

カメラによる文字認識のための付加情報の埋め込みに関する検討

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

[†]九州大学大学院システム情報科学研究科 〒812-8581 福岡市東区箱崎 6-10-1

^{††}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

^{†††}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

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

あらまし デジタルカメラの小型化, 高解像度化, また低価格化に伴う普及に伴い, 実環境での文字認識・文書画像解析への期待が高まりつつある. 本研究では, バーコードと同程度の精度で実環境中の文字/文書を認識することを目標として, 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ことを検討している. 本論文では, 撮影の際の射影変換に対応すべく, 文字画像に対し, 帯状のパターンからなる縞模様を埋め込む手法を提案する. この隣接する帯の幅から計算される複比は射影変換不変, すなわちどのような撮影角度でも常に一定である. 従って, 各文字種について, あらかじめ決まった帯の幅でパターンを埋め込んでおけば, それから計算される複比を識別の手がか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, 複比情報と文字形状情報を併用することで, 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わかった.

キーワード カメラ, 文字認識, 文書画像解析, 情報埋め込み, 複比

Data Embedding for Camera-Based Character Recognition

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

[†] Faculty of Information Science and Electrical Engineering, Kyushu University
Hakozaki 6-10-1, Higashi-ku, Fukuoka-shi, 812-8581 Japan

^{††} Graduate School of Engineering, Osaka Prefecture University Sakai-shi, Osaka, 599-8531 Japan

^{†††} Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan

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

Abstract In this paper, we propose a data embedding method for camera-based character recognition as easy and accurate as bar-code reading. Specifically, a horizontal stripe pattern, called cross ratio pattern, is embedded into a font image of a character category. Since the cross ratio obtained from the pattern is invariant to projective distortion, it always provides correct category candidates regardless of camera angle. Experimental results showed that (i) the cross ratio can be extracted from distorted character images with very high accuracy and thus (ii) the cross ratio can be used for enhancing recognition performance of conventional matching-based recognizers.

Key words camera-based character recognition, embedding information, cross ratio

1. はじめに

文字認識 (OCR) の対象は, 従来スキャナによって撮影された文字画像パターンであった. 非常に高い解像度で入力画像を得ることができる反面, 以下のような制約がある.

- 処理対象に関する制約 … スキャナのほとんどが紙に印刷された文字を対象としている. 従って, スキャナを使う限り, シーン中であって動かせない文字 (例えば看板) は, 認識対象にはならない.
- 装置に関する制約 … 小型化が進んだとは言え, 携帯で

きる小ささには達していない. 電源の問題もある.

- 認識範囲に関する制約 … スキャナは基本的に紙面全体をその入力対象とする. 認識したい個所を限定するには, 一度予備的なスキャンを行い範囲を確定しなくてはならない.

これに対し, 最近ではカメラを文字や文書の入力装置として捉える研究が多くなってきている [1], [2]. 文字認識や文書画像解析への入力手段としてデジタルカメラを捉えると, その携帯性やボタン一つの操作性は魅力的である. スキャナは使いにくいと感じる一般ユーザでも, デジタルカメラならば抵抗が少なく考えられる.

そこで本研究では、バーコードリーダーでバーコードを読み取るような手軽さ・精度を目標として、カメラによる文字・文書の認識を検討する。見方を変えると、人間にも読めるバーコードを目指すことも出来る。このような技術を確認することで、例えば、商品の値札からバーコードが不要になったり、画像中のテキストを手がかりとしたデータベース検索が可能になるなどの、様々な応用・発展が期待される。

反面、処理技術の立場から眺めると、デジタルカメラ画像中の文字・文書の認識・解析にはかなりの困難が伴うことも事実である。スキャナという2次元統制環境下で取得した画像と比べ、カメラ画像は3次元の非統制環境で取得されるものであり、またそれに含まれる文字・文書は多様なものとなる。このため、従来の文字認識の分野では起こりえなかった様々な問題点が生じてくる。具体的な問題点は以下の通りである。

(1) 撮像環境に起因する問題

- 幾何歪み、特に射影変換歪み
- 低解像度
- 非均質照明、ぼけ、手ぶれ

(2) 撮像対象そのものに由来する問題

- オクルージョン
- 複雑な背景
- 非平面上の文字・文書
- 飾り文字

これまでもそれぞれの問題に対し、様々な対処法が提案されている [1], [2]。例えば、幾何歪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]、飾り文字についてはマルチスケールでの細線化とぼかしを併用した構造抽出法 [7]、といった検討が為されている。これらの検討はいずれも、カメラで撮像された文字画像を、なるべく従来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言える。

本研究では、前出の目標を実現するために、これら従来の対処法とは異なったアプローチを採る。すなわち、文字パターン側を工夫して、実環境での機械可読性の向上を図ることとする。その第一段階として、本論文では、撮影時の幾何歪みに対応することを目的として、水平縞模様状のパターン(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。この縞目の間隔から複比 [8] を計算すると、複比は射影変換不変であるので、その文字をどのような角度で撮影したとしても、量子化等の影響がなければ、常に同じ複比を抽出できる。従って、各文字種について、あらかじめ決まった縞模様を持つ複比パターンを埋め込んでおけば、それから計算される複比を識別の手がかりとして認識時に利用できる。

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) の黎明期に既に為されている。すなわち、機械読み取りのために独特の形状を持たせたフォントである [9], [10]。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種類の間隔を持つ7本の垂直線分を用いて設計されており、

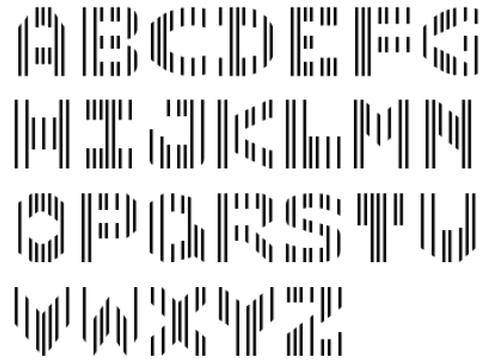


図1 MICR フォント C.M.C.7.

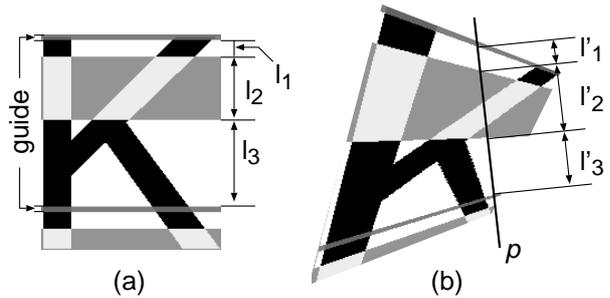


図2 (a) 複比パターンを埋め込んだ文字画像および (b) その射影変換例 ($\delta = 32$)。見易くするために、意図的に複比パターンが目立つような配色にしている。

これら6つの間隔が文字コードを定めている。本論文で提案する情報を埋め込んだ文字パターンは、こうした OCR/MICR フォントをカメラによる文字認識用に拡張したものと言える。

2. 複比を用いた文字への情報埋め込み

文字画像パターンに対し、ある複比 r を持つパターン(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、英語大文字を対象する場合、“A”, ..., “Z” の26カテゴリそれぞれに異なる複比を埋め込んでもよいし、いくつかのカテゴリに同じ複比を埋め込んでもよい。前者の場合、入力文字から複比を正しく抽出できるならば、特に文字画像情報を使わなくても文字を認識できることになる。以下本節では、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

埋め込みの形式については色々考えられるが、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縞を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、複比パターンを目立つように配色している。実際の配色については、6.2節で改めて議論する。

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、

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



図3 複比を埋め込んだ場合の文字パターン ($K = 26$) . これらに射影変換を施したものがテストパターンとなる .

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持つものであるが , ここで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 . 量子化誤差などのノイズが無ければ , 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 , 埋め込んだ時と同じ複比 r_k が求まる . 従って , 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となる .

実際には , 量子化の影響等で , 取り出した複比には誤差が入る . 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なった後 , 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 . 量子化誤差等の影響を最小化しようとすれば , 文献 [11] 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 ならびに l_1 は K 種の複比パターンにおいて共通とする . 従って , $l_2 + l_3 = L - l_1$ 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 .

$$\begin{cases} l_2 = \frac{(L - l_1 - 2\epsilon)(k - 1)}{K - 1} + \epsilon \\ l_3 = L - l_1 - l_2 \end{cases} \quad (2)$$

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 , それが取りうる区間を等間隔に $K - 1$ 分割した際の分割点および区間両端の計 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 が文字カテゴリ数 $|C|$ と同一の場合 (C は文字カテゴリの集合) , 各カテゴリ 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 , 漢字などカテゴリの多い文字集合を対象とすると , 限られた解像度の下で , 多種類の複比を定義することは難しいと考えられる .

より一般的な $K < |C|$ の場合には , 複比とカテゴリは一対多の関係となるため , 複比単独での認識結果には曖昧性が残る . 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のいずれかまではわからない . 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を 1 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には以下 2 点で異なっている . 第一は , 上述のように , 抽出された複比による認識結果には , 曖昧性が残るという点である . 第二は , 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 , 撮像時の射影変換歪みが大きいと , 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 , もしノイズが無ければ常に正しく抽出される . 実際にはノイズ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を幾つかに絞込み , 次に文字形状による認識結果により , カテゴリ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なうことになる .

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 ,

表 1 複比の単純割り当て (naive assignment) と最適割り当て (optimal assignment)

		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
$K = 4$	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
$K = 12$	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
$K = 20$	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

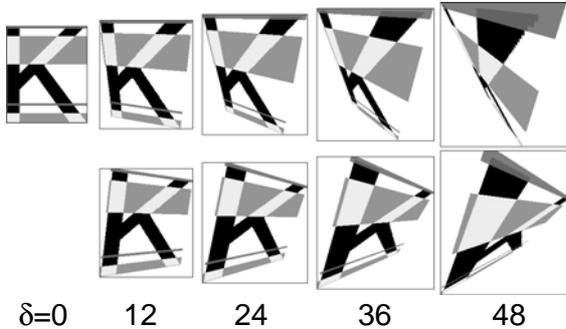


図 4 射影変換により生成されたテストパターンの例。

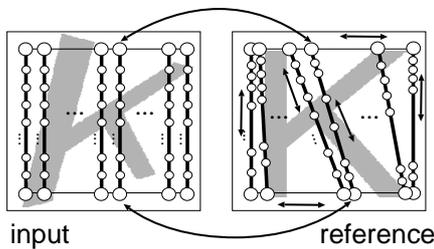


図 5 本実験で用いた弾性マッチング。

かつ (ii) C_k 中の正解カテゴリ以外のカテゴリの認識コストが 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、たとえ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、こうした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) の条件が成り立ち易くなるための工夫である。そのような割り当て $\{C_k | k = 1, \dots, K\}$ 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、それから互いに誤認識対となりにくいカテゴリを知得した後、なるべくそうしたカテゴリが一つの集合 C_k に含まれるように複比割り当て $\{C_k | k = 1, \dots, K\}$ を定めればよい。この割り当ての効果については 4.1.5 において議論する。文献 [12] 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毎で差異が大きく、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、さらに l_1 および ϵ については 15 画素とした。この条件の下、2.3 で述べた方法に従って r_k ($k = 1, 2, \dots, K$) を設定した。さらに、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 $c \in C_k$ のフォントに埋め込んだ。図 3 は、すべてのカテゴリ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 4 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歪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歪みを受けた文字パターンが含まれている。

以上のように、本実験で用いるテストパターンは計算機内で生成したものであり、実際にカメラによって撮像したものではない。従って、本テストパターンは、照明等の影響による濃淡値の変動を含んでいない。カメラを実際に用いた実験については、将来的な課題として、今後行う予定である。

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 2 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、2 パターンの最小距離を求める手法である。実際には 2 画像間の画素対応写像

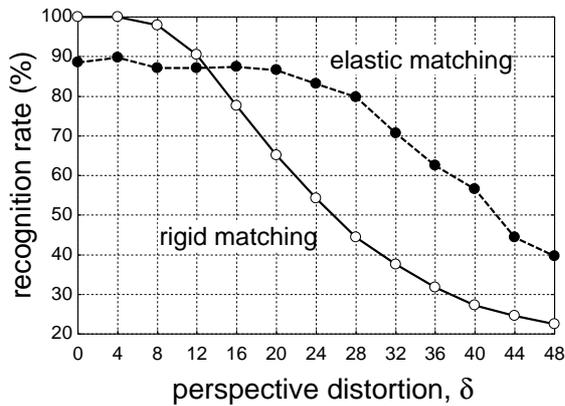


図 6 文字形状単独での認識率 (単純マッチングおよび弾性マッチングによる結果)。

(ワープ)の最適化問題として定式化される。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に示すような写像に基づく手法を用いた。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写像される。その際、各列の内部でも1次元的な非線形写像を許す。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、従って、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なった。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 複比の割り当て

集合 C_k の構成、すなわち複比 r_k をどのカテゴリ 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通りの割り当て方式を使用した。

- 単純割り当て方式 ... カテゴリ 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。

- 最適割り当て方式^(注1) ... 文字形状による認識手法の特性を考慮して、各カテゴリに複比を割り当てる方式[12]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り当ての手がかりとして用いる。例えば“H”という認識結果を与えるカテゴリが“H”と“N”であることが混同行列からわかれば、それら2カテゴリには、曖昧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、全カテゴリについて複比割り当てを定める。

表1に、前節で述べた弾性マッチング法を用いた際の、 $K = 4, 12, 20$ の場合の単純割り当てと最適割り当てを示す。いずれの K の時も、最適割り当てにより、“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カテゴリ間で誤認識が起りにくいことを示唆している。なお、最適割り当ての決定に際しては、 $\delta = 0 \sim 48$ すべてのデータを対象とした

(注1): 文献[12]のアルゴリズムでは大局的最適割り当てができる保証はなく、従って準最適割り当てと呼ぶ方が正確であるが、単純割り当てと対照するために、ここでは単に最適割り当てと呼ぶ。

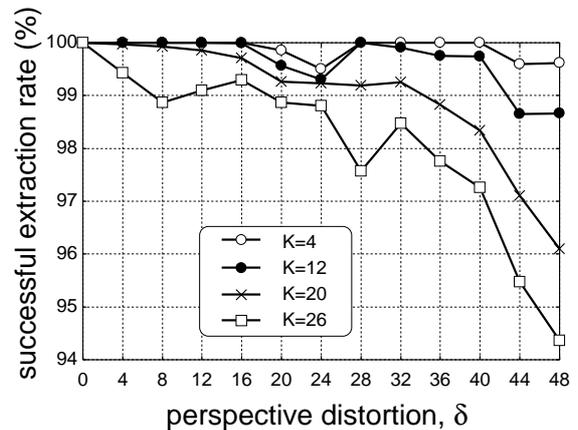


図 7 複比抽出精度。

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2つの手法(単純およ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なった。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わかる。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わかる。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値の画素値のみ特徴量とした非常に単純な距離評価を行なっている。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、合わせ過ぎによる誤認識が起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%程度に留まっているのは、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2)。従って、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、なら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

単純割り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、埋め込んだ複比 r_k は安定して抽出できていることがわかる。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

(注2): 特に $\delta \sim 0$ 付近については、合わせ過ぎにより入力パターンとの距離が0となったカテゴリが複数発生し、タイプレイクの結果、不正解カテゴリが選ばれてしまった場合が多かった。

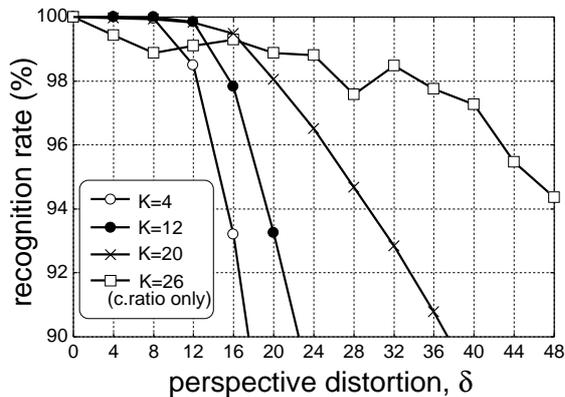


図 8 文字形状と複比の併用による認識率 1. 単純マッチングおよび単純割り当てを使用.

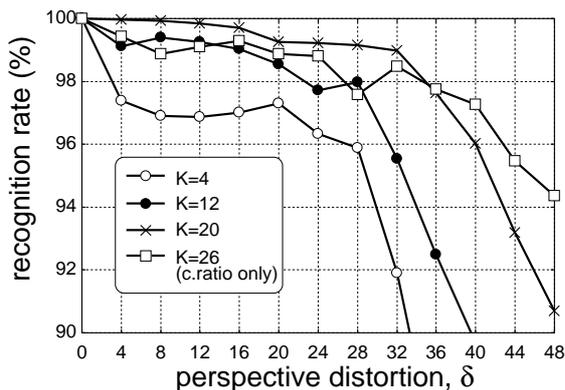


図 9 文字形状と複比の併用による認識率 2. 弾性マッチングおよび単純割り当てを使用.

桁程度の違いがある. 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当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.

また, 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歪みが高々 $\delta = 24$ 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わかる.

誤抽出原因としては, 量子化により複比パターンの各帯の幅 (l_1, l_2, l_3) に誤差が含まれたことが挙げられる. 実際, $K = 26$ の場合を例に採ると, r_k を $r_{k \pm 1}$ 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% は量子化の影響でガイドを捕捉できず, 複比パターンそのものが検出できなかった場合であった.

4.4 文字形状による認識結果と複比による認識結果の統合による認識実験

図 8 および図 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, いずれも単純割り当てを用いた場合の結果である.

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わかる. 例えば $\delta = 4$ の場合の弾性マッ

		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
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	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	241	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	26	0	0	230	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	250	0	0	0	0	0	0	6	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	0	182	0	0	0	0	0	64	10	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	42	0	0	0	0	0	0	0	0	0	0	214	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	252	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	6	0	0	0	0	0	0	0	0	0	0	24	0	226	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	58	0	0	0	0	0	0	0	0	0	0	0	0	194	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	113	0	0	0	0	0	0	0	0	140	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	81	175	0	0	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	10	0	0	0	0	0	0	0	0	0	0	0	1	0	51	0	0	203	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. アスタリスク “*” は 256 を意味する.)

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 2 つの可能性が考えられることがわかる. 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曖昧性が残ると言える. 一方, 単純割り当て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曖昧性は, 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. 同様の曖昧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 および図 9 を見ると, 射影歪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わかる. 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, 複比の数を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言える.

4.5 割り当て方式の影響

図 11 は, 割り当て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, 単純割り当てと最適割り当て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, 最適割り当てを用いることにより相当の改善が得られることがわかる. 特に $\delta \leq 20$ の場合, 単純割り当てした 12 種の複比に比べ, 最適割り当てした 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, 複比を割り当てる際には, 文字形状による認識手法の特性を把握

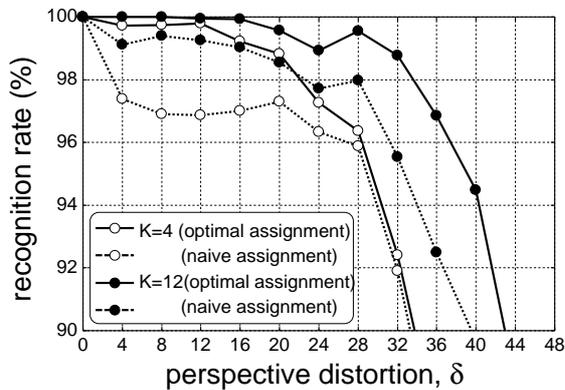


図 11 複比 r_k の単純割り当て (naive assignment) と最適割り当て (optimal assignment) による認識率の差異. 弾性マッチングを使用.

した上で, 各カテゴリに適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る.

5. 他の技術との関連

本研究の目標は機械可読性の高い文字パターンの実現であるが, 類似の目標を持って開発されたメディアとしては, 冒頭で述べた OCR/MICR フォント以外にも, バーコード [13] ~ [15], RFID (Radio Frequency Identification) [16], 電子透かし (例えば [17]), DataGlyph [18] が挙げられる. 以下, 本節では, これら従来メディアと本研究の関連について述べる.

5.1 バーコード, RFID, および電子透かしとの関連

バーコードは, 機械にデータを読み取らせる手段として最も普及しているものである. 通常の 1 次元バーコードに加え, 最近では QR コード [13] や Visualtag [14] など様々な 2 次元バーコードが開発されている [15]. いずれも印刷だけで実現可能であり, コストが低い点が特長として挙げられる. バーコード利用の中心は, ID 用データや情報コンテンツへのポイントの表現であるが, 2 次元バーコードではコンテンツを記録可能な数キロバイトといった大量のデータを表現することもできる. 可読性という点で見ると, 機械による可読性は高いが, 人間には読み取れない. また, 情報取得は一括であって一部を選択的に読むことはできない. 加えて, 読み取りは逐次的であって, 複数のバーコードがある場合には基本的には数だけ読み取りを繰り返す必要がある.

RFID はコピキタスコンピューティングなどの概念とともに注目を集める技術であり, バーコードの代替として捉えられることも多い. このメディアもバーコードと同様, 機械可読性は非常に高い反面, 人間がその情報を直接読むことはできない. また, 情報の部分的な読み取りも不可能である. 一方, バーコードとは異なり, 近くに複数の RFID が存在すれば, それらの情報を一括して読み取ることが可能である. しかし, このことは, 自分の意図していない情報が混入してしまう場合があることを意味する.

電子透かしは, 写真などの領域にデータを埋め込む技術である. 印刷の場合には数バイト程度のデータ量となるが, 外観を

ほぼ損ねずに情報を表現可能である. ただし, 埋め込みの存在は人間には知覚できず, どこを読めば情報が取得できるのかわからないため, 読み取りに問題が生じることも考えられる.

以上と比較して, 情報を埋め込んだ文字の最も大きな特徴は, 人間に可読であるという点 [19], [20] である. 他の情報表現形態では, どのような情報が取得されたのか, ユーザである人間が直接知ることはできないが, 文字は人間にも読めるため, これが可能である. すなわち, 「モノに情報を貼り付ける」だけではなく, 「貼り付けた情報を知る」ことが可能である. この特徴によって, 取得したい情報をユーザが選択することも実現できる. さらに文字は「人間が読むもの」なので外観を悪化させないという利点もある. 従って, 文字単独での機械可読性は他と比較して低いものの, 情報埋め込みにより可読性を向上できれば, 機械による文字の利用は十分促進されうると考えられる.

また, 文字形状単独でもある程度の機械可読性が確保される点も重要である. 従って, 情報を埋め込んである文字も, そうでない文字も, 区別無くカメラ経由での認識が可能である. もし, 情報が埋め込まれてなかったり, その抽出に失敗したとしても, 精度が下がるだけで, 常に何らかの情報を取得できることになる. これは電子透かしや RFID には無い特徴である.

5.2 DataGlyph との関連

残る DataGlyph [18] は, 外観への影響を極力排除した文字へのデータ埋め込み方式である. 具体的には, 文字図形を細かな「/」, 「\」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. 文字に直接データを埋め込むと言う点で本検討との共通点も多く, その用途など, 参考になる点も多い. しかし, 元来スキャナでの撮像を前提として設計されたものであり, カメラでの撮像を前提としたものではない. 従って, カメラで撮像された DataGlyph から情報抽出する場合は, 撮像時の歪みを補正しながら元の DataGlyph を推定するという方法しかない [21]. これに対し本研究では, カメラでの撮像を想定した上での情報埋め込みを検討するものであり, より直截なアプローチであると言える.

6. 今後の課題

本論文は, 文字パターンへの情報埋め込みに関する試みの第一歩であり, 多くの課題が残されている.

6.1 実環境下での実験

本論文では, 計算機内で射影歪みを施した文字パターンの認識および複比抽出実験を行なった. 4.1.3 で述べたように, この実験設定には, 実環境と異なる点がある. 今後は, 紙等に実際に印刷した文字をカメラで撮像し, それを対象とした認識・複比抽出実験が必要である. 特に, (i) 撮像時の各画素値の変動に対する頑強性, (ii) 実際に目的意識を持って撮影する場合に受ける射影変換の程度, を明らかにする必要がある.

6.2 複比パターンの設計

複比パターンの設計についても, 検討の余地が多く残されている. 本論文では簡単のために, l_1 を固定として l_2, l_3 を可変としたが, これら 3 つをすべてを可変にすることで, より設計の自由度を向上できる. さらに, l_1, l_2, l_3 の設定に際しては, ρ

イズ・量子化誤差による抽出誤差を考慮した上で、複比の誤抽出が最も少なくなるように考慮すべきと考えている [11]。

埋め込む情報量を増やすために、帯の数を増やすことで複数の複比を連続的に埋め込むことも考えられる [11]。水平縞模様と垂直縞模様も併用する、文字線部と背景部で別の複比を埋め込むといった発展形も考えられる。

関連して、背景デザインに影響を与えないために、文字線のみ複比パターンを埋め込むという検討も必要と思われる。この場合、例えば“S”のように曲線の多い文字については、直線を引いて複比を抽出するという方法だけでは不十分と考えられる。

本論文のシミュレーション実験では、テストパターン生成の際に画素値(濃淡値/色)の変化が無い場合、複比パターンの配色について特に配慮しなかった。しかし、実際には、画素値は撮像時に変化するため、隣接する帯の境界がなるべく明確になるような画素値とする必要がある。さらに人間には見えづらく、一方でカメラには見えやすいようなパターンであることが望ましい。この問題に関して、文献 [22] のアクティブビジョンに関する検討は参考になる。

以上に加え、射影変換不変量として、複比以外のものを用いる可能性も検討する必要があると思われる。

6.3 文字形状による認識手法に関する検討

文献 [12] で指摘されているように、複比のような付加情報を併用した場合、文字形状による認識率の優劣が、手法間で逆転する可能性がある。具体的には、その混同行列を見たとき、同一列内に非ゼロの要素が少ない(すなわち混同行列が疎であるほど)ほど情報付加の効果が高い。この場合は、最適割り当ての効果も併せて高くなり、少ない付加情報でも相当の認識率向上を見込める。逆に、たとえ全体の誤認識数が少ない認識手法であっても、同一列内に非ゼロの要素が多い(すなわち混同行列が疎でなくなるほど)、情報付加の効果は低くなる。従って今後は、単体での認識率は低くても複比の付加効果が高いような認識手法も開発する必要がある。

6.4 埋め込む情報

文字に埋め込む情報については、用途に応じて様々考えられる。前節の実験では、「人間が読むもの」と「機械が読むもの」を同一化すべく、各文字にそのカテゴリを表す複比を埋め込むことを試みた。しかし、こうしたアプローチ以外にも、文字カテゴリとは独立に情報を埋め込むことも考えられる。例えば、名刺の自分の名前部分に、URL を埋め込むといった応用が考えられる。

7. む す び

本研究では、実環境中の文字・文書をバーコードと同程度の精度で認識することを目指し、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ことを検討した。具体的には、ある複比を縞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。各文字種について、あらかじめ決まったパターンを埋め込んでおけば、認識時にはそれから計算される複比を識別の手がかりとして利用できる。シミュレーション実験の結果、複比情報と文字形状情報を併用することで、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わかった。

従来のパターン認識手法は、機械認識を前提とせずに生成されたパターン、すなわち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、研究開発されてきた。手書き文字認識はその一例である。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像装置がまさに遍在しており、パターンが機械認識される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、まさにその一例と考えている。今後も、前述したような課題を中心として、検討を進めていく予定である。

文 献

- [1] 黄瀬, 大町, 内田, 岩村: “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 (2005).
- [2] D. Doermann, J. Liang and H. Li: “Progress in camera-based document image analysis”, Proc. ICDAR’03, pp. 606–616 (2003).
- [3] S. Uchida and H. Sakoe: “A survey of elastic matching techniques for handwritten character recognition”, IEICE Trans. Info. & Syst., E88-D, To appear (2005).
- [4] 大谷, 塩: “情景画像からの文字パターンの抽出と認識”, 信学論 (D), J71-D, 6, pp. 1037–1047 (1988).
- [5] 仙田, 西山, 旭: “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 (2004).
- [6] 平山, 大町, 阿曾: “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 (2005).
- [7] S. Omachi, M. Inoue and H. Aso: “Structure extraction from decorated characters using multiscale images”, IEEE Trans. PAMI, 23, 3, pp. 315–322 (2001).
- [8] 佐藤: “コンピュータビジョン”, コロナ社, 東京 (1999).
- [9] The British Computer Society: “Character Recognition 1967”, Unwin Brothers Limited (1966).
- [10] 橋本: “文字認識概論”, オーム社 (1980).
- [11] 金澤, 松永, 金谷: “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 (1999).
- [12] 岩村, 内田, 大町, 黄瀬: “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) (2005).
- [13] <http://www.qrcode.com/>.
- [14] 磯, 倉掛, 杉村: “携帯電話機搭載カメラを用いた読取り条件にロバストな id 読取り方式”, 信学論 (D-II), J87-D-II, 12, pp. 2188–2197 (2004).
- [15] <http://www.adams1.com/pub/russadam/stack.html>.
- [16] 増井: “スマートタグ連載開始にあたって”, 情報処理, 45, 1 (2005).
- [17] 中村, 片山, 山室, 曾根原: “カメラ付携帯電話機を用いたアナログ画像からの高速電子透かし検出方式”, 信学論 (D-II), J87-D-II, 12, pp. 2145–2155 (2004).
- [18] D. L. Hecht: “Printed embedded data graphical user interfaces”, IEEE Computer, 34, 3, pp. 47–55 (2001).
- [19] 山田: “ユビキタス情報インタフェースのための文字・文書メディア認識・理解”, 信学技報, PRMU2003-229 (2004).
- [20] 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, 45, 9, pp. 923–927 (2004).
- [21] K. L. C. Moravec: “A grayscale reader for camera images of Xerox DataGlyphs”, Proc. of The British Machine Vision Conference, pp. 698–707 (2002).
- [22] 永井, 向川, 大田: “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 (2005).