

面積比による付加情報を用いた実環境文字認識

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

[†] 東北大学大学院工学研究科

^{††} 大阪府立大学大学院工学研究科

^{†††} 九州大学大学院システム情報科学研究院

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

あらまし デジタルカメラを入力デバイスとして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、および、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。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を認識する実験を行い、提案手法の有効性を確認する。

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

Camera-Based Character Recognition with Supplementary Information by Area Ratio

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

[†] Graduate School of Engineering, Tohoku University

^{††} Graduate School of Engineering, Osaka Prefecture University

^{†††} Faculty of Information Science and Electrical Engineering, Kyushu University

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

Abstract In order to achieve a highly accurate recognition of characters in a scene image with a digital camera, there are some attempts on offering supplementary information for recognition with a character image. The information should be human-friendly and robust against geometric distortions. In this paper, we propose a method of embedding information in a character pattern by designing a character pattern in two colors so that the information is embedded as the area ratio of the two colors. It is expected that the area ratio is correctly extracted even if the character pattern is affine-transformed since the area ratio is affine invariant. The effectiveness of the proposed method is confirmed by generating character patterns with the embedded information and carrying on an experiment of recognizing characters in a scene image taken by a digital camera.

Key words Camera-Based Character recognition, Information embedment, Supplementary information, Affine invariants, Area ratio

1. はじめに

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、一般に低解像度であり、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。認識を高精度化するための手法として、光学的歪みや幾何学的歪みをできる

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欲しい場合には、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取られることもある。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



図 1 看板中の文字

加情報を提示する方法が検討されている [6], [7]. すなわち, 文字画像を認識すると同時に付加情報の取り出しを行い, 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, および, 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編で表し, 編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概観を大きく変えるため, 人間の読み易さを若干損ねるという問題点がある.

本論文では, 文字パターンの概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め込み法を提案する. 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るようにデザインする. 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 1 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, デジタルカメラで取得した画像中の文字を認識する実験を行った. その結果, 埋め込まれた付加情報が正確に抽出できること, および, 付加情報を用いることで認識精度が大幅に向上することが確認された.

2. 面積比を用いた情報の埋め込み

2.1 埋め込み形態

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るようにパターンをデザインすることで情報を埋め込む. 図 2 に示す例のように, 様々な埋め込み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すでに行われており, 提案手法はその線幅や面積を変えるにすぎない. さらに, 文字は字種ごとに固

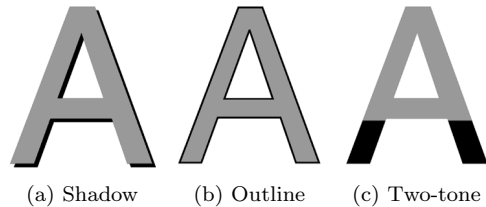


図 2 様々な埋め込み法

有の形を持つため, 図 2 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」面積比が存在すると考えられる. 従って, 複比パターンの埋め込みと比較してより自然な埋め込みが可能になると考えられる.

情報埋め込みの際は, 幾何学的変形や撮像系による誤差を考慮し, できるだけ誤りのない情報抽出が行えるように埋め込むべき情報を決定することが望まれる. 以下, 2.2 では幾何学的変形による量子化誤差をシミュレーション実験によって調査し, 2.3 ではカメラで撮像することによって生じる誤差を調査する. 2.4 では上述の字種ごとの自然な埋め込み量について考察し, 2.5 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. 埋め込みの形態としては図 2(a) に示すように, 影を付加する方法を使用した. 以下の説明で, 図 2(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 2(a) 全体を文字パターンと呼んで区別する.

2.2 幾何学的変形による埋め込み値の変動

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した. 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.

アルファベット大文字 26 文字を用い, Arial, Arial Black, Helvetica, Times の 4 つのフォントについて, 0.2 および 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. 元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, 得られた画像に対してアフィン変換を施す. 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. ただし, 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 3(b) ~ (e) に示す 4 種類の変換の合成変換と捉える. 図 3(a) が元の画像である. (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

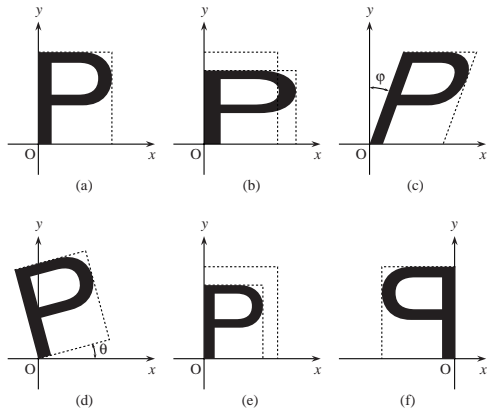


図3 様々な変換

るものとする．(c)は x 軸に沿った剪断変換であり，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，元の画像を r 倍する．なお，アフィン変換には図3(f)のような鏡像変換も含まれるが，ここでは考えない．

図3(b)～(e)の変換はそれぞれ以下の行列で表せる．

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

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

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

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

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．

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

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

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

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

式(2)～(5)のパラメータを変えて文字パターンの画像を変換する．具体的には，まず式(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

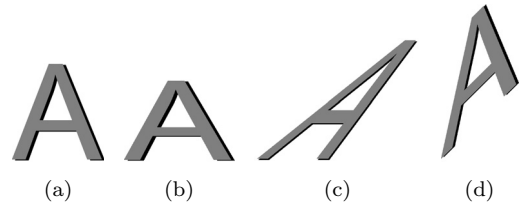


図4 アフィン変換を受けた画像

表1 パラメータと面積比

	(a)	(b)	(c)	(d)
α	1.0	1.2	0.8	1.2
φ	0.0	0.0	45.0	45.0
θ	0.0	0.0	0.0	45.0
β	1.0	1.0	1.0	1.0
r	0.200	0.199	0.201	0.200

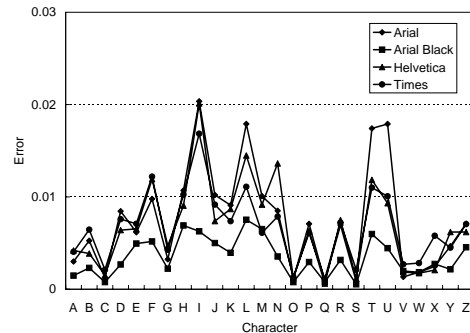


図5 アフィン変換による誤差

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 $([i'], [j'])$ は一対一対応ではないため，量子化誤差が発生する．

次に，変換された画像の面積比を求める．ただし，変換のパラメータは，アスペクト比の変換の α は $0.8 \leq \alpha \leq 1.25$ ，拡大・縮小変換の r は $0.8 \leq r \leq 1.25$ とし，回転変換の θ と剪断変換の φ については制限しなかった．図4にアフィン変換を受けた画像の例を示す．(a)が元の画像であり，高さが500画素の文字「A」(フォントはArial)の文字部分(灰色部分)に対して影の部分(黒色部分)を付加したものである．(b)～(d)は(a)の画像を表1に示したパラメータで変換した画像である．それぞれの灰色の画素の数に対する黒の画素の数の比を r とし表1に示してある．

埋め込んだ値と抽出された値との差(誤差)を求め，各字種の各フォントごとに，誤差が最大となるものを求めた．結果を図5に示す．紙面の都合上詳細は省略するが，シミュレーションの結果により，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．また，字種やフォントによって誤差が大きく異なることも確認された．

2.3 撮影された画像から取り出された値の誤差

2.2で作成した文字パターンを実際に印刷し，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．文字パターンの部分を赤，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



(a) Photographed image. (b) Result.

図 6 領域分割の例

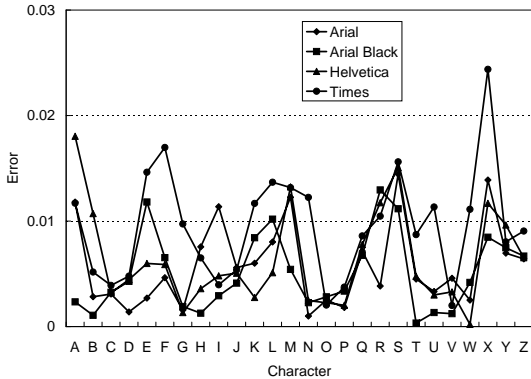


図 7 撮影した画像から抽出された面積比および埋め込んだ値との誤差

印刷した．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，太陽光下で撮影し，撮影された画像から各文字パターンを切り出した．画像サイズは 3072 × 2048 であり，各文字パターンの高さは 330 ~ 400 画素程度であった．

次に，切り出された文字パターンから面積比を求める訳であるが，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，撮影された画像に対して HSI 変換 [10] を行い，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 領域に分割する．そして，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．

実際の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．

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．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) と撮像することによる誤差 (図 7) の和を図 8 に示す．

2.4 文字パターン固有の面積比

2.1 で述べたように，字種ごとにある程度固有の「自然な」面積比が存在すると考えられる．そこで，前述の 4 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 2(a) の形態の埋め込みを行った場合の面積比について調べた．具体的には，2.2 と同様に「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

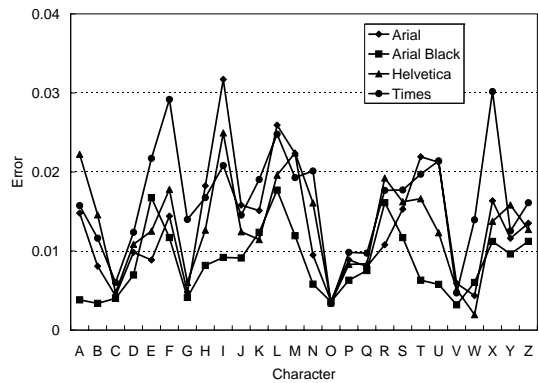


図 8 誤差の和

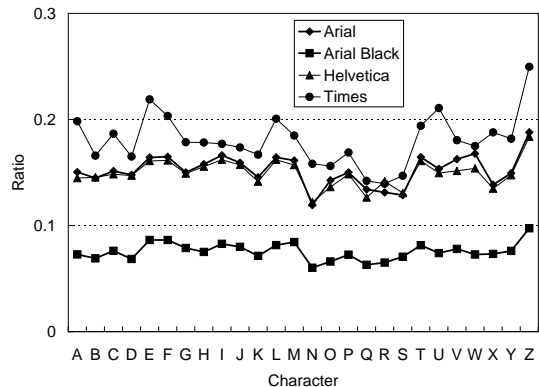


図 9 各文字の面積比

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．

2.5 埋め込み例

前述の 4 フォントについて，アルファベット 26 文字のそれぞれに対して情報を埋め込む．自然な埋め込みを行うために，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，5 ~ 6 文字を 1 つのクラスとし，クラスごとに小さい値を情報として埋め込む．

また，2.2 および 2.3 で考察した誤差の範囲を考慮する．図 5 および図 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 に示される誤差の和でも，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．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．実際に埋め込んだ値を表 2 に示す．Arial, Arial Black, Helvetica, Times の 4 つのフォントについて，表 2 に示す値を埋め込んだ画像を図 10 に示す．

3. 実 験

3.1 文字認識アルゴリズム

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

表 2 実際に埋め込んだ値

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

パターンを認識する実験を行った。ここでは図 11 に示すような単純な文字認識アルゴリズムを用いた。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。

3.1.1 正規化

最初に画像を正規化し、アフィン変換を補正する。正規化の手順を図 12 に示す。まず、文字パターンを表す画素(図 12(a))の連結成分毎に重心を計算し、重心位置に最小二乗法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 12(b))。次に、連結成分の行ごとに重心を求め、重心群の横座標と縦座標に最小二乗法を適用して直線を検出する(図 12(c))。ここで、傾きを表す直線が正しく検出できたかどうかを相関を用いて判断する。これを全ての連結成分に対して行い、連結成分ごとに求めた相関の最大値が 0.9 以上であれば、その相関が得られた直線の傾き角度を用いて傾きを補正する(図 12(d))。相関の最大値が 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。

3.1.2 類似度計算

類似度としては正規化相互相関 (Normalized cross correlation; NCC) を用いる。あらかじめ、各字種の平均文字パターンを求めて標準パターンとする。入力パターンと各字種の標準パターンとの NCC を計算する。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。

3.1.3 埋め込み情報の抽出

文字パターンの文字領域と影の領域の面積比を計算することで埋め込まれた情報を抽出する。

3.1.4 結果の統合

認識結果と抽出された情報から最終結果を決定する。ここでは単純に、取り出された情報が埋め込まれている字種の集合(表 2 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。

3.2 実験結果

2.5 で述べた方法で情報を埋め込んだ文字パターン(図 10)を用いて実験を行った。文字パターンを赤色、影を黄色とした。図 10 の各行をそれぞれ“ABCDEF”, “GHIJKLM”, “NOPQRST”, “UVWXYZ” という 4 つの文字列とみなし、認識を行った。各文字列の画像はデジタルカメラで 3 種類の角度から撮影した。文字画像の高さ、回転角度、傾き角度のおおよその値を表 3 に示す。また、実際の画像の例を図 13



(a) Arial



(b) Arial Black



(c) Helvetica



(d) Times

図 10 作成された画像の例

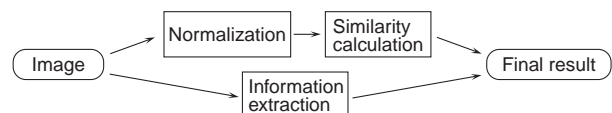


図 11 文字認識アルゴリズム

に示す。

各文字パターンを 3.1 で述べたアルゴリズムで認識した。認識精度を表 4 に示す。表中の従来法は埋め込み情報を用いない

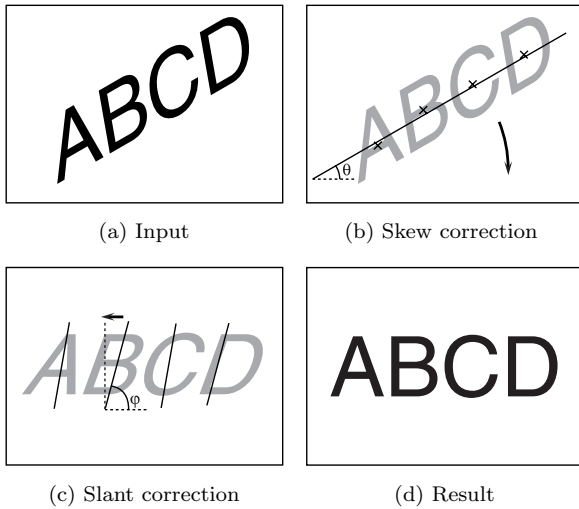


図 12 正規化

表 3 文字画像の高さ・回転角度・傾き角度

	高さ	回転角度	傾き角度
(a)	450	25°	115°
(b)	300	0°	0°
(c)	450	-25°	75°

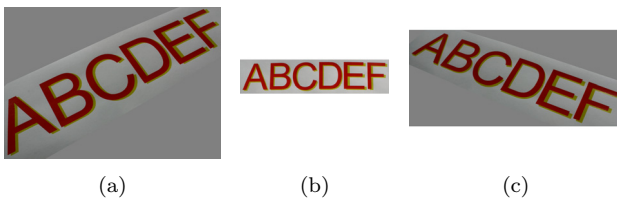


図 13 取得した画像の例

ときの結果である。なお、埋め込んだ情報そのものはすべて正しく抽出された。埋め込み情報を用いることで全てのフォントで精度が大幅に改善されたことから、提案手法の有効性が確認された。従来法における誤認識の主な原因は、今回用いた単純な正規化法ではアフィン変換のパラメータを正しく求めることができず、正規化が正しく行われなかったことである。提案手法では抽出した情報が効果的に働き、従来法と比較して認識精度が大幅に向上した。特に Times のパターンで、従来法と提案手法の違いが大きかった。

4. むすび

本論文では、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すでに行われており、提案手法はその線幅を変えるにすぎないため、自然な情報の埋め込みが可能となる。また、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、埋め込

表 4 認識率

フォント	従来法	提案手法
Arial	95%	97%
Arial Black	92%	95%
Helvetica	87%	90%
Times	73%	89%
Average	87%	93%

んだ情報の抽出と文字認識アルゴリズムを合わせて環境中の文字を認識する実験を行った。その結果、従来の文字認識アルゴリズムのみを用いる手法と比較して、大幅に認識精度を向上させることが可能であることが示された。

今回は付加情報として、クラス分けした文字の字種の情報を埋め込んだが、提案手法は情報の種類を制限するものではなく、他にも様々な使用方法が考えられる。例えば、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。また、文字切り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が判断できる。

また、埋め込んだ情報の抽出精度や作成される文字パターンの外観を多少犠牲にすれば、より多くの種類の情報を埋め込むことが可能になる。情報の抽出精度や外観を考慮して、より効果的な情報の埋め込み方法を検討することは今後の課題である。

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, March 2005.
- [2] D. Doermann, J. Liang and H. Li, “Progress in camera-based document image analysis,” Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), pp.606–616, August 2003.
- [3] X. Chen, J. Yang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” IEEE Trans. Image Processing, vol.13, no.1, pp.87–99 2004.
- [4] 柳詰進介, 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, February, 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, February, 2006.
- [7] 岩村雅一, 内田誠一, 大町真一郎, 黄瀬浩一, “情報付加による認識率 100%の実現—一人にも機械にも理解可能な情報伝達のために—,” 画像の認識・理解シンポジウム 2005 (MIRU2005) 論文集, pp.901–908, July 2005.
- [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*, Cambridge University Press, 2nd edition, 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.