

## 検出容易な文字パターン生成に関する検討

服部 亮史<sup>†</sup> 内田 誠一<sup>††</sup> 岩村 雅一<sup>†††</sup> 大町真一郎<sup>††††</sup> 黄瀬 浩一<sup>†††</sup>

<sup>†</sup>九州大学大学院システム情報科学府 〒819-0395 福岡市西区元岡 744

<sup>††</sup>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

<sup>†††</sup>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

<sup>††††</sup>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: †hattori@human.is.kyushu-u.ac.jp, ††uchida@is.kyushu-u.ac.jp, †††{masa,kise}@cs.osakafu-u.ac.jp,  
††††machi@ecei.tohoku.ac.jp

あらまし 情景画像中の文字領域を検出する処理に関して、検出が容易となる文字パターンの生成について検討する。従来の文字領域検出処理は、あらゆる文字パターンが共通して持つ普遍的な特徴を用いて文字領域と非文字領域を識別することで実現される。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の問題も多い。本研究では、検出が容易となる特徴を備えた文字パターンを生成するという視点からこの問題を捉える。具体的には、検出特徴空間において非文字分布から最も離れた文字分布内の点を、検出が容易なパターンと定める。この考え方により生成されるパターンは、検出特徴、文字分布、非文字分布、パターン間の距離の定義に依存して異なってくる。その一例として部分空間法を応用したパターン生成手法を提案し、また提案手法により生成された文字の検出容易性を確認した。

キーワード 文字領域検出, 文字生成, 部分空間法

## Conspicuous Character Patterns

Ryoji HATTORI<sup>†</sup>, Seiichi UCHIDA<sup>††</sup>, Masakazu IWAMURA<sup>†††</sup>, Shinichiro OMACHI<sup>††††</sup>, and  
Koichi KISE<sup>†††</sup>

<sup>†</sup> Graduate School of Information Science and Electrical Engineering, Kyushu University

<sup>††</sup> Faculty of Information Science and Electrical Engineering, Kyushu University Motooka 744, Nishi-ku,  
Fukuoka-shi, 819-0395 Japan

<sup>†††</sup> Graduate School of Engineering, Osaka Prefecture University Sakai-shi, Osaka, 599-8581 Japan

<sup>††††</sup> Graduate School of Engineering, Tohoku University 6-6-05 Aoba, Aramaki, Aoba-ku, Sendai-shi,  
980-8579 Japan

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

**Abstract** Characters in scene image are often hard to detect, i.e., not conspicuous. Thus, one of the main tasks for camera-based character recognition is the detection of characters in scene image. There are many past attempts for this difficult task. This paper investigates the essence of this task, that is, “what is conspicuous character images?” In order to have an example of the conspicuous character image, we use the relation between the subspace of non-character images and that of character images. Specifically, we try to select an image in the set of character images furthest from the subspace of non-character images.

**Key words** Character detection, Character synthesis, Subspace method

### 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、照明条件が悪く、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、今後様々なブレイ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が一般的に持つ複雑性などの普遍的特徴が用いられてきた [2] ~ [5]。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。

本研究では、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、環境中にある通常の文字を努力して検出するという従来のアプローチとは、逆転の発想に基づく。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。人間が試行錯誤的に生成する手段も考えられるが、本研究では、検出容易性の評価基準を定め、その基準を最もよく満たすパターンを自動生成するという、システムティックな方法を目指す。

本稿において、検出容易な文字パターンとは、何らかの特徴空間の中で、文字分布の中であって、かつ非文字分布から最も遠いパターンとして定義される。この定義には確定すべき要素が多い。すなわち、(i) 文字分布とは何か、(ii) 非文字分布とは何か、(iii) 非文字分布からの距離をどのように定義するのか、(iv) そしてどのような特徴空間で考えるのか、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。当然、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ことになる。例えば (iii) や (iv) については、実際に用いる文字検出法に依存する要素でもある。また (i) については、文字パターン形状 (色、表面のテクスチャなども含む) の扱い方にも依存するであろう。(ii) についても、何をもちて非文字とするかは、相当に難しい問題である。

このように様々に考えられる問題設定のうち、本論文ではその一例を検討する。具体的には、(i) をフォント画像の集合、(ii) を非文字画像の集合、(iii) を部分空間法に基づく距離尺度、そして (iv) を局所方向ヒストグラム特徴と定めて、検出容易な文字パターン生成を行う。今回の問題設定では、パターン生成というよりは、既存のフォント画像の中で最も検出し易いものを「選出」する処理になる。選出ではなく合成により新たに生成する手法については、今後の課題として 4. で述べる。

計算機にとって取り扱いやすいようにあらかじめパターン側を操作しておくという考え方は、既に文献 [6] ~ [8] においても議論されている。これらの文献では、検出のためではなく認識が容易となるようにデザインされたパターンが紹介されている。これに対し本報告では、検出・切り出しが容易となるようにパターン側を操作しておく (正確には検出容易なパターンを選出して利用する) 方法について論ずる。

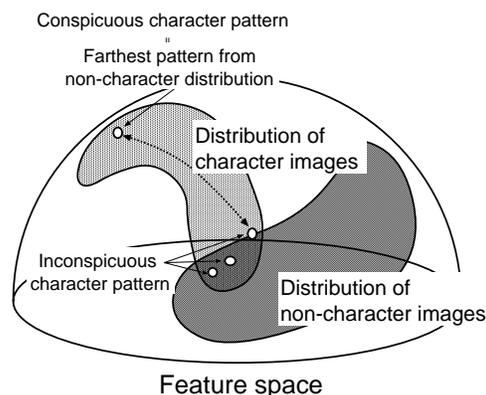


図 1 検出容易な文字生成の基本的考え方

## 2. 検出容易な文字パターンの選出

### 2.1 手法の概要

検出容易な文字パターンの自動生成のためには、パターンの形状と検出の容易さの関係を定義する必要がある。その基本的な考え方を図 1 に示す。このように本手法では、検出処理に用いる特徴空間において非文字分布から離れた位置に存在するパターンを、検出容易な文字パターンと考える。そのような文字パターンであれば、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。この場合の検出の容易さは非文字分布からの距離であり、その最大化を図ることになる。ただし、非文字分布から遠いというだけでは、人間にとって読みにくい文字が選出される可能性がある。加えて同じ距離の尺度が検出処理の際にも用いられることを考慮に入れると、文字分布から近いという評価も加える必要がある。

この考え方に基づき、本報告では、非文字/文字分布からの距離の算出に部分空間法を使用した方法について述べる。前述した 4 要素に当てはめると、(i) 文字分布としてフォント画像の集合、(ii) 非文字分布として情景画像の集合、(iii) 距離尺度として文字分布および非文字分布それぞれの部分空間とのなす角の比、(iv) 特徴として局所方向ヒストグラム特徴、となる。以下よりこれら 4 要素の詳細およびパターン生成の手順について述べる。

### 2.2 局所方向ヒストグラム特徴による特徴空間

前述のように、非文字分布からの距離を測る際の特徴としては、図 2 で示される局所方向ヒストグラム特徴 [9] を用いる。文字のエッジの局所的方向の分布で文字を表現するというこの特徴は、色や明るさで変化しにくい性質を持つ。同一のフォントでも様々な色で用いられるため、色 (輝度値) の変化に頑健な特徴により文字を表現するのは合理的と言える。

### 2.3 部分空間法を用いた距離尺度

本論文では、非文字分布からの距離の尺度、すなわち文字らしさの尺度の計算に部分空間法を用いる。部分空間法とは、パターンの分布を次元の小さい部分空間で近似し、その部分空間のみに着目して認識を行う手法である [10]。各クラスについて学習パターンを収集し、それらについて主成分分析を行い、得られた上位主成分を正規直交基底とした部分空間により、各ク

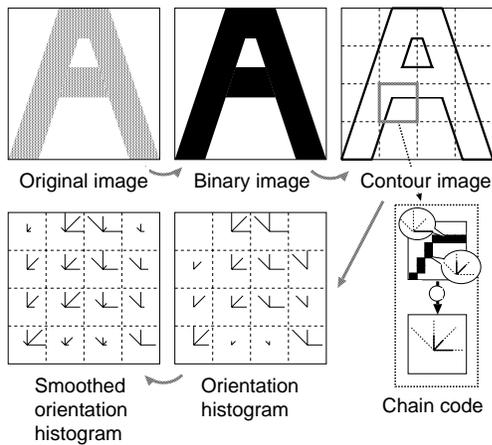


図 2 局所方向ヒストグラム特徴

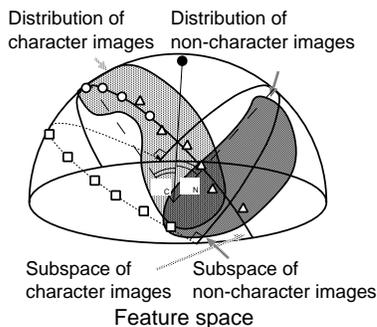


図 3 文字と非文字の部分空間

ラスの图案の分布を近似表現できる．部分空間との距離（なす角）を計算することで，分布からの距離を高速に求めることができる．

あるパターン  $p$  の非文字分布からの距離，すなわち文字らしさの評価値  $D_N$  は，非文字部分空間とのなす角（正準角） $\theta_N$  と，文字部分空間とのなす角  $\theta_C$  から求める．まず，パターン  $p$  と正規直交基底  $\phi_i (i = 1, 2, \dots, I)$  が張る  $I$  次元部分空間とのなす角  $\theta$  は以下の式で計算される．

$$\cos^2 \theta = \sum_{i=1}^I \frac{\langle p, \phi_i \rangle^2}{\|p\|^2 \|\phi_i\|^2} \quad (1)$$

ここでベクトル  $p$  ，および  $\phi_i (i = 1, 2, \dots, I)$  を単位ベクトルとすると，

$$\cos^2 \theta = \sum_{i=1}^I \langle p, \phi_i \rangle^2 \quad (2)$$

となる．そして文字らしさ  $D_N$  は，

$$D_N = \begin{cases} 0 & (\theta_N < \Theta) \\ \frac{\theta_N}{\theta_N + \theta_C} & (\theta_N \geq \Theta) \end{cases} \quad (3)$$

と計算される ( $0 \leq \Theta \leq \frac{\pi}{2}$ ) ． $\theta_N$  ,  $\theta_C$  は， $p$  と非文字部分空間， $p$  と文字部分空間 とのなす角をそれぞれ上式 (2) で計算したものである．

これは前述の通り「非文字分布から遠い」と「文字分布に近い」の両方を考慮に入れた尺度である．この尺度で選出，及び検出されるパターンは，図 3 においては白丸で表されるような，

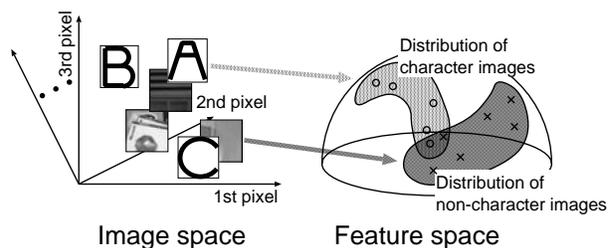


図 4 検出手順

文字分布内の非文字から遠い位置のパターンになると思われる．

単純に  $\theta_N$  を最大化するという尺度であれば，選出，検出されるパターンは文字分布から大きく離れたものになってしまう恐れがあり（四角のパターン）， $\theta_C$  を最小化するという尺度であれば検出の際に分布が重なっているところで誤検出の恐れがある（三角のパターン）．よって式 (3) で示される尺度を用いた．

## 2.4 文字分布と非文字分布

非文字分布は非文字画像の集合の為す低次元部分空間として定義される．本報告では，文字を含まない様々な情景画像を多数集めて非文字分布を形成する．

一方，文字分布はフォント画像の集合の為す低次元部分空間として定義される．この分布から離れると人間可読性が低くなり，文字らしくなくなると考えられる．実環境中の文字を意識するならば，フォント画像に対し実環境で起こり得る変動を施し，それらも含めて文字分布を定義してもよい．

## 2.5 文字パターンの生成

前節まで述べた問題設定において，検出容易な文字パターンを生成する．本報告においては，前述の通り，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こととする．この場合，全ての文字画像に対して，特徴抽出及び距離計算を行い，最も  $D_N$  が大きいフォントを選出すればよい．この作業はカテゴリ毎に行ってもよいし，全カテゴリを一括して行ってもよい．後者のほうが全カテゴリで同じフォントが利用されるため実動的であるが，今回は手法の性質を詳細に吟味するべく，前者を用いた．

以上の選出処理を，図 4 に示す．文字分布及び非文字分布は，局所方向ヒストグラム特徴空間内の部分空間として準備される．次に距離尺度  $D_N$  を基準として非文字分布から遠くかつ文字分布付近に位置するフォント画像を，やはり局所方向ヒストグラム特徴空間において選出する．

## 3. パターンの選出実験

本手法を用いて検出容易な文字パターンの選出を行った．以下ではパターン選出処理の結果，及び選出された文字パターンの性能評価の結果を示す．

### 3.1 実験条件

文字画像集合として 308 種の英語大文字フォント画像（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

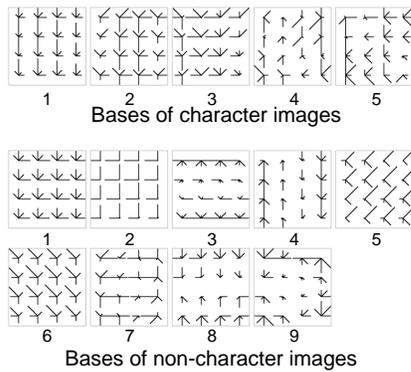


図 5 分布の上位基底

大まかなカテゴリを表す．今回の実験では一般文書画像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．一方の非文字画像としては，一般物体認識用の情景画像データベース (Caltch background database) から  $64 \times 64$  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 5 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  $\theta$  は実験的に 0.4 と定めた．

### 3.2 選出結果

文字らしさ  $D_N$  を計算し，各カテゴリで最も  $D_N$  が大きいパターンを選出し，検出容易な文字パターンを得た．選出された文字パターン，およびその特徴値を可視化したものを図 6 に示す．

同図 (a) が本手法による検出容易なパターンである．このように，全体的に丸い形状の文字が選択されていることが分かる．このようなパターンが選ばれた理由は，図 5 の非文字基底を見ることで以下のように理解できる．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角成分を持ち，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  $\theta_N$  が大きくなり， $D_N$  が大きくなったと考えられる．

図 6 (b) は検出困難なパターン，すなわち  $\theta_N$  が小さいパターンである．非文字の上位基底に近い，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分かる．

図 7 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 6 (a) で示したフォントだけではなく，類似の傾向を持つパターンであればよいと言える．

### 3.3 検出実験

生成された検出容易なパターン (図 6 (a)) 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 6 (b)) 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，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という方法で行った．なお  $D_N$  に関する閾値  $\epsilon$  は実験的に 0.6 と定めた．

検出結果を図 8 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と評価した．

図 8 (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分かる．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分かる．

図 8 (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分かる．またこの画像にはサイズの異なる検出容易なパターンが存在しており，サイズに依らず高い検出性能を持っていることが分かる．また図 8 (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．

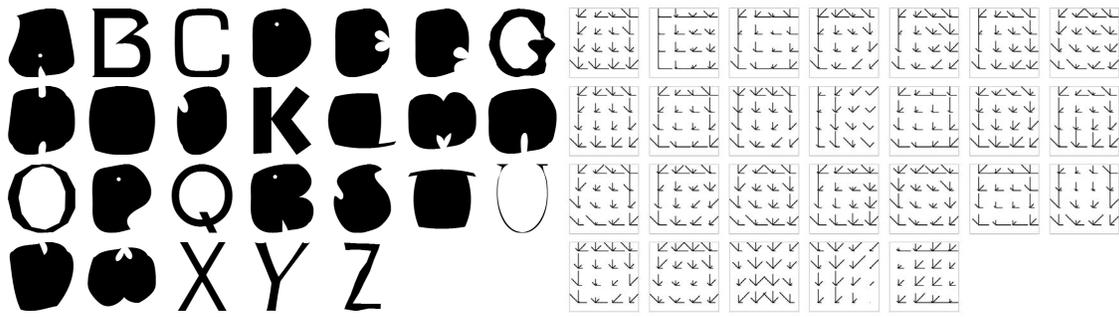
誤検出としては，縦方向の直線，または“X”や“O”に似た形状のオブジェクトが多い．図 8 (b),(c) でそのような傾向が確認できる．“X”や“O”に似た形状のオブジェクトは，検出容易な文字パターンと同様な特徴値を持っているため誤検出されている．また縦方向の直線は文字上位基底と似た特徴値を持つため，非文字部分空間とのなす角が閾値  $\theta$  を超えてしまうような特徴値であった場合は誤検出されてしまうと考えられる．

式 (3) で計算される文字らしさ  $D_N$  の有効性を確認するため，文字らしさの尺度を変更した上で，改めてパターン選出および検出実験を行った．なお検出の際の閾値は，誤検出の数がある程度以下になるように実験的に定めた．

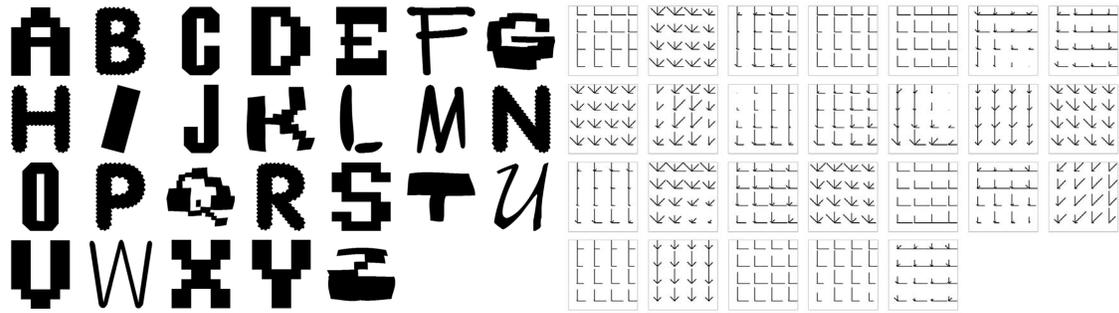
「 $\theta_N$  が大きいほど文字らしい」という尺度 (図 9 (a))，および「 $\theta_C$  が小さいほど文字らしい」という尺度 (図 9 (b)) でパターン生成と選出実験を行った．検出結果を見ると，ともに誤検出が多く，生成されたパターンが正しく検出されていないことが分かる．これらの手法では，非文字分布からの距離もしくは文字分布からの距離いずれかが用いていないため，文字分布と重なっている非文字画像の誤検出を防ぐことができない．よってこのように検出精度の低下が見られたと考えられる．

## 4. まとめと今後の課題

本研究では，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．具体的には，検出特徴空間において非文字分布から最も遠くなる文字分布内のパターンを検出容易な文字パターンとし，既存



(a)



(b)

図 6 本手法により選出された文字パターン



図 7 検出容易性の順列

のフォントから選出を行う。選出されるパターンは、文字分布、非文字分布、特徴空間、非文字分布からの距離の4つの要素の定義によって異なってくる。本報告ではその一例として、部分空間法による距離を用いて、検出容易な文字パターン選出処理を行った。実際に実験による評価を行った結果、選出されたパターンの検出容易性を確認した。

今後の拡張としては以下のようなアイデアが挙げられる。文字パターンの形状をなんらかのパラメータで制御し、あるパターンの検出容易性を形状パラメータを入力とする関数で表現することを考える。検出容易性を最大にするパラメータを求め、そのパラメータから文字を合成することで、検出容易な文字パターンを生成する。つまり、検出容易な文字パターンの生成という問題を、一種の最適化問題として考えることになる。本手法では、文字分布は与えられた文字画像の集合であったが、拡張手法では文字分布は文字形状パラメータ空間が生成する文字画像の集合となる。パラメータ空間を、画像空間を経て特徴空間に写像すると、多様体のような分布を為すことになる。すな

わち、この多様体状の文字分布内において非文字から最も遠い点を見つける問題として捉えることができる。

謝辞: 非文字画像の収集に協力して頂いた、東北大学阿曾研究室の松田友輔氏に感謝する。

#### 文 献

- [1] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “カメラを用いた文字認識・文書画像解析の現状と課題,” 信学技報, PRMU2004-246, 2005.
- [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,” Proceedings of 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,” CVPR 2004, vol. 2, pp. 366-373, 2004.
- [5] V. Wum, 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] 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, “カメラによる文

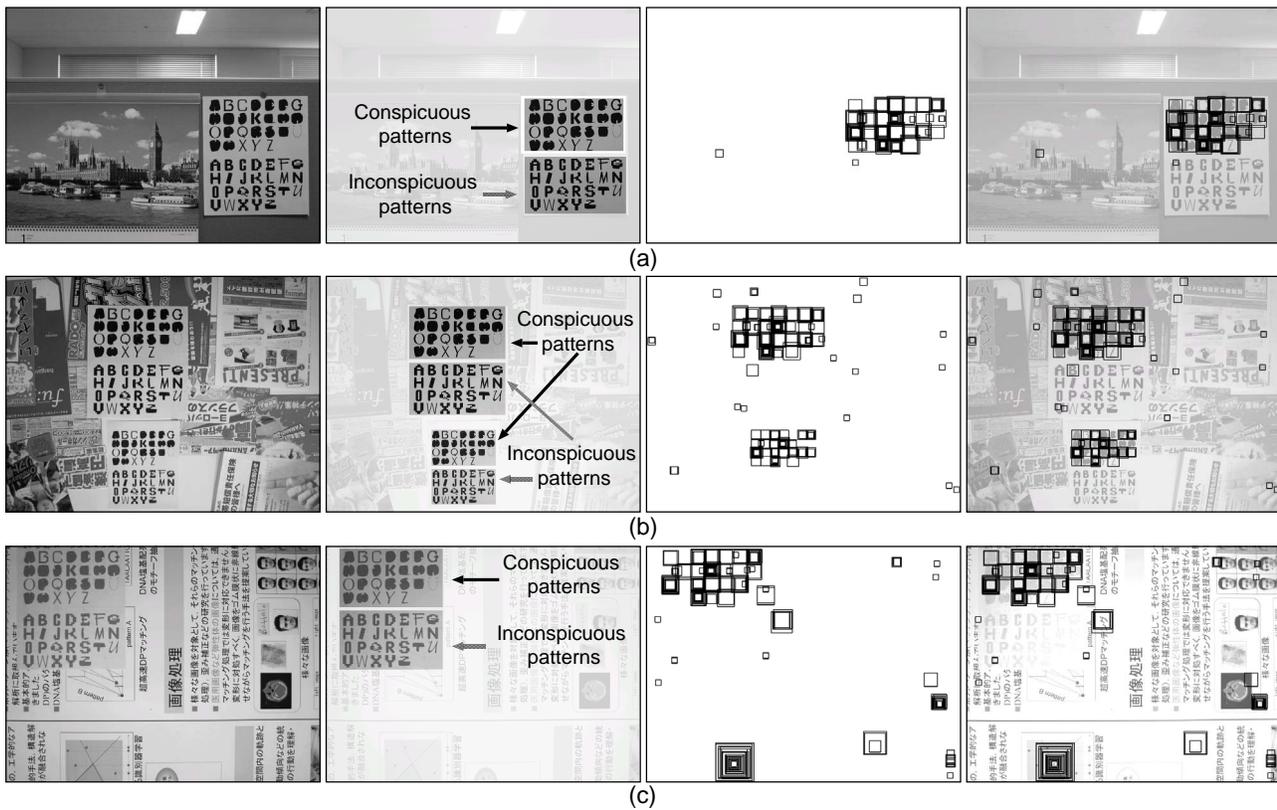


図8 検出結果

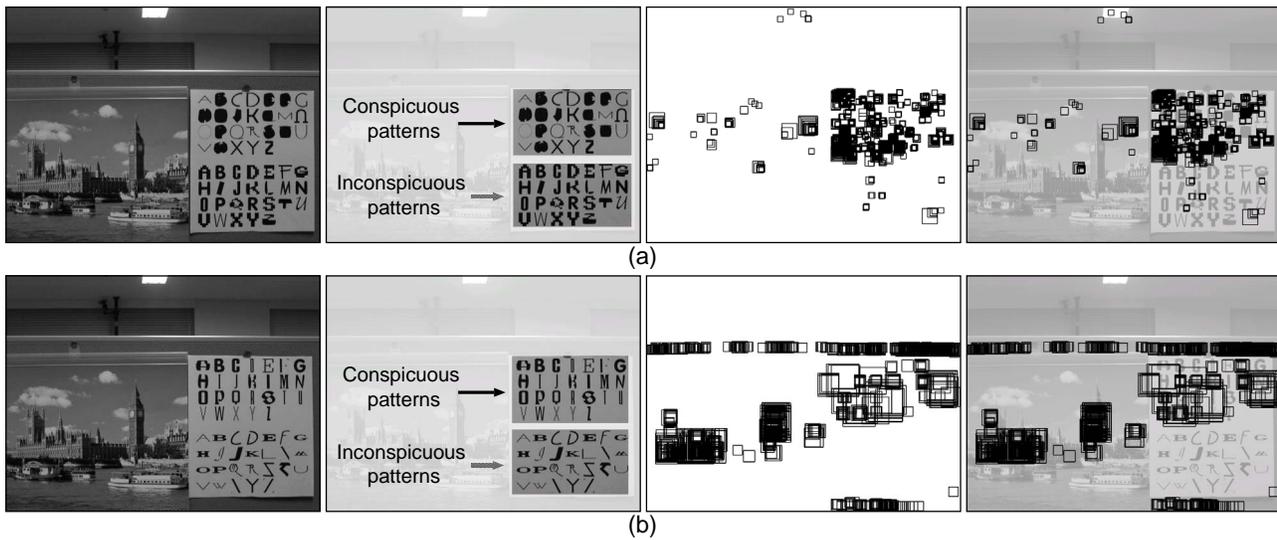


図9 評価尺度を変えた場合の検出結果

字認識のためのカテゴリー情報の埋込に関する検討,” 信学論 (D), vol. J89-D, no. 2, pp. 344–352, 2006.

- [7] 岩村雅一, 内田誠一, 大町真一郎, 黄瀬浩一, “付加情報を用いるパターン認識,” 信学論 (D), vol. J90-D, no. 2, pp. 460–470, 2007.
- [8] 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 (D), vol. J90-D, no. 12, pp. 3246–3256, 2007.
- [9] 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, Oct. 1994.
- [10] 石井健一郎, 上田修功, 前田英作, 村瀬洋, わかりやすいパターン認識, オーム社, 1998.