

検出容易な文字パターン

内田 誠一[†] 服部 亮史[†] 岩村 雅一^{††} 大町真一郎^{†††} 黄瀬 浩一^{††}

[†]九州大学大学院システム情報科学研究所 〒819-0395 福岡市西区元岡 744

^{††}大阪府立大学大学院工学研究科 〒599-8531 大阪府堺市学園町 1-1

^{†††}東北大学大学院工学研究科 〒980-8579 仙台市青葉区荒巻字青葉 6-6-05

E-mail: [†]uchida@is.kyushu-u.ac.jp, ^{†††}{masa,kise}@cs.osakafu-u.ac.jp, ^{††††}machi@ecei.tohoku.ac.jp

あらまし 環境画像からの文字領域検出(切り出し処理)に関して, 検出が容易となる文字パターンの合成について検討する. 文字領域検出には, 例えば文字パターンが一般的に持つ複雑性などの普遍的な特徴が用いられていた. しかし用いた特徴では検出できない例外的なパターンが存在したり, 検出対象が存在する環境によっては検出が極めて困難になるなど問題も多い. 本研究では, 検出処理に用いる特徴を備えた文字パターンを合成するという視点からこの問題を捉える. 具体的には, 文字パターンの形状とある検出器による特徴抽出結果の関係を明らかにし, その検出器における処理に最適な文字パターン形状を定量的な評価により求めるというアプローチを用いる. この一例として, 部分空間法を用いた文字量器検出に適した文字パターンを合成し, そのパターンの検出処理における有効性を確認した. キーワード 文字認識, 文字切り出し, 文字合成, 主成分分析

Designing Conspicuous Character Pattern

Seiichi UCHIDA[†], Ryoji HATTORI[†], Masakazu IWAMURA^{††}, Shinichiro OMACHI^{†††}, and Koichi KISE^{††}

[†] Faculty of Information Science and Electrical Engineering, Kyushu University, 819-0395 Japan

^{††} Graduate School of Engineering, Osaka Prefecture University, 599-8581 Japan

^{†††} Graduate School of Engineering, Tohoku University, 980-8579 Japan

E-mail: [†]uchida@is.kyushu-u.ac.jp, ^{†††}{masa,kise}@cs.osakafu-u.ac.jp, ^{††††}machi@ecei.tohoku.ac.jp

Abstract Detection of characters in scenery images is often a very hard problem. Although many researchers have tackled this difficult problem and achieved a good performance, it is still difficult to suppress many false alarms and/or missings. This paper investigates a conspicuous character pattern, which is a special pattern designed for easier detection. In order to have an example of the conspicuous character pattern, we select a character font with a larger distance from a non-character pattern distribution and, simultaneously, with a smaller distance from a character pattern distribution. Experimental results showed that the character font selected by this method is actually more conspicuous (i.e., detected more easily) than other fonts.

Key words Character recognition, Character detection, Character synthesis, Principal component analysis

1. はじめに

デジタルカメラなどで取得された実環境画像について, その画像中の文字を認識する技術が検討されている [1]. デジタルカメラを文字認識器への情報入力手段として用いることは, その携帯性や操作性, 自由度の面から有用であると考えられる. 一方, 実環境画像中の文字は低解像度であり, 照明条件が一定せず, 更に幾何変形を受けている場合があるため, 高精度な認識を実現するには, 今後様々なブレークスルーが必要とされる.

ところで, 認識処理の前には, 文字領域の検出・切り

出し処理が必要となる. ある意味, 認識処理よりも難しいこの検出処理には, 例えば文字パターンの複雑性などの特徴が用いられてきた [2] ~ [6]. しかしどのような特徴を用いても必ず例外的なパターンは存在しうる. また環境内には “Y” や “I” 状のエッジが多数存在し, これらと真の文字パターンとの識別も相当困難な問題である.

本論文では, 検出容易な文字パターンを生成することで, この困難な文字検出問題の抜本的解決を目指す. これは, 環境中にある一般的な文字パターンを努力して検出するという従来のアプローチとは, 逆の発想に基づく. QR コードは機械可読だけでなく検出容易なパターン

としてもデザインされているが、まさにそうした性質を持った文字パターンを生成する試みとも言える。

計算機にとって取り扱いやすいようにあらかじめ文字パターンに工夫を施すという考え方は、1960年代に開発されたOCR・MICRフォント[7]をはじめ、最近でも議論されている[8],[9]。これらはいずれも認識を容易とするための文字パターンの生成法に関する。これに対し本論文では、検出容易となるように文字パターンを工夫する方法について論ずる。

検出容易な文字パターン生成のアプローチには、(a) 人間による試行錯誤的な生成法、ならびに (b) 工学的的方法論によるシステムティックな生成法、の二つが考えられる。前者の利点として、生成プロセスが常に人間の監視下にあるために、人間にとって違和感の無い文字パターンが得られることが挙げられる。一方で欠点として、自動的すなわち工学的に実施される検出処理にとって、生成された文字パターンが本当に高い検出性能を持つか不明確という点が挙げられる。後者の利点・欠点は、ちょうど前者の正反対になる。すなわち、想定する検出処理にとって真に検出容易な文字を生成できるという利点と、生成された文字が人間に自然という保証が無いという欠点がある。

本論文では、検出の際の利点を評価し、(b) のアプローチを採用する。その結果、上述のように生成されたパターンの自然さについての問題が残るが、そもそもこれは感性的・主観的な問題である。実際、(a) のアプローチを採ったとしても、ある人による生成パターンが万人に受け入れられるという保証はない。それよりも、まずは (b) の枠組みの方法論を確立し、検出容易な文字の存在を実証するべきと考える。その上で必要ならば、この方法を道具として用いながら、(a) を実施することも考えられよう。

ところで以下の議論からも明らかになるように、(b) のアプローチを採ったとしても、具体的な文字生成の方法には、ほとんど無限とよってよい選択肢がある。そのことは、そもそも文字の検出方法が無限にあることから理解できる。このような様々な選択肢の中、本論文では最も一般的な考え方の一つとして、検出容易な文字パターンを「検出処理に用いる特徴空間において、文字分布内に存在し、かつ非文字分布から最も離れて存在するパターン」として定義する。

対象検出は、文字パターンに限らず、画像認識・理解の根本的な問題の一つである。例外的に顔検出が高い性能を得るに至っているが、それ以外のより一般的な物体検出においては、まだ未解決な部分が多い。これはそもそも物体のカテゴリをどのように既定するのかという問題と深く関わっている。もし物体のバリエーションが簡単なモデルで既定できれば、その検出は比較的容易であろうが、実際には、例えばライオンのバリエーションをモデル化するのは非常に難しい。

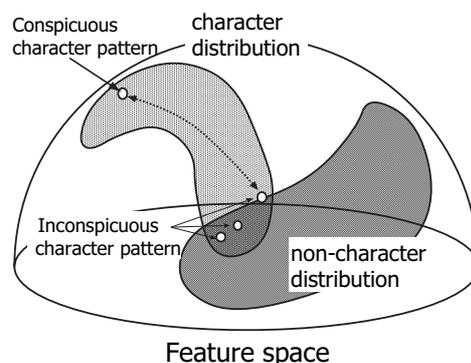


図1 検出容易な文字パターンの定義

こうした一般的な対象検出と文字検出の最大の違いは、文字は人工物であり、自由にデザインすることが可能であるという点である。すなわちライオンを検出容易とすべく加工することはできないが、文字パターンならばそれができる。このように文字という扱い易い対象を用いて対象検出の問題に取り組むことにより、より複雑な対象すなわち一般的な物体の検出問題にも何らかの知見・方法論を与えるのではないかと期待している。現在のパターン認識技術の多くが、文字認識研究の技術という事実も、この期待を高めてくれる。

2. 検出容易な文字パターン生成の基本方針

2.1 検出容易な文字パターンの定義

上述のように、検出容易な文字パターンを、「検出処理に用いる特徴空間において、文字分布内に存在し、かつ非文字分布から最も離れて存在するパターン」と定義する。この定義を図1に示す。同図の半球はノルム1のベクトルの集合が成す超球である。すなわち以下ではパターンを表す特徴ベクトルはノルム1に正規化されているものとする。同図中の検出容易な文字パターン(“conspicuous character pattern”)であれば、非文字との混同が少なく、高精度に検出できると思われる。このパターンは、文字分布から離れないという制約の下で非文字分布からの距離を最大化する問題の解として得られる。もしくは文字分布からの距離を最小化しながら、非文字分布からの距離の最大化する問題の解として与えられる。

この定義には確定すべき要素が多い。すなわち、(i) 文字分布、(ii) 非文字分布、(iii) 特徴空間、(iv) 分布からの距離尺度、について具体的に設定しなければ、検出可能な文字パターンを生成できない。別の言い方をすれば、これら4要素の設定の違いによって、生成される文字パターンも変わってくる。以下では、4要素それぞれの選び方について詳述する。

2.2 検出容易な文字パターンに関する4つの要素

2.3 文字分布

文字分布は、検出容易な文字パターン生成の際の学習

パターンとして機能する。文字分布として最も単純なものは、様々なフォント画像を収集して作成した文字画像集合である。実環境中で起こり得る様々な変動に頑健な文字パターンを生成したい場合、それらの変動を受けた文字画像も文字画像集合に含める必要もあると考えられる。

2.4 非文字分布

非文字分布も同様に、学習パターンとして機能する。「非文字とは何か」をどのように規定するかによって、様々なバリエーションが考えられる。要するに文字集合の補集合を考えればよいが、それは困難である。そのため、実際には数多くの非文字画像集試料を収集し、それから経験主義的に非文字画像集合を定めることになる。また、何らかの文字検出器を想定し、それにより誤検出されてしまった非文字画像のみを収集して用いることもできる。文字検出の際に問題になるのはそうした“きわどい”非文字画像であり、それらの利用は本手法において効果的であると考えられる。

2.5 特徴空間

特徴空間、すなわち検出のための特徴は、当然ながら非文字分布と文字分布を明確に分離できるものが望ましい。一つの例として色特徴を考えてみる。非文字空間は自然界に存在する色が為す空間なので、それから最も離れた(すなわち検出容易な)文字パターンは、自然界には存在しにくい色を持つものとなる。従って、検出容易な文字パターンは、例えばショッキングピンクで彩色されたものとなる。確かに、情景画像からショッキングピンクの領域を探すのは容易であり、検出されたものが文字である可能性も高いと思われる。実際、店舗ではアイキャッチャーとしてそうした蛍光色の文字が用いられることも多く、意識的・無意識的に検出容易な文字パターンが利用されている好例と言えよう。ただし、色特徴で全てが解決するわけではない。実際には蛍光色が使えるような場合は極めて限定されるためである。

そこで本論文では、文字の形状に着目した特徴を用いることを考えることとする。ただし、一言で文字形状特徴と言っても実に様々なものがある。線分幅が一定というような特徴から、局所的複雑性といった統計的な特徴などがあり、これらは実際、従来の文字検出法が着目していた特徴でもある。いずれの特徴を用いれば、より検出容易な文字パターンが生成されるかは、興味深い検討課題となる。

一つの重要な点は、これまでの文字検出法で利用されてきた文字形状特徴が最適とは限らない点である。従来の文字形状特徴は、文字パターン全般を対象としたものであった。これに対し、今回の試みでは、検出容易な文字パターンさえ検出できればよいので、文字形状特徴に対する要求も異なったものとなる。すなわち検出対象を

限定できる分、特に誤検出を極力排除できるような特徴を利用することも考えられよう。例えば、従来用いられていた局所的複雑性は、文字の多くを検出できる反面、木の葉のような細かいテクスチャにも反応し、誤検出を生んでいた。これは、木の葉に似たテクスチャを持つ文字までも対象としているために生じた誤検出とも言える。これに対し、検出容易な文字パターンを(ある特徴空間において)木の葉との混同が少ないものとして生成できれば、こうした誤検出は排除できる。このような特徴を見出すことは、大きな課題の一つである。

2.6 距離尺度

距離尺度、すなわち“文字らしさ”の尺度を定義することは、文字と非文字の識別境界を定義することに等しいと言える。従って距離尺度は特徴空間内でのパターンの分布の形状に応じて定める必要がある。この点において、前項の特徴の定義とも密接に関係する。

この距離尺度は文字検出ならびに文字生成の両方において用いられる。文字検出については、情景画像中の任意領域について、この距離尺度を計算し、その結果に対する閾値処理により、その領域が文字であるか否かを決定する。これが本手法が前提とする文字検出法であり、本手法で求める検出容易な文字とは、この検出法によって検出容易な文字となる。

文字生成の観点から、この距離尺度設定には注意が必要である。すなわち、非文字分布からの遠さだけを重視するような距離尺度では、文字分布から遠く外れたパターンを生成してしまうといった状況が起こり得る。このことは、人間可読性が著しく低い文字パターンが生成される可能性を意味している。この状況を回避するには、「非文字から遠い」ではなく、「非文字から遠くかつ文字に近い」という尺度を考えればよい。この尺度を用いれば、文字分布内に存在し、かつ非文字から遠いような、検出容易なパターンが生成されると考えられる。

3. 具体的な例

このように様々な考えられる問題設定のうち、本論文では一例として部分空間法を用いた生成手法に関して検討する。具体的には、文字分布および非文字分布を学習パターンから求められる部分空間で表現し、パターン生成に利用する。また、距離尺度として、文字部分空間からの距離と非文字部分空間からの距離両方を用いることでパターンの“文字らしさ”を評価し、文字らしいパターンを求める。

具体的には以下のように問題を設定する。

- (i) 文字分布: フォント画像集合が成す部分空間
- (ii) 非文字分布: 背景画像集合が成す部分空間
- (iii) 特徴空間: 局所方向ヒストグラム特徴 [10]
- (iv) 距離尺度: 文字分布、非文字分布それぞれの部分空間との為す角の比

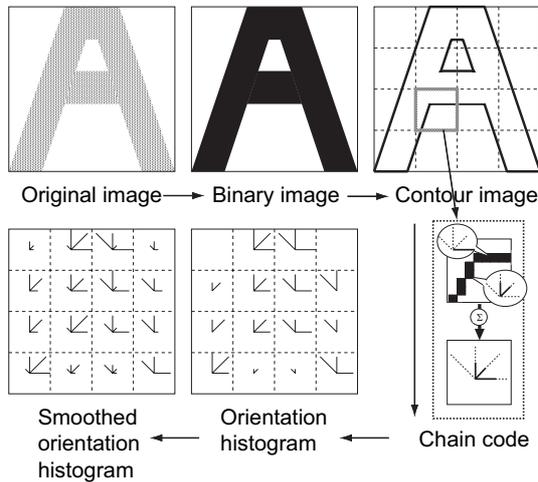


図2 局所方向ヒストグラム特徴

この問題設定に対する解空間，すなわちどのような文字を生成するかについても，実際には様々なバリエーションが考えられる．本論文では，その最も単純な形態として，文字分布として用いたフォント画像の集合から検出容易なパターンを“選出”することとする．フォント形状モデルなどを用いたパターンの生成については，今後の課題とする．

3.1 文字分布と非文字分布

今回の例では，文字分布をフォント画像集合の線形部分空間により定義する．従って，フォント画像の集め方によって，文字分布は変化することになる．後述の実験では，308種類のフォントセットにより，この文字分布を定義している．用途によってはより限定的なものとしてもよいし，より多くのフォントセットを集めてもよいだろう．部分空間の次元数については，累積寄与率に基づいて決定すればよい．

非文字分布も非文字画像集合の線形部分空間により定義する．真の非文字の分布をより適切に表現するため，文字以外の様々画像を多数集めて非文字画像集合とする^(注1)．

3.2 局所方向ヒストグラム特徴による特徴空間

文字形状特徴として，図2に示す局所方向ヒストグラム特徴[10]を用いる^(注2)．これは文字パターン認識に適

(注1): 2.4で述べたように，特定の文字検出法を想定し，それにより誤検出されたパターンだけを集めた“きわどい”非文字集合を作ることとも考えられる．実際，そうした非文字集合を用いて実験も行ったが，結論から言えば，今回の問題設定では，“きわどい”非文字に限らずに集合を生成したほうが検出性能が高かった．詳細は略するが，その理由は用いた距離尺度の性質から説明可能である．

(注2): もっと単純に，濃淡特徴を直接用いることも考えられる．しかし実際に実験を行ったところ，その性能は高くなかった．詳細は略するが，この場合，非文字部分空間は理論的に離散コサイン変換の低周波基底が成す空間となる[11]．その結果，高周波成分をなるべく多く持つ文字が選ばれるが，実際にはそこまで高周波成分を持った文字は少なく，従って文字分布と非文字分布の分離度が悪くなった．

した特徴として良く知られている．文字パターンの輝度値や色の変化に無関係であり，従って本論文で生成される検出容易な文字パターンは，どのように彩色されても，その検出性能にはほとんど影響しないことになる．また，文字認識に広く用いられているということは，それだけ様々な文字形状を表現するのに適しており，文字検出問題における様々なノイズに対する耐性も示唆する．その一方で，検出対象を限定できるという状況を積極的に活用した特徴にはなっていない点には注意が必要である．

図2に示すように，ある正方形領域の画像 X に対する局所方向ヒストグラム特徴は以下の手順で求められる．

(1) 画像 X を2値化する．2値化の手法としては，輝度値の変化に対して安定な結果を得られる大津の手法を用いた[12]．なお，文字パターン集合すなわちフォント集合は元々2値パターンであるため，この処理は不要である．

(2) 得られた2値画像を 4×4 の小領域に分割し，各領域内でエッジのヒストグラムを作成する．具体的には領域内の各エッジ点の方向を，水平方向，垂直方向，右上り斜め方向，左上り斜め方向の4方向に量子化し，各方向をピンとしたヒストグラムを作成する．この結果，特徴量は $4 \times 4 \times 4$ の64次元ベクトルとなる．

(3) 全ての小領域でヒストグラムを作成した後，各領域内のヒストグラム間の平滑化を行う．これは画像内のパターンの位置ずれの影響を軽減するための処理である．

(4) 最後に特徴ベクトルのノルムを1に正規化する．

3.3 部分空間を用いた距離尺度

正方形画像 X の文字らしさ D_N を， X が非文字部分空間と成す角(正準角) θ_N と， X が文字部分空間と成す角 θ_C から求める．具体的には，

$$D_N = \begin{cases} 0, & \text{if } \theta_N < \Theta, \\ \theta_N / (\theta_N + \theta_C), & \text{otherwise} \end{cases} \quad (1)$$

と定義する．ここで Θ は閾値であり，その導入意図については後述する．

この文字らしさ D_N は，非文字分布からより遠くかつ文字分布により近いほど大きい値を取るという所期の性質を持つ．単純に θ_N と θ_C の比を取った場合と性質はほとんど同じであるが，値域が $0 \leq D_N \leq 1$ となるため扱いやすいという利点がある．

非文字と文字の部分空間は部分的に重なっている場合があるため， X が非文字画像であったとしても，場合によっては θ_N と θ_C の両方が小さくなり，結果として D_N が大きくなってしまいう状況が有りうる．こうした状況を避けるべく， θ_N が閾値 Θ を下回る場合は D_N を強制的に0にする(すなわち非文字判定とする)という処理を加えている．ただし $0 \leq \Theta \leq \pi/2$ である．

以上は文字らしさの定義であったが，同時に本手法が

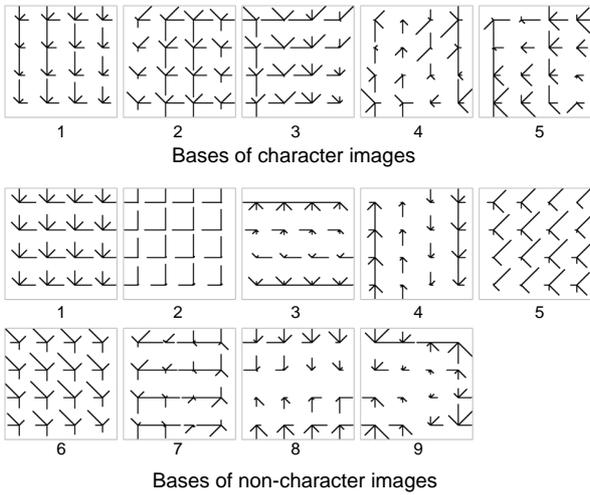


図3 文字分布および非文字分布の部分空間の上位基底

想定する文字検出法も規定する．すなわち，画像 X を情景画像中の正方形領域として，その X について求めた D_N がある閾値 ϵ 以上であれば，その領域 X を文字として検出する．文字らしさ D_N を基準として導出される検出容易な文字パターンとは，この文字検出法にとって容易なパターンとなる．

3.4 検出容易な文字パターンの選出法

以上で述べた問題設定において，検出容易な文字パターンの生成法について述べる．本節冒頭でも述べたように，本論文では，パターン生成の最も単純な形態として，文字分布として用いたフォント画像の集合から最も検出容易なパターンを“選出”することとする．すなわち，フォント画像集合の中から D_N が最も大きくなるものを選出する処理を持って生成に代える．

具体的には，全ての文字フォント画像を X として，特徴抽出及び正準角計算を行い，最も D_N が大きくなったフォントを，検出容易な文字パターンとして選出すればよい．本論文では，この選出をカテゴリ毎に独立に行った．従って，カテゴリ毎に違うフォントの文字が選ばれる．無論，全カテゴリで同一のフォントを選択してもよい．この場合，各フォントについて，全カテゴリ分の D_N の総和を計算し，その値が大きくなったフォントを選出すればよい．このほうが現実的であるが，今回は手法の性質を詳細に吟味するべく，カテゴリ毎の選出処理を採用した．

4. 実験

前節の方針による検出容易な文字パターンの選出を行った．以下では選出の結果，及び選出された文字パターンが本当に検出容易であるかどうかを評価した実験の結果を示す．

4.1 実験条件

文字分布を構成するために，308種の英語大文字フォ

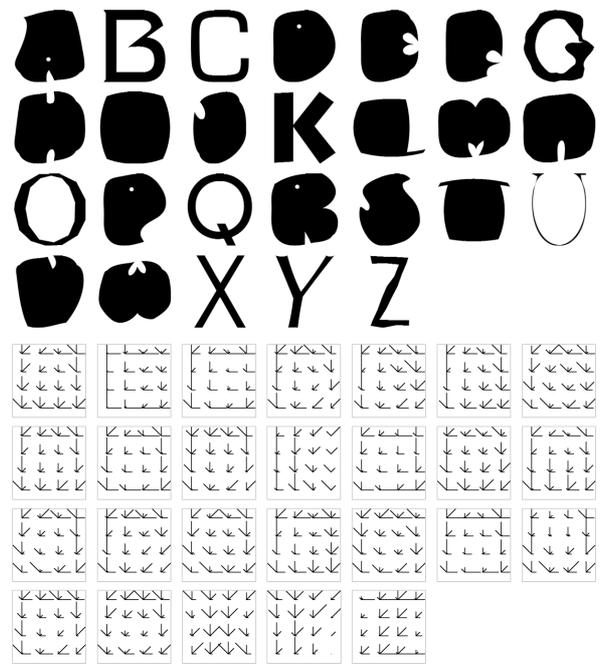


図4 選出された検出容易な文字パターン

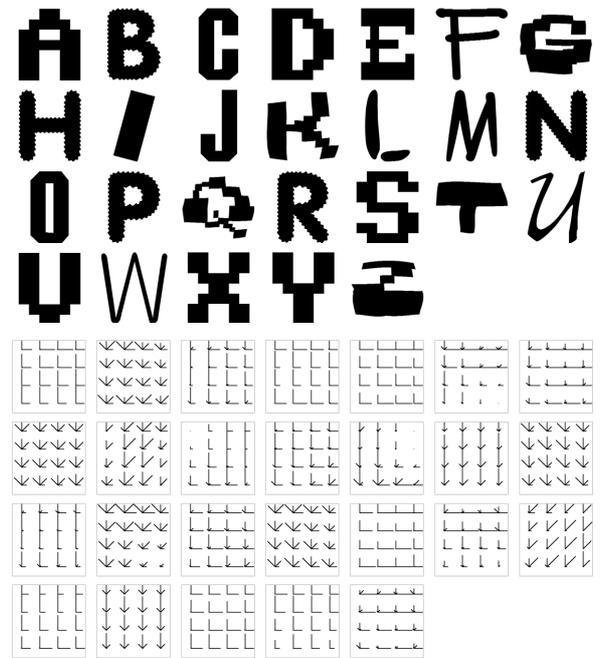


図5 検出困難な文字パターン

ント画像 (全部で $308 \times 26 \sim 8,000$ 枚) を用いた．これらのフォントは，各フォントファイルのヘッダに記されている PANOSE と呼ばれる値を参考に選択されたものである．PANOSE はフォントの形状を表す 10 桁の数値であり，1 桁目はフォントが属する大まかなカテゴリを表す．今回の実験では一般文書に用いられるフォントのカテゴリである“Latin text”に分類されているものを用いた．ただし，この PANOSE の値はあくまで主観で定められているようで，一般文書にはほとんど見られないようなフォントも混入していた．一方，非文字分布を構



図 6 検出容易な順に並べた結果

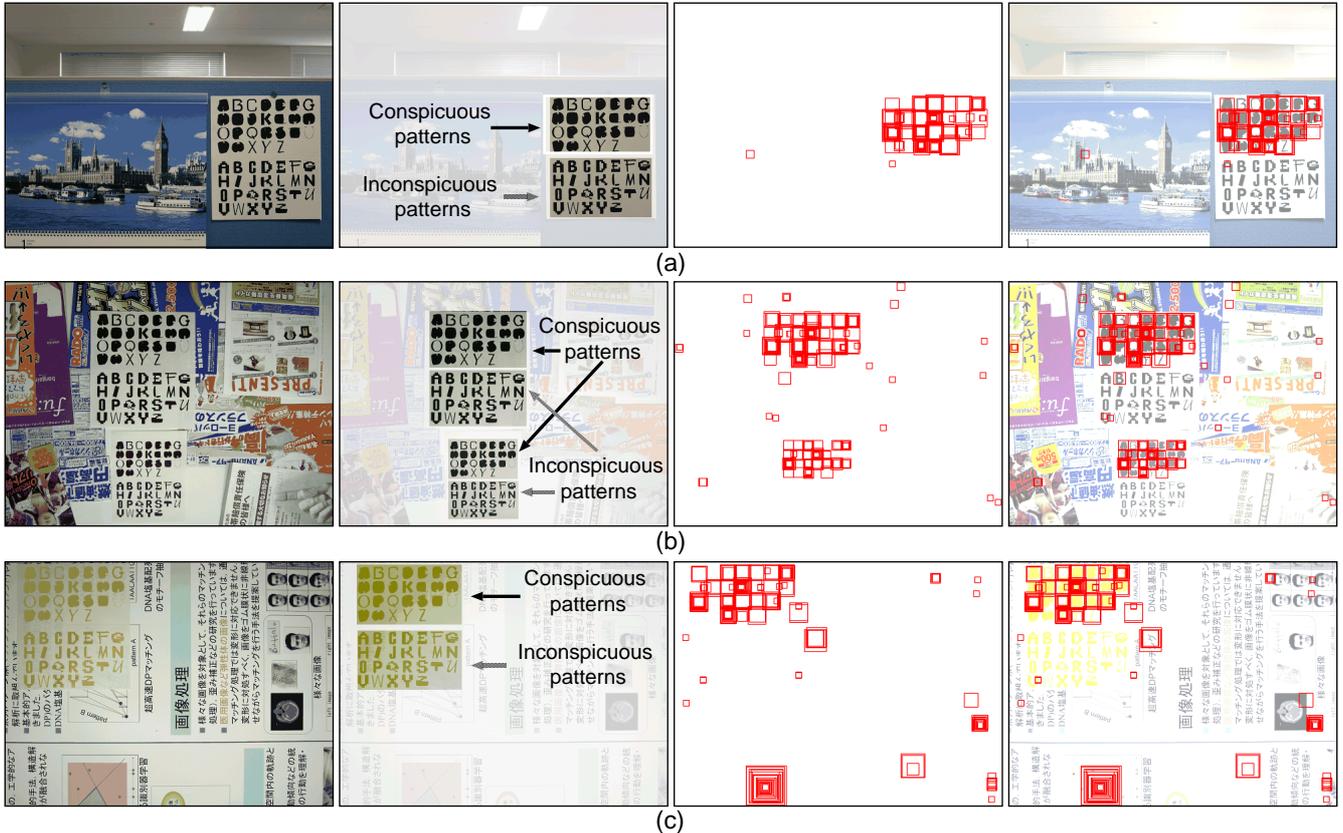


図 7 検出結果

成するために、一般物体認識用の情景画像データベース (Caltech background database) から 64×64 画素の部分領域を切り出した約 8 万枚の画像を準備した。

文字と非文字、それぞれの集合を局所方向ヒストグラム特徴の空間上で主成分分析し、部分空間を張る基底を求めた。基底は累積寄与率が 90% を超えるように上位から順に選んだ。その結果、文字部分空間は上位 5 基底、非文字部分空間は上位 9 基底を用いることとなった。図 3 はそれら基底ベクトルを可視化したものである。また文字らしさ D_N の評価の際に用いる閾値 θ は実験的に 0.4 と定めた。

4.2 選出結果

本手法により選出された検出容易なパターンおよびその特徴量を図 4 に示す。全体的に丸い形状の文字が選択されていることがわかる。このようなパターンが選ばれ

たのは、直感的には、自然界にはこうした丸い形状が存在しにくいということから理解できる。より詳細には、図 3 の非文字基底を見ることで理解できる。すなわち、丸い形状の文字は画像の周辺部に偏った方向成分を持ち、これは非文字分布の第 1 基底から離れた特徴である。また様々な向きの方向成分を持ち、この特徴は同じく第 2, 5 基底から離れている。更に画像周辺部を囲うような方向成分は第 3, 4 基底と離れた特徴である。これらの作用により、非文字部分空間とのなす角 θ_N が大きくなり、 D_N が大きくなったと考えられる。

本手法の考え方をを用いれば、検出困難なパターンを求めることもできる。図 5 は、 D_N が小さいパターンであり、非文字分布に近い検出困難なパターンである。非文字部分空間の上位基底に類似した、画像全体に均一な方向成分を持つようなパターンが選ばれていることがわかる。

図 6 は、308 種のフォントを各カテゴリ内で D_N が大きい順に並べたものである。2 番目、3 番目のフォントも、やはり上で述べたような非文字基底から離れた特徴を持つようなパターンになっている。つまり、検出容易なパターンは図 4 で示したフォントだけではなく、これらに類似の傾向を持つパターンであればよいと言える。

4.3 検出実験結果

生成された検出容易なパターン (図 4) を環境中に配置して撮影し、検出実験を行った。比較対象として検出困難なパターン (図 5) も同一画像内に含まれるように配置している。検出処理は、様々なサイズの正方形の窓で画像を網羅的に探索し、3.3 で述べたように窓内の文字らしさ D_N をパターン選出時と同様の基準で評価する処理に基づく。なお D_N による検出閾値 ϵ は、F 値を最大化する値であった 0.6 に実験的に定めた。すなわち $D_N \geq 0.6$ を満たす正方形領域を検出された文字領域とした。

検出結果を図 7 に示す。検出結果の画像は左から順に、元画像、選出したパターンの位置、検出結果である窓の位置とサイズ、窓情報と元画像を重ねたもの、となっている。文字パターンとほぼ同サイズで、かつパターンを中央付近に含む窓が正しい結果である。

図 7 (a) を見ると、検出困難なパターンは一つも検出されていないのに対し、検出容易な文字パターンのほとんどが正しく検出されていることがわかる。また、誤検出の数も低く抑えられている。このことから、パターンのデザインによって検出精度を向上させるという本手法の考え方が有効であることがわかる。

図 7 (b) のように複雑な背景であっても、検出容易なパターンの方が優位に検出されることがわかる。この画像にはサイズの異なる検出容易なパターンが 2 セット存在しており、サイズに依らず高い検出性能を持っていることがわかる。また図 7 (c) ではパターンの色を変えて検出を行っており、パターンの色 (輝度値) にも依らない検出性能を確認できる。

誤検出としては、縦方向の直線、または“X”や“O”に似た形状のオブジェクトが多い。図 7 (b),(c) でそのような傾向が確認できる。“X”や“O”に似た形状のオブジェクトは、検出容易な文字パターンと同様な特徴値を持っているため誤検出されている。また縦方向の直線は文字上位基底と似た特徴値を持つため、非文字部分空間との為す角が閾値 θ を超えてしまうような特徴値であった場合は誤検出されてしまうと考えられる。

4.4 定量的評価

検出結果の定量的評価は以下の手順で行った。まず、画像全体のうち、検出容易なパターンが存在する矩形領域を正解として与えておく。その面積を R とする。そして、本手法により検出された正方形領域群を求め、そ

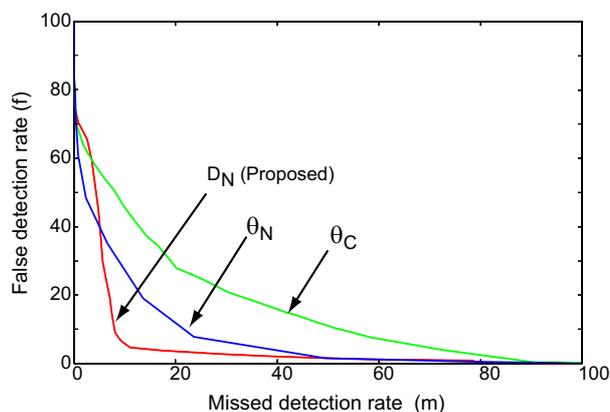


図 8 誤検出率 m と検出漏れ率 f の推移

れらの合計面積を V とする。その正解領域と検出された領域が重なった部分の面積を U 、検出された領域と重なっていない正解領域の面積を S 、正解領域外の検出された領域の面積を T とする。これらを用いて、検出漏れ率 $m = S/U$ と誤検出率 $f = T/(R - U)$ を計算し、評価に用いた。

本手法すなわち D_N による検出実験における検出漏れ率 m と誤検出率 f の関係を、図 8 に示す。同図は、検出の閾値 ϵ を D_N が取り得る値の範囲内で変化させながら検出を行い、各閾値での誤検出率 f と検出漏れ率 m をプロットしたものである。このように、適切な閾値を用いれば誤検出率と検出漏れ率を非常に低く保った高精度な検出が可能であることがわかる。

4.5 他の距離尺度による結果

異なる文字らしさの尺度を用いた実験も行った。具体的には、距離尺度を「 θ_N が大きいほど文字らしい」もしくは「 θ_C が小さいほど文字らしい」と考え、それぞれの場合についてパターン選出および検出処理を行った。なお、尺度が異なれば選出される文字も異なってくる。このため、本実験ではこれらの尺度下で最も検出容易とされた文字を図 7 と同環境中に配置して、検出実験を行った。

図 8 に、これら θ_N ならびに θ_C による検出実験の定量的評価結果を示す。いずれも本手法 (D_N) に比べると性能の劣化が見られる。これらの結果を再現率 $r = (U - S)/U$ と適合率 $p = (U - S)/V$ に基づく F 値 $= 2/(1/r + 1/p)$ により評価したところ、 D_N 、 θ_N 、 θ_C によってそれぞれ 0.73、0.39、0.60 が得られ、従って D_N が最も高い値、すなわち最も検出性能が高いことがわかる。非文字分布からの距離もしくは文字分布からの距離いずれかしか用いない場合、文字分布と重なっている非文字画像の誤検出を防ぐことができない。よってこのように検出精度の低下が見られたと考えられる。

5. まとめと今後の課題

本論文では、環境画像からの文字領域検出処理の高精

度化を目的として、検出容易な文字パターンの生成手法を提案した。検出法によって検出容易さが変わることからもわかるように、検出容易な文字パターンには様々な定義が考えられる。これに対し本論文では、その具体化例の一つとして「検出処理に用いる特徴空間において、文字分布内に存在し、かつ非文字分布から最も離れて存在するパターン」として検出容易な文字パターンと定義した上で、部分空間法で用いられる技法を利用して検討を行った。具体的には、特徴抽出に局所方向ヒストグラム特徴を用い、文字部分空間と非文字部分空間それぞれの距離から計算される“文字らしさ”の評価値が高いものを既存のフォントの集合から選出するというアプローチをとった。実験の結果、選出されたパターンは、その色やサイズによらず高精度に検出されることを確認した。

検出容易な文字パターンについての研究開発はまだその端緒にあり、今後の課題は数多く残されている。まず“選出”ではなく“合成”によりパターンを生成することが挙げられる。合成に際しては文字の形状モデルが必要になるが、幸いにして過去に様々な検討が為されている。すなわち (i) フーリエ記述子 [13], Active Shape Model [14], 幾何変換 [15] のようなパラメトリックなモデル, (ii) 線分系列 [16], [17], インクジェット列 [18], HMM などのストロークモデル, そして (iii) 摂動 [19] や GA [20], [21] による発見的手法, などがあり参考になる。

また、人間可読なパターンの生成が前提であるから、実際にはこのパラメータ空間の全体を考えるのではなく、ある程度制約された範囲に限定する必要がある。例えば、ストロークの太さもある程度以下にするべきである。例えば標準的な文字パターンから得られたパラメータから大きく変えたり、高次係数を無視したりすれば、文字として全く読めないパターンが生成されてしまう恐れがあるので [22], 注意が必要である。同様に、各カテゴリの分離度についても配慮する必要がある。

さらに、今回は局所方向ヒストグラム特徴による文字検出法を前提として議論したが、再三述べたように、文字検出法が変われば生成される文字も違うため、他の特徴ならびに検出法 (距離尺度) も多く試すべきである。本論文の考え方は一般的なものであり、他の検出法についても同じようなアプローチで検出容易な文字を生成できると考えられる。様々な検出法のうち、最も人間にとって自然で、かつ撮影時の様々な悪影響に対して頑強な文字が生成できるようなものを生成することが目標となる。

最後に、本検討に際して、一般的な文字検出問題さらには物体検出問題を常に視野に入れておくべきと考える。文字は人工物であり、自由にデザインすることが可能である。その性質を活かした本検討を出発点に、より複雑で一般的な対象の検出問題にも何らかの知見・方法論を与えることはできないだろうか。

- [1] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “デジタルカメラによる文字・文書の認識・理解,” 信学誌, vol. 89, no. 9, pp. 836-841, 2006.
- [2] J. Keechul, K. I. Kwang, and A. K. Jain, “Text information extraction in images and video: a survey, Pattern Recognition,” vol. 37, no. 5, pp. 977-997, 2004.
- [3] R. W. Lienhart and F. Stuber, “Automatic text recognition in digital videos,” Proc. SPIE, vol. 2666 (Image and Video Processing IV), pp. 180-188, 1996.
- [4] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” Proc. CVPR, vol. 2, pp. 366-373, 2004.
- [5] V. Wu, R. Manmatha, and E. M. Riseman, “An automatic system to detect and recognize text in images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 11, pp. 1224-1229, 1999.
- [6] H. Goto, “Redefining the DCT-based feature for scene text detection - Analysis and comparison of spatial frequency-based features,” Int. J. Doc. Anal. Recog., vol. 11, no. 1, pp. 1-8, 2008.
- [7] The British Computer Society, Character Recognition 1967, Unwin Brothers Limited, 1966.
- [8] 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, “カメラによる文字認識のためのカテゴリ情報の埋込に関する検討,” 信学論, vol. J89-D, no. 2, pp. 344-352, 2006.
- [9] 大町真一郎, 岩村雅一, 内田誠一, 黄瀬浩一, “実環境文字認識のための面積比による付加情報埋込,” 信学論, vol. J90-D, no. 12, pp. 3246-3256, 2007.
- [10] 若林哲史, 鶴岡信治, 木村文隆, 三宅康二, “特徴量の次元数増加による手書き数字認識の高精度化,” 信学論, vol. J77-D-II, no. 10, pp. 2046-2053, 1994.
- [11] A. K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, 1989.
- [12] 大津展之, “判別および最小 2 乗基準に基づく自動しきい値選定法,” 信学論, vol. J63-D, no. 4, pp. 349-356, 1980.
- [13] 森俊二, 坂倉柊子, 画像認識の基礎 (II), オーム社, 1990.
- [14] D. Shi, S. R. Gunn, and R. I. Damper, “Handwritten Chinese radical recognition using nonlinear active shape models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 2, pp. 277-280, 2003.
- [15] 塩野充, “非線形な幾何学的ひずみを用いた手書き風文字パターン生成の手法,” 信学論, vol. J74-D-II, no. 2, pp. 209-219, 1991.
- [16] 迫江博昭, “Rubber String Matching 法による手書き文字認識,” 信学技報, PRL74-20, 1974.
- [17] H. Nishida, “A structural model of shape deformation,” Pattern Recognit., vol. 28, no. 10, pp. 1611-1620, 1995.
- [18] M. Revow, C. K. I. Williams, and G. E. Hinton, “Using generative models for handwritten digit recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 6, pp. 592-606, 1996.
- [19] 石井健一郎, “変形文字パターン発生法とその応用,” 信学論, vol. J66-D, no. 11, pp. 1270-1277, 1983.
- [20] 坂野鋭, 木田博巳, 武川直樹, “遺伝的アルゴリズムによる文字識別系の解析,” 信学論, vol. J80-D-II, no. 7, pp. 1687-1694, 1997.
- [21] http://www.puls-der-zeit.de/genotyp/english/genotyp_en.htm
- [22] 鶴岡信治, 村瀬晶彦, 木村文隆, 横井茂樹, 三宅康二, “人間の字種識別基準を用いた自由手書き片仮名文字認識,” 信学論, vol. J68-D, no. 4, pp. 781-788, 1985.