

アフィン不変な文字認識手法とその高速化

堀松 晃[†] 丹羽 亮[†] 岩村 雅一[†] 黄瀬 浩一[†] 内田 誠一^{††}

大町 真一郎^{†††}

[†] 大阪府立大学大学院工学研究科 〒 599-8531 堺市中区学園町 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

^{††} 九州大学大学院システム情報科学研究所 〒 819-0395 福岡市西区元岡 744

E-mail: [†]{horimatsu,niwa}@m.cs.osakafu-u.ac.jp, [†]{masa,kise}@cs.osakafu-u.ac.jp,

^{††}uchida@is.kyushu-u.ac.jp, ^{†††}machi@ecei.tohoku.ac.jp

あらまし デジタルカメラによる文字認識手法には課題が多く存在する．その課題の一つとして，幾何学変換により文字が歪むことが挙げられる．この課題に対し，歪みを正規化して認識する手法が存在する．しかし，この手法では文字の回転方向が推定できないため，方向を全て試すなどの処理が必要となり処理時間が大きくなる．そこで，本稿では，アフィン不変な特徴量を用いて認識対象の字種を絞り込み，角度の推定を多段階で行うことで高速な認識を行うことを提案する．提案手法は，アフィン変換に対する頑健性を備え，認識時間を短縮できるという特徴を持つ．シミュレーションと実際に撮影した文字に対して実験を行った結果，歪んだ文字に対し高速に認識することを確認する．
キーワード 近似楕円，正規化，特徴ヒストグラム，段階的な枝刈り

Affine Invariant Character Recognition and Its Speeding Up

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

[†] Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuencho, Naka, Sakai, 599-8531 Japan

^{†††} Graduate School of Engineering, Tohoku University, 6-6-05 Aoba, Aramaki, Aoba, Sendai, 980-8579
Japan

^{††} Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motoooka, Nishi, Fukuoka, 819-0395 Japan

E-mail: [†]{horimatsu,niwa}@m.cs.osakafu-u.ac.jp, [†]{masa,kise}@cs.osakafu-u.ac.jp,

^{††}uchida@is.kyushu-u.ac.jp, ^{†††}machi@ecei.tohoku.ac.jp

Abstract There are many problems to realize camera-based character recognition. One of the problems is that characters in scenes are often distorted by geometric transformation. Some methods have been proposed to normalize the distortion. However, they cannot estimate the rotation angle of a character except examining all the possible angles. This consumes much time. Thus, in this paper, we propose a method to estimate the angle by pruning unpromising categories. It is realized by combining multiple classifiers based on affine invariant features. The proposed method is robust to affine transformation and can reduce computational cost. We confirm that the effectiveness of the proposed method by experiments.

Key words approximate ellipse, normalization, feature histogram, progressive pruning

1. ま え が き

近年，デジタルカメラや携帯電話などを用いたカメラベースの文字認識技術の需要が高まっている．これらの機器は数百

万～一千万画素の解像度を持っており，もはや「汎用画像取得機器」と呼んで差し支えない程の性能を持つ．可搬性に優れたこれらの機器を用いて情景中の文字が認識できれば，いつでも必要なときに利用することができるため，これを活かした様々

な応用が考えられる。例えば、外出先で店舗の看板や標識などを撮影することにより、関連情報を引き出すサービスであったり、外国語で書かれた文字をカメラで撮影し、他の言語に翻訳するといった翻訳システムが考えられる。また、もしカメラに写るあらゆる文字を認識できれば、その中から特に注意が必要な情報を選び出し、利用者に提示するシステムも考えられる。

これらを実現するためには、認識の頑健性と高速性を兼ね備えたカメラベースの文字認識技術が必要であり、様々な課題を解決しなければならない[1]~[3]。本稿ではこのうち、文字を斜めから撮影した場合に生じる射影変換などの幾何学的変換に焦点を当てる。

この問題に対処する方法は大きく二つの種類に分けることができる。1つ目の方法は、予め歪んだ文字画像を用意しておく、認識時に様々な画像と照合する方法である[4]。この方法では、様々な劣化を考慮するほど認識精度が向上すると考えられるが、認識精度を重視すれば処理に時間がかかってしまうというトレードオフの関係がある。そのため、いかに認識精度と処理時間を両立させるかが課題である。2つ目は、何らかの方法で画像の歪みを検出し、それを復元(正規化)する方法である。Leuら[5]と大田ら[6]はそれぞれ、図形を楕円で近似し、楕円を円に戻す変換を施すことにより、アフィン変換を受けて歪んだ図形を正規化する方法を提案している。この方法は図形の独立変換を補正できるが、図形の回転を正規化することができない。そのため、回転角度は別の方法で合わせる必要がある。単純に画像を回転しながら全ての角度の類似度を計算していたのでは処理時間が膨大となるため、短い時間で画像の回転角度を合わせられる手法が求められる。この問題に対しては文献[7]においてモーメントを用いた方法が試みられているが、求まる角度が変形に依存してしまうため、成功しているとは言い難い。

本稿では、前述の画像の正規化手法を用いた後、回転に不変な特徴を用いて字種の大分類を行い、回転角度の推定を高速に行う手法を提案する。これにより、アフィン変換を受けた文字画像を高速に認識することが可能になる。提案手法の特徴は、アフィン不変な特徴を用いた字種の絞り込みと、角度推定の多段階化である。提案手法で用いているアフィン不変な特徴は、アフィン不変な文字の大分類に有効でもあるため、1番目に述べた様々な劣化を考慮したマッチングにも有効である。

2. アフィン変換の正規化による基本的な認識方法

本節では、アフィン変換を受けた文字を正規化によって認識する基本的な方法について述べる。この方法は後述する実験において従来手法として参照する。

2.1 アフィン変換による歪みの正規化

アフィン変換を受けた図形の歪みを正規化する手法が提案されている[5],[6]。この手法の特徴は、図1のように図形を近似する楕円を求め、これを円に変形することで歪みを正規化することである。図形を近似する楕円とは、その図形と同じ共分散行列を持つ楕円である。以後、この楕円を近似楕円と呼ぶこととする。この近似楕円には、「ある図形をアフィン変換した図形

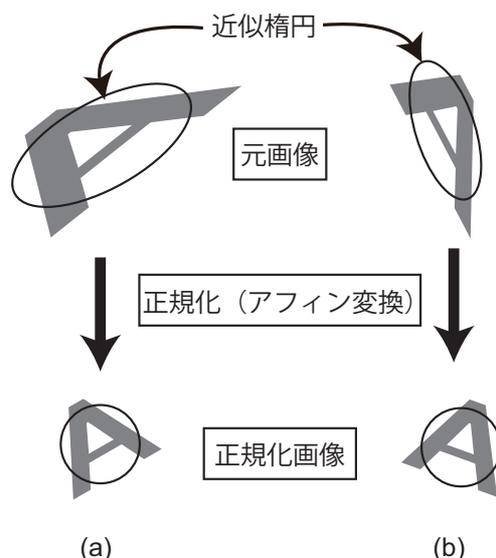


図1 アフィン変換による歪みの正規化

の近似楕円は、元の近似楕円をアフィン変換した楕円に等しい」という性質がある。楕円を円にする変換はアフィン変換であるため、図形の近似楕円を円にするアフィン変換を図形に対して施すことで、アフィン変換を受けた図形の歪みを正規化することが可能となる。以下では、図1に示すアフィン変換による歪みの正規化について述べる。

近似楕円を作成する方法を説明する。近似楕円は、その図形と同じ共分散行列を持つ楕円であるので、図形の共分散行列 C を求める必要がある。 C は重心まわりのモーメント $m_{00}, m_{02}, m_{11}, m_{20}$ を用いて次式のように表わすことができる。

$$C = \begin{bmatrix} m_{20}/m_{00} & m_{11}/m_{00} \\ m_{11}/m_{00} & m_{02}/m_{00} \end{bmatrix} \quad (1)$$

ただし、重心まわりのモーメントは文字の重心を原点とした場合、以下で定義される。

$$m_{ij} = \iint x^i y^j f(x, y) dx dy \quad (2)$$

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \text{文字領域 (c.f. 黒画素)} \\ 0 & \text{if } (x, y) \in \text{非文字領域 (c.f. 白画素)} \end{cases}$$

次に、共分散行列 C の逆行列 C^{-1} を用いて図形を正規化する。元の図形の x, y 座標を (x', y') 、アフィン変換後の図形の x, y 座標を (x'', y'') とおくと、図形の重心が $\begin{bmatrix} m_{10} \\ m_{01} \end{bmatrix}$ であり、重心は回転によって変化しないので、

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = C^{-1} \begin{bmatrix} x' - m_{10} \\ y' - m_{01} \end{bmatrix} + \begin{bmatrix} m_{10} \\ m_{01} \end{bmatrix} \quad (3)$$

となる。

なお、この正規化によって得られる画像の形状が、我々が普段見慣れているような文字と同じとは限らないことに注意が必要である。例えば「0」(ゼロ)や「O」(オー)のような楕円形の

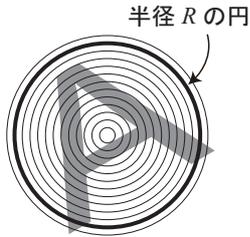


図 2 距離ヒストグラム

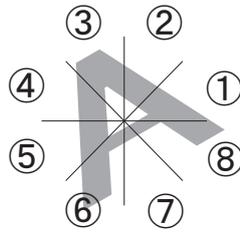


図 3 角度ヒストグラム

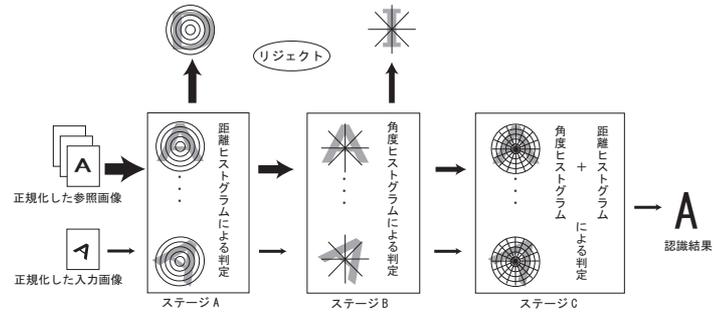


図 4 多段階の認識

文字画像を正規化すると近似楕円が円になるので、文字画像自体もほぼ円形になることは容易に想像できる。同様に、ゴシック体の「I」（アイ）は縦長の長方形であるため、正規化によって正方形になる。したがって、正規化画像による文字認識は、学習用の文字画像である参照画像と、認識対象の文字画像である入力画像をそれぞれ正規化してから行われるべきである。これらの処理については次節で述べる。

2.2 回転角度の推定と認識

前節の処理により、アフィン変換による歪みを正規化できた。しかし、図 1(a) と図 1(b) の回転角度が異なるように、この処理では回転の補正はできない。そのため、正規化された参照画像と入力画像の類似度が最も高くなる回転角度を探し出す必要がある。この処理を行う上で最も単純な方法は、片方の画像を一定角度ずつ回転していき、その都度画像の類似度を正規化相関などで調べて、類似度が最大になる角度を調べるというものである。この方法はすなわち、画像を、画像の中心を原点とする極座標 (r, θ) で表現し、角度 θ_0 毎にヒストグラムを作成しておき、ヒストグラム同士の照合を行うものである。このようにすれば、 $2\pi/\theta_0$ 回の照合でおおよその角度が判明するため、最後に θ_0 より小さい角度 θ_1 毎の詳細な照合を行えば良い。

3. 提案手法

前節で述べた基本的な認識方法を改良する形で提案手法を説明する。2.1 節で述べた正規化については、提案手法でも全く同じ処理を行う。一方、2.2 節の処理については、異なる処理を行う。

3.1 特徴ヒストグラム

処理の説明に入る前に、その準備として、後の処理で用いる 2 種類の特徴ヒストグラムを説明する。

3.1.1 距離ヒストグラム

文字の重心と各画素の距離の分布を表す距離ヒストグラムを作成する。 k 個のピンを持つ距離ヒストグラムの作成方法を以下に示す。

まず、2.1 節の正規化処理により、図 2 のように文字の近似楕円を半径 R の円に正規化する。ヒストグラムのピン数 k によって、円の中のピン数 n を

$$n = \begin{cases} k - 1 & \text{if } k \leq 9 \\ k - 2 & \text{if } k \geq 10 \end{cases} \quad (4)$$

と変化させる。

次に、半径 $\frac{i}{n}R$ の同心円を考える。ただし、 $i = 1, \dots, k - 1$

である。ヒストグラムは、画像の中心と画素との距離 r で構成される。文字領域の画素の r を調べ、

- $\frac{k}{n}R \leq r$ の場合、 k 番目のピン

• それ以外は、 $\frac{i-1}{n}R \leq r \leq \frac{i}{n}R$ を満たす i 番目のピンの度数を 1 度増加させるという処理を行いヒストグラムを作成する。最後に、ピンの総和が 1 になるように正規化する。この処理により、距離ヒストグラムが作成出来る。

3.1.2 角度ヒストグラム

各画素の角度の分布を表す角度ヒストグラムを作成する。図 3 に文字画像を放射状に $m = 8$ 分割した例を示す。 m 個のピンを持つヒストグラムを用意し、文字領域の各画素を放射状の各領域のピンに入れる。最後に、すべてのピンの総和が 1 になるように正規化する。この処理により、角度ヒストグラムが作成出来る。

3.2 多段階の認識

従来手法における 2.2 節の処理の代わりに行う処理を説明する。2.2 節では回転角度の推定において、大まかな角度合わせの処理と詳細な認識処理について述べた。提案手法では、この 2 種類の処理に、字種の絞り込みの処理を加えた 3 種類の処理を行う。説明の都合上、字種の絞り込みの処理をステージ A、大まかな角度合わせの処理（角度の絞り込み）をステージ B、詳細な認識処理をステージ C と呼ぶことにする。これらの 3 種類の処理は、図 4 に示すようにカスケード接続されている。多段階で認識・照合を行うことにより、入力画像とかけ離れた参照画像をリジェクトし、認識処理を高速化することが出来る。

3.2.1 字種の絞り込み（ステージ A）

この処理では、3.1.1 節で作成した距離ヒストグラムを使用して、字種の絞り込みを行う。この段階では角度に関する照合は行わない。

距離ヒストグラムの字種判別性能を調べるために、実験を行った。実験は数字 10 文字とアルファベットの大文字 26 文字、そしてアルファベットの小文字 a~z の 26 文字のうち輪郭が複数ある i と j を除いた 24 文字の、合計 60 文字を対象に行う。図 5 のように、文字を描いた一辺の長さが 96 ピクセルの画像を用意し、これらの画像に対して様々なアフィン変換を加える。ここで、アフィン変換行列 $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ を次のように分解する。



図 5 実験データ

$$T = L(\beta)R(\theta)S(\varphi)A(\alpha) \quad (5)$$

$$= \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & \tan \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & 1/\alpha \end{pmatrix}$$

$$\alpha = \pm \sqrt{\frac{a^2 + c^2}{ad - bc}} \quad (6)$$

$$\varphi = \tan^{-1} \frac{ab + cd}{ad - bc} \quad (7)$$

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

$$\beta = \pm \sqrt{ad - bc} \quad (9)$$

このうち拡大縮小のパラメータ β を除いた α, φ, θ の 3 パラメータを変化させてアフィン変換を加えた図形を実験に利用する。各パラメータを $\alpha = \{1, 2, 3, 4\}$, $\varphi = \{-1.0, -0.9, \dots, 1.0\}$, $\theta = \{0.0, 0.1, \dots, 1.4\}$ と変化させ、 $4 \times 21 \times 15 = 1260$ 通りのアフィン変換を加える。

$n = 5$ としたときの実験結果を図 6 に示す。(a) が 1 番目 (1 番内側), (b) が 4 番目, (c) が 5 番目のピンの値を表している。図は各字種に対して 1260 個のデータの分布を赤のヒストグラムで表している。ヒストグラムは値を縦方向、度数を横方向にとっている。ただし、値は平均 0、分散 1 に正規化して表記している。図から、値が字種によって異っていること、ヒストグラムの分布はピンによって異なることがわかる。したがって、字種を完全に分類するには至らないものの、ある程度の判別性能が期待できる。

3.2.2 角度の絞り込み (ステージ B)

この処理では、3.1.2 節で作成した角度ヒストグラムを使用して、字種の絞り込みを行う。提案手法の処理が従来手法と違うところは、従来手法では 1 種類の角度についてのみ絞り込みを行っていたが、提案手法では角度ヒストグラムのピンの数を変えて複数回の絞り込みを行うことである。

3.2.3 詳細な認識 (ステージ C)

最後の処理として、字種の認識を行う。この処理は、最も類似度が高い字種と角度の組を見つけるものである。ステージ B までの処理で字種と角度が絞り込まれるため、正解である可能性の高い組についてのみ照合を行う。照合には、距離ヒストグラムと 1 度刻みの角度ヒストグラムを同時に用いた。すなわち、これは極座標における画素毎の照合に当たる。

3.3 類似度の計算

本稿における類似度の計算には、Swain らが提案した類似度 [8] を用いる。 H_b を入力画像の b 番目のヒストグラム、 M_b を参照画像の b 番目のヒストグラム、ピンの数を d とすると、類似度 S_{HM} は以下の式で表される。

$$S_{HM} = \sum_{b=1}^d \min(H_b, M_b) \quad (10)$$



図 7 アフィン変換を受けた文字

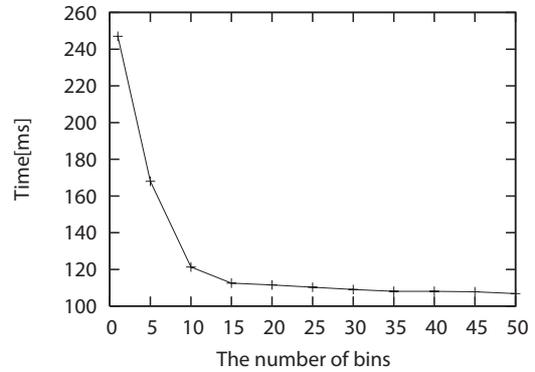


図 8 距離ヒストグラムのピン数と処理時間

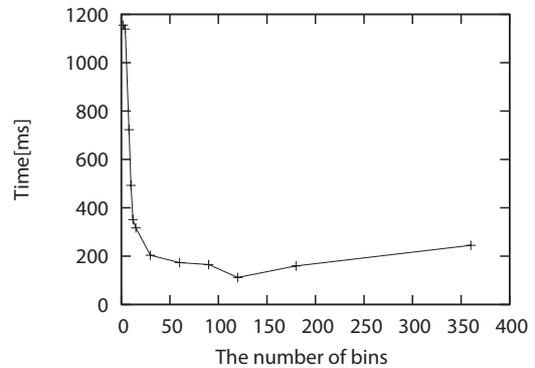


図 9 角度ヒストグラムと処理時間 (角度ヒストグラムが一段の場合)

表 1 形状が類似している文字

0 O o	6 9	l l	S s	V v
W w	X x	Z z	p d	q b

ヒストグラムは総和が 1 になるように正規化されているので、類似度の最大値は 1、最小値は 0 となる。距離ヒストグラム、角度ヒストグラムの類似度がそれぞれ閾値 T_i, T_a 以下であればそのヒストグラムに対応する字種や角度をリジェクトする。

4. 実験

提案手法の有効性を確認するために 3 種類の実験を行った。1 つ目は、アフィン変換された文字に対して、認識率と絞り込みによる処理時間の変化を調べた。2 つ目は、射影変換された文字に対して、頑健性の検証を行った。3 つ目は、実画像に対する認識性能を調べた。Arial フォントを用い、認識対象は英数字 62 字種から「i」と「j」を除いた 60 字種である。

4.1 アフィン変換を受けた文字の実験

シミュレーションによりアフィン変換を施した文字に対する認識率と絞り込みによる処理時間の検証を行った。文字サイズは 100 ポイントの文字を作成し、「0」のサイズが 51×98 ピク

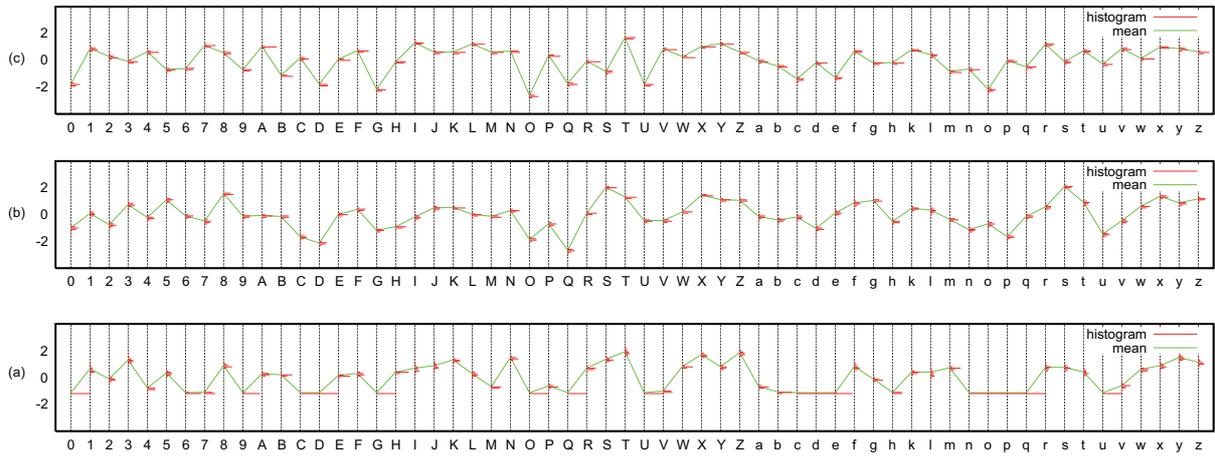


図 6 距離ヒストグラムの字種判別性能

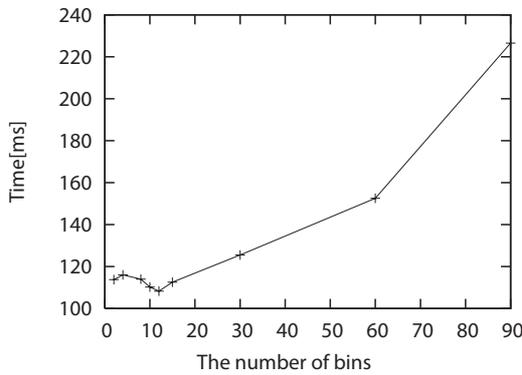


図 10 角度ヒストグラムの一段目のビン数と処理時間 (角度ヒストグラムが二段の場合)

セルとなるように変換して用いた。入力画像として、 150×150 ピクセルの画像の中心に参照画像を配置し、4 種類のアフィン変換を施した入力画像を計算機により生成した。例を図 7 に示す。(1) は左上と右下、(2) は右上と左下の x, y 座標をそれぞれ文字の中心に 25 画素変位させたものである。(3) は縦、(4) は横に $2/3$ 倍伸縮したものである。本実験と以降の実験では、表 1 に示す形状が類似している英数字は同一のものとして扱った。また、複数の輪郭を持つ「i」「j」は除いた。実験結果を順に説明する。

図 8 は、距離ヒストグラムのビン数と処理時間の関係を検証した結果である。ここで処理時間とは、文字が入力されてから結果を出力するまでの 1 文字当たりの平均である。ステージ B の絞り込みを 1 段、角度ヒストグラムのビン数を 120 とし、ステージ A の距離ヒストグラムのビン数を変化させ処理時間を計測した。縦軸が処理時間で、横軸が距離ヒストグラムのビン数である。実験結果より、距離ヒストグラムのビン数を増加させると処理時間が減少することが分かった。これは、ステージ A の処理により候補が絞り込まれるからである。認識率は 100% であった。

図 9 は、ステージ B を 1 段とした場合の角度ヒストグラムのビン数と処理時間の関係を検証した結果である。距離ヒストグラムのビン数は 20 とした。縦軸が処理時間で、横軸が角度ヒ

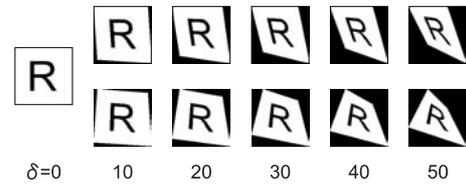


図 11 射影変換を受けた文字

ストグラムのビン数である。実験結果より、ビン数が 120 の時に 111ms となり、最も処理時間が小さくなることが分かった。これは、120 よりビン数を増加させるとステージ C での認識に時間がかかり、ビン数を 120 より小さくすると候補を十分に候補を絞り込めないためである。認識率は 100% であった。

図 10 は、ステージ B を 2 段とした場合の 1 段目の角度ヒストグラムのビン数と処理時間の関係を検証した結果である。距離ヒストグラムのビン数を 20、ステージ B の 2 段目のビン数を 120 に固定した。縦軸が処理時間で、横軸がステージ B の 1 段目のビン数となっている。実験結果より、ビン数を 12 とした場合に 108ms となり、最も処理時間が減少することが分かった。これは、12 よりビン数を 120 に近づけていくとステージ B の 2 段目の処理で行われる絞り込みに近くなるため、同じ処理を 2 度行うことになるからと考えられる。また、12 より小さくすると候補が十分に絞り込めていないため、処理時間が大きくなる。認識率は 100% であった。

以上の結果より、アフィン変換を受けた文字を正確に認識することが確認できた。各実験結果より、角度ヒストグラムによる絞り込みの回数を 2 段とし、角度ヒストグラムのビン数が 12、120 のとき最も高速であると分かった。また、予備実験により、3 段以降については逆に処理時間が大きくなることを確認している。これは、段数が 1 つ増えることによるオーバーヘッドが大きくなり、角度の絞り込みによる計算量削減効果を上回っていることが原因である。そこで、以後の実験では、ステージ B を 2 段とし、角度ヒストグラムのビン数を 12、120 とした。

4.2 射影変換に対する耐性の実験

シミュレーションにより射影変換を施した文字に対する認識

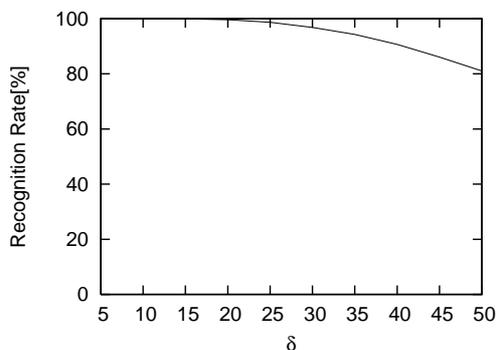


図 12 射影変換に対する結果

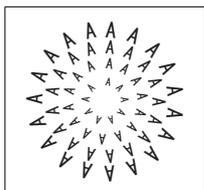


図 13 撮影対象

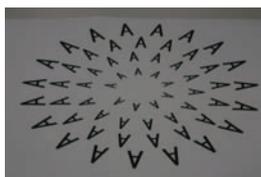


図 14 撮影画像

性能を検証する実験を行った。文字サイズは、100 ポイントの文字を作成し「0」のサイズが 35×55 ピクセルとなるように変換して用いた。入力画像として、図 11 に示すような文献 [9] で用いられている文字画像を用いた。この文字画像は、 150×150 ピクセルの画像の中心に参照画像を配置し、4 隅の x, y 座標を 0 または δ ($\delta = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$) だけ変位させた射影変換を施し、各 δ について $2^{4 \times 2}$ 個生成した画像である。

結果を図 12 に示す。 $\delta = 40$ までの射影変換に対し、90% を超える認識率を達成することができた。また、 $\delta = 50$ では、80% ほどの認識率となった。これは、 δ を大きくしていくと、アフィン変換に近似出来なくなるため、認識率が減少していくと考えられる。実験結果より、認識手法は射影変換にもある程度頑健であるといえる。

4.3 実画像に対する実験

デジタルカメラで撮影した英数字に対しての認識性能の検証を行った。参照画像はフォントを MS ゴシックフォントとした。文字サイズは、100 ポイントの文字を作成し「0」のサイズが 114×203 ピクセルとなるように変換して用いた。入力画像として、実画像を 2 値化し、連結成分を切り出したものを用意した。図 13 に示すような、A4 の用紙に中心から 36, 48, 60, 72 ポイントの文字を円状 5 に配置し、デジタルカメラで撮影したものが実画像である。実画像の例を図 14 に示す。デジタルカメラには Canon EOS 5D を用い、被写体までの距離を 60cm、紙面に対し 45 度の角度から室内で撮影した。画像サイズは 4368×2912 である。

この実画像から切り出した文字画像に対する平均認識率は 78.5% であった。主な誤認識は「7」と「L」、「0」と「Q」などで起こった。これは、射影変換による正規化の誤差、文字の劣化による影響であると思われる。

5. むすび

本稿ではアフィン不変な文字認識手法について検討し、アフィン不変なヒストグラムを用いた字種の絞り込みと、回転角度の高速な同定方法を提案した。提案手法の高速化は、複数の特徴を用いて字種と角度の絞り込みを多段階で行うものである。この手法を用いて、人工的にアフィン変換を施した文字、人工的に射影変換を施した文字、実際に撮影した文字に対して手法の有効性と高速性を検証した。

今後の課題としては、劣化文字への適用が挙げられる。文献 [10] では、デジタルカメラ撮影時に起こる光学的ぼけ、手ぶれ、低解像度化をモデル化して、人工的に劣化文字を生成している。このような手法はカメラベースの文字認識に非常に有効であると考えられる。そこで、劣化画像を用いて精度の向上を行うことを考えている。その際に問題となるのが、多量の学習サンプル (参照画像) である。このような状況では、提案手法のアフィン不変な特徴を用いて明らかに正解ではない文字を認識対象から除くことにより、処理時間を大幅に軽減することが可能であると考えられる。

また、複数の輪郭を持つ文字に対処することが挙げられる。輪郭単位で方向を算出し、輪郭の相対的な位置と方向を利用することで漢字などの認識を行うことができる。

謝辞 本研究の一部は、科学研究費補助金 若手研究 (B) 19700177 ならびに財団法人 大川情報通信基金の補助による。

文 献

- [1] D. Doermann, J. Liang and H. Li, "Progress in camera-based document image analysis," Proc. ICDAR2003, pp.606–616, 2003.
- [2] J. Liang, D. Doermann and H. Li, "Camera-based analysis of text and documents: a survey," International Journal on Document Analysis and Recognition, vol.7, no.2–3, pp.84–104, 2005.
- [3] 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.
- [4] 草地良規, 伊藤直己, 鈴木章, 荒川賢一, "画像インデクシングを目的としたテキスト領域抽出不要の景観中文字認識: 階層的認識辞書の適用による処理量の削減," 信学技報 PRMU2004-369, Oct. 2004.
- [5] J.-G. Leu, "Shape normalization through compacting," Pattern Recognition Letters, vol.10, no.4, pp.243–250, 1989.
- [6] 大田友一, 前信潔, 坂井利之, "透視投影されたテキストチャプターから平面の傾きを得る一方法," 1982年 情報処理学会コンピュータビジョン研究会資料, vol.16-2, pp.1–8, Jan. 1982.
- [7] W. Xiaohong and Z. Rongchun, "A new method for image normalization," Intelligent Multimedia, Video and Speech Processing, pp.356–359, May 2001.
- [8] M. Swain and D. Ballard, "Color indexing," International Journal of Computer Vision, vol.7, pp.11–32, 1991.
- [9] 根岸和也, 岩村雅一, 大町真一郎, 阿曾弘具, "孤立文字認識のための変形を許容する高速なテンプレートマッチング," 電子情報通信学会論文誌 D, vol.J90-D, no.2, pp.503–513, Feb. 2007.
- [10] 石田皓之, 高橋友和, 井手一郎, 目加田慶人, 村瀬洋, "携帯カメラ入力型文字認識におけるぼけやぶれに対処するための生成型学習法," 電子情報通信学会論文誌 D, vol.89, no.9, pp.2055–2064, Sept. 2006.