ユーザの簡易指定に基づく情景中の文字抽出と認識

張 暁暉[†] 長井隆行[†] 樽松 明[†]

情景画像中の文字を認識し，その文字情報に基づいた様々なコンテンツを利用することができれば，非常に利便性が高い．これを実現するための技術として解決しなければならない重要な課題は，複雑な情景画像中の所望の文字をいかに抽出し認識するかということである．本論文では，ユーザが簡易に指定した領域の情報を基に，正確な文字領域を抽出し認識する手法を提案する．これは，すべての領域を自動抽出しその後ユーザに所望の領域を選択させるよりも，あらかじめユーザに位置を指定させた方が計算量や抽出精度の点で有利であると考えられるためである．この際，ユーザが正確に領域を指定する必要がないことが，使いやすさのうえで重要であると考える．提案手法は，可変テンプレートをベースとして，ユーザが指定した初期領域と同じ性質を持つ最大の領域を抽出することで最終的な文字領域を抽出する．領域抽出後は，判別分析法による 2 値化，射影により各文字パターンを切り出し認識を行う．

Extraction and Recognition of Characters in Real Scenes by Use of User's Rough Marking

Xiao Hui Zhang,[†] Takayuki Nagai[†] and Akira Kurematsu[†]

It is very convenient for us to be able to access various information easily by recognizing text automatically in real scene images. Character detection and recognition in complex real scenes are key issues to be developed for such technology. In this paper, extraction and recognition method of characters in real scenes based on the initial region, which is the user's rough specification, is proposed. This semiautomatic method has an advantage over fully automatic extraction methods from the view point of computational complexity and extraction accuracy. For example, if there is a multiple number of character regions, the user might be required to specify the region, in which he is interested. Therefore it seems reasonable to prompt the user to specify his interesting region at first. In this case, however, it is very important to allow the user to specify the character region roughly. The proposed method is based on deformable template and extracts the largest region, which has the similar characteristics to the user specified initial region. Then, the extracted region is binarized by discriminant analysis. Finally, each character is extracted by a projection and recognized.

1. はじめに

現在スキャナをベースとした OCR は，非常に高いレベルにあり，様々な場面に応用されている．今後 OCR はより手軽に活用できるカメラをベースとしたものに移行していくものと思われる．このカメラベースド OCR の重要な応用としては，情景画像中の文字認識があげられる．情景画像中の文字認識が安定して行えるようになると，たとえば店の看板などを撮影し，その情報をネットワークを介して検索することや，外国人に対する翻訳，視覚障害者の支援，ロボットのナビゲーションなどを実現することが可能となる．しかし，情景中の文字を認識するためには，いくつかの困難な問題が存在する[1)]．1 つは，いかに複雑な情景中に存在する文字を抽出するかという問題である．もう 1 つは，抽出した文字の認識に関する問題である．実際，情景中の文字は，スキャナで取り込んだ文書中の文字と異なり，変形やノイズの影響を大きく受けることが多く，また十分な解像度が得られない場合もある．これら抽出と認識の問題は独立ではないが，本論文では文字を抽出しその後に認識を行う枠組みで処理を行うことを考え，特に前者の文字の抽出に重点を置いて議論する．また著者らは，CCD カメラ付きの携帯端末（PDA）を用いた文字情報検索（翻訳）システムを構築するという観点からこの問題に取り組んでいる[2)〜5)]．この場合 OCR の性能だけでなく，システム全体としての使い勝手の良さもアルゴリズムを考える

† 電気通信大学電子工学専攻
Department of Electronic Engineering, The University of Electro-Communications

うえで考慮すべき重要なポイントである．

従来，情景画像中の文字領域抽出法は多く提案されており，たとえば局所的 2 値化[7]，輝度コントラストと空間周波数に基づく手法[8]，色情報のクラスタリング[2),9)]，文字列や看板の縁によって生じる直線の利用[11)] などがあげられる．また，文字らしさの特徴量を設定し，多くの学習サンプルでニューラルネットワークなどの識別器を訓練することで文字領域を抽出する手法も提案されている[3),10)]．これらの手法はすべて自動的な文字抽出を目指しているが，実際の情景画像中には，ユーザが興味を持たない複数の文字領域が存在する場合がある．この場合，すべての領域を自動抽出し，その後ユーザに所望の領域を選択させるよりも，あらかじめユーザに位置を指定させた方が計算量や抽出精度の点で有利であると考えられる[12),13)]．実際，PDA などの携帯端末では，ペンを用いて画像中の領域を指定することが容易に行える．ただし，指定の際の負担を小さくするために，大雑把に指定した領域から正確な文字領域を抽出することが望まれる．文献 12)，13) では，ユーザが文字領域を指定する方法を採っているが，ある程度正確に文字領域を指定する必要がある．文献 14) では，ユーザが画面の枠内に読みたい文字列をとらえることで文字領域を指定する手法について述べている．しかしこの方法では，ユーザが画像中の文字位置を気にしながら撮像する必要があり，また仮に対象となる文字領域が複数箇所あり，それらが 1 枚の画像に収まる場合であっても，複数回撮像し直す必要がある．

本論文ではこうした観点から，ユーザが簡易に指定することで文字を抽出する手法を検討する．ここでは，ユーザが簡易に指定した領域内の性質と同じ性質を持つ最大の領域が真の文字領域であると考える．提案法は，可変テンプレート（Deformable Template）を用いてユーザが指定した初期領域から領域を拡張することで最終的な文字領域を抽出する．領域抽出後は，判別分析法による 2 値化，射影による文字の切り出しを経て各文字パターンの認識を行う．

本論文は，以下次のように構成されている．2 章ではまず，文字抽出の問題について考え，提案法の基本的な方針を述べる．3 章では，提案する文字領域の抽出法について，4 章では文字パターンの切り出しと認識手法について述べる．5 章は実験結果であり，最後に 6 章で本論文をまとめる．

2. 文字抽出の問題

文字を抽出するためには，「文字とは何か？」という問題を考える必要があると思われる．この問いに対するもっともな答えは，「読むことができ，それが意味をなすもの」であろう．しかし，文字抽出後に認識を行うという枠組みでこの定義を考えると，文字抽出のためにすでにその文字が認識されている必要があることになってしまい矛盾を生じる．もっとも，適当に画像を切り取り，その部分を文字認識し，たまたま意味をなした場合にその領域を文字とするということを繰り返すことも考えられるが，PDA などの限られたリソースで実現するのは難しいと思われる．一方，文字の信号としての一般的な性質（文字らしさ）を考えてみると，おおむねコントラストが強く，エッジが多く存在する（ある程度高い空間周波数を持つ）領域であるといえるであろう．しかし，このような情報だけから文字かそうでないかを判断することは簡単ではない．たとえば，横棒があった場合それが「一」なのか何か他の物体の一部なのかを判定することは不可能である．

この問題は，文字が基本的にはいくつかが集まることで意味をなすという点を考慮することが必要であることを示唆している．ここで，情景中の重要な文字情報の多くが看板中に存在することを考慮し，看板に注目する（ただし，文字が集まった似た性質の領域であれば必ずしも一般的な意味での「看板」である必要はなく，本論文ではそのような領域を含めて看板と呼ぶ）．つまり，文字 1 つ 1 つを抽出する前に看板を抽出することで，処理の対象をいくつかの文字が集まっているある局所的な領域に移すことを考える．これにより，その後の 2 値化やレイアウト解析などの処理が容易になることが期待できる．また看板は，1 つの意味のまとまりと考えられるため，このまとまりを発見できることは，その後に行う高次の処理にも役立つ可能性がある．しかし，逆に問題となるのは，抽出されなかった看板領域の文字がいっさい認識されないことである．よって，看板領域の抽出は非常に高い精度で行われる必要がある．ここで対象とする看板の性質についてまとめると次のようになる．

(1) ある程度同じ性質（色，空間周波数）を持つ塊である．

(2) 目立たせるために背景（周囲）と異なる色を持つ．

(3) 長方形などの比較的単純な形である．

本論文ではすでに述べたように，ユーザが指定した領域（初期領域）を拡張することで看板を抽出する．ただしユーザは正確に文字領域を指定するのではなく，所望の領域の一部に印をつけることで領域を指定する．

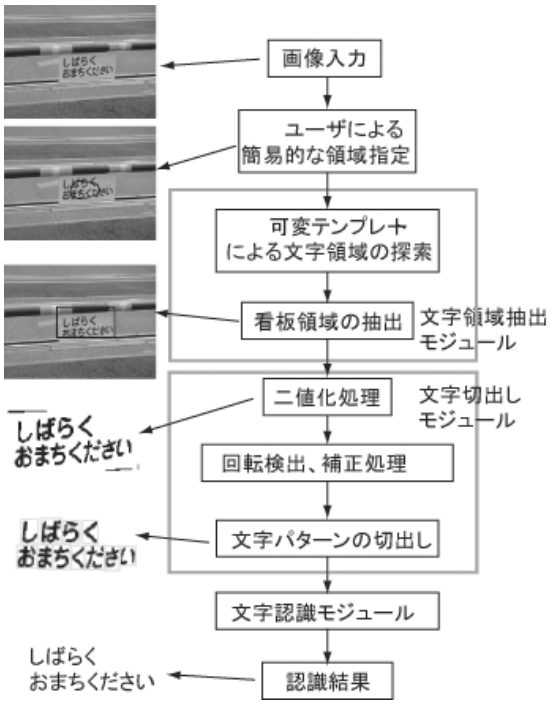

図 1 提案手法の処理の流れ
Fig. 1 Overview of the proposed method.

この際，ユーザの簡易的な指定領域が看板領域内にあり，大きくはみ出さないという前提を置く．したがってここでの問題は看板の性質 (1) を用いることで，初期領域内の性質と同じ性質を持つ最大の領域をいかに抽出するかということになる．次章では，このための可変テンプレートを用いた領域拡張法を提案する．

3. ユーザの簡易指定に基づく文字領域抽出

全体の処理の流れは，**図 1** に示すように，ユーザによる初期領域の指定，領域拡張による看板領域の抽出，文字の切り出し，文字認識の順となる．ここではまず看板領域の抽出について述べる．

3.1 エネルギー関数

提案法は可変テンプレートを用いて，ユーザ簡易指定領域中の情報に基づき領域の拡張を行う[4]．すなわち，ユーザ指定の初期領域の情報をリファレンスとして，拡張領域の情報とリファレンスを比較しながら最適領域を探索してゆく．ここで，画像中の任意の領域 R のエネルギー関数（対数尤度比）を，次のように定義する．

$$f(R) = \sum_{r \in R} \log \left(\frac{Pr(\boldsymbol{x}_r|\bar{T})}{Pr(\boldsymbol{x}_r|T)} \right) \qquad (1)$$

ただし $\boldsymbol{x}_r$，T，$\bar{T}$ はそれぞれ，画像中の位置 r における特徴量，所望の看板領域，画像全体から看板領域を除いた領域を表している．また，$Pr(\boldsymbol{x}_r|T)$ は $\boldsymbol{x}_r$ が看板領域に属する確率，$Pr(\boldsymbol{x}_r|\bar{T})$ は $\boldsymbol{x}_r$ が非看板領域に属する確率である．式 (1) は，非看板領域に属する確率と看板領域に属する確率の対数比を選択領域 R にわたり足し合わせたものなので，領域 R を変形することにより変動する．仮に確率分布 $Pr(\boldsymbol{x}_r|T)$，$Pr(\boldsymbol{x}_r|\bar{T})$ が既知であれば，$f(R)$ が最小となる領域 R が看板領域であると考えることができる．したがって，どのように $Pr(\boldsymbol{x}_r|T)$，$Pr(\boldsymbol{x}_r|\bar{T})$ を計算し，どのように $f(R)$ を最小とする領域 R を決定するかということが問題となる．

後者の問題は次節で考えることとし，ここでは前者の問題について述べる．ここで使うことのできる情報は，ユーザが概略的に指定した初期領域内の情報である．その領域を R_0 とすると，初期領域は看板領域をはみ出さないという前提を設けたので $R_0 \in T$ となる．そこでまず，R_0 内の情報を用いることで $Pr(\boldsymbol{x}_r|R_0)$，$Pr(\boldsymbol{x}_r|\bar{R}_0)$ を計算し，対数尤度比が最小となる領域 R_1 を決定する．この新たな領域 R_1 に対して再び $Pr(\boldsymbol{x}_r|R_1)$，$Pr(\boldsymbol{x}_r|\bar{R}_1)$ を計算し，対数尤度比が最小となる領域 R_2 を決定する．以上の手順を繰り返し，領域の変化が十分小さくなるか最大の繰返し数に達した場合に得られる領域 R_n を文字領域とする．したがって，n 回目の繰返しでは以下を最小化することになる．

$$f(R_n) = \sum_{r \in R_n} \log \left(\frac{Pr(\boldsymbol{x}_r|\bar{R}_{n-1})}{Pr(\boldsymbol{x}_r|R_{n-1})} \right) \qquad (2)$$

実際の $Pr(\boldsymbol{x}_r|R_n)$，$Pr(\boldsymbol{x}_r|\bar{R}_n)$ の計算にはそれぞれ，領域 R_n, 画像全体から R_n を除いた領域 $\bar{R}_n$ における特徴量のヒストグラムを用いる．つまり，

$$Pr(\boldsymbol{x}_r|R_n) = \frac{H_{R_n}(\boldsymbol{x}_r)}{N(R_n)} \qquad (3)$$

$$Pr(\boldsymbol{x}_r|\bar{R}_n) = \frac{H_{\bar{R}_n}(\boldsymbol{x}_r)}{N(\bar{R}_n)} \qquad (4)$$

となる．ただし $H_R(x)$，$N(R)$ はそれぞれ，領域 R における値 x の計数，領域 R 内の総画素数を表している．

特徴量としては HSV 色空間の色相 h_r および，輝度分解画像に対する 8×8 ブロックごとの空間周波数の重み付け平均 $\bar{w}$

$$\bar{w} = \sum_{k=0}^{4} \sum_{\ell=0}^{4} |F(k,\ell)| \sqrt{k^2 + \ell^2} \qquad (5)$$

を用いる．ただし，$|F(k,\ell)|$ は 2 次元フーリエ変換係数の絶対値である．平均の際の重みは，直流成分を原点とした場合の各周波数成分までの幾何学的な距離である．これは，文字領域が高域側に特徴が出ることを考慮したものである．また，8×8 のブロックは 4 画素ずつシフトさせることでブロック間に重なりを持たせている．したがって，空間周波数の特徴量は縦横に 1/4 ずつ縮小したものになるが，ここではこれを線形補間し 4 倍に拡大した w_r を各位置 r における特徴として用いる．空間周波数の特徴としては，DCT やウェーブレット変換などを用いることも考えられるが，特徴量のシフト不変性を考慮し，ここではフーリエ変換の振幅を用いる．また色相を用いることで，看板の抽出が影などの雑音の影響を受けにくくなると考えられる．ここで，$X(r)=[h_r\ w_r]$ とし，色相と空間周波数の独立性を仮定すると

$$\begin{aligned} f(R_n) &= \sum_{r\in R_n}\log\left(\frac{Pr([h_r\ w_r]|\bar{R}_{n-1})}{Pr([h_r\ w_r]|R_{n-1})}\right)\\ &= \sum_{r\in R_n}\log\left(\frac{Pr(h_r|\bar{R}_{n-1})}{Pr(h_r|R_{n-1})}\frac{Pr(w_r|\bar{R}_{n-1})}{Pr(w_r|R_{n-1})}\right)\\ &= \sum_{r\in R_n}\log\left(\frac{Pr(h_r|\bar{R}_{n-1})}{Pr(h_r|R_{n-1})}\right)\\ &\quad+\sum_{r\in R_n}\left(\frac{Pr(w_r|\bar{R}_{n-1})}{Pr(w_r|R_{n-1})}\right)\\ &= f_h(R_n)+f_w(R_n) \end{aligned} \tag{6}$$

と計算することができる．ただし，$f_h(R)$，$f_w(R)$ はそれぞれ色相に対するエネルギー関数，空間周波数に対するエネルギー関数である．

また，式 (3)，(4) の計算のためにはそれぞれを量子化する必要がある．量子化のレベルは実験的に次のように決定した．色相に関しては，0〜360 度を 120 段階に量子化する．空間周波数は，画像中の最大値を 1 と正規化し，0〜1 を 50 段階に量子化することとした．

3.2 領域探索アルゴリズム

本節では，エネルギー関数を最小とする領域の探索方法について述べる．可変テンプレート法では，領域の形状に関するパラメータを設定しそのパラメータを変化させることで探索を行う．問題は，どのようなパラメータを設定するべきかということである．この際，計算量の観点からするとパラメータは少ない方が望ましい．ここで看板領域の性質 (3) に注目すると，領域の形は長方形のようなある程度単純な形でよいと考えられる．しかし，仮に看板が単純な長方形であったとしても，撮像位置によっては回転や透視投影の影響により形が歪むことになる．こうした歪みの影響をどの程度考慮するかによって，設定すべき領域のパラメータ数が変化する．本論文では，以下 3 つの異なるレベルにおける探索アルゴリズムを提案する．

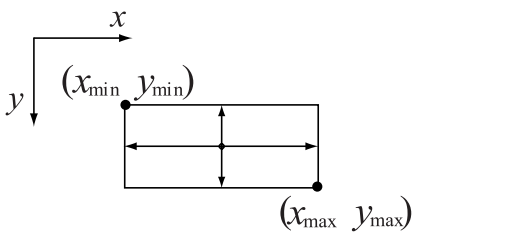

図 2 長方形探索のパラメータ
Fig. 2 Parameters for the rectangular search.

3.2.1 長方形探索

長方形探索では，初期領域から徐々に長方形の各辺を拡張してゆき，最終的にエネルギー関数を最小とする領域を文字領域とする．したがって，探索パラメータは**図 2** に示すように 4 つである．探索はまずステップサイズ s を決め，x 軸の正方向に拡張を行う．つまり x_{max} を $x_{max}+s$ としてエネルギーを計算し，エネルギーが減少した場合には，再びステップサイズ分だけ x 軸の正方向に拡張を行う．これをエネルギーが減少する間繰り返す．次に y 軸の正方向へ同様に拡張する．そして，x 軸の負方向，y 軸の負方向へと順に拡張する．以上の手順を変化がなくなるまで繰り返し，収束後はステップサイズを半減させて再び同様の手順を繰り返す．最終的には，ステップサイズが 1 で収束した時点で探索の終了となる．ステップサイズは実験的に決定することとするが，後に示す実験ではすべて 10 から始め順に，5，3，1 とした．**図 3** 上段に，実際の画像に対する領域探索結果を示した．これより，正面向きの看板に対しては有効に抽出できているが，傾斜した看板や，透視歪のある場合にはうまく抽出できないことが分かる．ただし，上段右から 2 番目のように少しの傾きであれば，看板は正確に抽出できないものの必要な文字はすべて抽出領域に含まれるため，文字認識には問題がない．実際にこの例では，すべての文字が正しく認識された．

3.2.2 回転探索

回転探索では，傾いた領域に対処するために回転角度 θ をパラメータに加える．したがって，パラメータは**図 4** に示す 5 つとなる．探索は，前述の長方形探索の 1 ステップに回転を加えることで行う．図 3 中段に

図 3 各探索法による看板抽出結果（上段は長方形探索，中段は回転探索，下段は 4 辺探索）
Fig. 3 Results of the signboard extraction by each region search method (upper: rectangular search, middle: rotational search, lower: four-side search).

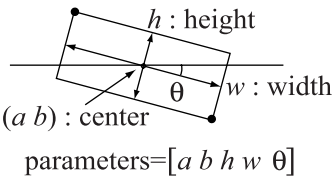

図 4 回転探索のパラメータ
Fig. 4 Parameters for the rotational search.

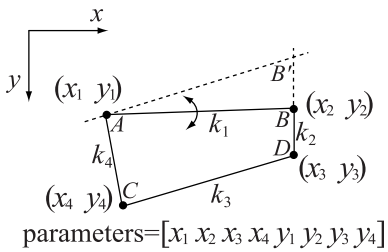

図 5 4 辺探索のパラメータ
Fig. 5 Parameters for the four-side search.

実際の画像に対する結果を示した．これより，傾斜した看板も正しく抽出できていることが分かるが，透視歪のある場合にはうまく抽出できないことが分かる．

3.2.3 4 辺 探 索

ここでは，透視歪のある領域が抽出できない問題を解決するため，四辺探索法を提案する．探索パラメータは，**図 5** に示すように 4 つの頂点座標 $A(x_1, y_1)$，$B(x_2, y_2)$，$C(x_3, y_3)$，$D(x_4, y_4)$ の 8 つである．探索は，四角形領域の各辺 (k_1, k_2, k_3, k_4) を一辺ずつ動かし，最終的に k_1，k_2，k_3，k_4 の 4 つの直線が囲む領域のエネルギーが最小となるとき，その領域を看板領域として抽出する．各辺の移動は，たとえば k_1 であればまず座標 A を中心として回転させることで行う．A を中心として回転した k_1 と，k_2 を延長したものとの交点を新たな座標 B' とし，四角形 $AB'CD$ 内のエネルギーが最小となるときの B' を求める．次に，B' を中心として再び k_1 を回転しエネルギーが最小となる k_4 との交点を新たな点 A' とする．以上の手順を各辺に対して繰り返すことにより探索を行う．

実験結果を図 3 下段に示す．これより，傾斜がありかつ透視投影された看板でも正しく抽出することができていることが分かる．この方法により，正面，傾き，透視領域などほぼすべての領域に対応できる．また，長方形でなくても四角形であれば基本的には対応可能である．ただしこの方法では処理が多少複雑となるため，処理時間が長方形探索に比べ約 3 倍程度長くなる．また，ここで述べた 3 つの探索法によって得られる領域がグローバルな解（全探索によって得られる最小の $f(R)$ を持つ領域 R）であることは保証されないことに注意が必要である．しかし，実際の情景画像を用いた結果からこのことがあまり問題とならないことが確かめられる．

図 6 ユーザによる領域の指定法．(a) 全体を丁寧に囲む指定，(b) 看板の一部を囲む指定，(c) 線による指定
Fig. 6 Region specification by user's marking. (a) specification by circling round the whole region carefully, (b) specification by marking up a part of the signboard, (c) specification by curved line.

図 7 看板の上下で文字の白黒が反転する例
Fig. 7 Examples of the signboard with characters of opposite binary level in its upper and lower parts.

3 つの探索法の内どれを用いるかは，使用できる計算リソースによるといえる．PDA などの携帯端末で実現するためには，当然，より少ない計算量で実現できる長方形探索が望ましい．また本論文で想定しているアプリケーションでは，ユーザが対象となる文字領域にカメラを向けることが前提であり，物理的に不可能な場合を除いては対象を正面から撮影することが可能である．また長方形探索の例で見たように，多少の回転があって看板領域が完全に抽出できない場合でも，抽出領域に文字がすべて含まれていればほとんどの場合は問題がない．したがって，回転や透視歪を考慮しないこともそれほど問題がないと考えられる．こうしたことから，以降の議論は長方形探索を用いることを前提として行い，実験においても特に断りのない限り長方形探索を用いることとする．

3.3 初期領域

ここでは指定の仕方を特に規定せず，ユーザが自由に指定できることとする．実際にユーザに領域を簡易指定してもらうと，**図 6** の 3 つのような指定の仕方に分類できることが分かった．初期領域は，この指定領域（実際には座標の集合）を囲む最小の矩形領域とした．

4. 文字パターンの切り出しと認識

前章の手法で抽出した看板領域から個々の文字パターンを切り出し，これらを文字認識することで最終的な結果を得る．本章では，領域からの文字パターンの切り出しおよび認識について述べる．

4.1 文字パターンの切り出し

抽出された領域から文字パターンを切り出すために，まず抽出領域を判別分析法を用いて 2 値化する．文献 7) のような動的閾値を用いることも考えられるが，ここでは看板領域のみを 2 値化する点を考慮し，単一の閾値を用いることとした．また 2 値化後は画素数の多い方を背景，少ない方を文字とする．

次に，ラベリングによる連結成分の分析を行い，小さな成分や大きすぎる成分をノイズとして除去する．ただし看板では，**図 7** に示すように上下で白黒が反転している場合があるため，大きい連結成分については，白黒を反転しその領域のみ再びラベリングとノイズ除去を行う．

その後，隣接差分法[4] で文字列の傾きを検出しその補正を行う．これは，次のブロックで行う射影による文字切り出しをより精度良く行うために，小さな角度を補正することを目的としている．隣接差分法では，2 値画像を q 度回転させ各軸に射影する．そして，水平方向への射影を $h_x(q)(x)$, 垂直方向への射影を $h_y(q)(y)$ とすると，次のような差分の自乗和を計算する．

$$\bar{h}(q) = \sum (h_x(q)(x) - h_x(q)(x+1))^2/\alpha_x(q) + \sum (h_y(q)(y) - h_y(q)(y+1))^2/\alpha_y(q) \quad (7)$$

ただし，$\alpha_x(q)$，$\alpha_y(q)$ はそれぞれ，軸を q 度回転さ

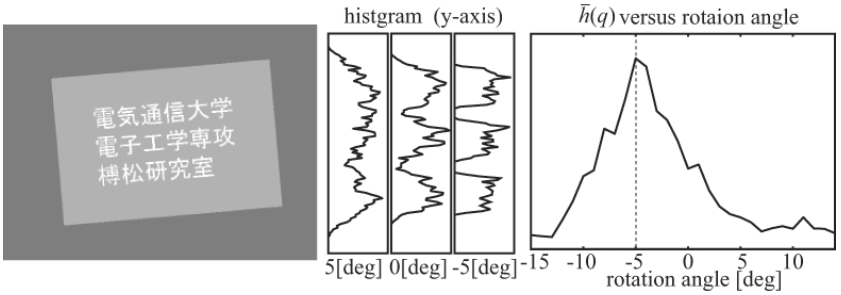

図 8 隣接差分による傾き角度の推定
Fig. 8 Estimation of the slant angle.

せた場合の x 軸射影の計数がゼロでない成分数，y 軸射影の計数がゼロでない成分数を表している．$\bar{h}(q)$ は一般に，q が文字列の方向と一致したときに最大値をとると考えられる．**図 8** は，左の画像を 2 値化し角度の推定を行った場合の例である．図の中央は 5 度，0 度，-5 度の y 軸射影を表しており，右図が回転角度に対して $\bar{h}(q)$ をプロットしたものである．実際，図左の画像は水平から左上を中心に -5 度時計回りに回転させたものであり，$\bar{h}(q)$ の値も -5 度で最大値をとっていることが分かる．ここでは，射影のための回転補正が目的であるので，-15 度から $+15$ 度までを 1 度ずつ $\bar{h}(q)$ を計算し，その最大値を求めその角度により回転を補正する．

このように回転補正された領域を水平，垂直方向へ射影し，その谷を探すことで文字候補ブロックを切り出す．ここでは縦書・横書が混在したようなレイアウトに対処するため，切り出された各ブロックの縦横のサイズをチェックし，それらが閾値以上の場合は，再び水平，垂直方向への射影による切り出しを繰り返す．この際，サイズの閾値は 30 ピクセルとした．このサイズによる制約は多くの過分割を防ぐためのものである．

さらに切り出した各候補ブロックをアスペクト比（高さ/幅）により合併，分割することで最終的に各文字を抽出する．これは，射影によるブロックの切り出しが完全ではなく，過分割や未分割が起こるためである．ブロックのアスペクト比が 2.0 より大きい場合はブロックが縦長であるため，横のブロックと合併するか縦に分割する．まず左右に同様な高さのブロックがある場合は，距離の近い方と合併する．ただし合併後のブロックのアスペクト比が，もとのアスペクト比に比べ 1 より遠ざかる場合は合併しないこととする．また，縦に同様の幅のブロックがある場合は，縦に分割する処理を行う．分割処理では，ブロックの高さを幅で割った値で高さを等分し，その等分した位置に近い x 軸射影における谷を最終的な分割位置として分割する．一方，ブロックのアスペクト比が 0.5 より小さい場合はブロックが横長であり，縦のブロックと統合するか横に分割する．手順は縦長の場合と同様であるが，縦と横が入れ替わる．

4.2 文字認識

本論文では，切り出された文字の認識に修正 2 次識別関数をベースとした手法を用いる[15]．特徴量としては，文字を 7×7 の 49 のブロックに分割し，各ブロックにおける輪郭線の方向ヒストグラムを 4 方向について計算し，ガウスフィルタでぼかしたもの計 196 次元と 1 次の peripheral 特徴を上下左右 4 方向各 15 次元分を用いた．したがって特徴量は合計 256 次元である．また，すべての特徴は 0.5 乗のべき変換を施した．学習には 56 種類の PC フォントを用いた（各 3,214 字種）．実際に情景画像から手動で切り出した文字を認識したところ，認識率は約 90%であった（評価総文字数：20,650，正解：17,264，同系文字：1,503）．また，学習に使用していない PC フォントでの認識率は 99%であった．誤認識した情景文字約 2,000 文字の原因を調べたところ，ノイズや欠損，回転，低解像度などの影響があることが分かった．ここでは特に回転や低解像度の文字に対応するために，±10 度回転させた文字セットと解像度を 1/2 に下げた文字セットを加えて学習したところ，認識率は約 93%まで向上した．つぶれや変形，ノイズをともなう文字の認識は文献 16), 17) などで論じられており，こうした手法を用いることでさらなる高精度化を図ることが可能である．また，文字画像の高解像度化により認識率が向上することも報告されており[18),19)] こうした手法の適用も今後の課題である．

5. 実　　験

5.1 データベース

情景中の文字認識を研究するうえで，文字を含む情景画像データベースが必要不可欠である．現在のとこ

図 9 不十分な拡張の例
Fig. 9 Examples of the insufficient region expasion.

ろ，共通に利用できる日本語（漢字・かな）のラベル付きデータベースがないため，著者らは看板（文字）を含む画像を収集し，正解ラベル（正解文字コードと座標）付けを行っている．画像は様々な情景のもの約 2,000 枚があり，今回の実験は，正面近くから撮影されたものの中から無作為に選んだ 100 枚を用いて行った．

5.2 実験結果

上述の画像に対し，被験者 10 名に実際に領域を指定してもらい，その情報に基づいて文字抽出を行った．画像のサイズは，すべて 320 × 240 である．この際，各人の目的に応じて認識させたい看板を自由に指定してもらったが，指定領域が看板から大きくはみ出さないようにという指示を与えた．また被験者には実験前に，実験データとは異なる画像数枚でシステムを使用してもらった．

実験の結果，看板領域の抽出率は 97.7%であった．ただし看板の抽出は，所望の文字がすべて含まれ正しい看板境界から 10 ピクセル以上はみ出さない場合を正解としている．比較のために，文献 2) の自動抽出手法を同じ 100 枚の画像に対して適用した．その結果，看板抽出率は 93%であり，提案手法とそれほど大きな差異は認められなかった．しかし，自動抽出の手法では文字領域ではない領域を看板として抽出してしまう誤抽出領域が 16 存在した．したがって，抽出精度（正しく抽出した領域数/抽出した全領域数）は 85.3%となり，こうした誤抽出領域が存在しない提案手法が精度の点で有効であることが分かる．

提案手法における抽出失敗は，領域の拡張が十分ではなくユーザの所望の文字が一部含まれない合計 23 枚であった．これらの不十分な拡張のうち最も多かった原因は，看板内の文字が複数の色であるにもかかわらずユーザがそのうちの 1 色のみ指定した場合であり，12 枚あった．また，指定領域が小さすぎるために十分に情報が得られず拡張が進まなかったものが 6 枚，ノイズの影響を受けて拡張が十分でなかったものが 5 枚であった．抽出失敗の例を**図 9** に示す．図の左は「土地」が緑,「建物」が青で書かれているが，ユーザが「建」にのみ印をつけたため，緑の部分には領域が拡張されなかった．ユーザが「建物」の部分だけを所望としていればこれで正解であるが，実際には看板全体を所望の領域としていたので，抽出失敗である．図 9 の中央は，初期領域が小さく，その領域に十分な色情報，周波数情報がなかったために拡張が進まなかったと考えられる．図 9 の右は，初期領域にノイズ（看板の地の部分の塗装がはげて白く細かいノイズとして現れている）が入り込みそのノイズが途切れる「止」の中央で拡張が止まっているのが分かる．ただしこれらの画像では，最大でも 3 回ユーザが指定し直すことですべて正しい領域を抽出することができた．

多くのユーザは，図 6 に示した看板の指定方法の中で，看板の一部を囲む指定法を用いていた．また，細長い看板に対しては線による指定が多く見られた．これら 2 つの指定方法と抽出結果の間には，特に強い相関関係は見られなかった．したがって，どのように指定するかということよりも，初期領域の看板領域に占める面積の割合が抽出の成否における重要な要因であるといえる．ただし，一度抽出に失敗した場合，ユーザは全体を丁寧に囲む指定法を用いる傾向があり，これが最終的にすべてを抽出できた要因である．

同様の条件で，回転探索，4 辺探索を用い看板抽出を行った．その結果，回転探索では抽出率が長方形探索と同じであったが，4 辺探索では抽出失敗が 18 枚に減り抽出率が 98.2%と若干の向上が見られた．一方計算時間に関しては，長方形探索の平均を 1 とした場合，回転探索は 1.5，4 辺探索は 3.2 であった．

また，長方形探索の結果に対する文字抽出・認識の結果は次のようになった．実験においてユーザが意図した看板領域には 9,382 文字存在し（ただし，高さもしくは幅が 10 ピクセル以下のものは対象外とした），そのうち 8,955 文字を抽出した．したがって，文字抽出率は 95.4%であった．正しく認識した文字は，7,973 文字あり，総文字数に対する認識率は，85.0%であった．以上の結果は，ユーザが指定し直さなかった場合であり，仮に指定をし直して看板領域が 100%抽出され

図 10 文字抽出・認識の例
Fig. 10 Examples of the character extraction and recognition.

たとすると，文字抽出率は 96.0%, 認識率は 85.3%となる．またいずれの場合も，誤抽出率（文字でないものを文字として抽出した率）は約 2%であった．文字抽出・認識の例を**図 10** に示す．図では上段が入力画像にユーザ指定領域，看板抽出結果を重ねたもの，下段は文字抽出結果と認識結果を示している．

抽出できなかった文字のほとんどは，看板領域がうまく 2 値化できなかったか，文字切り出しがうまくいかなかったためである．また誤認識に関しては，文字が小さくつぶれてしまった，2 値化の問題，ノイズ除去によって文字の一部が欠けてしまった，などが主要な原因であった．認識精度の向上には，単語辞書などの高次の情報を用いることが必要であると考えられる．

さらに，これまで述べた手法を小型のノートパソコン（CPU は crusoe 600 MHz）に実装し USB カメラを用いてテストしたところ，処理時間の平均は約 1.6 秒であった．

6. おわりに

本論文では，情景中の文字抽出と認識について述べた．提案法は，可変テンプレートを用いた領域拡張法をベースとしており，ユーザが簡易的に指定した初期領域の情報を用いることで文字領域を高い精度で抽出できる．被験者 10 名に対して行った実験の結果，領域の抽出率は 97.7%, 文字抽出率は 95.4%, 認識率 85.0%であった．また，自動抽出の手法と比較し，文字領域の抽出精度が大幅に向上することを示した．今後，実用化に向けてさらなる高精度化や PDA への実装などを行う予定である．また，こうした研究を進めるうえで重要である画像データベースの整備を進める必要がある．

謝辞 本研究を進めるにあたり，多くの有意義なご助言をいただいた，金子正秀教授ならびに日立製作所中央研究所知能システム研究部の諸氏に感謝する．

参 考 文 献

1) Fujisawa, H., Sako, H., Okada, Y. and Lee, S.-W.: Information Capturing Camera and Developmental issues, *Proc. ICDAR '99*, pp.205–208 (1999).
2) 傅，長井，金子，樽松：情景画像からの看板領域および看板文字の自動抽出，映像情報メディア学会誌，Vol.57, No.7, pp.819–828 (2003).
3) 長井，影広，金子，樽松：情景画像中の文字及び看板領域の抽出，信学技報，DSP2000-183, pp.103–108 (2001).
4) 張，長井，樽松：変形テンプレート法を用いた情景画像中の文字領域抽出，情報処理学会第 66 回全国大会予稿集（2），pp.399–400 (2004.3).
5) 五十嵐，張，長井，樽松：ビジュアルアテンションと領域拡張法を用いた情景画像の文字領域抽出，電子情報通信学会全国大会，p.206 (2004.3).
6) 後藤，阿曾：様々な画像に適用できる文字パターン抽出手法について〜サーベイ及び一構成例，信学技報，PRMU99-234 (2000.2).
7) 塩：情景中文字の検出のための動的 2 値化処理法，電子情報通信学会論文誌，Vol.J71-D, No.5, pp.863–873 (1988).
8) 劉，山村，大西，杉江：シーン内文字列領域の抽出について，電子情報通信学会論文誌，Vol.J81-D, No.4, pp.641–650 (1998).
9) 上羽，武田，岡田：色線処理によるカラー画像からの文字領域の抽出，信学技報，PRU94-28 (1994.9).
10) Li, H., Doermann, D.S. and Kia, O.: Automatic Text extraction and tracking in digital

video, Univ. Maryland, College Park, Tech. Report, LAMP-TR-028, CAR-TR-900 (1998).
11) Clark, P. and Mirmehd, M.: Reconising text in real scenes, *Internationa Journal on Document Analysis and Recognition*, Vol.4, pp.243–257 (2002).
12) 渡辺，岡田，金，武田：シーン中のテキスト翻訳，画像の認識・理解シンポジウム '98，pp.99–104 (1998).
13) 園，渡辺，岡田：カメラつき携帯電話を利用したシーン中の文字の認識と翻訳 TCMP: Translation Camera on Mobile Phone, 信学技報，TL2003-51, PRMU2003-237, pp.37–42 (2004.2).
14) 山田，仙田：携帯カメラを用いたユビキタス情報インタフェース，情報処理学会学会誌，Vol.45, No.9, pp.923–926 (2004).
15) 若林，鶴岡，木村，三宅：手書き文字認識における特徴量の次元数と変数変換に関する考察，電子情報通信学会論文誌，Vol.J76-D-II, No.12, pp.2495–2503 (1993).
16) 大町，阿曽：低品質文字認識におけるつぶれを動的に補正する部分空間法，電子情報通信学会論文誌，Vol.J82-D-II, No.11, pp.1930–1939 (1999).
17) 森，倉掛，杉村，塩，鈴木：背景・文字の形状特徴と動的修正識別関数を用いた映像中テロップ文字認識，電子情報通信学会論文誌，Vol.J83-D-II, No.7, pp.1658–1666 (2000).
18) 中里，長井，嶺，酒匂，槫松：サブバンド EHMM を用いた低解像度文字画像の高解像度化，第 1 回情報科学技術フォーラム FIT2002 (2002).
19) Baker, S. and Kanade, T.: Limits on Super-Resolution and How to Break Them, *IEEE Trans. PAMI*, Vol.24, No.9 (2002).

(平成 16 年 9 月 28 日受付)
(平成 17 年 9 月 2 日採録)

張　暁暉

平成 16 年電気通信大学大学院前期博士課程修了．平成 16 年情報処理学会第 66 回全国大会学生奨励賞受賞．在学中画像処理・認識に関する研究に従事．

長井　隆行（正会員）

平成 5 年慶應義塾大学理工学部電気工学科卒業．平成 9 年同大学大学院博士課程修了．博士（工学）．平成 10 年電気通信大学電子工学科助手．平成 16 年同大学大学院電子工学専攻助教授．平成 15 年カリフォルニア大学サンディエゴ校客員研究員．マルチメディア信号処理，知能システムに関する研究に従事．

槫松　明（正会員）

昭和 36 年早稲田大学理工学部電気通信学科卒業．同年 KDD 入社，研究所にてパターン認識，音声情報処理等の研究に従事．昭和 61 年 ATR 自動翻訳電話研究所社長．平成 5 年電気通信大学電子工学科教授．平成 9 年同大学大学院電気通信学研究科教授．平成 16 年同大学名誉教授．専門は音声認識，知的ヒューマンインタフェース，音声言語理解等．工学博士．