

# 複数フォント 3000 カテゴリを対象とした回転文字認識

馬場 悠太<sup>1,a)</sup> 長谷 博行<sup>1,b)</sup> 東海 彰吾<sup>1,c)</sup>

**概要:** 筆者らはこれまでパラメトリック固有空間法を用いた回転文字認識の研究を行ってきた。入力画像は固有部分空間に写像されるが、その際入力画像を一定角回転し複数の画像を生成してそれらを固有部分空間に投影することにより、単純に投影するよりも偶発的な誤分類を避けることができる。その結果、単一フォントの漢字を対象とした場合では 10 次元程度で 99% 以上の正読率が得られ、角度推定も高い精度であった。本報告ではこの方法を JIS 第一水準漢字、ひらがな、カタカナ全 3,133 カテゴリに適用し、認識実験を行った。学習文字は 10 度ごとに回転した 36 個のデータを用いて固有部分空間を生成し、テストデータは 7 度ごとに回転した総数 319,566 データで実験を行った。次に実際にカメラで撮影された文字画像には二次元的な回転だけでなく三次元的な回転が含まれていると考えられるため、三次元的に変形した文字画像についても同じ辞書構成でどの程度対応可能なのか実験を行った。さらに認識結果として回転角度を得ることができるのを利用して文書画像の回転角度推定を行った。最後に、現在のままでは 1 文字の認識を行うために 0.6[秒] 程度かかってしまっているため、認識に要する時間を短縮するために正しく認識される際の距離のデータを閾値に利用して計算時間を短縮した。

**キーワード:** 固有部分空間, 回転文字, 文字認識, JIS 第一水準漢字

## Rotated Character Recognition over 3,000 Categories with A Few Popular Fonts

**Abstract:** We proposed a rotated character recognition method using eigen-subspace characterized multiple projection and simple projection so far. In our previous work, a high recognition accuracy over 99.8% was achieved especially by multiple projections of the input rotated images. In this paper, a challenge of rotated character recognition for ideal character images of Japanese Kanji, Hiragana and Katakana with two popular fonts of Mincho and Gothic is presented. First, we obtained very high accuracy and high angle precision for 3,133 Japanese categories of two fonts. Second, this method were tested for not only 2D rotation but also 3D rotation of a character image. As the result, when the angles of X and Y axes were within 6 degrees, we found out that we can get more than 90% of recognition rate. Next, we experimented for camera images, then when the character size is over 45×45 pixels, it is found that we can get a high accuracy like the ideal character image. Next, we applied this method to a rotated document image with a complicated layout. As the result, we obtained a good estimation statistically by voting the recognized angle of a connected component. At last, although the computation time using all dimensions per a character was 0.6[s], the computation time was reduced by decreasing the number of categories using a distance threshold. An advantage of this method is to be obtained the category, the font and the angle at the same time.

**Keywords:** Eigen-space, Rotated character, Character recognition, Kanji

### 1. はじめに

回転文字認識は過去には多くの報告がある [1][2] が、近年

では村瀬らのパラメトリック固有空間法 [3] に基づき筆者らがやってきた二次元的な回転のみを対象にした研究 [4], さらに三次元的な回転をも対象とした研究 [5] が報告されている。いずれも英数字 62 文字種に対して認識実験を行っている。これら二つの研究は固有空間を用いる点において原理的に同じ手法である。

<sup>1</sup> 福井大学  
University of Fukui, Fukui, Japan  
a) y-baba@monju.fuis.u-fukui.ac.jp  
b) haseh@u-fukui.ac.jp  
c) tokai@u-fukui.ac.jp

本報告では文献 [4] の方法を用いて、膨大なカテゴリ集合からなる JIS 第一水準漢字、ひらがな、カタカナ全 3,133 文字種を対象とした文字認識実験を行った。

一般に、文字認識の研究では実世界からサンプルを収集し、カテゴリ内変動を吸収する特徴抽出あるいは認識アルゴリズムを開発し、認識実験を行い評価する。サンプル収集の代わりに公開された文字画像データベースを用いることもある。印刷漢字データベースでは過去に ETL2 がある [6]。このデータベースには 2,184 文字種が収納されており、特許広報と新聞記事から約 5 万字のデータが収集された。しかし、カテゴリ数が少なく、カテゴリによっては収集データ数が不揃いであり、2 データしかないカテゴリもある。そういう意味では ETL2 のデータは偏っている。確かに全てのカテゴリについて均一に実データを準備するというのは困難な作業である。実際の文字認識研究では、使用するデータや手法によって認識結果の捉え方を考えなければならない。すなわち、認識率が低下する原因として、画像ノイズ、フォント、特徴抽出法、認識法など多くの要因が考えられる。システム性能に関わるそれらの要因を極力少なくすることと、サンプル収集の容易さのため、本稿ではノイズフリーの文字画像データを扱う。このようなデータを取得することは現実には難しいが、画像入力時にはなるべく画像品質の良さを目指して条件を設定する。そのため、本研究による結果は、もっとも理想的な文字形状に近い実データが得られた場合のほぼ最良の値として本実験結果を参考にして頂くとよい。すなわち、本研究では画像サイズ縮小と次元制限による情報欠落と競合カテゴリのみが認識率低下の要因となっている。先の研究においても、まず理想文字データを用いてシミュレーションを行い [4]、その後、カメラ入力による実時間システムを構築してシミュレーション結果に近い結果を得ている [8]。さらに、日本語回転文字認識については実時間システムは構築できていないが、シミュレーション実験の結果に近い結果をカメラ入力された文字を対象にした実験を行い得ることができている [9]。

シミュレーション実験では自動生成した文字画像を用い、JIS 第一水準漢字、ひらがな、カタカナという膨大なカテゴリ集合に対し、画像平面上に対して任意の回転角の入力画像に対し、高精度な結果を得ることができた。さらに三次元的に回転した画像を用意して認識を行い、三次元的な変動が認識率にどう影響するかを確認した。つぎに、カメラで撮影された文字画像に対する認識実験では様々なサイズの文字画像を撮影し、サイズの違いが認識率に与える影響や、ある程度のサイズの文字画像が得られた際の認識に用いる次元による認識率の変化を調べた。さらに、認識結果として回転角度が得られることを利用して、文書画像から文字画像を抽出し認識を行うことによって傾いている文書画像の角度推定を行った。最後に認識にかかる時間

を短縮するために、認識を行う際に辞書として登録されている軌跡上の点との距離を計算しているが一定の距離値によって計算を打ち切ることによって認識にかかる時間の短縮を行い、その際の認識率と計算時間を調べた。

## 2. 漢字画像の生成と認識方法

本研究では漢字画像を自動生成したので、その方法と、文献 [4] で採用した学習方式、認識方式について述べる。

### 2.1 漢字画像データ生成

本研究ではソフトウェアフォントエンジンである FreeType ライブラリを用いて漢字画像を自動生成した [7]。このプログラムにより、フォントや文字の大きさを指定してビットマップ上に画像として文字を描画することができる。本研究では、ゴシック体、明朝体の 2 種類のフォントで JIS 第一水準漢字、ひらがな、カタカナ全 3,133 文字を 128×128 画素の 2 値画像として出力した。各カテゴリでは、この文字画像を 10 度ごとに回転し、最小正方形で切り取り 50×50 画素の 2 値画像として 36 個の回転文字画像を生成して学習に用いた。さらに、文献 [8] では、32×32 画素の文字画像を用いて英数字を対象とした認識実験を行ったが、高速化のため文字サイズを 8×8 画素 17 値画像に変換して実験を行い、認識率においてほとんど低下が見られないという結果を得ているので、本報告においても上下左右の隣接小領域で 2 画素重複させて 50×50 画素 2 値画像から 8×8 画素 65 値画像を生成した。漢字画像で 8×8 画素はかなり粗いように思えるが、その分 65 値にすることで情報を残している。この特徴抽出法はひとつの簡易な手段である。このことにより 2,500 次元から 64 次元となり格段に計算速度が上がる。

### 2.2 学習方法

例えば、50×50 の 2 値画像は 2,500 次元ベクトルで記述でき、各要素は 0 または 1 である。8×8 の 65 値画像では 64 次元ベクトルで記述でき、各要素は 65 値である。一般に画像データを  $f_{\theta(i)}^{(j,k)}$  と表記し、 $j$  はフォント、 $k$  はカテゴリ、 $\theta(i)$  は傾き  $\theta(i) = 10 \times i | i = 0, 1, \dots, 35$  を表す。

次に、各々のカテゴリにおいて 36×フォント数個の回転文字画像データを使って固有部分空間を作る。同一文字種の異なるフォントを他カテゴリと考える方法もあるが、それではカテゴリ数が 2 倍になり計算時間も 2 倍になるため、本研究では異なるフォントの同一文字種は統一的に扱う。共分散行列  $\Sigma^{(k)}$  は、

$$\Sigma^{(k)} = E_{j,i} \left[ \left( f_{\theta(i)}^{(j,k)} - m^{(k)} \right) \left( f_{\theta(i)}^{(j,k)} - m^{(k)} \right)^t \right] \quad (1)$$

で計算する。ただし、 $m^k$  は  $k$  番目のカテゴリの平均ベクトルである。共分散行列は次式を満足する。

$$\Sigma^{(k)}\phi = \lambda\phi \quad (2)$$

ただし、カテゴリ添字  $k$  は  $\lambda$  と  $\phi$  に対して省略している。共分散行列のランクは最大 64 であるから最大 64 個のゼロ以外の固有値を得る。固有値を  $\lambda_1, \lambda_2, \dots, \lambda_{64}$ , 対応する固有ベクトルを  $\phi_1, \phi_2, \dots, \phi_{64}$  とする。最初の  $n (\leq 64)$  個の固有ベクトルを用いてカテゴリ  $k$  の固有部分空間  $U_n^{(k)} = \{\phi_1, \phi_2, \dots, \phi_n\}$  を形成する。

次に、 $f_{\theta(i)}^{(j,k)}$  ( $i = 0, 1, \dots, 35$ ) の  $U_n^{(k)}$  上への投影点  $F_{\theta(i)}^{(j,k)}$  は  $U_n^{(k)\perp} (f_{\theta(i)}^{(j,k)} - m^{(k)})$  で計算でき、角度は連続的に変化するので投影点の集合  $\{F_{\theta(i)}^{(j,k)}\}$  は連続な軌跡を描く。得られた各フォント 36 点からの軌跡は適当な補間法により密に埋めることができる。本研究では周期スプライン補間法を用いて各フォント 360 点とした。補間点の角度は  $F_{\theta(i-1)}^{(j,k)}$  と  $F_{\theta(i)}^{(j,k)}$  間を 10 等分し整数角度を与える。故に、角度精度は 1 度単位となる。図 1 にカテゴリ “亜” のゴシック体と明朝体の三次元固有部分空間上の軌跡  $L_3^{(k)}$  を示す。

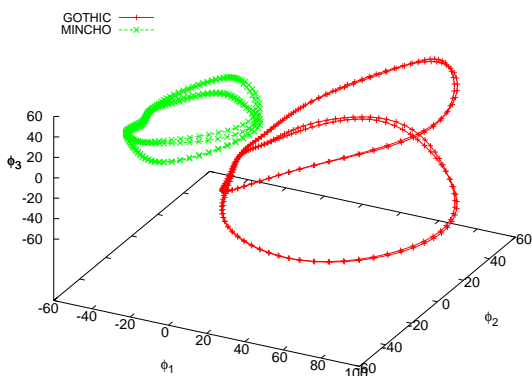


図 1 カテゴリ “亜” のゴシック体と明朝体の軌跡  
Fig. 1 Loci in different fonts of the category “亜”

## 2.3 認識方法

### 2.3.1 単純投影による認識

与えられた未知文字画像データ  $x$  を、すべての  $U_n^{(k)}$  ( $k = 1, 2, \dots, C$ ) 上に投影する。  $x$  のカテゴリ  $k$  の固有部分空間上への投影点  $X$  は  $X = U_n^{(k)\perp} (x - m^{(k)})$  で表される。  $X$  と  $L_n^{(j,k)}$  上の最小の距離の点を探すことによって照合は行われる。カテゴリ  $k$  の最小距離を  $d^{(j,k)}(X)$  とすると、認識結果  $(j^*k^*)$  は、

$$(j^*, k^*) = \arg \min_{j,k} \left\{ d^{(j,k)}(X) \right\} \quad (3)$$

と表すことができる。

推定角度は軌跡上の最近傍点によって与えられ、認識結果としてフォント、カテゴリ、角度を同時に得ることができる。単純投影の認識過程のイメージを図 2 に示す。

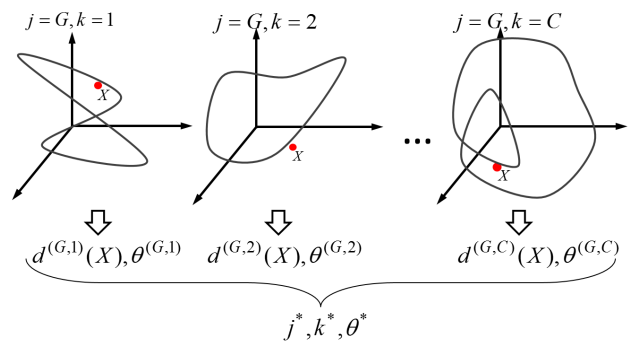


図 2 単純投影による認識過程

Fig. 2 Recognition scheme by simple projection

### 2.3.2 複数投影による認識

前述の方法では、未知パターンの固有部分空間上への投影点が偶然に他カテゴリの軌跡に近くなる場合が多いと考えられる。この偶然性による誤認識を防ぐため、認識の過程において、入力された角度を基準として、未知画像を均等に一定角度回転して複数の画像を作成し（元画像を含め、生成された個数を  $R$  と表記する）、それぞれを固有部分空間に投影する方法を考案し、英数字を対象とした実験で認識率の高精度化を実証している。例えば、 $R = 3$  では 120 度と 240 度回転した文字画像を生成する。先の研究 [4] では  $R$  を 1 から 5 まで変えて実験した結果、 $R = 3$  や  $R = 5$  の場合が  $R = 2$  または 4 の場合より良い性能を示した。なお、上記の単純投影は  $R = 1$  の場合である。

図 3 は  $R = 3$  の場合である。認識結果  $(j^*, k^*)$  はカテゴリ  $k$  同一フォントの軌跡上での 3 つの距離  $d_r^{(j,k)}(X)$  を平均してカテゴリ・フォントの距離とする。

$$(j^*, k^*) = \arg \min_{j,k} \left[ \frac{R}{E} \left\{ d_r^{(j,k)}(X) \right\} \right] \quad (4)$$

$E\{\cdot\}$  は平均値演算を表す。

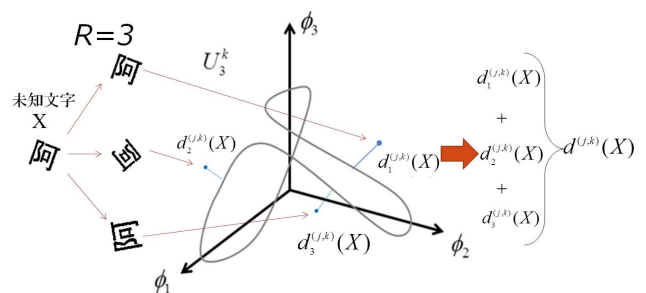


図 3 複数投影による認識過程

Fig. 3 Recognition scheme by multiple projections

## 3. 自動生成した文字画像に対する認識結果

認識実験の内容は、テストサンプルとして 7 度から 357 度まで 7 度ごとに回転した文字データを、各フォント、各カテゴリにつき 51 個、テストサンプルは全 319,566 個を

用意した. ゴシック体, 明朝体の2フォント3,133カテゴリを対象とした認識実験の認識に使用する次元に対する認識率の変化を図4に示す. 認識に全64次元を使用した際の認識率は99.84%となった. 単純投影による認識方法では低い次元において認識率が低くなっているが, 複数投影による認識を行うことによって低次元でも認識率を向上させることができている.

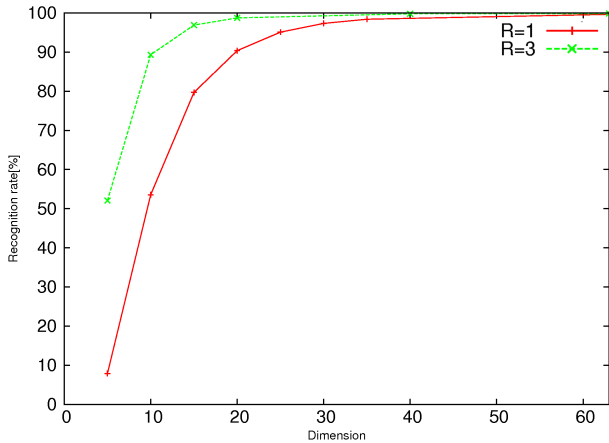


図4 認識に用いる次元に対する認識率

Fig. 4 Recognition rate by simple projection and multiple projection

図5に認識方法として単純投影による認識を用い, 認識に64次元すべてを使用した際のカテゴリが正しく認識された文字画像についての推定角度と真の角度との差のヒストグラムを示す. 図5からもわかるように, ほとんど真の角度と±1度以内で推定され非常に高い精度で推定されていることがわかる.

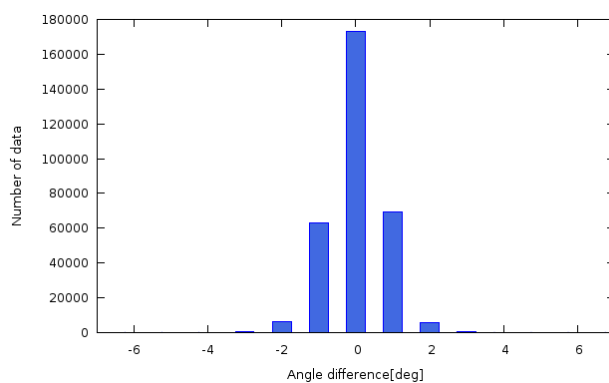


図5 角度推定の精度

Fig. 5 Precision of estimated angle errors

表1に全64次元を認識に使用した際の, 単純投影による認識 ( $R = 1$ ) と複数投影による認識 ( $R = 3$ ) における誤認識結果の例を示す. 表1の(G), (M)はそれぞれゴシック体と明朝体の2種類のフォントを表している. 誤認識の傾向として, 明朝体が同一カテゴリのゴシック体に誤認識

することが多く, その他の誤認識については入力文字と非常に似た形状の文字に誤認識される傾向にある.

表1 誤認識の例

Table 1 Examples of misclassification

Input Result	$R = 1$	$R = 3$
轟 (M) ⇒ 轟 (G)	35	22
銅 (M) ⇒ 鋼 (M)	22	5
膏 (M) ⇒ 膏 (G)	21	10
鷺 (M) ⇒ 鷺 (G)	17	5
園 (M) ⇒ 圃 (M)	14	3
驚 (M) ⇒ 驚 (G)	12	1
誓 (M) ⇒ 誓 (G)	10	4
筒 (M) ⇒ 簡 (M)	10	4
壇 (M) ⇒ 壇 (G)	9	0
問 (M) ⇒ 問 (M)	9	2
others	256	5

### 3.1 三次元変動に対する認識性能

ここまでのシミュレーション実験では理想画像を用い画像平面上のみの二次元的な回転をした文字画像を認識対象として実験を行ってきたが, 次のシミュレーション実験では三次元的に回転した文字画像を認識対象として認識実験を行った. テストサンプルは画像中心から下向きにX軸, 右向きにY軸を設定し, X軸, Y軸まわりに-15度から15度まで3度刻みで回転させた画像をテストサンプルとして用いて実験を行った. 辞書構成は変更せず, 二次元的な回転のみの画像で作成された辞書である. 図6に作成したテストサンプルの例を示す.

		x軸回転												
		-15	-12	-9	-6	-3	0	3	6	9	12	15		
y軸回転	-15	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	-12	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	-9	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	-6	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	-3	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	0	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	3	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	6	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	9	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
	12	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜
15	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	亜	

図6 作成したテストサンプルの例

Fig. 6 Example of 3D rotated images



全 64 次元で単純投影による三次元的な回転角度に対する認識率を図 7 に示す。図 7 は認識率が 50% を超えている場合に赤で色付けされ、50% から 100% まで 10% ごとに色が徐々に濃くなるように表示している。色の变化からもわかるように、三次元的な回転量が大きくなるほど認識率が低下している。X 軸まわりに  $\pm 3$  度、Y 軸まわりにも  $\pm 3$  度以内の変化であれば、99% 以上の認識率を得ることができ、X 軸まわりに  $\pm 6$  度、Y 軸まわりにも  $\pm 6$  度以内の変化では、90% 以上の認識率となった。

		x軸回転										
		-15	-12	-9	-6	-3	0	3	6	9	12	15
Y 軸 回 転	-15	9.88	17.93	27.82	36.63	43.69	45.51	43.83	36.86	27.89	19.60	12.18
	-12	16.32	30.56	46.22	59.16	68.18	73.07	69.97	59.02	45.45	29.53	17.71
	-9	23.02	42.98	63.09	79.24	87.44	90.22	88.63	80.25	63.32	42.95	25.16
	-6	27.08	53.37	77.22	91.06	96.46	97.77	96.95	91.57	78.09	55.60	31.30
	-3	28.43	58.82	85.23	96.22	99.14	99.61	99.31	96.24	85.80	61.99	34.87
	0	29.26	61.33	87.54	97.82	99.71	100.0	99.54	97.47	86.39	65.01	37.27
	3	28.01	61.37	87.20	97.55	99.60	99.66	99.21	96.49	87.49	65.26	39.27
	6	26.91	56.53	80.81	94.57	97.72	98.40	97.23	93.24	84.01	63.91	40.93
	9	24.13	47.98	71.38	85.95	92.26	93.86	91.80	86.76	75.85	58.55	38.67
	12	18.20	35.78	54.81	69.01	77.67	79.53	78.63	72.78	60.62	45.30	33.46
	15	12.02	22.31	34.94	45.26	53.00	54.94	53.41	48.52	39.39	30.62	32.88

図 7 各角度に対する認識率  
Fig. 7 Recognition rate by 3D rotated images

#### 4. カメラで撮影された文字画像に対する認識結果

3 章では、収集の容易性のため自動生成されたノイズフリーな文字画像を用いて認識実験を行ってきたが、本章ではゴシック体と明朝体の文字を印刷しカメラで撮影した文字画像テストサンプルとして用いた実験の結果を示す。

はじめに、撮影された文字画像のサイズによって認識率がどのように影響を受けるかを調べる実験を行った。実験内容としては、3,133 カテゴリの中から 200 カテゴリをランダムで選び印刷しカメラで距離を変えて撮影してテストサンプルを収集する。カメラで撮影された文字のサイズは様々であるが、認識を行う際は一度  $50 \times 50$  の二値画像に変換しその後  $8 \times 8$  の 65 値画像に変換し認識を行う。図 8 に文字画像のサイズに対する認識率の変化を示す。文字画像サイズは撮影された文字画像を最小正方形で切り出した際の一辺のサイズ (pixel) を表している。認識を行う際に  $50 \times 50$  の二値画像に変換してから  $8 \times 8$  の 65 値画像に変換しているため、3 章の理想画像と同等の認識結果を得る

には、撮影された文字画像のサイズが  $50 \times 50$  [pixel] 程度でないと単純投影 ( $R = 1$ )、複数投影 ( $R = 3$ )、画像相関による認識 ( $corr$ ) において高い認識率を得ることができていない。しかし、 $45 \times 45$  [pixel] 以上のサイズの文字画像ならば理想画像とほとんど変わらない認識率を得ることができている。

画像相関による認識 ( $corr$ ) とは、単純投影による認識結果の上位 10 位に入った軌跡上の点から近似画像を作成し、作成した近似画像と入力画像の間で相関係数を計算し相関の高いものを認識結果とする認識方法である。近似画像  $\overline{f^{(j,k)}}$  は式 (5) によって計算される。 $c_i$  は軌跡上の点の  $i$  次元座標、 $\phi_i^{(k)}$  はカテゴリ  $k$  の第  $i$  番目の固有ベクトル、 $m^{(k)}$  はカテゴリ  $k$  の平均ベクトルを表している。

$$\overline{f^{(j,k)}} = \sum_{i=1}^{64} c_i \phi_i^{(k)} + m^{(k)} \quad (5)$$

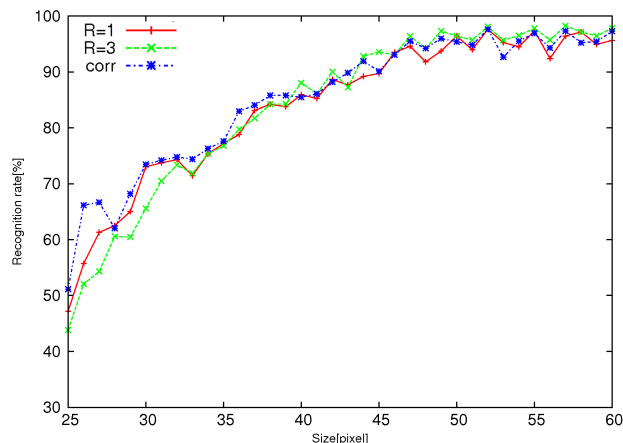


図 8 カメラ入力文字サイズに対する認識率  
Fig. 8 Recognition rate vs. character size

次に、撮影された  $45 \times 45$  以上  $55 \times 55$  [pixel] のサイズの文字画像を用いて、認識に使用する次元に対する認識率の変化を調べた。図 9 に上位  $n$  次元に対する認識率の変化を示す。低次元において画像相関による認識 ( $corr$ ) が最も認識率が高く、さらに単純投影よりも複数投影による認識の方が高い認識率を得ることができた。

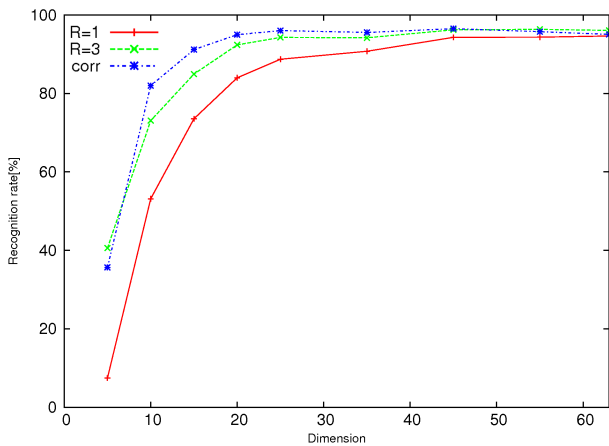


図 9 カメラ入力文字画像の次元に対する認識率  
Fig. 9 Recognition rate vs. dimension

## 5. 文書画像の角度推定

本研究の手法では認識結果として、カテゴリ、フォント、回転角度を同時に得ることができる。そのため、傾いた文書画像の角度を、文書に含まれる文字を認識した結果として得られる角度情報を利用して、統計的に推定することができる。文書画像の角度推定のプロセスは、はじめに文書画像を二値化し連結成分を抽出する。つぎに、抽出した連結成分の認識を行い、推定角度で投票を行うことによって文書画像の角度を推定する。この方法で用いる連結成分は文字の一部である可能性があるが、モフォロジー処理をすることにより一つの連結成分が一文字を構成する可能性を増やしている。カメラ画像に対する認識結果より、 $40 \times 40$ [pixel]以下の文字画像では誤認識される可能性が高いため、連結成分のサイズが $40 \times 40$ [pixel]以下の場合には文書画像の角度を推定に使用しない。

テスト文書の例を図 10 に示す。図 10 の画像サイズは  $3096 \times 4128$  画素であり、1文字のサイズはおおよそ  $55 \times 55$  画素程度である。図 10 の文書画像は縦書きと横書きが混在し、さらに写真など文字以外の要素が含まれている。図 11 に図 10 に対する角度投票結果を示す。最も多く投票された角度は明らかなピークを示し、この例では 267 度という結果になった。例に示した文書画像の他に 10 枚の文書画像についても実験を行い正しく角度を推定することができた。この方法により、レイアウト解析をしなくても、またそれが困難な場合でも文書画像の角度を推定することが示された。



図 10 角度推定文書画像の例  
Fig. 10 A test image

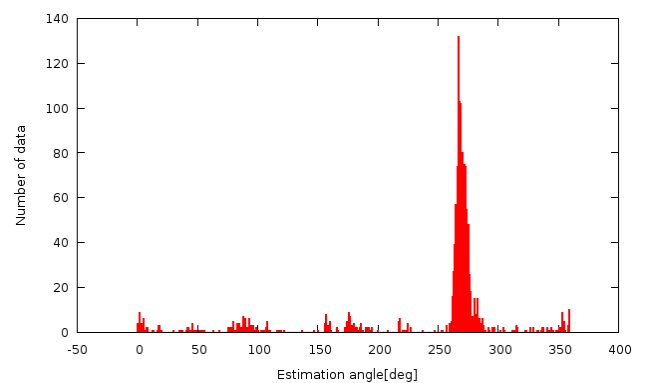


図 11 図 10 に対する角度推定結果  
Fig. 11 Angle estimation of a test image in Fig.10

## 6. 認識時間と高速化

現在のシステム構成では、3133文字種各2フォントデータ群に対し、1文字の  $8 \times 8$ [pixel]65値の画像の認識を行うために0.6[秒]程度かかっている。今後カメラ入力の実時間文字認識システムを目指す上では高速化しなければならない。現在は認識結果を得る際に、軌跡上のすべての点について64次元を用いて距離値を算出しているが、明らかに違う文字や、明らかに違う角度の場合、低い次元の段階でかなり大きな距離値となり64次元すべてを用いなくても除外することができると考えられるため、今回は閾値を設け距離値による計算の打ち切りを行うことによって認識の高速化をした。

設定する閾値の決定方法は理想画像に対する認識実験と同じように7度から357度まで7度ごとに回転させた画像を同一カテゴリの固有部分空間に投影し、同一フォントの

軌跡上の点との最短距離を計算し、その平均を元にして閾値を決定した。調べた結果、平均認識距離は 772.25 という結果となった。そして今回の実験では、閾値を 1000 から 6000 まで 500 ごとに変化させて 3 章で行ったのと同じテストサンプルを用いて最大 64 次元を使用して認識実験を行い認識率と認識時間の変化を調べた。計算を打ち切る閾値に対する認識率の変化を図 12 に、計算を打ち切る閾値に対する計算時間の変化を図 13 に示す。閾値が 2000 以下の場合には閾値を設けない場合に比べて大きく認識率が低下してしまうが、打ち切る閾値が 2000 以上の場合では、99% を超える認識率を得ることができた。さらに、閾値が 2000 から 6000 の間ではほとんど認識率に変化はなかった。計算時間については、閾値の値を 2000 で設定した際に一文字の認識にかかる時間が 0.12[秒] 程度と閾値を設定しない場合の約 5 分の 1 程度の時間で認識を行うことができた。閾値を 2000 に設定した際の次元に対する探索軌跡数のヒストグラムを図 14 に示す。低次元で探索軌跡数をかなり削減できていることがわかる。ここで、探索軌跡数とは、全 6266 軌跡 (カテゴリあたり 2 フォントの軌跡) から、残存している軌跡数のことである。

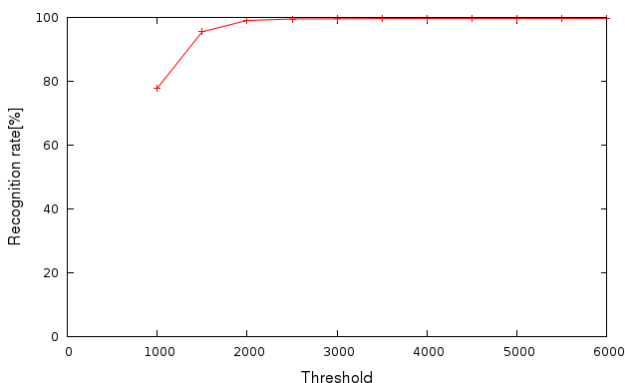


図 12 閾値に対する認識率の変化  
 Fig. 12 Recognition rate vs. Threshold

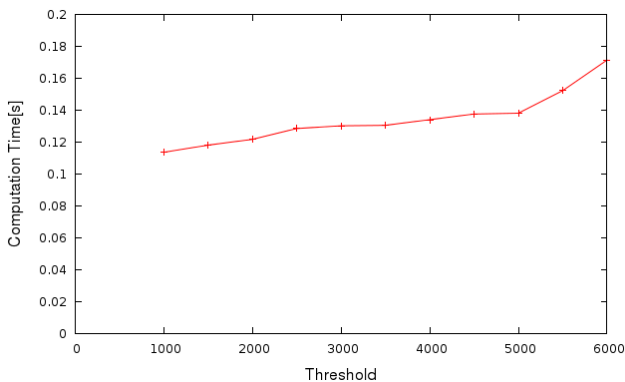


図 13 閾値に対する計算時間の変化  
 Fig. 13 Computation time vs. Threshold

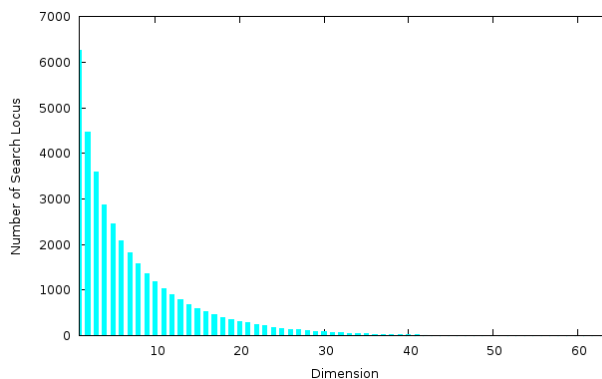


図 14 閾値 2000 の場合の次元に対する探索軌跡数

Fig. 14 Number of search locus vs. Dimension of Threshold with 2000

## 7. まとめと今後の課題

本報告では、パラメトリック固有空間法を用いた回転文字認識法をゴシック体と明朝体の 2 フォントの JIS 第一水準漢字、ひらがな、カタカナ全 3,133 カテゴリを対象に認識実験を行い実験結果をまとめた。

はじめに、自動生成した理想画像を用い二次元的な回転のみをテストサンプルに与え認識実験を行い 99.84% という高い認識率をえることができた。さらに複数投影による認識を行うことにより低い次元において単純投影による認識で他カテゴリの軌跡と偶然近くなる偶然性を抑え認識率を向上させることができることを示した。角度推定精度も推定角度が真の角度との差がほとんど  $\pm 1$  度以内と高い精度で推定できている。

つぎに、カメラで撮影された文字画像には二次元的な回転のみでなく、三次元的な回転も含まれていると考えられるので、二次元的な回転のみを考慮して作成した辞書構成においての三次元的な回転を含んだ文字画像に対する認識性能を確認した。その結果、X 軸まわりに  $\pm 3$  度、Y 軸まわりにも  $\pm 3$  度以内の変化であれば、99% 以上の認識率を得ることができ、X 軸まわりに  $\pm 6$  度、Y 軸まわりにも  $\pm 6$  度以内の変化では、90% 以上の認識率となった。よってカメラで撮影する際に認識させたい文字をなるべく正面から撮影するように注意をすれば、十分に対応できると考えられる。

さらに、実際にカメラで撮影された文字画像に対する認識実験を行い、取得された文字画像のサイズに対する認識性能を調べ、サイズに対する性能の結果によって  $45 \times 45$  [pixel] 以上の文字画像であれば理想画像による認識率とあまり変わらない結果を得ることができたため、そのサイズの文字画像を用いてカメラで撮影された文字画像について認識に使用する次元に対する認識率の変化を調べる実験を行った。その結果、低次元において画像相関による認識 (*corr*) が最も認識率が高く、さらに単純投影よりも複数投影によ

る認識の方が高い認識率を得ることができた。

本手法の応用として、角度が得られることを利用して、二値化した文書画像中から連結成分を抽出し、それらの角度を統計的に集計することにより文書画像の角度推定ができることを示した。

最後に、認識の際に計算する距離値を、閾値によって打ち切ることによって計算時間を短縮した。閾値を設定しない場合は一文字の認識に0.6[秒]程度かかっていたが、閾値を2000に設定し計算の打ち切りを行った場合、認識率は99%をこえる認識率を保ったまま認識にかかる時間を約5分の1の0.12[秒]程度に短縮することができた。

今後の課題としては、より多くのフォントに対応することや実時間文字認識システムを構築するためにさらに高速化を行う必要がある。さらなる回転文字認識の性能向上について考えていきたい。

## 参考文献

- [1] 解秋生, 小林彬, “平行移動・スケール変換・回転変換に不変なパターン認識系の構成法”, 計測自動制御学会論文集, 27.10, 1167-1174, 1991.
- [2] S.Sato, S.Miyake, H.Aso, ”Evaluation of Two Neocognitron-type Models for Recognition of Rotated Patterns”, ICONIP 2000, WBP-04, pp295-299, (2000).
- [3] 村瀬 洋, シュリー K. ナイヤー, “2次元照合による3次元物体認識-パラメトリック固有空間法-”, 電子情報通信学会論文誌, J77-D-II, No.11, pp.2179-2187, 1994.
- [4] 篠川敏行, 長谷博行, 角谷浩, 米田政明, “パラメトリック固有空間法による回転文字の認識”, 画像電子学会誌, vol.33, no.6, pp.1123-1131,(2004.11).
- [5] 成田了, 大山航, 若林哲史, 木村文隆, “3次元回転不変文字認識”, MIRU2011, OSI-4: 26-32, (2011).
- [6] <http://projects.itri.aist.go.jp/etlcdb/etln/etl2/etl2.htm>
- [7] Shukun NING, Hiroyuki HASE, Shogo TOKAI, ”Kanji Character Image Generation for Character Recognition”, 2012 Joint Conference of Hokuriku Chapters of Electrical Societies, F-56, 2012.
- [8] 山本将史, 佐藤詩織, 長谷博行, 東海彰吾, “固有部分空間を用いたリアルタイム回転文字認識”, 電子情報通信学会パターン認識・メディア理解研究会, PRMU2010-246, 55-60, 2011-03.
- [9] Yuta Baba, Hiroyuki Hase, Shogo Tokai, ”Rotated Character Recognition and its Properties”, ITE Trans. on MTA, vol.3, no.1, pp.22-29, (2015).