

実環境文字認識のための面積比による付加情報埋込

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

Supplementary Information Embedment with Area Ratio for Camera-Based Character Recognition

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

あらまし デジタルカメラを入力デバイスとして実環境中の文字を高精度に認識するために、文字画像と同時に認識補助のための付加情報を提示する方法が検討されている。付加情報は、人間にとって自然な形で提示されること、及び、幾何学的変形に対してロバストに抽出できることが要求される。本論文では、これらの要求を満たす手法として、面積比を利用した付加情報提示手法を提案する。すなわち、文字パターンを2色で印字することを前提とし、それぞれの色の領域の面積比を特定の値とするようにデザインする。具体的には、文字に影を付加したり輪郭線を別の色とする。これらは文字パターンのデザインとして既に行われており、提案手法はその線幅や面積を変えるにすぎない。したがって、提案手法は様々な用途に広く応用することが可能である。面積比はアフィン変換に不変であり、アフィン変換を受けた環境においても誤りなく抽出されることが期待される。実際に付加情報を埋め込んだ文字パターンを作成し、デジタルカメラで撮影された画像中の文字パターンから付加情報を抽出する実験を行い、提案手法の有効性を確認する。また、付加情報を用いて文字を認識する実験を行い、認識精度が向上することを確認する。

キーワード 実環境文字認識, 情報埋込, 付加情報, アフィン不変量, 面積比

1. ま え が き

デジタルカメラにより取得された画像から、実環境中の文字を認識する手法が検討されている [1], [2]。このような画像は光学式スキャナで取得した画像と異なり、照明が非均一であり、幾何学的な変形を受けているため、高精度な認識を実現するのは非常に困難である [2]。認識を高精度化するための手法として、光学的ひずみや幾何学的ひずみをできるだけ取り除く方法 [3] や、劣化をシミュレートする劣化モデルを用いて劣化した文字パターンをそのまま認識する方法 [4] 等が検討されている。

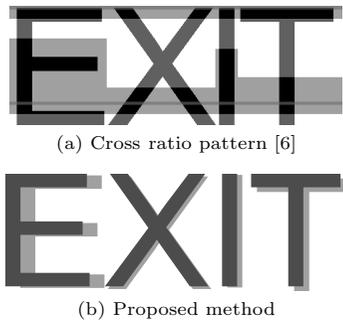
一方、情報提示側が積極的に文字情報を取得してほ

しい場合には、OCR フォントや MICR フォント [5] のようにフォントそのものの形状を工夫することで誤りのない認識を実現する方策がとられることもある。この考え方をカメラによる文字認識に適用する試みとして、文字画像と同時に認識補助のための付加情報を提示する方法が検討されている [6], [7]。すなわち、文字画像の認識と同時に付加情報の取出しを行い、両方の結果を統合することで、より高精度な認識を実現するものである。付加情報は、人間にとって自然な形で提示されること、及び、様々な幾何学的変形に対してロバストに抽出できることが要求される。カメラで撮影された画像は、一般に射影変換を受けるため、射影変換を受けても変わらない量（射影不変量）の一つである複比 [8] を用いてカテゴリー情報を文字パターンに埋め込む方法が提案されている [6]。すなわち、文字パターンを複比パターンと呼ばれるしまで表し、しまの間隔の複比の値をある特定の値とすることで情報を埋め込む。文献 [6] では文字のカテゴリー情報を埋め込むことで認識精度が向上することが示されている。しかし、複比パターンは文字パターンの外観を大きく

[†] 東北大学大学院工学研究科, 仙台市
Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan

^{††} 大阪府立大学大学院工学研究科, 堺市
Graduate School of Engineering, Osaka Prefecture University, Sakai-shi, 599-8531 Japan

^{†††} 九州大学大学院システム情報科学研究院, 福岡市
Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, 819-0395 Japan

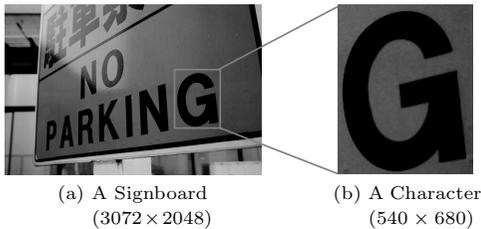


(a) Cross ratio pattern [6]

(b) Proposed method

図 1 生成される文字列

Fig. 1 Generated strings.



(a) A Signboard
(3072 × 2048)

(b) A Character
(540 × 680)

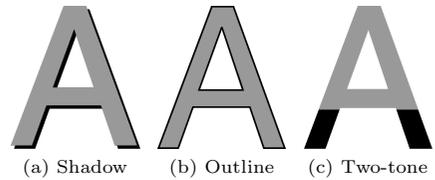
図 2 看板と看板中の文字

Fig. 2 A signboard and a character on the signboard.

変えるため、文字列として不自然になり、人間の読みやすさを損ねるといった問題点がある。例えば、文字列“EXIT”を文献[6]の手法で生成したものを図1(a)に示す。個々の文字間に不連続な背景の変化が生じていることが分かる。

本論文では、文字パターンを外観をできるだけ損ねずに安定した情報抽出を可能とする方法として、2領域の面積比を用いた付加情報の埋込法を提案する。すなわち、文字パターンを2色で印字し、それぞれの色の領域の面積比を特定の値とするようにデザインする。文字列“EXIT”を本論文で提案する手法で生成したものを(フォントはHelvetica, 付加情報は表3に示す)を図1(b)に示す。図1(a)と比較して、提案手法では自然な文字列の生成が可能であることが分かる。

2領域の面積の比はアフィン変換を施しても変化しないアフィン不変量である[9]。前述のようにカメラで撮影された画像は一般には射影変換を受けるが、図2の例に示すように、個々の文字の領域に限定すれば近似的にアフィン変換とみなすことが可能な場合も多く、アフィン不変量を用いることで実用的には十分な精度で情報の抽出が行えるものと考えられる。実際に付加情報を埋め込んだ文字パターンを作成し、デジタルカメラで撮影した画像中の文字パターンから付加情報



(a) Shadow (b) Outline (c) Two-tone

図 3 様々な埋込法

Fig. 3 Various embedment methods.

を抽出する実験を行い、提案手法の有効性を確認する。また、付加情報を用いて文字を認識する実験を行い、認識精度が向上することを確認する。

2. 面積比を用いた付加情報の埋込

文字パターンを2色を用いて印字し、それぞれの色の領域の面積比を特定の値とするようにパターンをデザインすることで付加情報を埋め込む。図3に示す例のように、様々な埋込の形態が考えられる。(a)は影の部分とそれ以外の部分の面積を利用したもの、(b)は輪郭線の面積とそれ以外の部分の面積を利用したもの、(c)はツートンカラーのフォントのそれぞれの色の部分の面積を利用したものである。

文字に影を付加したり輪郭線を別の色とすることは文字パターンのデザインとして既に行われており、提案手法はその線幅や面積を変えるにすぎない。更に、文字は字種ごとに固有の形をもつため、図3のどの形態で埋め込むにしても、字種によってある程度固有の自然な面積比^(注1)が存在すると考えられる。したがって、複色パターンの埋込と比較してより自然な埋込が可能になると考えられる。

3. で文字パターンの設計例を示すが、今回は埋込の形態として、図3(a)に示すような影を付加する方法を使用する。ただし、提案手法は図3(b)、図3(c)やその他文字パターンを2色で印字する方法に対して一般的に適用できる。以下の説明では、図3(a)の灰色の部分文字画像あるいは文字部分、黒色の部分を影の画像あるいは影の部分、図3(a)全体を文字パターンと呼んで区別する。

3. 文字パターンの設計

本章では、付加情報を面積比として埋め込んだ文字パターンの設計例を述べる。まず3.1で、カメラで撮

(注1): 文字は字種ごとに固有の形をもつため、例えば図3(a)においてすべての字種で同一の影幅に設定したとしても、それぞれ面積比は異なる。これを本論文では字種の自然な面積比と呼ぶ。

影された文字パターンから情報を抽出する際に生じ得る誤差について調査する．3.2 では2. で述べた字種固有の自然な面積比について考察し，3.3 ではこれらの結果を踏まえて実際に情報を埋め込んだ文字パターンを作成する．

3.1 誤差の調査

実環境中の文字パターンをカメラで撮影した場合，カメラの位置と姿勢によって幾何学的変形が生じ，また，撮像系によるぼけが生じるため，埋め込んだ値と抽出される値には差が生じる．この差のことを本論文では誤差と呼ぶ．

埋め込むべき付加情報の値は，生じ得る誤差を考慮し，できるだけ誤りのない情報抽出が行えるように決定することが望まれる．幾何学的変形と撮像系の両方によって生じる誤差を求めるためには，本来は幾何学的変形のパラメータを変化させ，各パラメータ値で実際に画像の撮影を行うことが必要になる．しかし，そのためにはパラメータごとに文字画像に対するカメラの位置と姿勢を求め，カメラをその位置と姿勢になるよう制御して撮影する必要がある，これを正確に行うことは非常に困難である．そこで，本論文ではその代わりに，

- アフィン変換を施すことによる誤差
- 撮像系によって生じる誤差

の和を求めることで，幾何学的変形のパラメータを変化させた場合の実環境での誤差の近似値を求める．

3.1.1 アフィン変換のパラメータ

アフィン変換は6個のパラメータで表せるが，平行移動のパラメータを除いた

$$P(a, b, c, d) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (1)$$

による変換のみを考える．ここでは，実際に文字パターンを撮影する状況を想定し，現実的なパラメータでの変換のみを考えるために，アフィン変換を図4(b)~(e)に示す4種類の変換の合成変換ととらえる．図4(a)がもとの画像である．(b)がアスペクト比を変える変換であり， x 軸方向に α 倍， y 軸方向に $1/\alpha$ 倍するものとする．(c)は x 軸に沿ったせん断変換であり，傾き角度を φ とする．(d)は回転変換であり，回転角度を θ とする．(e)は拡大・縮小変換であり，もとの画像を β 倍する．なお，アフィン変換には図4(f)のような鏡像変換も含まれるが，ここでは考えない．

図4(b)~(e)の変換はそれぞれ以下の行列で表せる．

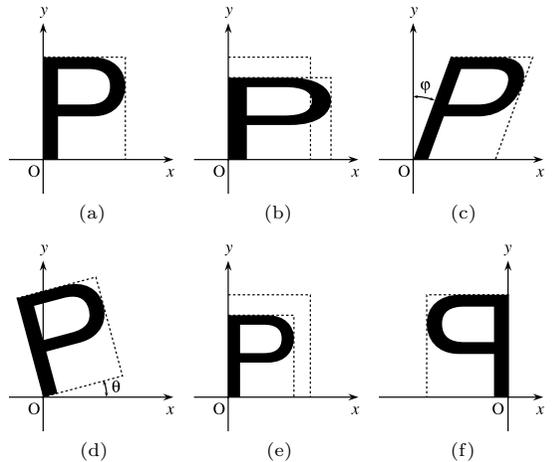


図4 様々な変換

Fig.4 Various transformations.

$$A(\alpha) = \begin{pmatrix} \alpha & 0 \\ 0 & 1/\alpha \end{pmatrix} \quad (2)$$

$$S(\varphi) = \begin{pmatrix} 1 & \tan \varphi \\ 0 & 1 \end{pmatrix} \quad (3)$$

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (4)$$

$$L(\beta) = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \quad (5)$$

これらを組み合わせることで，鏡像変換を除く任意のアフィン変換を表現できる．実際，

$$P(a, b, c, d) = L(\beta)S(\varphi)R(\theta)A(\alpha) \quad (6)$$

を解くことにより， $ad - bc > 0$ であれば，以下のようによりパラメータが定まる．

$$\alpha = \sqrt{\frac{a^2 + c^2}{ad - bc}} \quad (7)$$

$$\theta = \cos^{-1} \frac{a}{\sqrt{a^2 + c^2}} \quad (8)$$

$$\varphi = \tan^{-1} \frac{ab + cd}{ad - bc} \quad (9)$$

$$\beta = \sqrt{ad - bc} \quad (10)$$

3.1.2 アフィン変換による誤差

アフィン変換を施した場合に，量子化誤差により埋

め込んだ値がどのように変化するかを調べるために、情報を埋め込んだ文字パターンにアフィン変換を施した結果得られる画像について面積比を計算する。面積比は理論的にはアフィン不変であるが、実際には後述のように量子化の影響で誤差が生じる。

誤差の調査用に、アルファベット大文字 26 文字を用い、Arial, Arial Black, Helvetica, Times の四つのフォントについて、0.2 及び 0.3 の 2 種類の値を埋め込んだ文字パターンを作成した。すなわち、各字種ごとに 8 種類の文字パターンが存在する。具体的には、まず、各フォントにおける「A」の文字の高さが 500 画素になるように調整した文字画像を準備した。もとの文字画像を右方向に 1~120 画素、下方向に 1~30 画素の範囲で 1 画素単位で移動した画像を影の画像とし、面積比を計算して埋め込むべき値と最も近くなる移動量を用い、情報を埋め込んだ文字パターンを作成した。

次にこれらの画像を用いて、アフィン変換を施した文字パターンの画像を生成する。そして、生成された文字パターンの画像から面積比を求める。具体的には、式 (6) の行列 P を用い、変換後の画像の i 行 j 列の画素の位置に対応する変換前の画像の座標 (i', j') を以下の式によって求める。

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = P^{-1} \begin{pmatrix} i \\ j \end{pmatrix} + \begin{pmatrix} i_0 \\ j_0 \end{pmatrix} \quad (11)$$

ただし、 $(i_0, j_0)^t$ は $i' \geq 0, j' \geq 0$ となるように定める。そして、変換前の画像の $[i']$ 行 $[j']$ 列の画素値を変換後の画像の i 行 j 列の画素値とする。 (i, j) と $([i'], [j'])$ は 1 対 1 対応ではないため、量子化誤差が生じる。

式 (2)~(5) のすべてのパラメータを変化させて埋め込んだ値と抽出された値との誤差を求める。すなわち、式 (11) によってアフィン変換を施した文字パターンの画像を生成し、生成された画像の文字部分と影の部分の画素数の比を求める。これを抽出された値とみなし、埋め込んだ値 (0.2 または 0.3) との差を計算する。そして、8 種類の文字パターンのうち誤差が最大となるものを求めた。変換のパラメータは、アスペクト比の変換の α と拡大・縮小変換の β は 0.8, 1.0 及び 1.25 とし、回転変換の θ と剪断変換の φ については範囲は制限せず 5° 刻みとした。 α と β をこのような範囲に制限した理由は、現実的にはアスペクト比が大幅に変わるような撮影は行わないこと、及び情報を

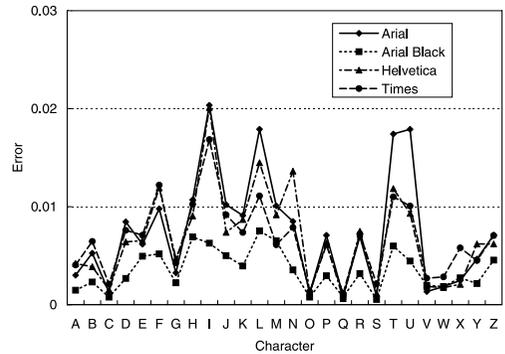


図 5 アフィン変換による最大誤差
Fig.5 Maximum error in the area ratio caused by various transformations.

抽出するためには撮影された文字画像にある程度の大きさが必要なこと (4. で確認する) である。

結果を図 5 に示す。図から「I」や「L」のような直線の多い字種は誤差が大きく、「O」や「Q」のように曲線の多い字種は誤差が小さいことが分かる。また、フォントを構成する線幅の大きい Arial Black の誤差が小さいことも確認された。このように字種やフォントによって誤差には差があるものの、その多くは 0.01 以下、最大でも 0.02 程度であることが分かった。

なお、0.2 を埋め込んだ場合と 0.3 を埋め込んだ場合の平均誤差はいずれも 0.007 であり、埋め込む値による影響はそれほど大きくないといえる。0.4 以上でも同様と考えられる。

なお、図 5 はすべてのパラメータを変化させた場合の結果であるが、一つのパラメータのみ変化させ、他のパラメータを固定することで各パラメータごとの誤差の特性が分かる。詳細は付録に記すが、回転や傾きといった変換と比較して、拡大・縮小による影響が非常に大きいことが分かった。

3.1.3 撮像系によって生じる誤差

3.1.2 で作成した文字パターンと同じものを印刷し、デジタルカメラで撮影した画像から面積比を抽出した場合の埋め込んだ値との誤差を調べた^(注2)。文字部分を赤、影の部分を黄色とし、アルファベット 26 文字を A4 判 1 枚に収まる大きさで印刷した。印刷した色を CMYK で表したものを表 1 に示す。各画像がデジタルカメラのほぼ正面になるように配置し、太

(注2): 文字パターンは、インクジェットプリンタ (キヤノン PIXUS iP8600) を用いて高品位専用紙に印刷した。デジタルカメラとしては、キヤノン DS6041 を用いた。4. の実験でも同じものを用いる。

表 1 印刷した色
Table 1 Colors used for printing.

	Cyan	Magenta	Yellow	Black
文字部分	1%	96%	91%	0%
影の部分	3%	2%	76%	0%

表 2 領域分割
Table 2 Region division.

	$S < 20$	$20 \leq S < 25$	$25 \leq S$
$0^\circ \leq H < 35^\circ$	背景	文字と背景の境界	文字部分
$35^\circ \leq H < 45^\circ$		文字と影の境界	
$45^\circ \leq H < 90^\circ$		影と背景の境界	影の部分
$90^\circ \leq H < 300^\circ$			
$300^\circ \leq H < 360^\circ$		文字と背景の境界	文字部分

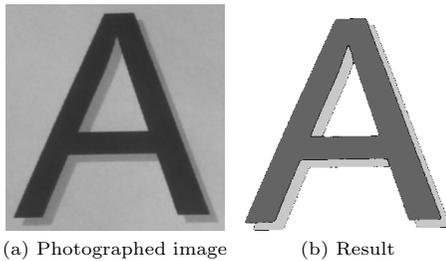


図 6 領域分割の例
Fig. 6 An example of region division.

陽光下で撮影し、撮影された画像から各文字パターンを人手により切り出した。

デジタルカメラで撮影した画像では、文字部分と影の部分、文字部分と背景、影の部分と背景の境界付近は中間色となっているため、まず撮影された画像に対して HSI 変換 [10] を行い、彩度 S と色相 H を用いて、文字パターンを文字部分、影の部分、背景、文字部分と影の部分の境界、文字部分と背景の境界、影の部分と背景の境界の 6 領域に、表 2 の規則に従って分割する。また、面積比を計算する際には、境界の部分の面積は境界の両側の領域の面積に等分配する。画像サイズは 3072×2048 であり、各文字パターンの高さは $330 \sim 400$ 画素程度であった。

実際に撮影された画像から文字パターンの領域を抽出し、領域を分割した例を図 6 に示す。(a) が撮影した画像で、(b) が領域分割の結果である。(b) では、濃い灰色の領域が文字部分、薄い灰色の領域が影の部分、白色の領域が背景、黒色の領域が 3 種類の境界を表す。前述のように、面積比を計算する際には、黒色の部分の面積は両側の領域の面積に等分配する。

各文字パターンの画像から求めた面積比と埋め込んだ値との誤差を図 7 に示す。横軸が字種を表す。図 5

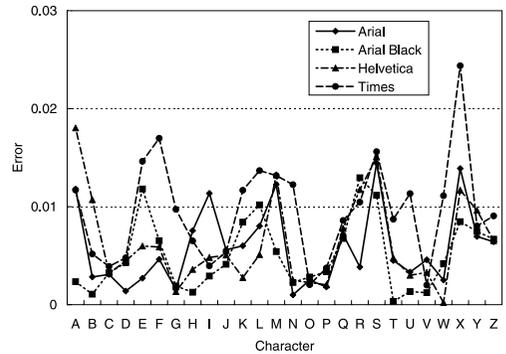


図 7 カメラで撮影することによる誤差
Fig. 7 Error in the area ratio extracted from the photographed image.

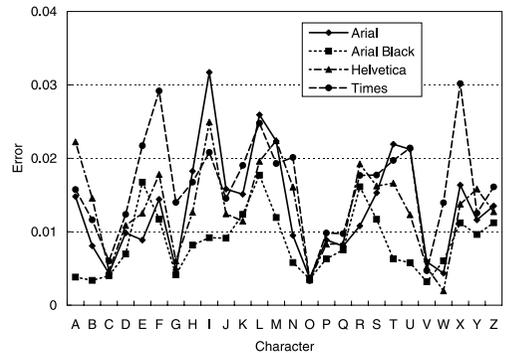


図 8 誤差の和
Fig. 8 Summation of errors.

と同様に字種やフォントによって誤差には差があるが、最大でも 0.025 以下であることが分かった。また、アフィン変換による誤差 (図 5) とカメラで撮影することによる誤差 (図 7) の和を図 8 に示す。ここで、図 7 の文字パターンの大きさは図 5 の文字パターンの大きさ (高さ 500 画素) よりも小さいため、図 8 は高さが 500 画素の文字パターンの誤差を厳密に示すものではない。しかし、4. の実験からも明らかのように撮影することによる誤差は文字パターンが大きいほど小さくなるため、図 8 は誤差の上限値を示すものといえる。提案手法で必要なのは抽出時に誤りが生じないような埋込値の差分であり、すなわち、図 8 で示される誤差の上限値が分かればよい。

図 5 と図 7 を比較してみると、誤差の大きい字種が異なっている。例えば「I」はアフィン変換による誤差 (図 5) は大きいですが、カメラで撮影することによる誤差 (図 7) は比較的小さい。「S」は逆である。したがって、図 8 の誤差はそれほど大きくならず、多くの

字種で 0.025 以下に収まっていた。

なお、文字部分が赤、影の部分が黄色という色の組合せは、後の実験でも用いる。この組合せを用いた理由は以下のとおりである。

- 看板は、色彩の視認度を考慮してデザインされることが多い。白地の看板を想定した場合、代表的な色の中で黄色は最も視認距離が短く、人間には視認されにくい色であることが知られている [11]。したがって、黄色の影を用いることは、提案手法で人間にとって若干不自然な影の付け方がされたとしても、人間には不自然さが知覚されにくいという長所がある。

- HSI 表色系の色相を考えると、赤 (0°) と黄色 (60°) は比較的近い値となっている。したがって、影を黄色とした場合、文字部分の色として赤を用いることは、他の色 (青など) を用いた場合と比較して色の分離が難しく、面積比の正確な抽出が困難であると考えられる。本論文では、このように比較的抽出が困難な状況でも正しく情報抽出ができるような条件を求めるところを一つの目的としている。

3.2 字種固有の面積比

2. で述べた字種固有の自然な面積比について調査するため、アルファベット大文字 26 文字を用い、Arial, Arial Black, Helvetica, Times の四つのフォントのアルファベット大文字 26 字種について図 3 (a) の形態の埋込を行った場合の面積比について調べた。「A」の高さが 500 画素となるように大きさを調整した文字画像を用い、文字画像を右方向に 10 画素、下方向に 10 画素ずらした画像を影の画像とした場合の、文字部分と影の部分の面積比を求めた。

結果を図 9 に示す。図から分かるように、フォントによって若干のばらつきはあるものの、あるフォントで面積比が大きい字種は他のフォントでも大きく、あるフォントで面積比が小さい字種は他のフォントでも小さいという傾向があることが分かる。

3.3 情報を埋め込んだ文字パターン

前述の 4 フォント (Arial, Arial Black, Helvetica, Times) について、アルファベット 26 字種のそれぞれに対して情報を埋め込む。3.1 で考察した誤差の範囲を考慮して、埋め込まれた情報がほぼ誤りなく抽出できるような埋込を行う。図 5 及び図 7 における誤差の最大値はいずれも 0.025 以下であり、図 8 に示す誤差の和でも、全体の 96% が 0.025 以下であることから、埋め込む値の間にはその倍の 0.05 の差を設ける。アルファベット 26 文字に対して異なるカテゴリー情

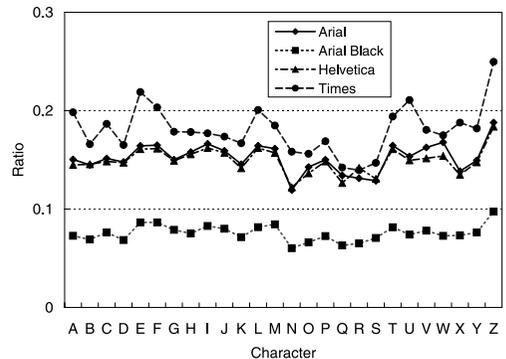


図 9 各文字の面積比

Fig. 9 Area ratio of each character.

表 3 実際に埋め込んだ付加情報の値

Table 3 Embedded values for supplementary information.

字種	値
N, O, Q, R, S	0.20
B, D, K, P, X	0.25
A, C, G, H, Y	0.30
J, M, U, V, W	0.35
E, F, I, L, T, Z	0.40



図 10 作成された画像の例

Fig. 10 Examples of generated character images.

報を埋め込むことは困難であるため、5-6 文字を一つのクラスとし、クラスごとに埋め込む値を定める。ここでは、埋め込む値の最小値を 0.20 とし、0.05 刻みで 0.40 までの値を埋め込むものとする。

また、自然な埋込を行うために、図 9 から、字種ごとに各フォントにおける面積比を平均し、平均値が小さい順にソートする。そして、その値が小さい順に小さい値を情報として埋め込む。実際に埋め込んだ値を表 3 に示す。Arial のそれぞれの字種に表 3 に示す値を埋め込んだ画像を図 10 に示す。

4. 実験

提案手法の有効性を確認するために、情報を埋め込

んだ文字パターンを撮影し、埋め込まれた情報を抽出する実験を行った。3.3で述べた方法で情報を埋め込んだ、Arial, Arial Black, Helvetica, Timesの四つのフォントの文字パターンを用いた。

なお、今回の実験では、個々の文字からの情報抽出精度に主眼を置くため、文字列の切出しそのものは誤りなく行えるよう、白地の背景のみを用いた。文字切出しは、背景以外の部分を抽出することで行った。すなわち、表2によるしきい値処理を行い、文字部分と影の部分の合わせた領域に対し、連結成分を1文字として外接枠を文字パターンとして抽出した。

すべての文字を用いて情報抽出を行うため、図10の各行をそれぞれABCDEF, GHIJKLM, NOPQRST, UVWXYZという四つの文字列とみなして印刷したものを、角度を3通り、距離を5通り変え、デジタルカメラで撮影した^(注3)。すなわち、文字列ごとに15枚の画像を撮影した。それぞれの角度において得られた文字画像の、回転角度と傾き角度のおおよその値(人間が目視で測った角度の平均値を5°単位に量子化したもの)を表4に示す。また、実際に撮影した画像の例と、それぞれの角度(回転角度 θ 及び傾き角度 φ)を図11に示す。それぞれの距離において得られた文字画像の高さの平均値と標準偏差を表5に示す。

表4 文字画像のおおよその回転角度と傾き角度
Table 4 Approximate skew angle and slant angle.

	回転角度	傾き角度
角度1	-25°	15°
角度2	0°	0°
角度3	25°	-15°

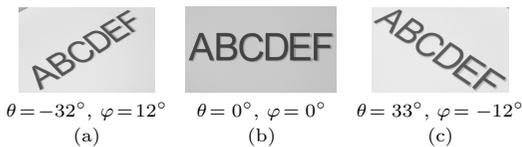


図11 取得した画像の例とそれぞれの角度
Fig. 11 Example of acquired images and angles.

表5 文字画像の高さと情報抽出率
Table 5 Heights of characters and accuracy of extracting embedded information.

	文字高さ		抽出率
	平均	標準偏差	
距離1	110	6.8	31.1%
距離2	186	11.2	91.0%
距離3	254	21.6	100%
距離4	320	31.8	100%
距離5	418	90.6	100%

各文字画像から面積比を計算し、その値と表3の5個の値を比較して、差が最も小さいものを文字画像から抽出された情報とした。埋め込んだ情報が正しく抽出された率を表5に示す。表から、文字の高さが250画素程度あれば、十分高い精度で埋め込まれた情報が抽出できることが分かる。また、180画素程度の比較的小さな画像からも、9割程度の精度で情報抽出が可能であることが確認された。しかし、更に小さい110画素程度の画像では、情報抽出精度は非常に低いものとなった。このことから、提案手法で情報抽出を高精度で行うには、少なくとも180画素程度の大きさが必要であるといえる。なお、実際にデジタルカメラで撮影された文字画像がどの程度の大きさになるのかを示すために、図2には630万画素(3072×2048)のデジタルカメラで実際の看板の文字を撮影した場合の画像の大きさを示してある。図2(b)の高さは680画素であり、正規化した後の文字の高さは640画素であった。したがって、提案手法により高精度に情報を抽出できることが期待される。1000万画素等更に解像度の高いカメラの存在を考えると、面積比による情報の埋込は実用範囲の広い手法といえる。

5. 文字認識への適用

本論文で提案する情報埋込手法を認識補助に用いた場合の有効性を確認するために、図12に示す単純な文字認識アルゴリズムを用いる。まず、認識対象の文字パターンを正規化した後、全字種の標準パターンとの類似度を計算する。同時に、埋め込まれた情報を抽出する。最後に、類似度と抽出された情報を組み合わせて字種の判別を行う。それぞれの処理については以下で述べる。

5.1 正規化

最初に画像を正規化し、アフィン変換を補正する^(注4)。正規化の手順を図13に示す。まず、文字部分(影の部分は含まない)の画素(図13(a))の連結成分ごと

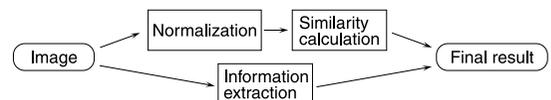


図12 文字認識アルゴリズム
Fig. 12 Character recognition algorithm.

(注3): カメラを移動し、カメラ中心と紙中心を合わせて撮影を行った。
(注4): ここでの補正は、5.2における類似度計算のために行う。付加情報の抽出にはアフィン変換を補正する必要はない。

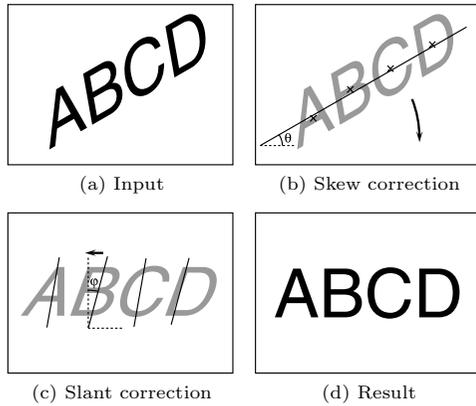


図 13 正規化
Fig. 13 Normalization.

に重心を計算し、重心群に線形回帰分析を適用して文字列の並びを表す直線を抽出する。その直線により回転角度を求め、画像を回転させることで画像の角度を補正する (図 13 (b))。

次に、連結成分ごとに、 i 行目の重心 (連結成分内で i 行目にある点の重心) m_i を求め、これらの重心群 $\{(m_i, i)\}$ を近似する直線を線形回帰分析により求める (図 13 (c))。傾きを表す直線が正しく検出できたかどうかは $\{m_i\}$ と $\{i\}$ の相関係数を用いて判断する。相関係数の最大値が 0.9 以上であれば、最大値が得られた連結成分から求めた直線の傾き角度を用いて傾きを補正する (図 13 (d))。相関係数の最大値が 0.9 未満であれば、傾き角度が正しく推定できなかったとみなし、傾き角度の補正は行わない。

5.2 類似度計算

類似度としては文字部分の画像と標準パターンとの正規化相互相関 (Normalized Cross Correlation: NCC) を用いる。あらかじめ各字種ごとに、4 フォントすべての影がないパターンを用いて平均文字パターン (濃淡画像) を求め、標準パターンとする。求めた NCC を降順にソートすることで、類似度計算による候補字種を順位付きで求める。

5.3 付加情報の抽出

文字パターンの文字領域と影の領域の面積比を計算し、埋め込まれた付加情報を抽出する。面積比の求め方は 3.1.3 で述べた方法と同じである。

5.4 結果の統合

類似度計算の結果と抽出された情報から最終結果を決定する。正規化処理によって回転角度と傾き角度が

表 6 認識率
Table 6 Recognition rates.

フォント	従来法	提案手法
Arial	98.7%	99.0%
Arial Black	98.1%	98.4%
Helvetica	94.2%	95.5%
Times	76.6%	86.9%
平均	91.9%	95.0%

正しく求まっていれば、文字パターンが正しく認識され、しかも NCC が大きな値になると考えられるので、NCC があらかじめ定めたしきい値 t より大きければ類似度計算で得られた 1 位候補を最終結果とする。一方、NCC がしきい値 t 以下ならば、得られた情報が埋め込まれている字種の集合 (表 3 のいずれかの行の文字群) の中で類似度の最も大きいものを最終結果とする。しきい値は、予備実験により $t = 0.6$ とした。

5.5 文字認識精度

4. で撮影した文字パターンを上述のアルゴリズムで認識した。ただし、4. の結果から提案手法を適用するためにはある程度の大きさの文字パターンが必要であることが分かったため、表 5 における距離 1 の画像は除き、距離 2~距離 5 の画像のみを用いた。フォントごとの認識率を表 6 に示す。表中の従来法は付加情報を用いないときの結果である。付加情報を用いることですべてのフォントで精度が改善されたことから、提案手法を文字認識に用いることの有効性が確認された。

提案手法によって精度が改善されたのは、今回用いた単純な正規化法及び文字認識手法では正しく認識できなかった文字パターンについても、提案手法では付加情報を併用することにより正しく認識できたことによる。正規化の精度を、回転角度と傾き角度が正確に補正された (いずれも 1° 以内) 文字パターンの割合と定義すると、正規化の精度は 42% であった。すなわち、今回用いた単純な正規化法では、文字の回転角度と傾き角度を正確に補正することは困難であった。また、角度が正しく補正された文字パターンのみについて認識率を見てみると、Times 以外では 100%、Times は 76% であった。Times については、他のフォントと比較して形状が複雑であり、今回用いた単純な文字認識手法では十分な精度が得られなかったのに対し、付加情報の抽出は文字パターンの形状に依存しないため、付加情報を併用することにより認識精度を向上することが可能となった。Times 以外のフォントについては、

単純な文字認識手法でもある程度の精度が得られているが、正規化が正しく行われなために誤認識した文字パターンが存在した。付加情報の抽出には正規化が必要ないため、付加情報を併用することにより精度を向上することが可能となった。

6. む す び

本論文では、実環境中の文字をデジタルカメラを入力デバイスとして高精度に認識するために、文字画像と同時に認識補助のための付加情報を提示する手段として、面積比を利用した手法を提案した。すなわち、文字パターンを2色で印字し、それぞれの色の領域の面積比を特定の値とするようにデザインする。文字に影を付加したり輪郭線を別の色とすることは文字パターンのデザインとして既に行われており、提案手法はその線幅を変えるにすぎないため、自然な情報の埋込が可能となり、また、様々な用途に広く応用することが可能である。更に、面積比はアフィン変換に不変であり、アフィン変換を受けた環境においても誤りなく抽出されることが期待される。

実際に付加情報を埋め込んだ文字パターンを作成し、デジタルカメラで撮影された画像中の文字パターンから付加情報を抽出する実験、及び付加情報を用いて文字を認識する実験を行った。その結果、文字パターンがある程度大きければ埋め込まれた付加情報を高精度に抽出できること、また、付加情報を認識の補助に用いることで認識精度が向上することが確認された。

今回は付加情報として、クラス分けした字種の情報を埋め込んで実験を行ったが、提案手法は情報の種類を制限するものではなく、ほかに様々な使用方法が考えられる。例えば、文献[12]では、字種ごとに異なる付加情報を埋め込んだ場合の実験例が示されている。また、文字認識の補助のために用いることができる情報として、字種の情報のほかにフォントの種類も考えられる。更に、文字切出しの問題を解決する手段としても適用できる。個々の文字画像を正確に切り出すことは困難な問題であり、特に分離文字の統合については、認識を併用する以外に本質的な解決策はないが、分離文字の個々の部分に対して同じ値を埋め込んでおくことにより、別々の文字なのか分離文字の個々の部分なのか判断できる。どのような情報を埋め込むことが認識の高精度化に最も有効なのかを検討していくことは今後の課題である。

また、既存のフォントを用いた場合、埋め込んだ付

加情報の抽出精度と作成される文字パターンの外観の自然さはトレードオフの関係にあると考えられる。付加情報の抽出精度や外観を考慮したより効果的な情報の埋込方法を検討することや、情報埋込に適したフォントの開発を行うことも今後の課題である。

謝辞 本研究の一部は栢森情報科学振興財団の助成を受けて遂行された。

文 献

- [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, Sept. 2006.
- [2] J. Liang, D. Doermann, and H. Li, “Camera-based analysis of text and documents: A survey,” Int. J. Document Analysis and Recognition, vol.7, nos.2-3, pp.84-104, July 2005.
- [3] X. Chen, J. Yang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” IEEE Trans. Image Process., vol.13, no.1, pp.87-99, 2004.
- [4] 柳辻進介, 高橋友和, 目加田慶人, 井手一郎, 村瀬 洋, “携帯デジタルカメラにより撮影された動画からの低解像度文字認識”, 信学論(D), vol.J89-D, no.2, pp.323-331, Feb. 2006.
- [5] British Computer Society, Character Recognition 1967, Unwin Brothers Limited, 1966.
- [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(D), vol.J89-D, no.2, pp.344-352, Feb. 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, Feb. 2007.
- [8] 佐藤 淳, コンピュータビジョン—視覚の幾何学, コロナ社, 1999.
- [9] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed., Cambridge University Press, 2004.
- [10] 高木幹雄, 下田陽久(監修), 新編画像処理ハンドブック, 東京大学出版会, 2004.
- [11] 大智 浩(編), 商業と色彩, 修道社, 1956.
- [12] S. Omachi, M. Iwamura, S. Uchida, and K. Kise, “Affine invariant information embedding for accurate camera-based character recognition,” Proc. 18th International Conference on Pattern Recognition (ICPR2006), vol.2, pp.1098-1101, Aug. 2006.

付 録

アフィン変換のパラメータごとの誤差

式(2)~(5)に含まれるパラメータのうち一つのパラメータのみを変化させ、他をデフォルトの値($\alpha = \beta = 1.0$, $\varphi = \theta = 0^\circ$)に固定した場合の誤差(3.1.2で述べた量子化誤差で、8種類の文字パ

ターンのうち最大のものを)を図 A・1 に示す. 文字パターンの大きさ等, 他の条件は 3.1.2 と同様である. なお, 参考のため, α と β については, 3.1.2 での範囲よりも広い範囲での値を示してある. 26 字種すべての結果を表示するのは現実的ではないので, 26 字種の平均値, 最大値, 最小値を示してある. また, 平均値よりも大きい値のみを母集団とした場合の標準偏差を平均値に加えた値を点線で, 平均値よりも小さい値のみを母集団とした場合の標準偏差を平均値から引いた値を一点鎖線で示してある.

(1) アスペクト比の変換

式 (2) を用い, α を $0.2 \leq \alpha \leq 5.0$ の範囲で様々な変えた場合の誤差を図 A・1 (a) に示す. 横軸が α を表す. 図より, α が 1 から離れるに従って誤差が急激に増大していることが分かる. また, 字種により誤差の値に大きく差があった. 「S」や「O」など曲線から構成される字種は誤差が小さく, 最大でも 0.005 程度であった. 一方「I」「J」「L」のような直線部分を含む字種は誤差が大きかった.

(2) せん断変換

角度 φ で画像をせん断変換した場合の結果を図 A・1 (b) に示す. 横軸が φ を表す. 図から, わずかでも傾斜させると値が急激に増加するが, $\varphi = 5^\circ$ 程度で頭打ちになることが分かる. この場合も, 誤差が大きいのは「I」や「L」のような直線部分を含む字種であった. しかし, 誤差は最大でも 0.005 程度であり, アスペクト比の変換と比較して影響が小さいことが分かる.

(3) 回転変換

画像を θ だけ回転させた場合の結果を図 A・1 (c) に示す. 横軸が θ を表す. わずかでも回転させると値が急激に増加するが, $\theta = 5^\circ$ 程度でいったん頭打ちになる. そして, $\theta = 45^\circ$ 付近で再び急激に増加する. しかし, 最大でも 0.01 程度であり, やはりアスペクト比の変換と比較して回転変換の影響は小さいことが分かる.

(4) 拡大・縮小変換

画像を β 倍に拡大・縮小させた場合の結果を図 A・1 (d) に示す. 横軸が β を表す. β が大きくなっても誤差はそれほど増加しないが, β が小さくなると誤差が非常に大きくなる. これは, 文字パターンを縮小することによる量子化誤差の影響が非常に大きいことを表している.

(平成 19 年 1 月 10 日受付, 6 月 6 日再受付)

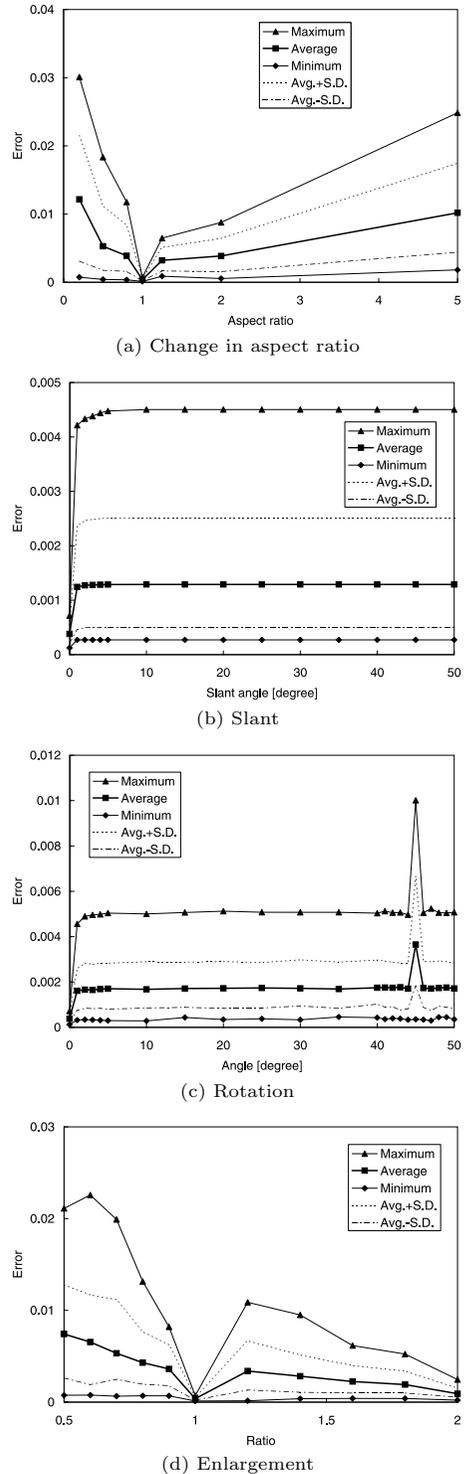


図 A・1 様々な変換による誤差
Fig. A-1 Error in the area ratio by various transformations.



大町真一郎（正員）

昭 63 東北大・工・情報卒．平 5 同大学院博士課程了．同年同大情報処理教育センター助手．平 8 同大工学部助手．平 11 同大学院工学研究科助教授．現在，准教授．博士（工学）．その間，平 12～13 米国ブラウン大学客員准教授．パターン認識，コンピュータビジョン，並列処理，文字認識システムの開発などの研究に従事．平 19 MIRU 長尾賞，ICDAR2007 The Best Paper Award 各受賞．IEEE，情報処理学会，人工知能学会，Pattern Recognition Society 等各会員．



岩村 雅一（正員）

平 10 東北大・工・通信卒．平 15 同大学院博士課程了．同年同大学院工学研究科助手．平 16 阪府大大学院工学研究科助手．現在，助教．博士（工学）．パターン認識，コンピュータビジョンに関する研究に従事．平 18 本会論文賞，ICDAR2007 The Best Paper Award 各受賞．IEEE，情報処理学会，Pattern Recognition Society 各会員．



内田 誠一（正員）

平 2 九大・工・電子卒．平 4 同大学院修士課程（情報）了．セコム（株）勤務を経て，現在，同大システム情報科学研究院知能システム学部門教授．博士（工学）．画像パターン・時系列パターンの解析・認識に関する研究に従事．平 15 本会 PRMU 研究奨励賞，平 18 MIRU 長尾賞，ICDAR2007 The Best Paper Award 各受賞．IEEE，日本音響学会，情報処理学会，映像情報メディア学会各会員．



黄瀬 浩一（正員）

昭 61 阪大・工・通信卒．昭 63 同大学院博士前期課程了．同年同大学院博士後期課程入学．平 2 阪府大・工・電気助手．現在，同大学院工学研究科教授．博士（工学）．平 12～13 ドイツ人工知能研究センター客員教授．文書画像解析，情報検索などの研究に従事．平 18 本会論文賞，ICDAR2007 The Best Paper Award 各受賞．情報処理学会，人工知能学会，電気学会，IEEE，ACM 等各会員．