

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

内田 誠^{†a)} 岩村 雅一^{††} 大町真一郎^{†††} 黄瀬 浩一^{††}

Category Data Embedding for Camera-Based Character Recognition

Seiichi UCHIDA^{†a)}, Masakazu IWAMURA^{††}, Shinichiro OMACHI^{†††}, and Koichi KISE^{††}

あらまし 本研究は、バーコードと同程度の精度で三次元実環境中の文字パターンを認識することを目標としている。実環境中の文字パターンは、撮影状況により様々なひずみ、例えば射影変換ひずみを受ける。このため、通常の文字認識手法の延長線上でこの目標を達成しようとしても、相当の困難が予想される。そこで本論文では、文字そのものに機械可読性を補強するような情報を埋め込む方式を検討する。具体的には、文字画像に対し、しま模様状のパターンを埋め込む。このパターンを構成する各しまの幅から計算される複比は、文字パターンがどのように射影変換ひずみを受けてたとしても常に一定値となる。したがって、カテゴリと複比の値をあらかじめ対応づけておけば、抽出された複比を識別の手掛りとして認識時に利用できる。シミュレーション実験の結果、複比と文字形状情報を併用することで、射影変換ひずみを受けても非常に高い認識精度が得られることが分かった。

キーワード カメラ, 文字認識, 文書画像解析, 情報埋込, 複比

1. ま え が き

本論文は、三次元実環境中の文字パターンをバーコードと同程度の精度で認識することを目標とした研究に関する。文字パターンの入力装置としてはデジタルカメラを想定する。その携帯性やボタン一つ操作性は、スキャナは使いにくいと感じていたユーザにも受け入れられやすいと考えられる。実環境中において動かせない文字（例えば看板）を認識対象とできる点も魅力的である。

反面、処理技術の立場から眺めると、実環境中の文字認識には様々な問題がある。例えば、撮影状況に起因した様々なひずみがある。すなわち、幾何ひずみ（特に射影変換ひずみ）、低解像度、非均質照明、ぼけ・手ぶれ、といったひずみを受けた文字は、従来の OCR

用に開発された認識手法では手に負えない。

こうした撮影状況に起因したひずみへの対処法は、最近数多く提案されている [1], [2]。例えば、幾何ひずみに対しては各文字を対象とした弾性マッチング法 [3] や文書全体を対象とした dewarping 手法、非均質照明に対しては局所 2 値化 [4] ~ [6] といった検討がなされている。これらの検討はいずれも、実環境中の文字をなるべく従来の OCR の枠組みで認識するための、特徴抽出や前処理に関する工夫であるといえる。

本研究では、冒頭で述べた目標の達成のために、これら従来研究とは異なった方針をとる。すなわち、機械可読性を補強するような付加情報を文字パターンそのものに埋め込んでおくことを考える。言い換えると、実環境での機械認識を容易なものとする工夫を文字パターン側に施すことを考える。

この方針に従って目標を達成するための第 1 段階として、本論文は次の 2 点について論ずる。

- 第 1 に、射影変換ひずみに影響されない付加情報の埋込・抽出手法を提案する。具体的には、水平しま模様状のパターン（以下、複比パターン）を埋め込んだ文字パターンを提案する。このしま目の間隔から計算される複比 [7] は射影変換不変となり、したがって、どのような角度から撮影した文字パターンであっ

[†] 九州大学大学院システム情報科学研究所, 福岡市
Faculty of Information Science and Electrical Engineering,
Kyushu University, Fukuoka-shi, 812-8581 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

^{†††} 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

a) E-mail: uchida@is.kyushu-u.ac.jp

ても、離散化誤差等の影響がなければ、常に同じ複比を抽出できる。このとき、複比の値とカテゴリを対応づけておけば、文字パターンに射影変換不変なカテゴリ情報を埋め込んだことになる。

● 第2に、こうして抽出された複比と、通常行われる形状による文字認識の結果を統合することで、射影変換ひずみを受けたとしても高精度な文字認識が可能なことを実証する。特に、その際、カテゴリと複比の対応付けの方式によって認識精度が大きく左右されるという性質についても注目する。

以上のように本論文では、特に射影変換ひずみに着目し、非均質照明によるひずみ等は考慮しない。これに応じ、後述する評価実験も、計算機内で射影変換を施した文字パターンを対象としたシミュレーション実験となっている。したがって、ひずみの様子に実環境との多少の乖離があることは間違いない。更に、複比パターンについても、なるべく目立たず、かつ実環境下で安定して抽出されるものになるためには、検討の余地が大いに残されている。しかし、今回のシミュレーション実験により、埋込に関する様々な知見が得られたのも事実である。実環境下での実験は、これらの知見を踏まえた上で、今後進めていくこととする。

機械可読性を高めるために、何らかの情報を文字に埋め込むという試みは、OCR や MICR (magnetic ink character recognition) の黎明期に既に行われている。すなわち、機械読取りのために独特の形状をもたせたフォントである [8], [9]。図1は C.M.C.7 と呼ばれる MICR フォントである。各フォントは、広狭2種類の間隔をもつ7本の垂直線分を用いて設計されており、これら六つの間隔が文字コードを定めている。より最近の関連研究として、DataGlyph [10], [11] がある。これは文字パターンを細かな「/」「\」からなるテクスチャパターンで構成するというものである。本論文で提案する文字パターンは、こうした OCR/MICR フォントや DataGlyph を、実環境中での文字認識用

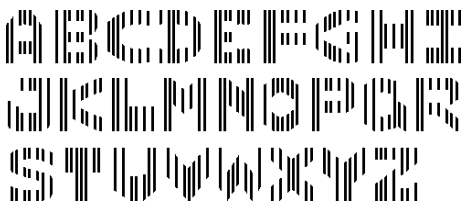


図1 MICR フォント C.M.C.7
Fig.1 MICR font called C.M.C.7.

に拡張したものといえる。

2. 文字への複比埋込と抽出

文字画像パターンに対し、ある複比 r をもつパターン (模様) を埋め込むことを考える。埋め込む複比の値と文字カテゴリの対応付けには様々な形態が考えられる。例えば、各カテゴリに異なる複比を埋め込んでもよいし、複数のカテゴリに同じ複比を埋め込んでもよい。前者の場合、複比さえ正しく抽出できれば、文字形状を使わなくても識別できることになる。以下本章では、これらいずれの場合も含む一般的な場合について議論する。

2.1 複比パターン

埋込の形式については様々考えられるが、ここでは文字線と背景を併せた文字画像全体に、5本の帯からなる平行しまを1セットとした模様 (以下、複比パターン) を埋め込むことを考える。図2(a)は複比パターンを埋め込んだ文字の例である。ここでは、説明の都合上、目立つように配色した複比パターンを用いている。

5本の帯のうち、最初と最後の帯は複比パターンの存在範囲を表すガイドである。これら2本のガイドに挟まれた残る3本の帯により埋め込む複比を表す。これら3本の帯の幅を l_1, l_2, l_3 とすると、このパターンによる複比は、

$$r = \frac{(l_1 + l_2)(l_2 + l_3)}{l_2(l_1 + l_2 + l_3)} \quad (1)$$

で表される。帯の幅 l_1, l_2, l_3 を変えることで、埋め込む複比の値を制御できる。

複比は本来連続的な値をもつものであるが、ここで

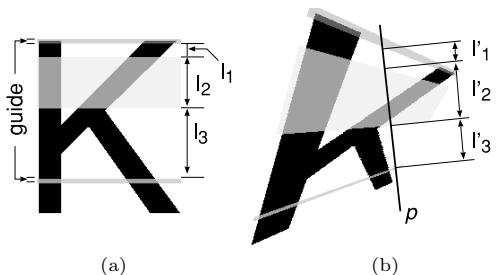


図2 (a) 複比パターンを埋め込んだ文字画像及び (b) その射影変換例 ($\delta = 32$)。見やすくするために、意図的に複比パターンが目立つような配色にしている

Fig.2 (a) A character pattern "K" where a cross ratio pattern is embedded. (b) Projective distortion ($\delta = 32$).

はそれを量子化した K 通りの値 r_k ($k = 1, 2, \dots, K$) を考える．値 r_k の具体的な設定法については、2.3 で述べる．

2.2 複比の抽出方式

複比 r_k を埋め込んだ文字画像から複比を抽出する際は、それを横断するような直線を引き（図 2(b) 中の直線 p ）、その線上のガイドに挟まれた区間について、帯の幅（同図 l'_1, l'_2, l'_3 ）を求め、 $l_i = l'_i$ ($i = 1, 2, 3$) として式 (1) により複比を計算すればよい．これら帯の幅に誤差がなければ、どのように直線を引いても、またどのように射影変換を受けても、埋め込んだときと同じ複比 r_k が求まる．したがって、この複比パターンを埋め込んだ文字画像をいかなる角度から撮影しても、原理的に常に同じ複比が取り出せることなる．

実際には、デジタル画像特有の離散化誤差等によって観測される帯の幅 l'_1, l'_2, l'_3 には誤差が入り、したがってそれらから計算される複比にも誤差が入る．その対処法として、ランダムに P 本の直線を引き、それぞれで求めた P 個の複比の値から、真値を推定するといった処理を行う．具体的には、まず、 P 本の線それぞれについて上述の方法で複比 r を求め、その複比 r に最も近い r_k を選択する．この選択処理を全 P 本の直線について行った後、最も多く選択された r_k を、その文字パターンに埋め込まれた複比とする．

2.3 複比パターンの設計

複比パターンの各帯の幅を変えることで、様々に複比の値を制御することを考える．離散化誤差等の影響を最小化しようとするれば、文献 [12] のような理論的解析が必要と思われるが、現時点では単純に以下のような手順で設定した．

複比パターンからガイドを除いた部分の幅を L とする．すなわち $L = l_1 + l_2 + l_3$ とする．この L 並びに l_1 は K 種の複比パターンにおいて共通とする．したがって、 $l_2 + l_3 = L - l_1$ を満たす l_2, l_3 の組を K 通り設定することで r_k ($k = 1, 2, \dots, K$) を定めることとする．

具体的には、次式を満たす l_2, l_3 により r_k を定める．

$$l_2 = \frac{(L - l_1 - 2\epsilon)(k - 1)}{K - 1} + \epsilon$$

$$l_3 = L - l_1 - l_2 \quad (2)$$

ここで ϵ は l_2, l_3 の最小幅を定める正定数である．上式により、 l_2, l_3 の境界点は、それが取り得る区間を等間隔に $K - 1$ 分割した際の分割点及び区間両端の計

K 点に設定される．

3. 文字認識結果と複比抽出結果の統合

2. の冒頭で述べたように、埋め込む複比の種類 K が文字カテゴリー数 $|C|$ と同一の場合 (C は文字カテゴリーの集合)、各カテゴリー c にそれぞれ複比 r_c を埋め込んでおけば、文字形状を用いなくても、抽出した複比のみを用いて識別可能となる．しかし、漢字などカテゴリーの多い文字集合を対象とすると、限られた解像度のもとで、多種類の複比を定義することは難しい．

より一般的な $K < |C|$ の場合には、複比とカテゴリーは 1 対多の関係となるため、複比単独での認識結果にはあいまい性が残る．具体的には、第 k 複比 r_k を埋め込んだ文字カテゴリーの集合を C_k ($C_k \subset C$) とすると、ある入力文字から複比 r_k が抽出されても、その文字が C_k に含まれるカテゴリーのいずれかまでは分からない．そこで、文字形状を用いた認識結果（すなわち通常の文字認識の結果）と複比による認識結果とを統合することで、最終的な認識結果としてカテゴリーを一つに絞り込むことを考える．

Voting をはじめ、複数の識別器の統合方式にも多くの形態があるのと同様、この統合についても様々な方式が考えられる．ただし、一般的な識別器統合と今回の統合は以下 2 点で異なっている．第 1 は、上述のように、抽出された複比による認識結果には、あいまい性が残るという点である．第 2 は、複比による認識精度（すなわち複比の抽出精度）の方が、文字形状による認識精度に比べて圧倒的に高いという点である．後述の実験結果からも明らかになるように、撮影時の射影変換ひずみが大きいと、文字形状による認識率は著しく低下する．一方、射影変換に対する不変量である複比は、もし離散化誤差や輝度値変化がなければ常に正しく抽出される．実際にはこれら離散化誤差等の影響で多少の誤抽出が発生するものの、文字形状による認識精度に比べると、複比の抽出精度は依然として相当高い．

以上の点を考慮し、本論文では、抽出された複比によりカテゴリーをいくつか絞り込み、次に文字形状による認識結果により、カテゴリーを唯一に決定することとする．具体的には、入力文字からまず複比 r_k を抽出し、それは正しいものと考えて、 C_k のうちで文字形状による認識コスト（距離）が最も低かったカテゴリーを最終的な認識結果とする．なお、 $K = |C|$

の場合、文字形状を用いずに複比だけで認識を行うことになる。

この方式では、文字形状での認識の結果、正解カテゴリーの認識コスト s が高かったとしても、(i) 複比 r_k が正しく抽出され、かつ (ii) C_k 中の正解カテゴリー以外のカテゴリーの認識コストが s より大きければ、正しい認識結果が得られる。逆に、複比抽出が失敗すると、文字形状による認識で正解が得られていたとしても誤認識に転じてしまう。しかし、前述のように複比の抽出精度は高いので、この改悪は少ないと考えられる。

同じ複比 r_k が割り当てられるカテゴリーの集合 C_k は、文字形状による認識において、互いに誤認識となりにくいカテゴリーで構成されることが得策である。これは、上記 (ii) の条件が成り立ちやすくなるための工夫である。そのような割当 $\{C_k | k = 1, \dots, K\}$ を構成するためには、まず文字形状による認識の際に用いる手法の誤認識特性を把握し、互いに誤認識となりにくいカテゴリー対を知得した後、なるべくそうしたカテゴリー対が一つの集合 C_k に含まれるように複比割当 $\{C_k | k = 1, \dots, K\}$ を定めればよい。この割当の効果については 4.5 において議論する。文献 [13] では、この点に関するより一般的な議論がなされている。

4. シミュレーション実験

4.1 準備

4.1.1 フォント画像

複比を埋め込む対象となる文字画像として、フォント “Arial” の英語大文字 26 画像を用いた。このフォント原画像の縦画素数は、各カテゴリーで多少異なり、最小で 194、最大で 212、平均 196 であった。一方、横画素数はカテゴリーごとで差異が大きく、最小で 52 (“I”), 最大で 251 (“W”), 平均 170 であった。今回の実験において、フォントの種類は複比の抽出精度には影響しない。これは、複比パターンを文字線だ

けでなく背景部にも埋め込んでいるためであり、すなわち複比パターンを文字形状とは無関係に抽出できるためである。今後、文字線上にのみ複比パターンを埋め込むといった検討を行えば、フォントによって複比抽出精度にも多少の差異が現れるものと考えられる。

4.1.2 複比パターンの設計

複比パターンについては、ガイドの幅を 5 画素、複比パターン全体の幅 L を 150 画素とし、更に l_1 及び ϵ については 15 画素とした。この条件のもと、2.3 で述べた方法に従って r_k ($k = 1, 2, \dots, K$) を設定した。更に、後述する方式に従い C_k を定めた後、この r_k による複比パターンをカテゴリー $c \in C_k$ のフォントに埋め込んだ。

いずれの画像も 256 レベルの濃淡画像である。文字線及び背景の濃淡値をそれぞれ 0 と 255、それらと複比パターン (幅 l_2 の帯) が重なった部分の濃淡値はそれぞれ 150 と 240 とした。また、文字線及び背景にガイドが重なった部分の濃淡値はそれぞれ 100 と 200 とした。なお、後述するように今回の実験ではこれら濃淡値に変動は生じないため、互いに異なる値である限り、他の値に設定しても全く同じ実験結果が得られる。図 3 は、すべてのカテゴリーに異なる複比パターンを埋め込んだ場合 (すなわち $K = |C| = 26$) の文字画像である。

4.1.3 射影変換によるテストパターン生成

複比パターンを埋め込んだフォント画像について、その四隅の x, y 座標をそれぞれ $\pm\delta$ ($\delta = 0, 4, 8, \dots, 48$) 画素だけ変位させて射影変換を施し、各 δ について $2^4 \times 2 = 256$ 個のテストパターンを生成した。図 4 はその一例である。このように、 δ が 36 程度まで大きくなると、非現実的なレベルまでひずんだ文字画像が発生する。すなわち、今回のテストパターンには、意図的に撮影された場合よりも強くひずみを受けた文字パターンが含まれている。



図 3 複比を埋め込んだ場合の文字パターン ($K = 26$)

Fig. 3 Reference character images painted by different cross-ratio patterns.

4.1.4 文字形状による認識手法

先に 3. で述べたように、一般的な $K < |C|$ の場合には、認識結果を確定するために、何らかの文字形状による認識手法（すなわち通常の文字認識手法）が必要となる。任意の手法が利用できるが、本実験では次の二つを使用した。

- 単純マッチング: 標準パターンと未知入力パターンを単純に重ね合わせて、両文字画像パターンのマッチングコスト、すなわち距離を求める手法である。

- 弾性マッチング: 直感的には、一方の文字画像パターンをゴム膜的に変形させながら重ね合わせる手法である。弾性マッチングには様々なものがあるが [3]、ここでは図 5 に示すような手法を用いた。入力パターンの各列が、その隣接関係を保ちながら、傾いた直線として標準パターン上に対応づけられる。その際、各列の内部でも一次元的な伸縮を許す。詳細は略すがこの写像全体の最適化には動的計画法 (DP) に基づく

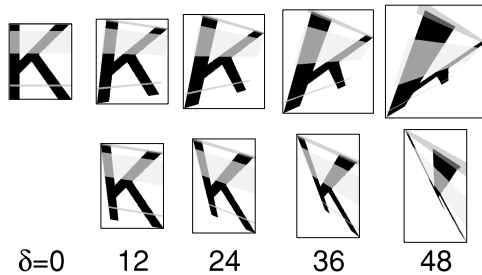


図 4 射影変換により生成されたテストパターンの例
Fig. 4 Test patterns.

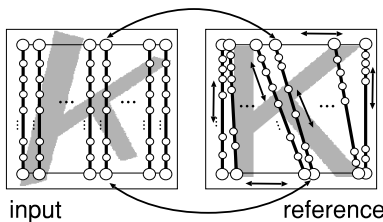


図 5 本実験で用いた弾性マッチング
Fig. 5 Elastic matching technique employed.

アルゴリズムが利用できる。この弾性マッチング法の自由度は射影変換より高く、したがって、原理的には任意の射影変換を補償可能である。

いずれの手法においても、前処理として文字大きさの線形正規化を行った。また各画素の特徴量は、非常に単純に、1 (文字線)、0 (背景) と設定した。

4.1.5 複比の割当

集合 C_k の構成、すなわち複比 r_k をどのカテゴリー c に割り当てるかについては、3. の最後でも述べたように、最終的な認識性能を大きく左右する要因として非常に重要である。本実験では次の 2 通りの割当方式を使用した。

- 単純割当方式: カテゴリー c に $k = ((c - 1) \bmod K) + 1$ を満たす複比 r_k を割り当てる方式である。

- 最適割当方式: 文字形状による認識手法の特性を考慮して、各カテゴリーに複比を割り当てる方式 [13] である。この方式では、文字形状による認識手法の混同行列を複比割当の手掛りとして用いる。例えば “H” という認識結果を与え得るカテゴリーが “H” と “N” であることが混同行列から分かれば、それら 2 カテゴリーには、あいまい性を避けるために異なる複比を割り当てる。この方針に従い、全カテゴリーについて複比割当を定める。

表 1 に、前項で述べた弾性マッチング法を用いた際の、 $K = 4, 12, 20$ の場合の単純割当と最適割当を示す。いずれの K のときも、最適割当により、“C” と “V” には同じ複比が割り当てられる。このことは、それら 2 カテゴリー間で誤認識が起こりにくいことを示唆している。なお、最適割当の決定に際しては、 $\delta = 0 \sim 48$ すべてのデータを対象とした認識結果から得られた混同行列を用いた。

4.2 文字形状単独での認識結果

埋め込まれた複比情報を用いずに、4.1.4 で述べた二つの手法 (単純及び弾性マッチング法) を用いて、文字形状単独での認識実験を行った。標準パターンとしては、射影変換を受けていないフォント画像そのも

表 1 複比の単純割当と最適割当

Table 1 Naive and optimal assignments of cross ratios to 26 classes.

K		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
4	naive	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
	opt.	1	1	1	1	2	3	3	2	1	2	1	2	3	4	4	2	2	4	1	4	2	1	1	4	3	2
12	naive	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2
	opt.	1	2	3	4	5	6	7	5	4	8	2	9	6	10	11	8	9	10	1	11	12	3	4	11	7	12
20	naive	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6
	opt.	1	2	3	4	5	6	7	8	9	10	11	12	6	13	14	15	16	17	18	14	19	3	9	20	7	19

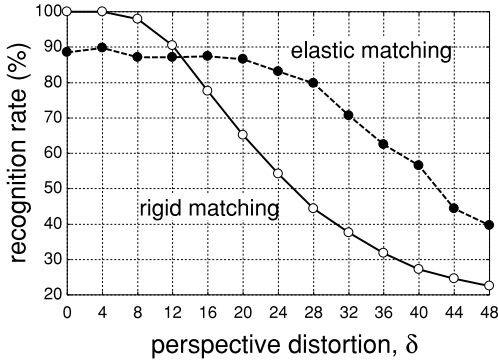


図 6 文字形状単独での認識率 (単純マッチング及び弾性マッチングによる結果)

Fig. 6 Recognition rates attained by using shape similarity alone.

の(すなわち図 3 の文字パターン)を用いた。図 6 に射影変換の程度 δ による認識率の変化を示す。単純マッチングは射影変換に非常に敏感であり、変形量 δ が少しでも大きくなると、認識率は急激に低下することが分かる。一方、弾性マッチングには変形吸収能力があるため、 $\delta = 28$ 程度まで一定した認識率を維持できていることが分かる。しかしそれ以上の変形が加わると、単純マッチング同様、認識率の低下が見られる。

この結果から、文字形状単独での認識だけでは、実環境においてバーコード並みの精度を達成することは困難と予想される。もちろん、改善の余地は多く残されている。例えば、今回は 2 値の画素値をのみ特徴量とした非常に単純な距離評価を行っている。また、マッチング法についても、今回利用した弾性マッチングの自由度は射影変換よりも高いので、合わせすぎによる誤認識が起こっている可能性が高い。実際、弾性マッチングの最高認識率が 90%程度にとどまっているのは、この単純な特徴量と高すぎる自由度の相乗的影響により、誤認識を多発しているためである(注1)。したがって、より洗練された特徴量やマッチング手法を用いれば、認識率を現状より相当改善できるものと考えられる。しかしながら、この実験結果の傾向、並びに今回の実験があくまで計算機内でのシミュレーションである点を考慮すると、文字形状単独での認識性能には、やはり限界があると思われる。

4.3 複比抽出精度

単純割当てに従って複比を埋め込んだテストパターンについて、それらからの複比抽出精度を実測した。結果を図 7 に示す。相当の射影変換を受けた場合でも、

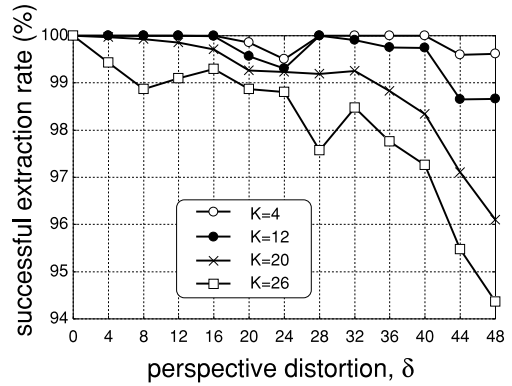


図 7 複比抽出精度

Fig. 7 Extraction accuracy of cross ratios.

埋め込んだ複比 r_k は安定して抽出できていることが分かる。前節で述べた文字形状単独での認識結果と単純に比較すると、その誤認識の数には 1~2 けた程度の違いがある。特に、 $K = 4$ や $K = 12$ の場合の精度は、バーコード精度の認識性能を目指すにあたって複比パターンが有用であることを十分に示唆するものである。

同図における $K = 26$ の結果は、複比単独での文字認識結果に相当する。意図して撮影する場合の射影変換ひずみがたかだか $\delta = 24$ 以下であるとすれば、本論文で用いたような単純な複比パターンでも、文字形状を用いずに 98%以上の認識率が得られることが分かる。

誤抽出原因としては、デジタル画像特有の離散化誤差により複比パターンの各帯の幅 (l'_1, l'_2, l'_3) に誤差が生じたことが挙げられる。実際、 $K = 26$ の場合を例にとると、 r_k を r_{k+1} として誤抽出した場合が全誤抽出の約 85%であった。より深刻な誤抽出 ($r_{k \pm \Delta}, \Delta \geq 2$) は約 10%であった。残る約 5%は幅が細くなったガイドを離散化誤差の影響で捕そくできず、その結果複比パターンを検出できなかった場合であった。

4.4 文字認識結果と複比抽出結果の統合結果

図 8 及び図 9 に、文字形状単独による認識結果と複比による認識結果を統合した場合の認識率を示す。前者は単純マッチングを用いた場合、後者は弾性マッチングを用いた場合である。また、いずれも単純割当てを用いた場合の結果である。

(注1): 特に $\delta \sim 0$ 付近については、合わせすぎにより入力パターンとの距離が 0 となったカテゴリーが複数発生し、タイプレイクの結果、不正解カテゴリーが選ばれてしまった場合が多かった。

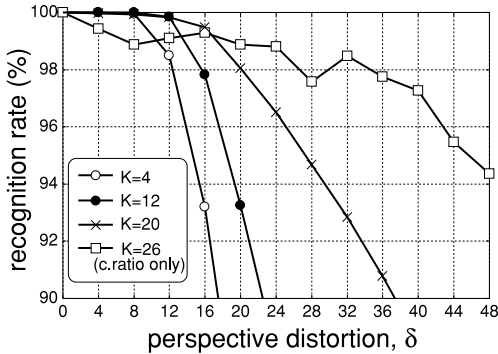


図 8 文字形状と複比の併用による認識率 1. 単純マッチング及び単純割当を使用
 Fig. 8 Recognition rate by rigid matching and cross ratio.

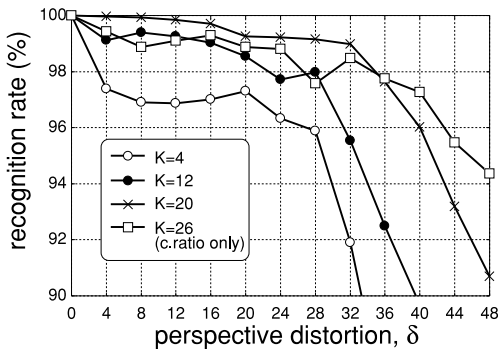


図 9 文字形状と複比の併用による認識率 2. 弾性マッチング及び単純割当を使用
 Fig. 9 Recognition rate by elastic matching and cross ratio.

両図から、形状による文字認識結果と複比抽出結果を併用することで、前者単独での結果(図6)に比べ、認識率を大幅に向上できることが分かる。例えば $\delta = 4$ の場合の弾性マッチングの結果を見ると、文字形状単独での認識率が 89.8% だったのに対し、併用した場合は $K = 4$ で 97.4%, $K = 12$ で 99.1%, $K = 20$ で 99.97% まで改善している。

この改善理由を、 $\delta = 4$ のテストパターンを文字形状単独で認識した際の、弾性マッチングの混同行列(図10)を用いて考察する。この混同行列から、例えば形状により“H”に認識されたパターンの真のカテゴリとしては、“H”と“N”の二つの可能性が考えられることが分かる。すなわち、文字形状単独で認識では、これら2カテゴリの判別にはあいまい性が残るといえる。一方、単純割当方式に従って複比が埋め込まれていたとすると、 $K = 4, 12, 20$ の場合いずれも、“H”と“N”にはそれぞれ異なる複比が割り当てられる(表1参照)。結局、文字形状単独での認識で残っていたあいまい性は、複比により解消され、“H”、“N”は正しく認識できることになる。同様のあいまい性解消が他のカテゴリ間について生じたため、大幅な精度向上が得られたものと考えられる。

再び図8及び図9を見ると、射影ひずみが大きくない場合($\delta \leq 16$ 程度)、複比単独による認識結果($K = 26$)よりも、文字形状による認識を併用した結果の方が、多くの場合高い認識率を得ていることが分かる。複比自身の抽出精度は文字形状による認識精度に比べて非常に高いものの、完全にカテゴリを絞り込めるほど多くの複比を埋め込むと、複比自体の抽出精度が悪化する(図7)。結局、今回の実験の設定では、複比の数ある程度少なめにしてその抽出精度を確保しておいた方が得策であったといえる。

4.5 割当方式の影響

図11は、割当方式の影響を示す結果である。複比の種類 K を4と12とした場合それぞれについて、単純割当と最適割当を用いた結果について示している。文字形状による認識法としては弾性マッチングを用いている。

この結果から、最適割当を用いることにより相当の改善が得られることが分かる。特に $\delta \leq 20$ の場合、単純割当した12種の複比に比べ、最適割当した4種の複比の方が高い認識率が得られている点は注目に値する。このように、複比を割当の際には、文字形状による認識手法の特性を把握した上で、各カテゴリに適

		recognition result																										
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
L B P U L	A	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	B	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D	14	0	0	-	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	E	0	25	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	-	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	I	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	J	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	K	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	L	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0
	N	0	0	0	0	0	0	1	2	0	0	0	0	6	4	-	0	0	0	0	0	0	0	0	0	0	0	0
	O	0	0	0	4	2	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
	P	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0
	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0
	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0
	S	0	4	0	0	0	5	8	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0
	U	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	-	3	0	0	0	0
	V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	8	1	0	0
	W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0
	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0
	Y	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
	Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

図 10 弾性マッチングの混同行列 (変形量 $\delta = 4$ のテストデータを対象とした場合。対角成分は各行の和が 256 になることから算出可能)
 Fig. 10 Confusion matrix of elastic matching.

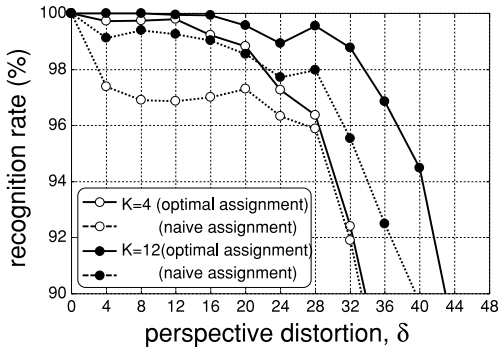


図 11 複比 r_k の単純割当 (naive assignment) と最適割当 (optimal assignment) による認識率の差異. 弾性マッチングを使用

Fig. 11 Naive assignment versus optimal assignment.

切な複比を割り当てることで、大幅な効率化を図ることができ.

5. むすび

本論文では、実環境中の文字パターンをバーコードと同程度の精度で認識することを目標として、文字に機械可読性を補強するような情報を埋め込むことを検討した。具体的には、射影不変量である複比をしま模様状のパターンとして文字に埋め込むことを提案した。あらかじめ複比とカテゴリーに対応を定めておけば、この複比を抽出することで、識別時に付加情報として利用できる。シミュレーション実験の結果、複比と文字形状を認識時に併用することで、射影変換を受けても非常に高い認識精度が得られることが分かった。

本研究には多くの課題が残されている。第 1 には、既に 1. でも述べたように、射影変換以外のひずみ、例えば非均質照明による濃淡レベルのひずみに対する本手法の頑強性を評価する必要がある。そのためには、紙等に実際に印刷した文字をカメラで撮影し、それを対象とした認識・複比抽出実験が必要である。第 2 に、複比パターンのデザインに関しても、離散化誤差の影響を少なくする工夫 [12] や、目立たなくする工夫 (例えば [14]) が必要であろう。第 3 に、文字形状に関する認識手法について、単体での認識率は低くても複比の付加効果が高い手法 (具体的には混同行列が疎になる手法) [13] を開発する必要がある。第 4 に、埋め込んだパターンを手掛りに積極的に射影変換を補償するという方針も併用できるのではないかと考えている。

文字認識を含め従来のパターン認識の研究は、人間可読性だけが低いパターンを何とかして機械認識する

ことを目指してきた。一方、昨今は、ユビキタスコンピューティングに関する研究の隆盛が暗示するように、ビデオカメラ、携帯カメラなどの撮影装置がまさに遍在しており、各種パターンを機械認識すべき機会が多くなっている。こうした状況を考えると、人間可読性と機械可読性の両方が高いパターンを生成し、それを人間と機械の両者を仲立ちするメディアとして活用することが今後重要になってくるものと思われる。本論文で提案した文字パターンは、まさにその一例となり得る。今後も、上述の課題を中心として、検討を進めていく予定である。

謝辞 本研究の一部は公益信託マイクロソフト知的財産研究助成基金研究助成並びにマツダ研究助成の支援を受けた。

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,” Proc. IC-DAR’03, pp.606–616, 2003.
- [3] S. Uchida and H. Sakoe, “A survey of elastic matching techniques for handwritten character recognition,” IEICE Trans. Inf. & Syst., vol.E88-D, no.8, pp.1781–1790, Aug. 2005.
- [4] 大谷 淳, 塩 昭夫, “情景画像からの文字パターンの抽出と認識” 信学論 (D), vol.J71-D, no.6, pp.1037–1047, June 1988.
- [5] 仙田修司, 西山京介, 旭 敏之, “携帯カメラによる日本語文字認識の手法と実現” 信学技報, PRMU2004-124, Dec. 2004.
- [6] 平山勝裕, 大町真一郎, 阿曾弘具, “カラー情報を利用した情景画像中の文字列の高精度抽出” 信学技報, PRMU2004-247, March 2005.
- [7] 佐藤 淳, コンピュータビジョン, コロナ社, 1999.
- [8] The British Computer Society, Character Recognition 1967, Unwin Brothers Limited, 1966.
- [9] 橋本新一郎, 文字認識概論, オーム社, 1980.
- [10] D.L. Hecht, “Printed embedded data graphical user interfaces,” Computer, vol.34, no.3, pp.47–55, March 2001.
- [11] K.L.C. Moravec, “A grayscale reader for camera images of Xerox DataGlyphs,” Proc. British Machine Vision Conference, pp.698–707, Sept. 2002.
- [12] 金澤 靖, 松永 力, 金谷健一, “複比による識別のための最良マーカーパターンの設計: 理論解析” 情処研報, 99-CVIM-115-13, March 1999.
- [13] 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 — 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 (MIRU2005), IS2-85, July 2005.

- [14] 永井悠文, 向川康博, 大田友一, “パターン埋め込みによる動的シーンへの適応的投影” 信学技報, PRMU2004-159, Jan. 2005.

(平成 17 年 6 月 15 日受付, 9 月 16 日再受付)



内田 誠一 (正員)

平 2 九大・工・電子卒・平 4 同大大学院修士課程(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助教授・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・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.



岩村 雅一 (正員)

平 10 東北大・工・通信卒・平 15 同大大学院博士課程了・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, 現在に至る・博士(工学). パターン認識に関する研究に従事・IEEE, 情報処理学会, Pattern Recognition Society 各会員.



大町真一郎 (正員)

昭 63 東北大・工・情報卒・平 5 同大大学院博士課程了・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 同大大学院工学研究科助教授, 現在に至る・博士(工学). その間, 平 12~13 米国ブラウン大学客員助教授・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・IEEE, 情報処理学会, 人工知能学会, Pattern Recognition Society 各会員.



黄瀬 浩一 (正員)

昭 61 阪大・工・通信卒・昭 63 同大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・平 5 同情報助手, 平 8 講師, 平 11 助教授を経て, 平 17 同大大学院工学研究科教授, 現在に至る・博士(工学). その間, 平 12 年~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・情報処理学会, 人工知能学会, 電気学会, IEEE, ACM 等各会員.