社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

特徴点探索による孤立文字認識

根岸　和也† 　岩村　雅一†† 　大町真一郎† 　阿曽　弘具†

† 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
†† 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
E-mail: †{negishi,machi,aso}@aso.ecei.tohoku.ac.jp, ††masa@cs.osakafu-u.ac.jp

あらまし　画像に含まれている文字の情報は，非常に利用価値の高いものであるが，その文字が時計盤のように孤立している場合は，従来手法では切り出すことができないため，有効に利用することができない．本稿では，このような孤立文字に対して，文字の持つ本質的な特徴を利用して，切り出しと認識を同時に行う手法を提案する．具体的には，特徴点として角と曲線部分を検出し，使用する．特徴点の探索には，文字のエッジの方向を利用した．これらの特徴点の探索結果を総合的に判断することで文字の切り出し箇所を決定する．また，従来法では参照する画像と認識したい文字のフォントが異なる場合は，認識率が低下してしまう問題が存在したが，実験により提案手法がこの問題に有効であることを確認した．
キーワード　孤立文字，特徴点，エッジの方向

Isolated Character Recognition by Searching Feature Points

Kazuya NEGISHI†, Masakazu IWAMURA††, Shinichiro OMACHI†, and Hirotomo ASO†

† Graduate School of Engineering, Tohoku University
6–6–05 Aoba, Aramaki, Aoba-ku, Sendai-shi, 980–8579 Japan
†† Graduate School of Engineering, Osaka Prefecture University
1–1 Gakuencho, Sakai-shi, Osaka, 599–8531 Japan
E-mail: †{negishi,machi,aso}@aso.ecei.tohoku.ac.jp, ††masa@cs.osakafu-u.ac.jp

Abstract　A character information in an image is valuable. However, if a character is isolated as in the face of a clock, it is not easy to extract the character region. We propose a method to identify an isolated character by extracting feature points of characters. We use corners and curves as the feature points. Using edge direction of each point of an image, these feature points are extracted, and the position of a character is determined. The effectiveness of the proposed method is shown by by experiments.
Key words　isolated character, feature point, edge direction

1. まえがき

画像に含まれている文字情報は，非常に利用価値の高いものである．特に車の自動運転化に関する技術や，ロボットビジョンの分野において，画像中の文字の認識は必要不可欠な技術である．しかしながら，切り出された活字文字に対する認識精度は実用上十分であるが，それを支える文字切り出し技術には制約が多い．従来の切り出し手法では，文字列 (行) を手がかりとし，文字列を抽出した後に個々の文字に分割する [1] ことが一般的である．従って，文字列を成さない「孤立した文字」を切り出すことは容易でない．また，数式認識においては，数式構造の複雑さから文字列の抽出は難しく，従来法の適用は困難である [2]．さらに，接触文字と呼ばれる，2 文字以上が接触した文字の認識も問題となっている．

そこで，我々はアクティブ探索法 [3] を利用して「孤立した文字」の切り出しと認識を同時に行う手法を提案した [4] [5]．この手法は，従来法では認識が難しかった接触文字に対しても適用でき，参照する文字と認識したい文字が同一フォントの場合は高い認識精度が得られたが，フォントが異なる場合には必ずしも高い認識精度が得られなかった．

そこで本論文の目的は，文献 [5] で提案した手法を，フォントが異なる場合にも認識出来るように改良し，高い認識精度を持つ切り出し手法を得ることである．本論文で提案する手法は，文字の持つ本質的な特徴を探索し，検出された箇所を元に，文字の切り出し位置を決定する．特徴としては「角」と「曲線」を用いる．これらは本質的な情報と考えられ，特徴はフォント

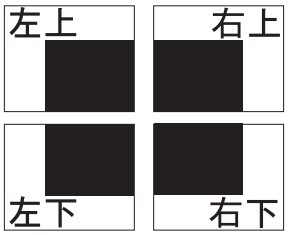

図 1 「角」の特徴点

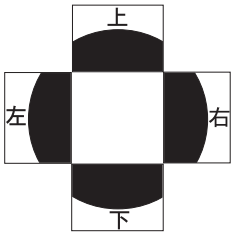

図 2 「曲線」の特徴点

が異なっても失われることは少ないため，フォントが異なる場合に対して，非常に効果があると考えられる．

2. 特徴点を探索するための準備

2.1 入力画像と参照画像

本手法では，切り出したい文字を含む画像を入力画像，個々の文字の画像を参照画像と呼ぶ．参照画像としては，対象文字である英数字 62 文字をあらかじめ用意して置く．

2.2 エッジの方向を利用したヒストグラム

本手法では，画像から Sobel Filter を用いてエッジを抽出し，エッジ方向を横方向と縦方向のエッジの強度から算出し，刻み幅 θ[rad] で量子化する．ただし，雑音の影響を排除するために，エッジの強度が一定値以下の画素についてはエッジを持たない画素と判断する．そして，部分領域を設定し，部分領域に含まれるそれぞれのエッジ方向の画素の数とエッジを持たない画素の数を要素とするベクトルを考え，ベクトルの要素の総和が 1 になるように正規化したものを特徴量とする．エッジの方向の刻み幅が θ であるから，エッジを持たない画素の数と合わせて部分領域の特徴量の次元数は $\frac{2\pi}{\theta}+1$ となる．この特徴量を，以下エッジヒストグラムと呼ぶ．

2.3 特 徴 点

文字認識で使われる特徴点としては，端点・分岐点・交点・屈折点等が検討されている [6]．本論文では文字の判別に有効であり，探索にも適していると考えられる「角」と「曲線」を用いる．角として「左上」「左下」「右上」「右下」の 4 種類 (図 1)，曲線として「上」「左」「右」「下」の 4 種類 (図 2) の計 8 種類の特徴点を用いる．これらの特徴点のエッジヒストグラムを，図 1 と図 2 の画像からあらかじめ作成しておく．

フォントの異なる文字の特徴点も柔軟に検出するためには，可能な限り θ を大きくしたほうがよい．一方，複雑な構造を持つ特徴点を検出するためには，θ を小さくする必要がある．そのため，特徴点の複雑さに応じて θ を変え，角は $\theta=\pi/2$，曲線は $\theta=\pi/8$ としてエッジヒストグラムを構築する．

(a) Arial に対して，左上の角を抽出した時　(b) Franklin に対して，左上の角を抽出した時

図 3　斜体字に対する角の検出の比較

具体的には「角」の場合，角度の範囲を $[-\frac{\pi}{4},\frac{7}{4}\pi)$ とし，

$$\left[-\frac{\pi}{4}+\frac{\pi}{2}n,\frac{\pi}{4}+\frac{\pi}{2}n\right) \qquad 0\leqq n\leqq 3 \tag{1}$$

の各区間に含まれる角度を持つ画素の数を数えてエッジヒストグラムとする．区間を上記のように広く設定することで斜体字に対して有効である様子を図 3 に示す．これは，左上の角の特徴を検出した結果である．それぞれの文字の左上のグレーの小領域が検出箇所である．この結果より，斜体字に対しても角を検出できていることが分かる．

「曲線」の場合，角度の範囲を $[-\frac{\pi}{16},\frac{31}{16}\pi)$ とし，

$$\left[-\frac{\pi}{16}+\frac{\pi}{8}n,\frac{\pi}{16}+\frac{\pi}{8}n\right) \qquad 0\leqq n\leqq 15 \tag{2}$$

の各区間に含まれる角度を持つ画素の数を数えてエッジヒストグラムとする．

2.4 類 似 値

特徴点のエッジヒストグラムと，探索領域 (適切なサイズの部分領域) 内のエッジヒストグラムがどれだけ似ているかを示す尺度として，Swain らが提案した類似値を用いる [7]．H を入力画像または参照画像中の探索領域内のエッジヒストグラム，M をある特徴点のエッジヒストグラム，Q をエッジヒストグラムの次元数とすると，H と M の類似値 S_{HM} は以下の計算によって求められる (H_i と M_i はそれぞれ H，M の要素である)．

$$S_{HM}=\sum_{i=1}^{Q}\min(H_i,M_i) \tag{3}$$

2 つのエッジヒストグラムは，共に和が 1 になるように正規化されているので，類似値の最大値は 1 ，最小値は 0 である．類似値が一定値以上の箇所を特徴点 M の検出箇所とする．

3. 孤立文字の切り出しおよび認識

3.1 参照画像からの特徴点の検出

まず，参照画像からすべての特徴点を検出し，各特徴点の種類と位置を求める．これを特徴位置表現と呼ぶ．位置の情報は，参照画像の左上の位置から検出した特徴点の位置へのベクトルで表す (図 4 参照)．このベクトルを投票ベクトルと呼ぶ．投票

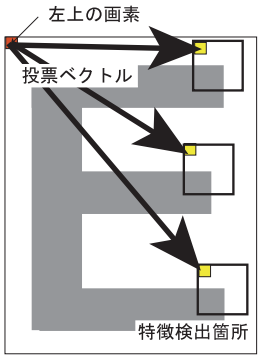

図 4　参照画像からの特徴点検出

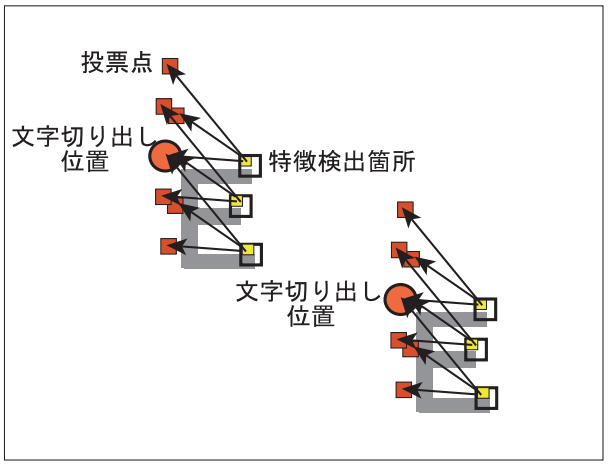

図 5　入力画像からの特徴点検出と投票

ベクトルは以下のように定める．まず，参照画像の左上の画素を原点とし，そこから右に x 軸，下に y 軸を取る．文字 k の参照画像から検出された i 番目の特徴点の個数を $M_i^{(k)}$ とおく（$i = 1 \sim 8$）．検出された特徴点 i に番号をつける．j 番目の特徴点 i を検出した位置 $(x_{ij}\ y_{ij})$ が投票ベクトル $\mathbf{f}_{ij}^{(k)}$ である．こうして得られた投票ベクトルのすべての組（式 (4)）が，文字 k の特徴位置表現である．

$$\mathbf{f}_{ij}^{(k)} = (x_{ij}, y_{ij}) \qquad 1 \leqq i \leqq 8 \qquad 1 \leqq j \leqq M_i^{(k)} \tag{4}$$

参照画像から検出された特徴点の全個数を $F_S^{(k)}$ とおくと，

$$F_S^{(k)} = \sum_{i=1}^{8} M_i^{(k)} \tag{5}$$

と表すことができる．この値は **3.2** で用いる．

3.2　文字切り出し

文字切り出しでは，まず入力画像から 8 種類の特徴点をすべて検出する．そして，検出された各特徴点から，特徴位置表現を用いて投票を行う（図 5 参照）．入力画像から検出された特徴点 i の位置が $\mathbf{x}_i$ のとき，投票ベクトル $\mathbf{f}_{ij}^{(k)}$ に対して次のように投票点 $P_{ij}^{(k)}$ を定める．

$$P_{ij}^{(k)} = \mathbf{x}_i - \mathbf{f}_{ij}^{(k)} \quad 1 \leqq j \leqq M_i^{(k)}, k : \text{すべての文字} \tag{6}$$

これは，検出された特徴点をもとに，文字の左上の点らしい場所に投票することを意味する．

次に投票点付近での投票方法を具体的に述べる．入力画像内の各点 (x, y) における文字 k の特徴点 i の投票値を $v_{ij}^{(k)}(x, y)$ とおく．あらかじめ全ての (x, y) について $v_{ij}^{(k)}(x, y) = 0$ と初期化しておく．類似値が閾値 T_1 以上の特徴 i の位置 $\mathbf{x}_i$ に対して，$P_{ij}^{(k)}$ を中心とする半径 r 画素以内の円に含まれる領域，すなわち

$$|(x, y) - \mathbf{P}_{ij}^{(k)}| \leqq r \tag{7}$$

を満たす箇所 (x, y) に投票値 1 を与える（$v_{ij}^{(k)}(x, y) = 1$ とする）．これは，投票が行われた箇所に再度投票を行っても，投票値は 1 であることを意味する．この領域を設けることで，フォントが異なる時の投票点の位置ずれを許容し，柔軟な文字の切り出しを可能にする．

全ての特徴点の探索が終了した後，次のように $V^{(k)}(x, y)$ を求める．

$$V^{(k)}(x, y) = \sum_i \sum_j v_{ij}^{(k)}(x, y) \tag{8}$$

この $V^{(k)}(x, y)$ は，座標 (x, y) が探索中の文字画像の左上の点である可能性の高さを表している．$V^{(k)}(x, y)$ の最大値は $F_S^{(k)}$ であるから，文字を切り出す際の閾値 T_2 を，パラメータ α を用いて式 (9) で定義する．

$$T_2 = \alpha F_S^{(k)} \qquad 0 \leqq \alpha \leqq 1 \tag{9}$$

この閾値を用いて，文字 k の切り出し箇所を，

$$V^{(k)}(x, y) \geqq T_2 \tag{10}$$

を満たす箇所 (x, y) として選定する．α の値を大きくすることは，参照画像の文字の特徴点により近い箇所だけを選ぶことになる．

4.　実　　験

4.1　特徴点検出

提案した手法が文字画像のどの部分の特徴点を検出するのかを調べる実験を行った．フォント Arial の『0』『E』『M』『S』から 8 種類の特徴点を検出した結果を示す．特徴点検出に用いる探索領域の大きさは任意に変更することができるが，今回の実験では 5×5 ピクセルの領域を用いた．探索領域を画像の左上から順に走査し，それぞれの箇所で特徴点との類似値を計算し，閾値以上の類似値になった箇所を特徴点の検出箇所とする．類似値の閾値 T_1 を 0.75 とした．

結果を図 6 ~ 図 9 に示す．図 6 は『0』に対して，図 7 は『E』に対して，図 8 は『M』に対して，図 9 は『S』に対して特徴点を検出した結果である．類似値の閾値 T_1 を越えた箇所を検出箇所とし，探索領域と同じ大きさの矩形で囲って示している．グレーレベルで白に近い方が，それぞれ「角の左上」「角の右上」「曲線の上」「曲線の左」である．

特徴点の検出位置が塗りつぶされているのは，角や曲線の付近において，類似値が T_1 を越えた箇所をすべて検出している

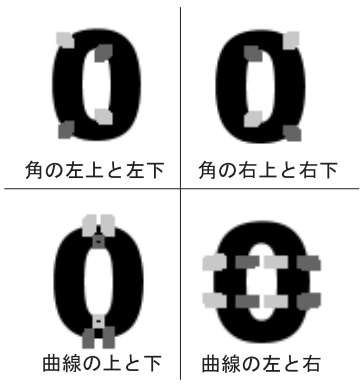

図 6 『0』の特徴点検出箇所

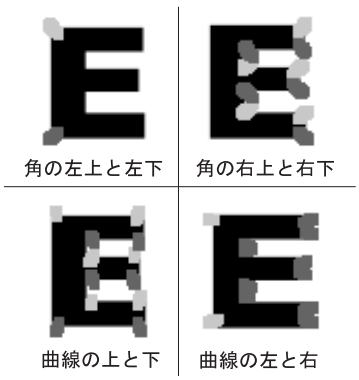

図 7 『E』の特徴点検出箇所

図 8 『M』の特徴点検出箇所

からである．また，角の特徴点を検出した時，角以外の箇所も多数検出されているが，参照画像と入力画像で同様の点が検出されているため，文字の特定には使用できる．

『0』などの曲線の特徴が，上の曲線の中心からずれたところで検出されているのは，直線と類似しないように 5×5 の探索領域より大きい領域からエッジヒストグラムを作成したからである．

4.2 認識実験

孤立した文字を含む画像に対して，以下に示す 3 通りの認識実験を行った．

（1）参照画像と入力画像とでフォントが異なる文字に対する認識実験

（2）実画像に対する認識実験

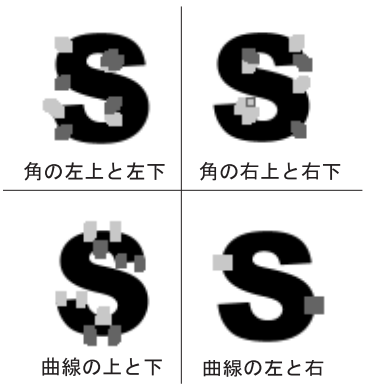

図 9 『S』の特徴点検出箇所

図 10 correct:参照画像に『A』を使用

（3）接触文字に対する認識実験

この実験では，低解像度の画像に対しても良好な検出結果を得るために，特徴の検出を行う前にあらかじめ平滑化フィルタを参照画像，入力画像の両方に掛けて置いた．

本手法は，エッジを持たない画素を除いてエッジヒストグラムを構築するので，探索領域中にわずかにエッジを持つ画素が含まれている場合でも，特徴を検出してしまうことがある．しかし，実際にエッジを持たない画素が多数を占めている場所の検出箇所は特徴点とは言い難い．そのため，特徴の検出において，探索領域に含まれるエッジを持たない画素の割合が 7 割以上含まれている場合は，その箇所での特徴検出は行わない．閾値 T_2 以上の得票数を得た領域を文字の切り出し箇所とする．

実験の評価方法について述べる．得られた結果を correct，match，miss の 3 通りで評価する．correct は探索している文字と同じ文字のみを切り出した場合で，文字の切り出しと認識を同時に行うことに相当する．図 10 はその例で，参照画像に『A』を用いた時に切り出しおよび認識が正しく行われている．match は探索している文字も切り出したが，他の箇所も切り出している場合で，文字の切り出しと認識の大分類の相当する．図 11 はその例で，参照画像に『C』を用いた時に切り出しはできているが認識はできていない．miss は探索している文字を切り出すことが出来なかった場合である．図 12 はその例で，参照画像に『K』を用いたが正しく切り出せなかった．

4.2.1 フォントの異なる文字に対する認識実験

フォント Arial の文字を参照画像とし，太文字である Arial-

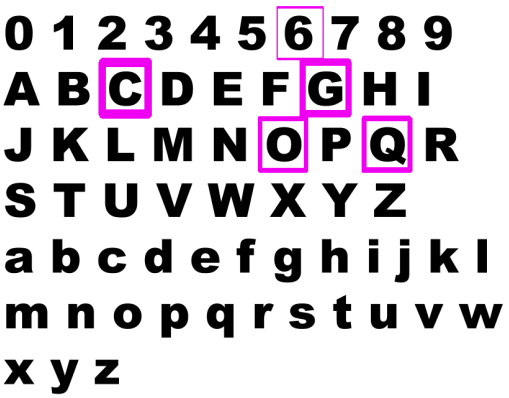

図 11　match: 参照画像に『C』を使用

図 12　miss: 参照画像に『K』を使用

フォント		correct	match	miss
ArialBlack	個数	20	41	1
	割合	32%	53%	2%
Franklin	個数	20	39	3
	割合	32%	63%	5%
丸ゴシック体	個数	24	20	16
	割合	39%	36%	26%

表 1　フォントの異なる文字に対する認識実験

Black，斜体字である Franklin，丸みを帯びている丸ゴシック体を認識対象として認識実験を行う．なお，参照画像と認識対象の文字のフォントが同じ時の認識率は 100%であった．それぞれの画像には，英数字 62 文字が含まれている．実験では，特徴を検出する時の類似値の閾値 T_1 は 0.75 とし，文字として切り出す時の投票数の閾値 $T_2 = 0.8F_S^{(k)}$ とした．認識結果を表 1 に示す．

実験の結果を見ると，認識まで正確に行えている correct の割合は 1/3 程度であるが，参照画像と同じ文字を切り出すことが出来た割合 (correct と match の割合の和) で考えると，AriaBlack と Franklin に関しては，95%以上の切り出し精度であり，非常に高い結果になっている．丸ゴシック体に関しては，約 75%の切り出し精度であった．match の場合他の文字を含んではいるものの，探索中の文字を含んでいるため，後処理を行うことで，正しい認識が行えるものと考えられる．これにより提案手法はフォントが異なる場合でも，非常に有効であると考えられる．

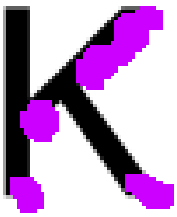

図 13　切り出しの失敗例:角の右下の特徴を検出

時計の文字盤		correct	match	miss
	個数	9	1	0
	割合	90%	10%	0%

表 2　時計の文字盤 10 文字に対する認識率

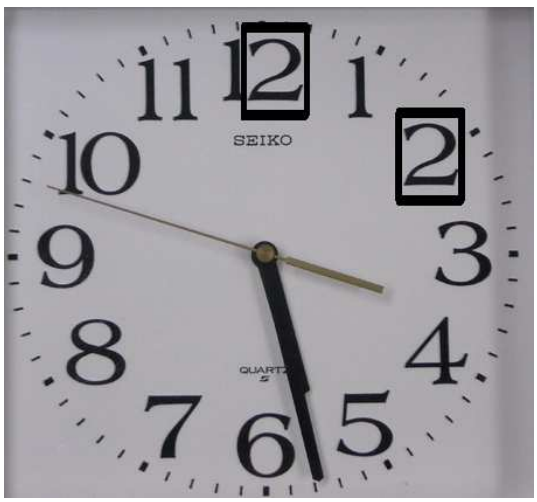

図 14　掛け時計に対する認識結果

次に切り出しが出来なかった原因を考える．切り出しが出来なかったのは ArialBlack が『K』，Franklin が『5』『7』『K』であった『K』の文字から「角の右下」の特徴点を検出した場合の検出箇所を図 13 に示す．本手法は，検出した特徴点の総個数に対する割合を閾値 T_2 として，文字の切り出し位置を決定しているために，図 13 のように，斜め直線を全て検出してしまう場合は，文字の切り出し精度の低下を招く．

また，丸ゴシック体においては，認識も切り出しも出来なかった文字が約 26%であった．精度が低くなってしまった理由として，文字を切り出す際の閾値を高く設定してしまったことが原因と考えられる．今回は閾値を $T_2 = 0.8F_S^{(k)}$ としたが，$T_2 = 0.75F_S^{(k)}$ とした時は，約 85%の文字は正しく切り出せた．このことは，文字毎に適切な閾値を設定することにより解決できると考えている．

4.2.2　実画像に対する認識実験

時計の文字盤を含む画像を用いて認識実験を行った．参照画像は Arial の数字の『0』~『9』の 10 文字である．特徴を検出する時の閾値 T_1 は 0.75 とし，得票割合の閾値は $T_2 = 0.85F_S^{(k)}$ とした．結果を表 2 に示す．また，参照画像に『2』を用いた時の認識結果を図 14 に示す．match となったのは『0』で，それ以外の 9 字種は correct となった．

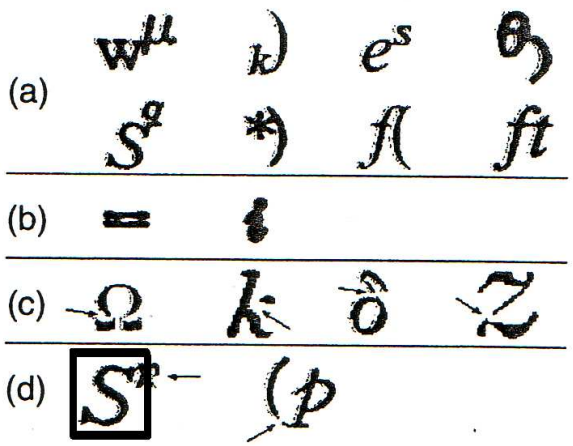

図 15　接触文字に対する認識結果:参照画像に『S』を使用

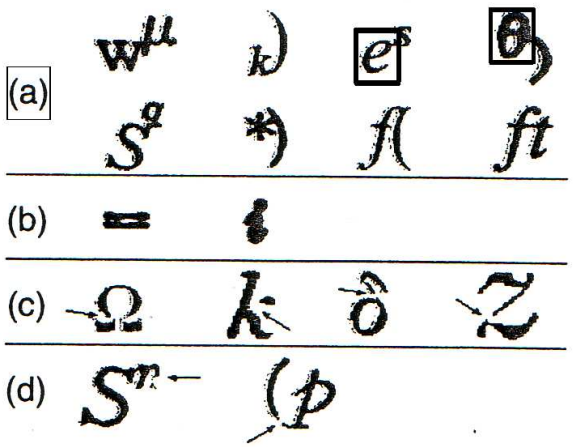

図 16　接触文字に対する認識結果:参照画像に『e』を使用

『1』と『2』に関しては文字が画像中に複数個存在するが，いずれも対象となる文字を全て切り出し，他の文字を一切切り出していなかったので，評価は correct とした．

実験の結果より，実画像に対しても本手法は非常に高い精度で切り出し・認識できていることがわかる．特に数字が複数ある『1』や『2』に関しても，全ての箇所で切り出しをすることが出来ていた．これは，従来法では出来ない切り出しであり，本手法が孤立文字の切り出しに非常に強いことが分かる．また，類似文字を出力してしまったのは『0』であり，この時他に『6』と『9』も切り出しをしてしまった．しかし，大分類という概念で考えると，十分な切り出し精度を持っていると言える．

4.2.3　接触文字を含む画像に対する認識実験

接触文字を含む画像 [2] を用いて認識実験を行う．実験で使用した参照画像は，Arial の『S』である．特徴を検出する時の閾値 T_1 は 0.75，文字を切り出す時の投票数閾値は，$T_2 = 0.8F_S^{(k)}$ とした．本手法は文字のサイズの拡大にも十分に拡張可能であるが，まだ未対応であるため，今回の実験では参照画像の大きさを認識対象に合わせることで対応した．認識結果を図 15 に示す『S』が正しく切り出せていることが分かる．また，図 16 は参照画像に『e』を用いた場合の結果で，match の評価の例である．実験の結果より，本手法は従来法では切り出しが困難である接触文字にも，精度の高い切り出しを実現しており，有効性が確認された．

今回の実験では，大きいサイズの『S』に参照画像の大きさを合わせたために，小さいサイズの『S』の切り出しは出来なかった．そのため，今後は文字の拡大と縮小を考えて切り出し行う必要があると考えられる．

5. ま と め

本論文では，エッジの方向に応じた特徴量 (エッジヒストグラム) を用いることで，文字の持つ本質的な特徴と考えられる角と曲線を検出し，それらの組を文字の特徴位置表現として用いることで，切り出しと認識を同時に行う手法を提案した．提案手法は，柔軟に角や曲線を検出することが出来るために，フォントの異なる文字に対しても有効に切り出しと認識がでる．実験により手法の有効性を示した．

今後の課題として，今回の実験では参照画像の拡大にまだ未対応であったので，動的にサイズを変化させることが挙げられる．また，切り出しができたものの認識精度が低くなってしまった原因の一つとして考えられるのは，文字を切り出す際の投票数に対する閾値の選び方である．そのため，閾値を文字やフォントに合わせて動的に変化させる手法の開発が課題である．

文　献

[1] 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抽出,” 信学論 D-II, Vol.J80-D-II, No.6, pp.1617–1626 1997.

[2] 内田誠一，野村明弘，鈴木昌和，“数学文書データベースの解析，” 信学技報，PRMU2003-234，2004．

[3] 村瀬 洋，V.V. Vinod，“局所色情報を用いた高速物体探索 — アクティブ探索法— ，” 信学論 D-II， Vol.J81-D-II，No.9，pp.2035–2042，1998．

[4] 根岸和也，岩村雅一，大町真一郎，阿曽弘具，“切り出しを同時に行う孤立文字認識法，” 平成 16 年度電気関係学会東北支部連合大会, 2H-13, 2004.

[5] 根岸和也, 岩村雅一, 大町真一郎, 阿曽弘具, “部分領域の探索による孤立文字認識, ” 第 3 回情報科学技術フォーラム (FIT2004) 一般講演論文集 第 3 分冊，I-018, pp.37–38, 2004.

[6] 森 健一 監修，“パターン認識,” 電子情報通信学会，1988.

[7] M.J. Swain and D.H. Ballard, “Color indexing,” International Journal of Computer Vision, Vol.7, pp.11–32, 1991.