

研究速報

傾斜文字認識のための正規化方法

志久 修[†] (正員) 手島 裕詞[†] (正員)内田 誠一^{††} (正員)

Character Normalization Method for Recognizing Slanted Characters

Osamu SHIKU[†], Yuji TESHIMA[†],
and Seiichi UCHIDA^{††}, Members[†] 佐世保工業高等専門学校, 佐世保市National Institute of Technology, Sasebo College, 1-1
Okishin, Sasebo-shi, 857-1193 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, 744 Motooka, Nishi-ku, Fukuoka-shi,
819-0395 Japan

DOI:10.14923/transinfj.2017JDL8007

あらまし 傾斜した単文字を認識するための正規化方法を提案する。提案法では、文字に外接する平行四辺形を正方形にすることで、文字の傾斜を正規化する。提案法を前処理とする文字認識システムは、傾斜した文字をそのまま学習させた文字認識システムよりも、傾斜した文字を高精度に認識できることを示す。

キーワード 傾斜文字, 射影ひずみ, 正規化, 文字認識

1. まえがき

カメラを用いた文字認識では、文字を斜めから撮影するとき生じる射影ひずみへの対処が課題の一つとなっている [1]。射影ひずみを受けた文字列には、図 1(a) に示すように、文字列の回転、文字の傾斜、文字サイズの変動が生じる。もし文字列抽出がある程度適切に行われ、その結果として図 1(b) に示すように文字列が水平になるように回転補正されたとしても、依然として文字列には水平方向に傾斜 (せん断) が残ってしまう。この傾斜は、図 1(b) のように文字列全体で一様ではなく、文字の位置ごとに異なっている。ただし、一文字程度の幅で見ると、傾斜はほぼ一様と近似できるものとして、本論文では一様に傾斜している

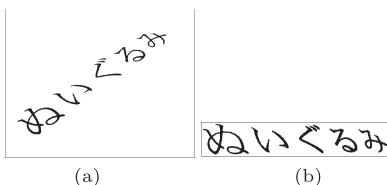


図 1 射影ひずみと回転補正

Fig. 1 Perspective distortion and rotation correction.

単文字を対象とする。

傾斜した文字を認識するために、文字列全体の傾斜を補正することで個々の文字を正立させる方法 [2]~[5] がある。これらの方法では文字の縦ストロークを極力垂直にすることが前提であるため、縦ストロークがない場合に文字を正立できない場合がある。したがって、正立できなかった文字は、傾斜したままの形状で認識する必要がある。変形した単文字の認識法として、不変量を用いる方法 [6], [7], 部分空間法と高速フーリエ変換を用いる方法 [8], 変形文字を全て学習する方法 [9], [10] などがある。方法 [6]~[8] では変形した英数字等を高精度に認識できるが、日本語のように字形が複雑で、かつ字種が多い文字に対する性能は明らかにされていない。方法 [9], [10] では変形文字の学習により特徴ベクトルの分布が広がり、これらの分布を表現するために固有ベクトルの数が増え、照合時間及び辞書容量が大きくなることが考えられる。

本論文では、一様角度で傾斜した単文字を認識するための正規化方法を提案する。提案法では、傾斜した文字を傾斜角にかかわらず同じような形状に整える。提案法を前処理とする文字認識システムを構築し、傾斜した文字の認識実験を行う。実験により、提案法を前処理とする文字認識システムは、傾斜した文字を全て学習 [9], [10] させた文字認識システムよりも、高い認識率が得られることを示す。なお本実験では、文字認識システムの識別器として部分空間法及び畳み込みニューラルネットワーク (CNN と呼ぶ) の 2 種類を採用する。

2. 傾斜正規化法

2.1 処理の概要

単文字の傾斜ひずみを正規化する方法を提案する。提案法では、図 2 に示すように、単文字画像 (図 (a)) から文字に外接する平行四辺形 (図 (b)) を求め、その平行四辺形が正方形になるように座標変換を行う (図 (c))。対象とする単文字は、図 1 のように回転補正された文字列から正確に切り出されたもので、傾斜角は一様とする。



図 2 正規化処理の例. (a) 文字画像 (b) 外接平行四辺形 (c) 正規化画像

Fig. 2 Normalization process.

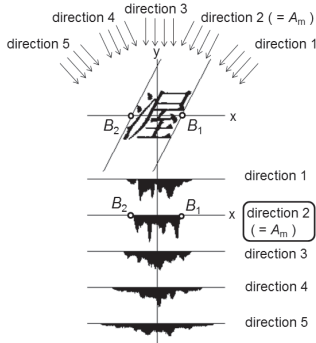


図 3 方向ごとのヒストグラム
Fig. 3 Histograms by directions.

2.2 外接平行四辺形の抽出

提案法では文字の傾斜角にかかわらず適切な外接平行四辺形を抽出する必要がある。そこで、平行四辺形を、単文字画像の投影ヒストグラムから求める。図 3 に、様々な方向からの x 軸に対する投影ヒストグラムを示す。ここで x 軸は水平軸（文字列の方向）である。基本的アイデアは、以下のとおりである。まず各方向の投影ヒストグラムにおいて、投影値の 2 乗の合計（山の面積と呼ぶ）と投影値が 0 より大きい要素の総数（山の幅と呼ぶ）を求める。その上で、面積と幅の比が最大となる投影方向を文字の傾斜角とする。この傾斜角と山の左右の端点から平行線を求める。文字の上下に接する平行線は、同様に y 軸への投影ヒストグラムから求める。

具体的な平行線の抽出手順を以下に述べる。本方法では投影ヒストグラムをハフ変換 [11] により求める。

(1) 左右の平行線の抽出：単文字画像の中心点を原点とし、全ての図形画素の座標 (x_i, y_i) ; (i は図形画素の番号) を、式 $B = -y_i A + x_i$ により $A - B$ 平面に投票する。ここで A と B は、単文字画像上の直線の y 軸に対する傾きと x 切片である。図 4 に、図 3 の文字画像に対する $A - B$ 平面を示す。 $A - B$ 平面において任意の傾き A をもつ 1 ラインが、方向 A から x 軸に対する投影ヒストグラムになる（図 4 の濃淡が、投影ヒストグラムの投影値を表している）。そこで、傾き A を変え 1 ラインごとに山の面積と幅の比を求めていく。そして、山の面積と幅の比が最大となる傾き A_m を平行線の傾き、 A_m での山の両端点 B_1, B_2 を平行線の x 切片とする。図 4 に $A - B$ 平面から求めた傾き A_m 、 x 切片 B_1, B_2 の 2 点を示す。これらの 2 点が図 3 の平行線に対応する。

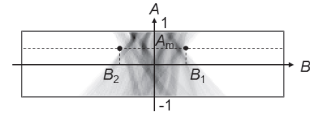


図 4 ハフ平面 ($A - B$ 平面)
Fig. 4 Hough plane ($A - B$ plane).

(2) 上下の平行線の抽出：左右の平行線と同様にして求める。すなわち図形画素 (x_i, y_i) を、 x 軸に対する直線の傾きを a 、 y 切片を b とする式 $b = -x_i a + y_i$ により $a - b$ 平面に投票する。そして $a - b$ 平面における投票値の山から平行線の傾きと y 切片を求める。

傾き A 及び a の量子化数は、 A は $-1.0 \sim 1.0$ まで 0.05 刻み、 a は $-0.02 \sim 0.02$ まで 0.01 刻みと設定する。これは図 1 のように文字列を行方向に回転補正した場合、切り出された単文字は、文字の左右方向の傾斜は大きく、上下方向の傾斜は本来は 0 になるべきだが多少の誤差が生じる、とそれぞれ仮定して設定している。

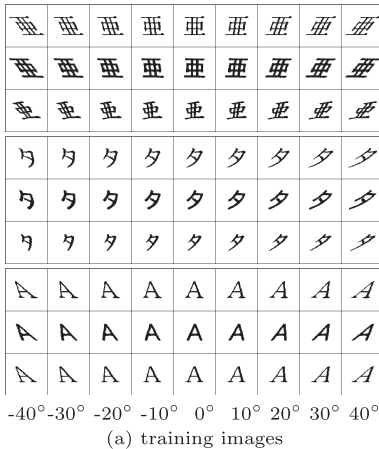
3. 実 験

提案法を前処理とする文字認識システム（提案システムと呼ぶ）を構築し、傾斜した文字の認識実験を行う。比較のため、文字の大きさだけを正規化する文字認識システム（従来システムと呼ぶ）を用いる。文字認識システムの識別器には部分空間法と CNN の 2 種類を採用する。識別器の学習と評価に用いる傾斜した文字画像は、data augmentation により、正立した単文字画像を水平方向に傾斜させて生成する。すなわち提案システムでは傾斜を正規化した文字画像を識別器に学習させ、傾斜を正規化した文字画像を認識対象とする。一方、従来システムでは傾斜したままの文字画像を識別器に学習させ、傾斜したままの文字画像を認識対象とする。

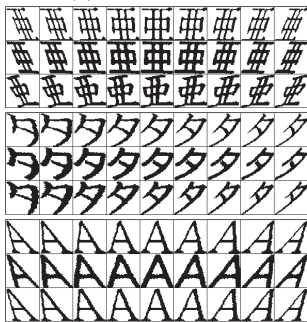
3.1 実験データ

実験では、3,169 字種（漢字、かな、カナ、英数字）を対象とする。識別器の学習のため、1 字種当り 270 個の文字画像を用意する。これらはスキャナ（400dpi, 2 値）で入力した 1 字種当り 30 個の文字画像に、 y 軸に対し -40° から 40° まで 10° 刻みで 9 種類のせん断変形を加えて作成する。30 個の内訳は、3 フォント（明朝体、ゴシック体、楷書体） $\times 5$ サイズ（6,8,10,12,14 ポイント） $\times 2$ 種類のレーザープリンタである。

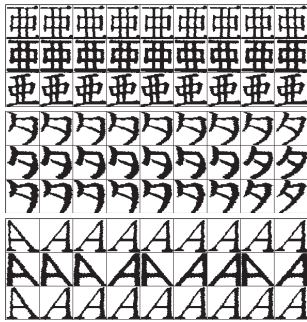
学習用の文字画像を正規化し、識別器の学習に用いる。正規化サイズは 64×64 画素とする。図 5 に学習



(a) training images



(b) normalized (conventional system ; 64 × 64)



(c) normalized (proposed system ; 64 × 64)

図5 学習用の文字画像と正規化画像
Fig.5 Training images and normalized images.

用の文字画像 (図(a)) と正規化画像 (図(b), (c)) を示す. 従来システムでは大きさの正規化だけなので, 正規化後の文字にせん断変形が残っていることが分かる. 一方, 提案システムでは正規化後の文字が本来の形状と異なる場合があるが, せん断角度によらずほぼ同じような形状あるいは少数の形状に揃っていることが分かる.

評価用の画像として, 1 字種当り 27 個の画像を用意する. これらは学習用と同じ手順で作成するが, 学習

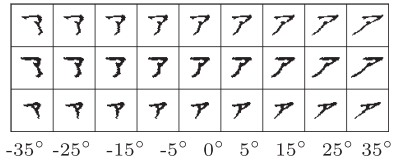


図6 評価用の文字画像
Fig.6 Test images.

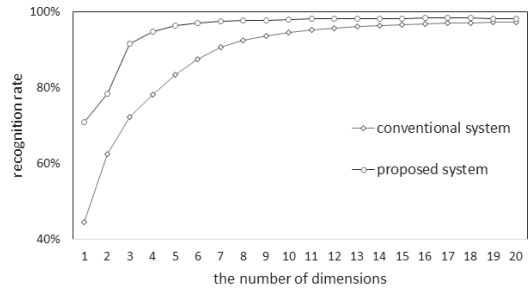


図7 部分空間の次元数と認識率の関係
Fig.7 The relation between the number of dimensions and recognition rate.

用とは別に印刷した 6 ポイントの 3 フォント文字を複写機で品質を落とし, 更にせん断角度を正立 (せん断角度 0°) 以外は異なる角度 (-35° から 35° まで 10° 刻み) にしている. 図6に評価用の文字画像を示す.

3.2 識別器

識別器として, 方向線素特徴量と部分空間法を組み合わせた方法 [12] 及び CNN を採用する. 部分空間法の辞書は, 正規化した学習用の文字画像から方向線素特徴量を抽出し作成する. CNN は正規化した学習用の文字画像をそのまま学習させる. CNN の構成は, 畳み込み層を C, プーリング層を P, 全連結層を F と表すと, 入力側 (64 × 64 画素) から, C(64 × 64 × 20)-P(60 × 60 × 20)- C(30 × 30 × 40)-P(26 × 26 × 40)-F(13 × 13 × 40)-F(500)-F(3169) である. 活性化関数は ReLU, 学習は誤差逆伝搬法で行う. 学習の反復回数は 50 回とする.

3.3 実験結果

(1) 部分空間法における次元数と認識率

部分空間法を識別器としたときの部分空間の次元数 (固有ベクトル数) と評価用の文字画像の認識率の関係を調べた. 図7に結果を示す. 提案システムでは, 従来システムに比べ, 少ない次元数で高い認識率が得られていることが分かる. この理由は, 傾斜を正規化した画像 (例えば図5(c)) から抽出された特徴ベクトルの分布は, 傾斜した文字画像 (例えば図5(b)) から

抽出された特徴ベクトルの分布よりも小さくなり、少ない次元数で分布を表現できるためと考えられる。部分空間法では照合時間及び辞書容量が次元数に比例するため、少ない次元数で高い認識率が得られる提案システムが有利であるといえる。

(2) 角度ごとの認識率

部分空間法と CNN をそれぞれ識別器としたときの評価用の文字画像の角度ごとの認識率を調べた。部分空間法の次元数は、図 7 より認識率がそれぞれほぼ飽和する提案システムは 8、従来システムは 15 とした。また従来システムでは正立した文字だけを識別器に学習させた場合も調べた。表 1 に角度ごとの認識率（第 3 位累積認識率）を示す。表より、両識別器において、従来システムでは正立文字だけではなく傾斜した文字も学習させることで、傾斜した文字の認識率が向上していることが分かる。同じく両識別器において、提案システムでは角度によらずほぼ一定の認識率になっているのに対し、傾斜文字を学習させた従来システムでは部分空間法においては傾斜が大きくなると認識率が低下し、CNN においては全ての角度で認識率が低くなっていることが分かる。これより提案法を前処理とする文字認識システムが、傾斜した文字の認識に有効であることが分かる。

(3) 提案システムでの誤認識例

図 8 に、傾斜文字を学習させた従来システムでは正しく認識でき、提案システムでは誤認識した例を示す。図 8(a) は誤認識の例、図 8(b) はその字種の学習画像の例をそれぞれ示す。図 8(a), (b) とともに文字画像（外接平行四辺形）及び正規化画像を示す。図 8(b) の正規化画像は同じような形状であるのに対し、図 8(a) の正規化画像は図 8(b) と異なる形状になっていることが分かる。すなわち、図 8(a) のような形状は学習しておらず、更に図 8(a) に類似する字種「久」が存在したために、誤認識したと考えられる。なお、異なる形状に正規化された理由として、「火」のように同程度の面積（黒画素数）で構成された異なる方向のストロークをもつ文字では、つぶれ等によりストロークの黒画素数が増減することで、提案法で採用した文字の傾斜角（投影ヒストグラムの面積と幅の比が最大になる方向）が変わってしまったためと考えられる。ちなみに、図 5(c) の「A」は同程度の 2 方向のストロークをもつため、字形によって二つの形状の正規化画像となるが、両方の形状を学習していることと、どちらの形状にも類似する文字がないことから、「A」は全て正

表 1 角度ごとの認識率

Table 1 The relation between shear angles and recognition rate.

角度	部分空間法 (次元数)			CNN		
	従来 (15)		提案 (8)	従来		提案
	正立	傾斜		正立	傾斜	
-35°	1.45 (3.60)	94.48 (98.66)	97.55 (99.53)	2.64 (5.40)	93.40 (98.22)	96.08 (98.97)
-25°	11.50 (21.59)	96.30 (99.36)	97.43 (99.51)	10.27 (18.64)	93.99 (98.39)	95.88 (99.04)
-15°	74.17 (88.99)	97.21 (99.57)	97.75 (99.61)	47.71 (64.15)	93.75 (98.40)	96.10 (99.18)
-5°	98.31 (99.89)	97.84 (99.71)	97.78 (99.59)	91.00 (96.68)	93.80 (98.47)	96.12 (99.11)
0°	99.02 (99.94)	98.17 (99.77)	98.00 (99.63)	94.27 (98.43)	94.30 (98.67)	96.68 (99.27)
5°	98.62 (99.93)	97.88 (99.68)	97.92 (99.54)	91.85 (97.19)	93.97 (98.44)	96.12 (99.17)
15°	74.54 (88.53)	97.45 (99.55)	97.91 (99.56)	45.41 (62.43)	93.80 (98.49)	96.22 (99.13)
25°	11.24 (21.14)	96.56 (99.33)	97.77 (99.60)	8.66 (16.08)	93.80 (98.30)	95.67 (98.98)
35°	1.17 (2.99)	94.23 (98.39)	97.77 (99.67)	2.60 (5.08)	92.81 (97.94)	96.14 (99.04)



(a) 誤認識 (上: 文字画像, 下: 正規化画像)



(b) 学習画像 (上: 文字画像, 下: 正規化画像)

図 8 誤認識例（「火」→「久」）と学習画像

Fig. 8 Examples of mis-recognition and training images.

しく認識できていた。

その他の誤認識例として、同形文字 (c, C 等)、類似文字 (ば, ぱ等) 及び文字のつぶれがあった。同形文字と類似文字については、表 1 の 3 位累積認識率がほぼ 99% 以上であるため、上位候補としては認識できていることが分かる。

4. む す び

本論文では、傾斜した文字を認識するための正規化法を提案した。提案法を前処理とした文字認識システムは、傾斜した文字をそのまま学習させた文字認識システムよりも、傾斜した文字を高精度に認識できることを示した。

提案法は単文字ごとに正規化を行ったが、一般的に一つの文字列内では個々の文字の傾斜角は急激に変わ

らないので、この制約条件を設けることにより正規化の精度向上が考えられる。今後の課題は、前述した文字列の制約について検討し、更に実際にカメラで撮影した射影ひずみや湾曲などの変形をしている文字列の文字認識に、提案法を適用 [13], [14] することである。

謝辞 本研究は、科学研究費補助金（課題番号：15K00258）によって行われたものです。

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] G.K. Myers, R.C. Bolles, Q.T. Luong, J.A. Herson, and H.B. Aradhye, “Rectification and recognition of text in 3-D scenes,” *Int. J. Document Analysis and Recognition*, no.7, pp.147–158, July 2005.
- [3] 萩原敏浩, 渡部広一, 河岡 司, “ひずみを有する活字文字列における形状補正手法,” 信学論 (D-II), vol.J86-D-II, no.2, pp..262-271, Feb. 2003.
- [4] 林 信吾, 志久 修, 玉井敬一, 福本博文, 黒田英夫, “活字文字列に生じる傾斜ひずみの補正方法,” 信学論 (D), vol.J90-D, no.2, pp.600–603, Feb. 2007.
- [5] C. Merino-Gracia, M. Mirmehdi, J. Sigut, and J.L. Gonzalez-Mora, “Fast perspective recovery of text in natural scenes,” *Image Vis. Comput.*, vol.31, no.10, pp.714–724, Oct. 2013.
- [6] L. Li and C.L. Tan, “Recognizing planar symbols with severe perspective deformation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.32, no.4, pp.755–762, April 2010.
- [7] 堀松 晃, 岩村雅一, 黄瀬浩一, “特徴点の一意決定性を利用した高速なアフィン不変文字認識,” 信学論 (D), vol.J98-D, no.4, pp.651–660, April 2015.
- [8] 張 放, 菅谷至寛, 大町真一郎, 阿曾弘具, “部分空間法と高速フーリエ変換を用いた傾斜文字認識,” 画電学誌, vo.39, no.5, pp.598–605, May 2010.
- [9] 成田 了, 大山 航, 若林哲史, 木村文隆, “3次元回転不変カメラベース文字認識,” 電学論 (C), vol.133, no.4, pp.876–882, April 2013.
- [10] 草地良規, 鈴木 章, 伊藤直己, 荒川賢一, 安野貴之, “景観画像中の文字候補群によるインデクシング及び検索技術,” 信学論 (D), vol.J90-D, no.9, pp.2562–2572, Sept. 2007.
- [11] 松山隆司, 輿水大和, “Hough 変換とパターンマッチング,” 情報処理, vol.30, no.9, pp.1035–1046, Sept. 1989.
- [12] 孫 寧, 安倍正人, 根元義章, “改良型方向線素特徴量および部分空間法を用いた高精度な手書き文字認識システム,” 信学論 (D-II), vol.J78-D-II, no.6, pp.922–930, June 1995.
- [13] 志久 修, 林 信吾, 手島裕詞, “2段階の正規化を用いた射影歪みを受けている文字列の認識,” 画像関係学会連合会第3回秋季大会, PC10, Nov. 2016.
- [14] 志久 修, 手島裕詞, “カメラ入力文字認識のための文字正規化方法,” 平成 28 年度電気学会九州支部沖縄支所講演会, OKI-2016-38, Dec. 2016.
(平成 29 年 2 月 27 日受付, 6 月 2 日再受付,
6 月 16 日早期公開)