論　文

階層型識別器を用いた情景画像からの文字抽出手法

山口　拓真[†*]　　丸山　稔[†]

Character Extraction from Natural Scene Images by Hierarchical Classifiers

Takuma YAMAGUCHI[†*] and Minoru MARUYAMA[†]

あらまし　本論文では，階層的な識別器を用いて情景画像から文字を抽出する手法について述べる．階層的な識別器には輝度ヒストグラムの形状を特徴とした識別器とSVMの2種の識別器を用いる．最初の段階では，非線形SVMと比べ識別能力は劣るが，処理は高速である輝度ヒストグラムの形状を特徴とする識別器を用いて情景画像中の文字候補領域を高速に限定する．次の段階では，非線形SVMを用いて最終的な結果を求める．非線形SVMは計算コストが高いため，Haar Waveletを用いて特徴量の削減を行うとともに，少量のサポートベクタでもとのSVMを近似した近似RBFも構築し高速化を行った．提案手法の有効性を示すために実験を行ったところ，単純にSVMを適用した場合に比べより良い識別率が得られ，計算時間においても非常に改善された．

キーワード　文字領域抽出，輝度ヒストグラム，Haar Wavelet，近似RBF

1. まえがき

近年，ディジタルビデオやディジタルカメラなどのディジタル映像機器が非常に発達した．それらの機器はまた携帯電話やPDAなどの端末にも組み込まれるようになり，使用者の周辺環境を容易に画像として取り込むことが可能となっている．このような映像機器の発達に伴い，文字認識の対象も変化・拡大していくことが予想される．従来，書籍や伝票などをスキャナで取り込み，それらの文字を識別していたのに対し，現在では自然情景画像中の文字の認識が注目を集めている[1]．自然情景中の文字としては例えば看板上の文字などが挙げられるが，それらの文字の認識が可能となれば，幅広い応用が期待できる．これまでに研究されてきたスキャナ画像を対象とした文字認識技術は，情景画像中の文字に対しても非常に有効であるとは考えられる．しかし，情景画像中の文字は，低解像度であったり，ぼやけていたりもするため，それらの技術をそのまま適用することが困難である場合も多いと思われる．また，従来の対象画像と比べ自然情景画像中の背景は多種多様であり，文字の抽出において困難を伴うことが予想される．そこで，本論文では自然情景中から文字候補領域をできるだけ高速かつ高精度に抽出することを考える．

情景画像からの目的物体の抽出の研究としては，顔，歩行者，車の検出等の研究が活発に行われている[2],[3]．また，文字を対象とした自然情景画像中の認識システムに関する研究も既にいくつか報告されている[4],[5]．FujisawaらはICC（Information Capturering Camera）と呼ばれるカメラシステムを提案している[4]．これはビデオカメラやディジタルカメラを用い，リアルタイムにカメラに写る画像からテキストを抽出し翻訳して表示するというものである．その他，情景画像中の地名の抽出・看板上の電話番号認識のための数字識別システムの研究などもある[6],[7]．自然情景中の文字認識を行うためには通常処理対象となる文字領域の抽出が行われる．

自然情景中の文字領域の抽出に関する研究についてもいくつか報告されている[8]～[10]．松尾らは明度を用い複数枚の分解画像を作成し，その分解画像内における閉領域の外接方形の形状や配置により文字列を抽出している[8]．劉らは空間周波数が高く輝度コントラストが大きい領域を文字候補領域として抽出し，外接方形のサイズや形状などを用いて文字列を抽出している[9]．携帯性の高い情報機器を用いて自然情景中の文字抽出を行う際には，高い抽出精度とともに，高速性

† 信州大学工学部情報工学科，長野市
Dept. of Information Engineering, Shinshu University, 4-17-1 Wakasato, Nagano-shi, 380-8553 Japan
* 現在，メディアドライブ株式会社

も要求される．対象クラスの画像からの高速抽出の研究には，Heisele らの階層型 SVM を用いた顔画像抽出の研究などがある [2]．これは識別能力には劣るが計算コストの低い線形 SVM を多重に用いて顔領域の限定を行っていき，最終段階で計算コストは大きいが識別能力は優れている非線形 SVM を用いて顔の検出を行うことで，高速性を実現するものである．このような階層型手法は，文字抽出においても有効であると考えられる．

文字抽出は文字認識過程における前処理に相当するため，文字抽出手法には，高い識別能力とともに，高速であることが求められる．本論文では，様々な応用例を通じて高い識別能力を有することが確認されている非線形 SVM を用いて文字抽出を行うが，カーネル評価に時間を要するため，単純に非線形 SVM を適用しただけでは非常に計算時間がかかるという問題がある．本論文では，非線形 SVM を用いた文字抽出を高速に行うために，輝度ヒストグラム形状に基づく識別と非線形 SVM を組み合わせた階層型識別器による手法をまず提案する．ここで輝度ヒストグラムによる識別とは文献 [9] で述べられている手法で，ヒストグラムの形状（単峰性・多峰性）に基づいて文字・非文字の判別を行うものである．本論文ではこの輝度ヒストグラム形状に基づく識別手法と非線形 SVM を階層型に組み合わせることにより，SVM だけの場合と比較して識別能力が同等以上で高速な文字抽出が可能であることを示す．更に，この提案手法の更なる高速化のために，学習により得られる SVM のサポートベクタ集合から少数をサンプリングしてもとの SVM を近似する手法を提案し，その有効性を示す．

2. 識別器の構成

2.1 識別対象領域の設定

本論文では固定サイズのウィンドウを用いて 24 bit カラー画像を走査し，ウィンドウ内の部分画像が文字か非文字であるかを判定する．図 1 に識別対象データの例を示す．本論文ではウィンドウサイズは 32×32 ピクセルに固定する．本研究においては対象文字サイズは 20×20 以上であると仮定している．この仮定は対象文字撮影位置等により，常に成立するとは限らない．しかしながら現在のカメラの性能を考慮すると，ズーミング機能などにより，この仮定を満足させることはさほど困難ではなく実用上は問題ないものと考える．対象となる文字色・文字種・字体などを区別する

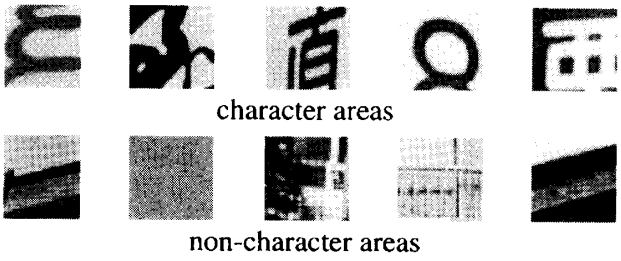


図 1 識別対象データ例
Fig. 1 Sample input images.

手法も有効であると考えられるが，本研究ではこれらを区別せずに取り扱っている．また，学習や検証に用いるラベル付けされたデータを作成する際の，文字領域か非文字領域かの判定は人間の判断に基づいて行った．文字領域か非文字領域かの判断については，筆者らが 32×32 のウィンドウのみを見た際に明らかに文字または文字の一部として読み取れる場合に，文字領域として判断しデータの作成を行った．

2.2 階層型識別の構成

本研究においては識別を高速に行うために階層的な識別器を用いる．通常画像中においては文字領域の占める割合はそれほど大きくなく，非文字領域の占める割合が非常に多いものと考えられる．したがって，まず文字候補領域を高速に限定し，後段の処理で精度の高い識別を行う階層型の識別器を構築することにより，精度と速度の両立が可能になることが期待できる．そこで本手法では，階層型の識別器の最初の段階には，識別能力は非線形 SVM には劣るが，処理は非常に高速な輝度ヒストグラムを用いた識別を行う．この識別器では文字領域に対して高い識別率を得られるため，ほとんどの文字領域を残したまま非文字領域の削減が行える．

次の段階では，識別能力に優れているが計算コストの高い SVM を用いた識別を行う．ここでは前段階で削除しきれなかった非文字領域の削減を行う．非文字領域の識別率を向上させることにより，文字領域の識別率が若干低下する可能性があるものの，実際の自然情景画像では文字領域よりも非文字領域の方が多いはずであるため，非文字領域の識別率を向上させることは，情景画像に適用した際のパフォーマンスの向上につながる．

3. 輝度ヒストグラムによる識別

一般的に情景画像中に存在する文字情報は看板上のものであると考えられる．看板上の文字は利用者にとっ

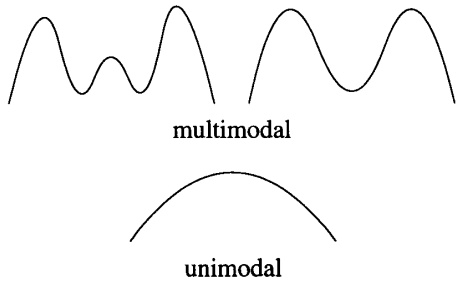


図 2　輝度ヒストグラムの形状
Fig. 2　Intensity histogram form.

て読み取りやすいように考慮されているはずである．よって，文字部分と文字背景部分とでは輝度の差が明確であると考えられる．そのような領域では，輝度ヒストグラムが山の数が 2〜3 の多峰性（multimodal）になり，文字領域以外では単峰性（unimodal）になることが多い（図 2）．非文字領域が多峰性になることはあるが，看板上の文字は多くの場合において，文字部分と文字背景部分の輝度差が明確であるといえるため，文字領域が単峰性であることは考えにくい．よって，入力データの輝度ヒストグラムが単峰性である場合は非文字領域と識別し，SVM での識別は行わない．そうすることによって，計算コストが高い SVM で識別するデータ数を減少させることができ，計算時間を節約することが可能である．

3.1　多峰性・単峰性の判別手法

ここで輝度ヒストグラムの形状の判別手法について述べる．以下に処理の流れを示す．

1. 入力画像の輝度ヒストグラム $h[i](0 \leq i \leq 255)$ を求める
2. 移動平均法を用いて $h[i]$ の平滑化を行う
3. $h[i]-\theta > 0 \quad \Rightarrow \quad b[i] \leftarrow 1$
   $h[i]-\theta \leq 0 \quad \Rightarrow \quad b[i] \leftarrow 0$
   ここで，$\theta = 4$(画素数の平均)
4. 1 次元配列 $b[i]$ の “1” の連結領域個数が 2〜3 ⇒ 多峰性
   それ以外 ⇒ 単峰性

まず，入力画像から以下の式を用いて輝度値を求め，

$$I = R \times 0.3 + G \times 0.59 + B \times 0.11 \qquad (1)$$

輝度ヒストグラムを作成する．ここで，I は輝度値，$R \cdot G \cdot B$ はそれぞれ赤・緑・青の成分の値を示している．その後，移動平均法による平滑化を行う．平滑化を行わなければ，ヒストグラムに多くの小さな

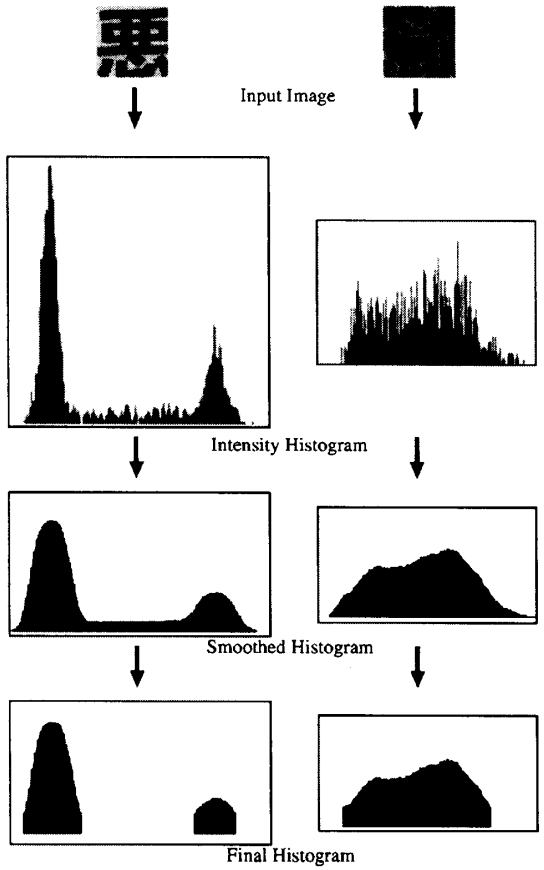


図 3　多峰性・単峰性の判別
Fig. 3　Differentiating multimodal and unimodal.

山が存在することから，ヒストグラムの形状は常に多峰性になってしまうからである．平滑化後，各々の輝度レベルにおいて存在画素数があるしきい値に満たない場合は，その輝度レベルには画素は存在しないこととする．本研究ではしきい値を画素数の平均（$32 \times 32\,\text{pixel}/256 = 4\,\text{pixel}$）に設定し実験を行った．画素数の平均をしきい値とすることで，階層型識別器の第一段階として適度な結果を得られることが予備実験において示されたため，本研究ではこのしきい値を用いている．参考までに **3.2** においてしきい値 θ を変化させた際の結果も示す．最終的なヒストグラムに連結領域がいくつ存在しているかで，多峰性・単峰性の区別を行う．処理の例を図 3 に示す．最終的に連結領域が二つ存在している図 3 左では多峰性と識別され，連結領域が一つである図 3 右では単峰性と識別される．

3.2　輝度ヒストグラムによる識別結果

表 1 にヒストグラムによる識別結果を示す．単峰性・多峰性の判別の際のしきい値を本研究では $\theta = 4$ としているが，参考までに 0.5 倍，2 倍 ($\theta = 2, \theta = 8$)

表 1 ヒストグラムによる識別結果
Table 1 Results by histogram classifier.

		識別率	
		文字領域(%)	非文字領域(%)
$\theta = 4$	データ 1	95.9	71.4
	データ 2	95.7	71.0
$\theta = 2$	データ 1	92.2	73.0
	データ 2	92.7	72.0
$\theta = 8$	データ 1	72.5	80.1
	データ 2	69.9	80.1

にした場合の結果も示している．これはデータとして文字領域 2250・非文字領域 2250 の合計 4500 のデータを用いて実験を行った．後述する最終段階での識別器では学習が必要なため，これらのデータを学習データとテストデータに分けることとなる．そこで，ここでもデータ 1 とデータ 2 に 2 分割して実験を行った．データ 1 を後述する識別器の学習データ，データ 2 をテストデータとして用いる．データ 1 とデータ 2 は質的に差異はないと考えられる．なお，実験結果の信頼性を上げるために，データの分割をランダムに 10 通りにわたって行い，識別実験を繰り返した．表 1 では，その平均値を示している．ヒストグラム法により，文字領域を高精度で検出できることが分かる一方，非文字領域の識別率は 70%前後であり，全非文字領域中の約 30%が排除できない．通常画像中に占める非文字領域数は多数となるため，ヒストグラム単体では満足のいく文字抽出結果は得られないことをこの実験結果は示している．しかしながら，単体では十分な性能を有するとはいえないものの，ヒストグラムに基づく識別は高速に実現可能でありほとんどの文字領域を残したまま，非文字領域の約 70%を排除できる．ヒストグラム法は，後段の識別の処理対象を高速に限定する手法として十分な性能を備えていると考えられる．

4. 非線形識別関数を用いた文字領域判定

4.1 Haar Wavelet を用いた特徴量削減

提案手法の最初の段階では輝度情報に基づいた識別器を用いた．本手法においては，次の段階では，良好な識別能力を有することが確認されている非線形カーネルに基づく SVM を用いて識別を行う．

SVM は Vapnik らによって提案されたマージン最大化に基づく識別手法である [11]．今，データ $\{\boldsymbol{x}_i\}$ $(i = 1, \cdots, N)$ とそのラベル $y_i \in \{+1, -1\}$ が与えられたものとし，$\phi : X \longrightarrow F$ を入力空間 X から特徴空間 F への写像とする．SVM は，特徴空間 F における線形識別関数で，マージンを最大化するものとして定義でき，以下の式で与えられる．

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i < \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}) > -b \qquad (2)$$

ここに，$< \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) >$ はデータ $\boldsymbol{x}, \boldsymbol{y}$ を写像後の特徴空間における内積である．この内積を与える関数はカーネル関数 k と呼ばれ，これにより SVM は以下のように表せる．

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) - b \qquad (3)$$

係数 $\{\alpha_i\}$ は二次計画問題を解いて得られる．非零の係数 α_k に対応するデータ $\boldsymbol{x}_k$ はサポートベクタと呼ばれる．

本研究においては，各ウィンドウ内の画像情報を入力として識別を行う．このとき，入力ベクトルとして，走査ウィンドウ内のピクセルを各成分とするベクトルを用いると，R・G・B の 3 要素をもつ 32×32 ピクセルの場合，ベクトルの次元は 3072 となってしまう．SVM を識別関数として用いる場合，識別関数の値を決定するためには，入力 $\boldsymbol{x}$ とすべてのサポートベクタ $\boldsymbol{x}_i$ についてカーネル関数の値 $k(\boldsymbol{x}, \boldsymbol{x}_i)$ の評価を行う必要がある．非線形カーネル関数としては，RBF，多項式などがよく用いられる．これらの場合，カーネル評価のための計算量は入力ベクトルの次元に比例して増大することになるため，高次元ベクトルをそのまま入力に用いるのは適当ではない．

そこで計算コストを削減するためにここでは Haar Wavelet を用いて Wavelet の係数のスパース化を行い，特徴量を削減する．

本研究ではまず入力画像を Haar Wavelet に基づく表現に変換し，その係数の絶対値が上位 5%未満の係数は 0 とすることによってスパース化を行った [12]．その後，カラー画像をグレースケール画像に変換するのと同様の計算式を用いて，次元を $\frac{1}{3}$ に圧縮した．本研究で用いたデータの場合，この手法により，結果的に Wavelet の係数の非零要素は全係数中の約 2%に減少した．

4.2 近似 RBF を用いた文字抽出

非線形カーネルに基づく SVM は高い識別能力を有するものの，サポートベクタが多数になった場合，識別に要する時間が掛かるという欠点がある．本研究においては，データを Haar wavelet を用いてスパース

化し，カーネル評価の高速化を図っているものの，サポートベクタ数だけカーネル評価を行う必要があるため，識別にかかる時間は依然として大きなものになってしまう．識別関数の計算時間は，カーネルの評価回数，すなわちサポートベクタ数に比例する．このため，計算時間を短縮するためにはサポートベクタを低減することが考えられる．

SVM は特徴空間 F における線形識別関数であり，

$$f(\boldsymbol{x}) = <\varphi, \phi(\boldsymbol{x})> - b \tag{4}$$

と表すことができる．ここに

$$\varphi = \sum_{i=1}^{m} \alpha_i y_i \phi(\boldsymbol{s}_i) \tag{5}$$

であり，$\boldsymbol{s}_i$ $(i = 1, \cdots, m)$ はサポートベクタである．計算時間短縮のため，Burges らはサポートベクタより少数の $\{\boldsymbol{p}_\ell\}$ を用いて，

$$\hat{\varphi} = \sum_{\ell=1}^{n} \beta_\ell \phi(\boldsymbol{p}_\ell) \in F \tag{6}$$

を考え，

$$<\varphi - \hat{\varphi}, \varphi - \hat{\varphi}> \tag{7}$$

を最小化するように $\{\beta_\ell\}$, $\{\boldsymbol{p}_\ell\}$ を求める手法を提案している [14].

このとき，本研究の対象となる問題の場合，$\{\boldsymbol{p}_\ell\}$ をも変動させて (7) を最小化させたとき得られる解がスパースな表現をもつことはほとんど期待できず，カーネル評価の計算時間の増大を招く．そこで，以下ではベクトル集合 $\{\boldsymbol{p}_\ell\}$ はスパース表現をもつ学習データの中からサンプリングし，固定することを考える．

今，固定ベクトル集合 $\{\boldsymbol{p}_\ell\}$ $(\ell = 1, \cdots, n)$ をとり，サポートベクタを $\{\boldsymbol{s}_i\}$ $(i = 1, \cdots, m)$ とおく $(n \ll m)$．また，$k(\boldsymbol{p}_i, \boldsymbol{p}_j)$ を第 (i, j) 要素とする $n \times n$ 行列を K_0, $k(\boldsymbol{p}_i, \boldsymbol{s}_j)$ を第 (i, j) 要素とする $n \times N$ 行列を K_{sv}, $\boldsymbol{\alpha} = (\alpha_1 y_1, \cdots, \alpha_m y_m)^T$ とおくと，(7) を最小化するパラメータ $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)^T$ は次式で与えられる．

$$\boldsymbol{\beta} = K_0^{-1} K_{sv} \boldsymbol{\alpha} \tag{8}$$

$\varphi, \hat{\varphi}$ により得られる識別関数をそれぞれ $f, \hat{f}$ とおくと，$f = \sum \alpha_i y_i k(\boldsymbol{x}, \boldsymbol{s}_i) - b$, $\hat{f} = \sum \beta_\ell k(\boldsymbol{x}, \boldsymbol{p}_\ell) - b$ であるから，(8) は，中心点 $\{\boldsymbol{p}_\ell\}$ 上で，$f(\boldsymbol{p}_\ell) = \hat{f}(\boldsymbol{p}_\ell)$ を満たすようにパラメータを決定したものにほかならず，他のデータ領域において f と $\hat{f}$ が近くならないおそれがある．そこで，本研究では (7) に代わり，データが発生する領域における $\varphi, \hat{\varphi}$ の差を小さくするために

$$\begin{aligned} &\int <\varphi - \hat{\varphi}, \phi(\boldsymbol{x})>^2 w(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int (\sum_i \alpha_i y_i k(\boldsymbol{s}_i, \boldsymbol{x}) - \sum_\ell \beta_\ell (\boldsymbol{p}_\ell, \boldsymbol{x}))^2 w(\boldsymbol{x}) d\boldsymbol{x} \\ &\rightarrow \min \end{aligned} \tag{9}$$

を満たす $\hat{\varphi}$ を求めることを考える．

今与えられている学習データのみを用いて (9) を近似的に実現するために，ここでは，

$$\hat{f}(\boldsymbol{x}) = \sum_{\ell=1}^{n} \beta_\ell k(\boldsymbol{p}_\ell, \boldsymbol{x}) + \gamma \tag{10}$$

とし，

$$\sum_{i=1}^{N} w_i (f(\boldsymbol{x}_i) - \hat{f}(\boldsymbol{x}_i))^2 \tag{11}$$

を最小化する．ここに w_i は各データ点の発生確率に対応する重みである．このとき，マージン最大化と同様の考え方により，近似 SVM $\hat{f}$ のマージンの逆数に相当する $\boldsymbol{\beta}^T K_0 \boldsymbol{\beta}$ を追加した，以下のような評価関数の最小化を行う．

$$\sum_{i=1}^{N} w_i (\hat{f}(\boldsymbol{x}_i) - f(\boldsymbol{x}_i))^2 + \lambda \boldsymbol{\beta}^T K_0 \boldsymbol{\beta} \tag{12}$$

ここに λ は定数である．

第 (i, j) 要素が $k(\boldsymbol{x}_i, \boldsymbol{p}_j)$ であるような $N \times n$ 行列を K，各データの重み w_i を要素とする対角行列を $\Lambda = \mathrm{diag}(w_1, \ldots, w_N)^T$ とおくと，(12) を最小にするパラメータ $\boldsymbol{\beta}$, γ は，以下の線形方程式を解いて得られる．

$$\begin{aligned} &\begin{pmatrix} K^T \Lambda K + \lambda K_0 & K^T \boldsymbol{w} \\ \boldsymbol{w}^T K \sum_i w_i & \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} \\ &= \begin{pmatrix} K^T \Lambda \boldsymbol{f} \\ \boldsymbol{w}^T \boldsymbol{f} \end{pmatrix} \end{aligned} \tag{13}$$

ただし，$\boldsymbol{w} = (w_1, \cdots, w_N)^T$ である．

表 2 単独の識別器による識別結果
Table 2 Results by simple classifier.

	学習データ	SV 数	識別率	
			文字領域(%)	非文字領域(%)
ヒストグラム	—	—	95.7	71.0
線形 SVM	2250	—	45.1	75.9
非線形 SVM (RBF)	2250	1403	91.1	81.3

5. 実験結果

5.1 単一の識別関数による文字抽出

本論文で提案した手法の性能を検証するために，実データを用いて識別実験を行った．まず，非線形 SVM の識別能力を見るために，文字，非文字領域それぞれ 1125 からなる 2250 領域の画像を学習データとして SVM の学習を行った．なお，本研究では SVM の学習に SVM^{light} を利用した [13].

このとき SVM のカーネルには以下に示される RBF (Gaussian) カーネルを用い，

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{y}||^2}{2\sigma^2}\right) \tag{14}$$

非線形カーネルの種類及びパラメータに関しては，予備実験において一番高い識別精度が得られた値 (RBF, $C = 10, 2\sigma^2 = 1000$) に設定した．

得られた識別関数を用いて学習データとは異なるテストデータ (文字 1125 領域，非文字 1125 領域からなる 2250 領域) を用いて識別実験を行った．表 2 に単一種類の識別器による実験結果を示す．この識別実験においては全学習データを用いて学習した SVM を用いた．

実験より，明るさのヒストグラムに基づく手法は，単独で用いるには能力が不十分であるものの，明らかな非文字領域を除外するのに効果的であることが分かる．非線形 SVM (RBF) は，ヒストグラム，線形 SVM に比べて良好な性能が期待できるものの，サポートベクタの数が 1403 個と多数になるため非常に計算コストが高い．

5.2 ヒストグラムと SVM に基づく文字抽出

ヒストグラムを用いた識別関数を第一段に用い，文字領域として検出されたデータに関してのみ SVM を適用する階層型の処理の実験を行った．本手法においては，SVM はヒストグラムを通過したデータに対してのみ適用される．そこで，ヒストグラムと組み合わせて用いる SVM の学習の際には，ヒストグラムによる識別過程を通過するデータに対して学習を行った．

表 3 SVM の学習結果
Table 3 Results of training SVM.

	学習データ数	SV 数
学習データすべて	2250	1403
ヒストグラム通過データ	1401	1173

表 4 ヒストグラムと非線形 SVM を組み合わせた識別結果
Table 4 Results by histogram+SVM(RBF) classifier.

学習データ数	SV 数	識別率	
		文字領域(%)	非文字領域(%)
1401	1173	91.3	86.1

表 3 に SVM の学習結果を示す．

学習データが減少したことによりサポートベクタの数が若干減少したが，依然としてサポートベクタの数は非常に多い状態であることには変わりはない．単独 SVM の場合に比べて若干の速度向上をもたらすものとは考えられるが，依然として計算時間の面では望ましくないことが分かる．表 4 にヒストグラムと非線形 SVM を組み合わせた場合の識別実験結果を示す．識別結果に関しては，単独で非線形 SVM を学習して適用する場合に比べて文字領域に関しては同等，非文字領域に関しては識別能力が向上しており，ヒストグラムと SVM の階層処理に基づく手法は有効であることが分かる．

5.3 ヒストグラムと近似 RBF に基づく文字抽出

識別計算時間の高速化を図るために，**4.2** で提案した近似 RBF に基づく識別手法の検証実験を行った．学習のためのデータとしては，ヒストグラムに基づく識別関数を通過したデータのみを用いた．提案手法においては，スパースな表現を有する中心点を確保するために，学習データから少数データのサンプリングを行い，これらを RBF の中心点として用いる．本手法においてはサンプリング方式としてはランダムサンプリングと k-means [15] に基づく手法を用いた．このとき，k-means を適用後，得られた各クラスタ中心に最も近いデータをクラスタ中心に代わるものとしてサンプリングした．

なお，本手法においてはヒストグラムに基づく識別

表 5　提案手法での識別結果
Table 5　Results by proposed method.

方式	データサンプリング手法	SV 数	重み	識別率	
				文字領域(%)	非文字領域(%)
式 (8) に基づく近似手法	random	25+25	—	94.0	76.5
	k-means	25+25	—	94.4	77.8
提案手法	random	25+25	1:1	93.7	78.7
	k-means	25+25	1:1	93.2	81.5
	random	25+25	1:3	88.7	84.4
	k-means	25+25	1:3	88.9	86.0

表 6　画像 1 枚当りの平均計算時間
Table 6　Average computation time per an image.

非線形 SVM	ヒストグラム+非線形 SVM	ヒストグラム+近似 RBF
2.12 s	0.86 s	0.27 s

関数は，文字候補領域の検出，明らかな非文字領域の削除のために使用される．このため，ヒストグラム通過後の学習データ中に含まれる文字データ数（1076）と非文字データ数（326）は大きく異なっており，単純にランダムサンプリング等を適用した場合には固定中心点としてサンプリングされる文字データと非文字データの数は大きく異なることが予想される．そこで，本手法においては，与えられた学習データを文字データと非文字データにまず分割し，各分割データごとに指定された個数の固定中心をサンプリングする方式を採った．

実験においては，文字，非文字のデータ集合からそれぞれ 25 個ずつ計 50 個のデータのサンプリングを行い，これを固定して評価関数 (12) を最小化するパラメータ $\{\beta_\ell\}$, γ を決定した．このとき，各データの重み w_i を一様とすると，学習データの性質により前述のように文字データと非文字データの数に偏りがあるため，これが近似 RBF の性能に影響を与え，非文字データに関する識別能力の低下を招くおそれがある．自然情景中に含まれる文字領域と非文字領域では，非文字部の方が多数である場合が多いことが予想される．また，顔や自動車などの検出問題とは異なり，文字部に関しては，たとえ文字列中の一部が検出できず欠落が生じても，単語辞書などの知識利用により回復できる可能性が高い．以上より，ここでは非文字部の検出能力を向上させるために学習データ中の文字領域数（1076）と非文字領域数（326）を考慮して，文字領域と非文字領域に対する重みの比を 1：3 とするように w_i を決定した．

提案手法による識別実験結果を他の重み，式 (8) に基づく手法と比較して表 5 に示す．結果はいずれもヒストグラムに基づく識別と組み合わせた階層型の処理の場合の識別率である．サンプリング法に k-means，重みに 1：3 を採用した提案手法により，文字部の識別率は非線形 SVM に基づくものに若干劣るものの，非文字部の識別率ではむしろ勝り，カーネルの評価時間は $50/1173 \approx 0.04$ と 25 倍高速であるような識別関数が構築できたことが分かる．本論文で提案した手法は輝度ヒストグラムの形状に基づく識別関数と RBF を階層型に組み合わせた識別器であり，カーネルの評価時間の高速化がそのまま全体の処理時間の高速化となるわけではなく，画像中に含まれる文字領域数などによっても左右されるが，カーネル評価の高速化により全体の評価時間も高速化されることは明らかである．画像ごとの文字抽出処理全体がどの程度高速化されるかを見るために行った実験結果を表 6 に示す．表 6 は画像 10 枚を用いてクロック周波数 1 GHz のマシンを使用して実験を行った際の画像 1 枚当りの処理時間の平均値を示している．この表からヒストグラムを用いた階層型識別器では，単純に非線形 SVM を適用するのに比べ約 2.5 倍高速化されており，近似 RBF を用いて更に高速化を行ったものだと約 8 倍の高速化に成功していることが分かる．また，高速化がされたにもかかわらず識別率は単純に非線形 SVM を適用した場合に比べ同等以上の結果が得られていることが，表 2，表 4，表 5 より分かる．本論文では処理時間の高速化のために，Haar Wavelet に基づくスパース表現を採用しているが，この実験結果は本論文で提案した階層型識別器，近似 RBF 法を用いることにより更に処理の高速化が可能になることを示している．

ヒストグラム法を情景画像に適用した際の実験結果を図 4 に，提案手法を適用した際の結果を図 5 に示

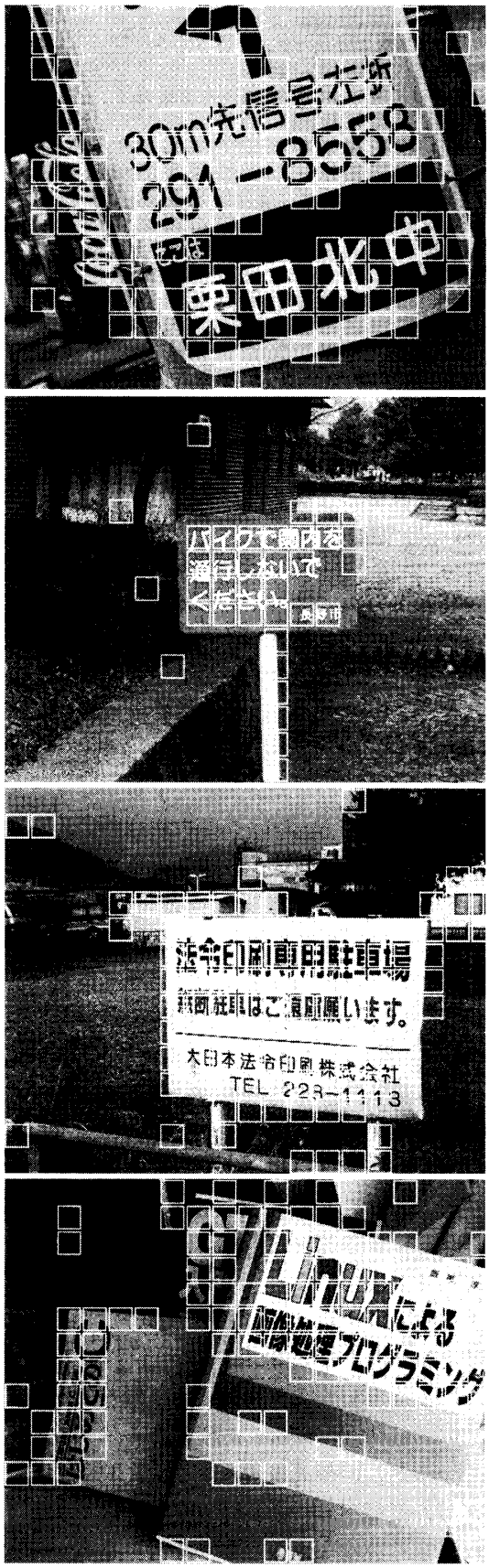


図 4　ヒストグラム単体での文字抽出結果と抽出領域数
Fig. 4　Resuls of character extraction from scene images with histogram classifier.

図 5　情景画像からの文字抽出結果と抽出領域数
Fig. 5　Resuls of character extraction from scene images by the proposed method.

す．実験に用いた入力画像は 640 × 480 の 24 bit のカラー画像である．その入力画像を 32 × 32 の領域に分割し，その各々の領域が文字領域であるのか非文字領域であるのかを提案手法により識別した．実験結果画像中で四角で囲まれている領域が文字領域と識別された領域である．

表 1～表 5 の実験において，信頼性を高めるために，データ集合をランダムに学習データとテストデータに分割し，学習と識別を繰り返した．本論文ではこの過程を 10 回繰り返し（10 度の Cross Validation），その平均値を実験結果としている．

6. むすび

本論文では，階層的な識別器を用いて情景画像から文字を抽出する手法を提案した．識別には輝度ヒストグラムの形状に基づいた識別器と非線形 SVM の 2 種を用いた．最初の段階では，識別能力は単独で十分なほど高くないものの計算コストが非常に低いヒストグラムに基づく識別器を用いて，明らかに非文字領域である部分を削除し情景画像の領域の限定を行った．次の段階では，非線形 SVM を用いて更に非文字領域を削除し最終的な結果を求めた．非線形 SVM は識別能力は高いが計算時間の点で問題がある．それを克服するために，Haar Wavelet を用いた特徴量の削減を行った．更なる高速化のために，サポートベクタ数の多い SVM を少量の中心点を用いて近似した近似 RBF も提案した．実験の結果，提案手法により，識別能力は非線形 SVM を用いた場合と同等に保ちながら，大幅な計算時間の向上が可能であることを示した．本論文で提案した文字候補領域の抽出後の，欠落部の復元を含む文字列抽出，文字列認識が今後の課題である．

文　献

[1] D. Doermann, J. Liang, and H. Li, "Progress in camera-based document image analysis," Proc. ICDAR, pp.606–616, 2003.

[2] B. Heisele, T. Serre, S. Prentice, and T. Poggio, "Hierarchical classification and feature reduction for fast face detection with support vector machines," Pattern Recognit., vol.36, no.9, pp.2007–2017, 2003.

[3] P. Viola and M. Jones, "Robust real-time face detection," Proc. ICCV, pp.1254–1259, 2001.

[4] H. Fujisawa, H. Sako, Y. Okada, and S.-W. Lee, "Information capturing camera and developmental issues," Proc. ICDAR, pp.205–208, 1999.

[5] J. Gao, J. Yang, Y. Zhang, and A. Waibel, "Text detection and translation from natural scenes," Technical Report CMU-CS-01-139, School of Computer Science, Carnegie Mellon University, 2001.

[6] T. Yamaguchi and Y. Nakano, "Extraction of place-name from natural scenes," Proc. IWFHR, pp.239–243, 2002.

[7] T. Yamaguchi, Y. Nakano, M. Maruyama, H. Miyao, and T. Hananoi, "Digit classification on signboards for telephone number recognition," Proc. ICDAR, pp.359–363, 2003.

[8] 松尾賢一，上田勝彦，梅田三千雄，"適応しきい値法を用いた情景画像からの看板文字列領域の抽出，" 信学論（D-II），vol.J80-D-II, no.6 pp.1617–1626, June 1997.

[9] 劉　詠梅，山村　毅，大西　昇，杉江　昇，"シーン内の文字列領域の抽出について，" 信学論（D-II），vol.J81-D-II, no.4, pp.641–650, April 1998.

[10] H. Li and D. Doermann, "Automatic identification of text in digital video key frames," Proc. ICPR, pp.129–132, 1998.

[11] V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.

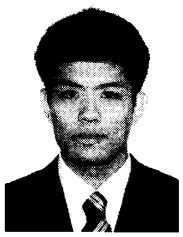
[12] C.E. Jacobs, A. Finkelstein, and D.H. Salesin, "Fast multiresolution image querying," Proc. SIGGRAPH'95, pp.277–286, 1995.

[13] T. Joachims, Making Large-Scale SVM Learning Practical, Advances in Kearnel Methods, Chapter 11, MIT Press, 1999.

[14] C.J.C. Burges, "Simplified support vector decision rules," Proc. ICML, pp.71–77, 1996.

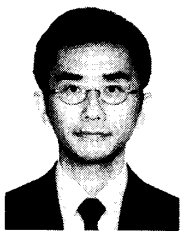
[15] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification Second Edition, John Wiley & Sons, 2000.

（平成 16 年 4 月 1 日受付，12 月 27 日再受付）



山口　拓真

平 14 信州大・工・情報卒．平 16 同大大学院修士課程了．パターン認識の研究に従事．現在，メディアドライブ（株）勤務．



丸山　稔　（正員）

昭 57 東大・工・計数卒．同年三菱電機（株）入社，先端技術総合研究所勤務．平 2～3 マサチューセッツ工科大学人工知能研究所客員研究員．平 8 信州大学工学部情報工学科助教授．工博．三次元物体認識，学習等の研究に従事．情報処理学会，IEEE，ACM 各会員．