.3, pp.337-353



- 16) Zhu, Y., Tan, T., and Wang, Y. (2001): Font recognition based on global texture analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.23, No.10, pp.1192–1200
- 17) Tan, T.N. (1998): Rotation invariant texture features and their use in automatic script identification, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.7, pp.751–756
- 18) Morris, R.A. (1992): Classification of digital typefaces using spectral signatures, *Pattern Recognition*, Vol.25, No.8, pp.869–876
- 19) Busch, A., Boles, W.W., and Sridharan, S. (2005): Texture for script identification, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.27, No.11, pp.1720–1732
- 20) Koyama, J., Kato, M., and Hirose, A. (2007): Handwritten character distinction method inspired by human vision mechanism, *Proc. Int. Conf. on Neural Information Processing (ICONIP) 2007 Kitakyushu, FMF-5*
- 21) Koyama, J., Kato, M., and Hirose, A. (2008): Distinction between handwritten and machine-printed characters without extracting characters or text lines, *Proc. IEEE World Congress on Computational Intelligence (WCCI) 2008 Hong Kong, NN1075*
- 22) Koyama, J., Kato, M., and Hirose, A. (2008): Local-spectrum-based distinction between handwritten and machine-printed characters, *Proc. IEEE Int. Conf. on Image Processing (ICIP) 2008 San Diego (to be presented)*
- 23) ETL character databases, *Advanced industrial science and technology*, <http://www.is.aist.go.jp/etlcdb/>