

## 段階的な枝刈りによるアフィン不変な文字認識

非会員 岩村 雅一\* 非会員 堀松 晃\* 非会員 丹羽 亮\*  
正員 黄瀬 浩一\* 非会員 内田 誠一\*\* 正員 大町真一郎\*\*\*

### Affine Invariant Character Recognition by Progressive Removing

Masakazu Iwamura\*, Non-member, Akira Horimatsu\*, Non-member, Ryo Niwa\*, Non-member,  
Koichi Kise\*, Member, Seiichi Uchida\*\*, Non-member, Shinichiro Omachi\*\*\*, Member

(2010年5月17日受付, 2011年1月10日再受付)

Recognizing characters in scene images suffering from perspective distortion is a challenge. Although there are some methods to overcome this difficulty, they are time-consuming. In this paper, we propose a set of affine invariant features and a new recognition scheme called “progressive removing” that can help reduce the processing time. Progressive removing gradually removes less feasible categories and skew angles by using multiple classifiers. We observed that progressive removing and the use of the affine invariant features reduced the processing time by about 60% in comparison to a trivial one without decreasing the recognition rate.

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

**Keywords:** approximate ellipse, normalization, feature histogram, progressive removing

#### 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, 文字を斜めから撮影した場合に生じる射影歪み, ほげや解像度の低下など, 様々な課題を解決しなければならない<sup>(1)~(3)</sup>。本論文ではこのうち, 文字を斜めから撮影した場合に生じる射影歪みに焦点を当てる。

この問題に対処する方法は大きく二つに分けることができる。一つ目の方法は, 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<sup>(4)</sup>。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るというトレードオフの関係があり, 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。二つ目は, 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<sup>(5)</sup>と大田ら<sup>(6)</sup>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倍を補正できるが, 図形の回転を正規化することができないため, 回転角度は別の方法で求める必要がある。このとき, 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に時間がかかってしまうため, 小さい処理コストで画像の回転角度を求められる手法が必要となる。この問題に対して, 文献(7)ではモー

\* 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  
Graduate School of Engineering, Osaka Prefecture University  
1-1, Gakuen-cho, Naka-ku, Sakai 599-8531, Japan

\*\* 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  
Faculty of Information Science and Electrical Engineering,  
Kyushu University  
744, Motooka, Nishi-ku, Fukuoka 819-0395, Japan

\*\*\* 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  
Graduate School of Engineering, Tohoku University  
6-6-05, Aoba, Aramaki-Aza, Aoba-ku, Sendai 980-8579,  
Japan

メントを用いた方法が試みられているが、成功しているとは言い難い。

本論文では、二つ目の方法において課題であった回転角度の推定を行い、アフィン変換を受けた文字画像を高速に認識する手法を提案する<sup>†</sup>。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あるため、一つ目に述べた認識時に様々な画像と照合する方法にも有効と考えられる。

## 2. アフィン変換による歪みの正規化

本論文で用いる図形の歪みを正規化する手法<sup>(5)(6)</sup>について述べる。この手法が想定している状況は、文字等の図形が3次元空間中の2次元平面に書かれていて、その図形を任意の位置に置かれたカメラで撮影するというものである。このようにして得られた図形には射影歪みが生じるのだが、3次元座標を求めることなく、得られた2次元画像だけで射影歪みを補正する。その際、図形に生じる射影歪みをアフィン歪みで近似し、図形の歪みを正規化する。

この手法の特徴は、Fig. 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。

近似楕円を作成する方法を説明する。近似楕円は図形と同じ共分散行列を持つので、図形の共分散行列  $C$  を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。

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

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。

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

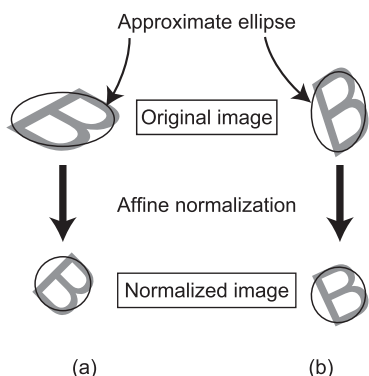


Fig. 1. Affine normalization.

<sup>†</sup> 提案手法はアフィン歪みにのみ対応しており、射影歪みを受けた画像に対してはアフィン歪みに近似して対処する。

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \text{character region} \\ 0 & \text{if } (x, y) \in \text{non-character region} \end{cases}$$

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  $(x_0, y_0)$ 、アフィン変換後の図形の  $x, y$  座標を  $(x_1, y_1)$  とおくと、図形の重心が  $\begin{pmatrix} m_{10} \\ m_{01} \end{pmatrix}$  であり、重心は回転によって変化しないので、

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = C^{-1} \begin{pmatrix} x_0 - m_{10} \\ y_0 - m_{01} \end{pmatrix} + \begin{pmatrix} m_{10} \\ m_{01} \end{pmatrix} \dots\dots\dots (2)$$

となる。

この正規化は直交基底の軸方向にのみ伸縮処理を施すため、アフィン変換のうち、シアーを完全に復元することはできない。そのため、本論文では (2) 式の正規化処理を2回続けて行うことで、この問題を緩和する。

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」(大文字のオー) のような楕円形の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」(大文字のアイ) は縦長の長方形であるため、正規化によってほぼ正方形になる。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。

また、元の図形は随分異なっているのに、正規化することによって類似する字種が存在することにも注意が必要である。例えば、「6」と「9」など、通常のOCRで誤認識することはあまり多くないと考えられる字種が、フォントによっては上記の正規化で類似した形状に変換される。すなわち、本論文のようにアフィン不変なテンプレートマッチングを行う場合には、本質的に認識不可能な字種が存在することに注意されたい。

## 3. 提案手法

本章では、アフィン歪みの正規化後に効率的に字種と角度の照合を行う提案手法について述べる。

**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** 距離ヒストグラムは、文字の重心と文字領域の各画素の距離の分布を表すヒストグラムである。Fig. 2(a) はビン数が6の場合の例である。 $k$  個のビンを持つ距離ヒストグラムは次のように作成する。まず、2節の正規化処理により、Fig. 2(a)のように文字の近似楕円を半径  $R$  の円に正規化する。ヒストグラムのビン数  $k$  によって、半径  $R$  の円の内側のビン数を  $k-1$ 、外側を1とする。次に、半径  $\frac{i}{k-1}R$  の同心円を考える。ただし、 $i = 1, \dots, k-1$  である。画像の中心と文字領域の各画素と

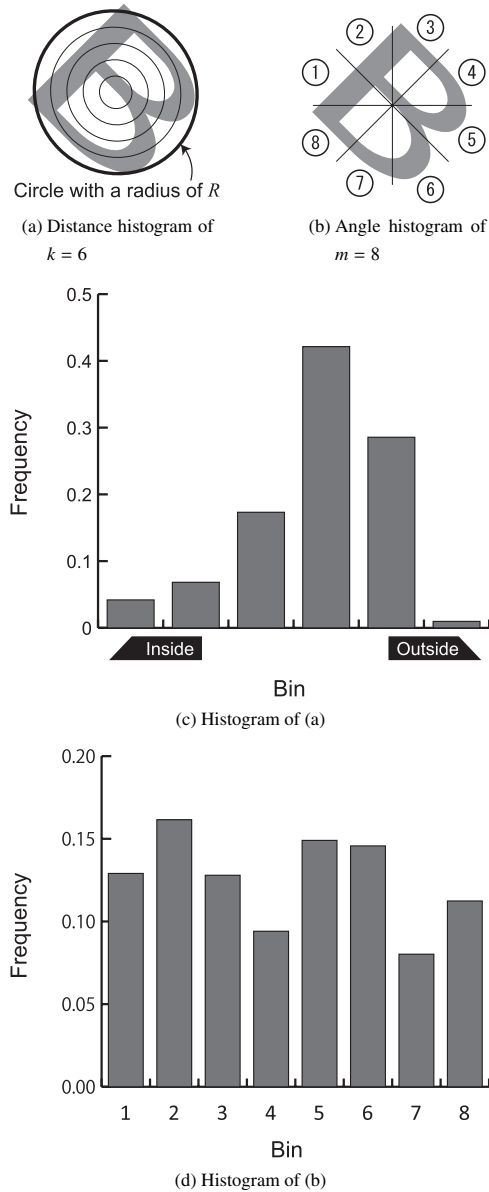


Fig. 2. Histograms.

の距離  $r$  を調べ、以下の処理を行う。

- $r \geq R$  のときは円の外側 ( $i = k$ ) に対応するので、 $k$  番目のビンの度数を 1 増加させる。
- それ以外は、 $\frac{i-1}{k-1}R \leq r < \frac{i}{k-1}R$  を満たす  $i$  番目のビンの度数を 1 増加させる。

この処理を全ての画素に行い、最後にビンの度数の総和が 1 になるように正規化する。Fig. 2(c) に実際のヒストグラムの例を示す。

〈3・1・2〉 角度ヒストグラム 角度ヒストグラムは、画素の角度の分布を表すヒストグラムである。Fig. 2(b) に文字画像を 8 分割した例を示す。 $m$  個のビンを持つ角度ヒストグラムは次のように作成する。正規化した文字画像の重心を原点とする極座標系  $(r, \theta)$  を考え、画素の角度  $\theta$  に基づいて  $[\theta/m]$  番目のビンの度数を 1 増加させる。この処理を全ての画素について行い、最後に全てのビンの度数の総和が 1 になるように正規化する。Fig. 2(d) に実際のヒストグ

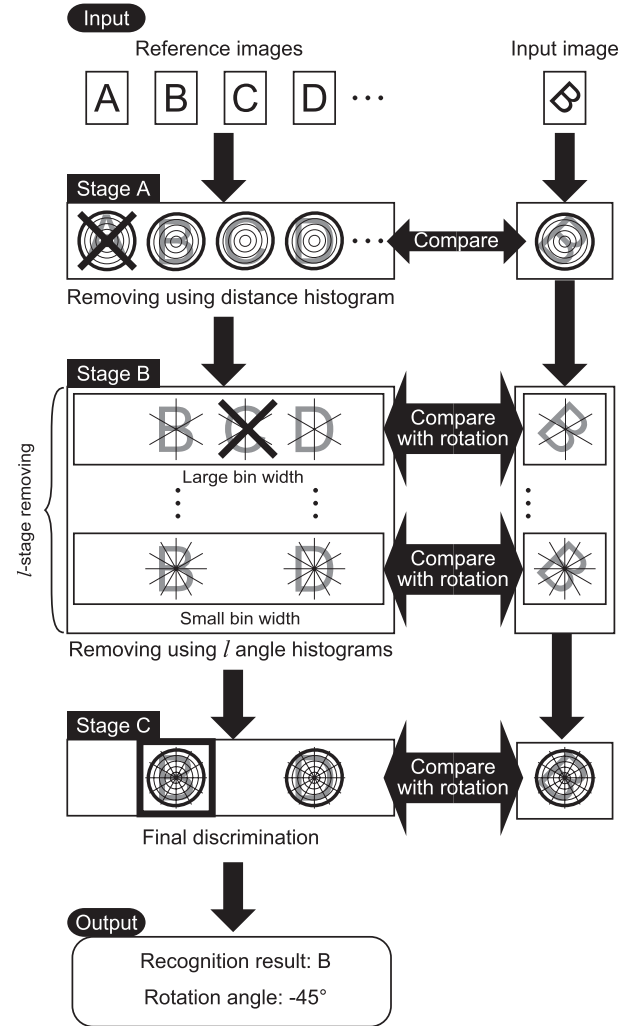


Fig. 3. Progressive removing.

ラムの例を示す。

〈3・2〉 類似度 次節で述べる処理では Swain らが提案した類似度<sup>(8)</sup>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、ビンの数を  $N$  とすると、類似度  $S_{HM}$  は以下の式で表される。

$$S_{HM} = \sum_{b=1}^N \min(H_b, M_b) \dots \dots \dots (3)$$

ヒストグラムは総和が 1 になるように正規化するので、類似度の最大値は 1、最小値は 0 となる。

〈3・3〉 段階的な枝刈り 提案する効率的な字種と角度の照合方法について述べる。提案手法は Fig. 3 に示すように、カスケード接続された 3 種類の処理で構成される。説明の都合上、字種の絞り込みをステージ A、角度合わせ (角度の絞り込み) をステージ B、詳細認識をステージ C と呼ぶことにする。上記の絞り込みは、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以下の字種や角度を候補から除く処理である。提案手法は多段階での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早い段階で候補から取り除き、認識処理を高速化する。

〈3・3・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閾値を 0.9 とした。

〈3・3・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。この処理の特徴は、1 種類の角度ヒストグラムで絞り込みを行うのではなく、角度ヒストグラムのピンの数を変えながら複数回の絞り込みを行う点である。この絞り込みの回数を以下では  $l$  で参照する。

ここで角度ヒストグラムの閾値について注意しなければならないのは、角度のずれによる類似度の低下具合が字種によって異なることである。例えば、「O」の場合は正規化するとほぼ円形であるため、角度のずれによる類似度の低下は比較的小さい。一方、「N」の場合はある程度回転すれば類似度が大きく低下する。類似度の低下の様子は字種毎に異なるため、閾値として一定値を用いることは得策でない。そこで本論文では、枝刈りによって正解を候補から取り除かない程度の閾値を字種毎に学習し、適応的に設定する。具体的には、角度ヒストグラムの 1 つのピンの角度が  $\theta$  であるとき、参照画像を  $\pm[\theta/2]$  回転させた画像と参照画像の類似度を計算する。角度を  $\pm[\theta/2]$  としたのは、 $\pm[\theta/2]$  のずれは許容されるべきであり、近似的に考えれば、許容範囲内で類似度が最も低下するのは  $\pm[\theta/2]$  回転させたときと考えられるからである<sup>†</sup>。そして、2 つの類似度の小さい方を角度ヒストグラムの閾値とする。ただし、角度ヒストグラムの類似度は回転角のずれに対して単調減少する訳ではなく、あくまでも目安であることと、実画像の認識においてはノイズの影響等による類似度の低下も考えられることから、上記の閾値にある定数 (0~1) を乗じて条件を緩和する。この定数の値は予備実験により定め、0.9 とした。

〈3・3・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、字種と角度の組から最も類似度が高いものをみつける処理である。照合には、文字領域を同心円状と放射線状に格子状に区切ったヒストグラムと〈3・2〉の類似度を用いる。このヒストグラムは、前述の距離ヒストグラムと角度ヒストグラムとは別のものである。本論文では、同心円状に 5 段階、放射線状に 360 段階 (1 度刻み) 区切り、合計 1800 個のピンを持つヒストグラムを作成した。

## 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。認識対象として英数字 62 字種を使用した。これ以降、画像の例として、62 字種のうち「B」のみを示す。正誤判定の際は字種が正解であれば正解とし、角度は考慮していない。以下の実験では AMD Opteron 2.8 GHz を搭載した

<sup>†</sup> 類似度の低下は角度に関して単調減少ではないため、必ずしも  $[\theta/2]$  回転させたときの類似度が回転角度が  $0 \sim [\theta/2]$  の区間での最小類似度になるとは限らない。これが「近似的に」とした理由である。



Fig. 4. Affine distorted character images.

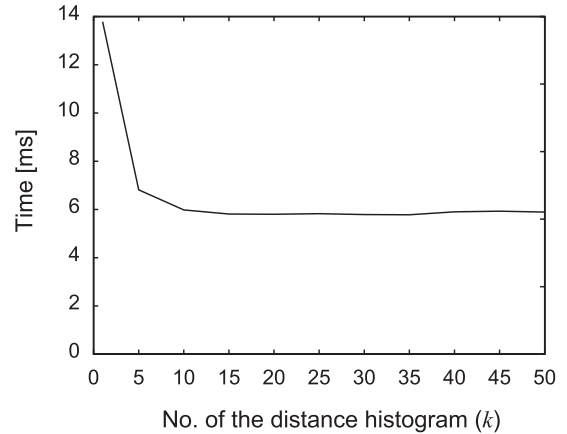


Fig. 5.  $k$  vs average processing time and recognition rate;  $l = 1$  and  $m_1 = 72$  were used.

計算機を使用した。

〈4・1〉 アフィン変換を受けた文字に対する頑健性と枝刈り効果 計算機上でアフィン変換を施した文字に対する認識率と平均処理時間の検証を行った。平均処理時間とは、文字が入力されてから結果を出力するまでの 1 文字当たりの処理時間である。参照画像と入力画像の両方で MS ゴシックフォントを使用した。参照画像は「O」の大きさが  $35 \times 62$  ピクセルであった。入力画像は Fig. 4 に示すように、10 種類のアフィン変換を参照画像に施して作成した。

Fig. 5 はステージ A の距離ヒストグラムのピン数  $k$  を変化させたときの平均処理時間と認識率である。横軸が距離ヒストグラムのピン数  $k$ 、縦軸が 1 文字当たりの平均処理時間と認識率である。ステージ B の絞り込み段数を  $l = 1$ 、角度ヒストグラムのピン数を  $m_1 = 72$  とした。その際、Fig. 6 に示す誤認識が生じた。これらはアフィン変換を受けるとほぼ同形になるため、アフィン不変な識別器での認識は困難と考えられ、実際に誤認識が生じた。そこで本節の実験では Fig. 6 に示す字種を同一クラスと考え、これらの字種同士の誤認識を誤認識と見做さない。すなわち、「v」を「V」と認識しても正解であるとする。これを Configuration 1 と呼ぶこととし、本節の実験で用いる。平均処理時間は  $k = 35$  のときに最も少なく、6.78 ms であった。提案手法で  $k = 1$  かつ  $l = 1$  の場合は、アフィン不変量である距離ヒストグラムによる枝刈りを行わず、また角度の枝刈りが 1 回のみの自明な手法であり、平均処理時間は 13.8 ms であった。したがって、提案手法でアフィン不変な距離ヒストグラムを使用することによって、処理時間を自明な手法から大幅に削減できることが確認できた。

Fig. 7 はステージ B の段数を  $l = 1$  とし、角度ヒストグ

vV zZ L7 nu 96 00o pd 11

Fig. 6. Misrecognized characters occurred in the experiment of Fig. 5. These characters, vV, zZ, L7, nu, 96, 00o, pd and 11, are treated as the same classes in the experiments in <4.1> because they are difficult to distinguish for affine invariant recognizers. We call this "Configuration 1".

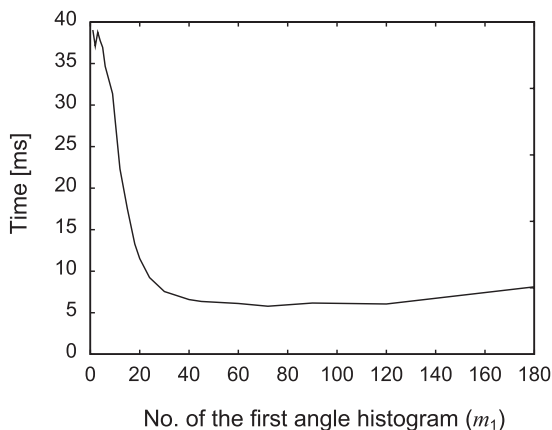


Fig. 7.  $m_1$  vs average processing time and recognition rate;  $k = 35$  and  $l = 1$  were used.

ラムのビン数  $m_1$  を変化させたときの平均処理時間と認識率である。横軸が角度ヒストグラムのビン数  $m_1$ , 縦軸が1文字当たりの平均処理時間と認識率である。距離ヒストグラムのビン数は  $k = 35$  とした。認識率は  $m_1 = 2$  の場合は 95.8%であったが, それ以外は 100%であった。平均処理時間については,  $m_1 = 72$  の時に最小値 5.78 ms をとった。 $m_1$  が小さいときはステージ B での識別能力が小さいため, ステージ B ではほとんど候補から除かれず, 多数の候補がステージ C で処理されたためと考えられる。逆に  $m_1$  が大きいときは多数の候補がステージ B で処理されたが, この処理には時間がかかるため, 全体の処理時間を押し上げたと考えられる。最適な  $m_1$  の値はこれらのトレードオフで決まったと考えられる。

Fig. 8 はステージ B の段数を  $l = 2$  とした場合の平均処理時間である。図には認識率が 100% となった  $m_1, m_2$  の組み合わせから平均処理時間が少ない順に 10 組を示す。距離ヒストグラムのビン数は  $k = 35$  とした。最も平均処理時間が少なかった  $(m_1, m_2) = (30, 120)$  の組み合わせでは平均処理時間を 5.46 ms (自明な手法の約 4 割) に減少することができた。このことから, ステージ B を  $l = 1$  とした場合よりも  $l = 2$  とすることで平均処理時間が減少することが確認できた。なお,  $l = 3$  の場合は  $(m_1, m_2, m_3) = (3, 30, 120)$  のときに平均処理時間が 5.60 ms となり,  $l = 1$  の場合を下回ったが,  $l = 2$  の場合を上回っている。これは段数を増やすことによるオーバーヘッドが処理軽減効果を上回ったためと考えられる。

以上より, 提案手法がアフィン変換を受けた文字を正確かつ高速に認識できることが確認できた。以降の実験では,

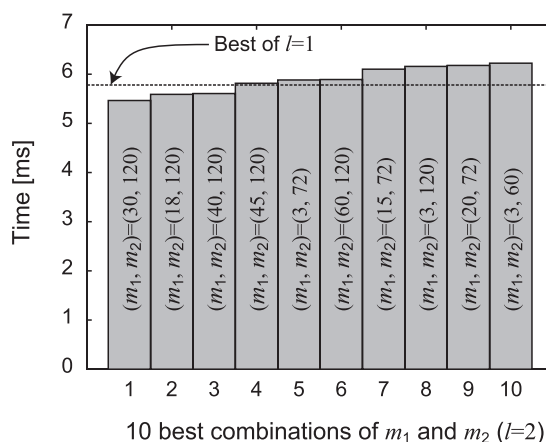


Fig. 8.  $(m_1, m_2)$  vs average processing time;  $k = 20$ ,  $l = 2$  and  $m_2 = 120$  were used.

これまでの実験での最良パラメータを使用する。すなわち, 距離ヒストグラムのビン数は  $k = 35$ , 角度ヒストグラムによる絞り込みの回数は  $l = 2$ , 角度ヒストグラムのビン数は  $(m_1, m_2) = (30, 120)$  である。

なお, 本節の実験で Arial フォントを使用した場合には, 字形の違いから認識不可能な字種が若干異なるものの, MS ゴシックと同様の結果が得られた。

**<4.2> 射影変換に対する頑健性** 計算機上で射影変換を施した文字に対する認識性能を検証する実験を行った。射影変換を受けると, 2章で求めた楕円の重心がずれると考えられるが, この問題に対する特段の対処をしなかった。参照画像は前章と同じものを使用した。入力画像は,  $150 \times 150$  ピクセルの画像の中心に参照画像を配置し, 画像の 4 隅の  $x, y$  座標を 0 または  $\delta$  ピクセルだけ変位させた射影変換を施し, 周囲の余白を取り除いた。このようにすることで, 各  $\delta$  について  $2^8$  個の画像を生成した。射影変換の程度を表す変数である  $\delta$  は  $\delta = 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$  とした。作成した入力画像を Fig. 9 に示す。また, 前節と同じ Configuration 1 で認識実験を行った結果を Fig. 10 の実線で示す。認識結果を解析したところ, 誤認識のほとんどは Fig. 11 で下線を引いた文字の組であり, これらは大文字と小文字が類似しているために認識が困難であった。そこで, Fig. 6 の Configuration 1 の他に, Fig. 11 に示す Configuration 2 でも認識実験を行い, Fig. 10 の破線で表した。その結果, どちらの Configuration でも, 全体的な傾向として  $\delta$  の増加に伴って認識率は低下すること,  $\delta = 50$  において Configuration 1 を用いた場合は誤認識が約 96.5%程度, Configuration 2 を用いた場合は誤認識が約 99%程度に留まっていることがわかった。したがって, 提案手法はある程度までの射影変換に対しては比較的頑健であることが確認できた。

**<4.3> 実画像に対する認識性能** 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。参照画像と入力画像の両方で MS ゴシックを使用した。参照画像は前節までと同じである。入力画像は, まず Fig. 12 に

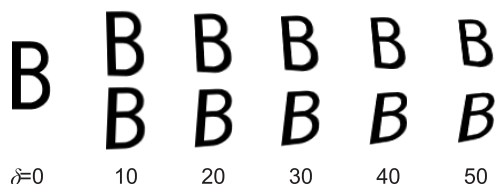


Fig. 9. Examples of perspectively distorted character images.

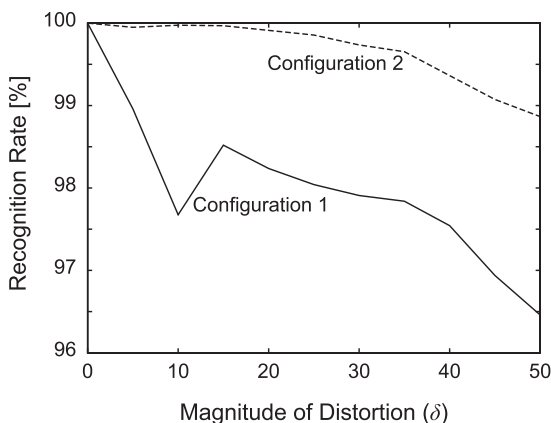


Fig. 10.  $\delta$  vs recognition rate for perspectively distorted character images.

vV zZ L7 nuU 96 00o pd 1I  
bq xX cC sS wW

Fig. 11. Misrecognized characters occurred in the experiment in  $\langle 4 \cdot 2 \rangle$ . As in Fig. 6, these characters, vV, zZ, L7, nuU, 96, 00o, pd, 1I, bq, xX, cC, sS and wW, are treated as the same classes in the experiments in  $\langle 4 \cdot 2 \rangle$  and  $\langle 4 \cdot 3 \rangle$ . We call this “Configuration 2”. Newly added characters from Configuration 1 are underlined.

示すように中心から 36, 48, 60, 72 ポイントの文字を円状に配置し, A4 の用紙に印刷し, デジタルカメラで撮影した (Fig. 13 参照)。そして, 画像の 2 値化と連結成分の切り出しを手動で行い, 入力画像とした。デジタルカメラには Canon EOS 5D を用いて室内で撮影した。撮影した画像の大きさは  $4368 \times 2912$  ピクセルであった。Configuration 1 と 2 を用いた実験結果を Table 1 に示す。平均処理時間は Configuration や角度に依らず約 10 ms であり, 前節までの計算機上で作成した画像の認識結果と同程度であった。また, 認識率は Configuration 1 では 97% 以上, Configuration 2 では 99% 以上であった。この結果より, 提案手法は実画像に対しても頑健かつ高速に認識可能であることが確認できた<sup>†</sup>。

<sup>†</sup> 前章の人工データを用いた実験とは実験条件が異なるため, 本章の実画像の認識率が前章の人工データを必ずしも下回らない。今回の場合, 量子化誤差と, 認識サンプルに加えた変形の大きさが原因として考えられる。前者については, 実画像の文字の大きさが人工データの 3 倍程度あるため, ヒストグラムを作成する際に量子化誤差の影響を受け難くなった可能性がある。後者については, 前章の人工データには本章の実画像よりも大きな変形を受けたものが存在する。

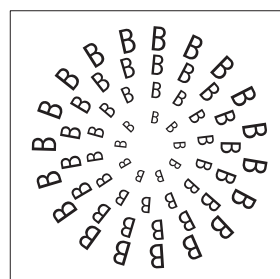


Fig. 12. An image for printing.

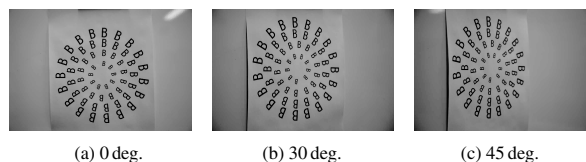


Fig. 13. Pictures of Fig. 12 each of which was printed on an A4 paper and then taken by a digital camera from a different angle for creating test images.

Table 1. Recognition results of characters in real images.

| Configuration | Angle (deg.) | Recog. Rate (%) | Time (ms) |
|---------------|--------------|-----------------|-----------|
| 1             | 0            | 99.8            | 9.88      |
|               | 30           | 97.6            | 10.2      |
|               | 45           | 97.5            | 9.87      |
| 2             | 0            | 99.9            | 9.87      |
|               | 30           | 99.8            | 10.2      |
|               | 45           | 99.4            | 9.82      |

## 5. むすび

本論文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, 計算機上でアフィン変換や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確認した。

今後の課題としては, 高速な文字の切り出し方法の確立と, アフィン不変を受けた場合に区別できない「6」と「9」などの判別が挙げられる。後者に関しては, 一文字単位の認識では区別できないが, 周囲の文字の情報を使用することで判断可能と考えられる。

## 謝辞

本研究の一部は, 科研費補助金若手研究 (B) (19700177, 21700202) の補助による。

## 文献

- (1) D. Doermann, J. Liang, and H. Li: “Progress in camera-based document image analysis”, Proc. ICDAR 2003, pp.606–616 (2003)
- (2) J. Liang, D. Doermann, and H. Li: “Camera-based analysis of text and documents: a survey”, IJ DAR, 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 (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, PRMU2004-89 (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ターンから平面の傾きを得る一方法」, 情報処理学会コンピュータビジョン研資, 16-2, pp.1-8 (1982)
  - (7) W. Xiaohong and Z. Rongchun: “A new method for image normalization”, Proc. 2001 Int'l Symposium on Intelligent Multimedia, Video and Speech Processing, pp.356-359 (2001)
  - (8) M. Swain and D. Ballard: “Color indexing”, IJCV, Vol.7, pp.11-32 (1991)

岩村 雅一 (非会員) 1998年東北大・工・通信卒業。2003



年同大大学院博士課程修了。同年同大大学院工学研究科助手。2004年阪府大大学院工学研究科助手。2011年同准教授。博士(工学)。パターン認識, 文書画像解析, 物体認識, 情報検索などの研究に従事。2006年電子情報通信学会論文賞, 2007年 IAPR/ICDAR Best Paper Award, 2010年 IAPR Nakano Award (最優秀論文賞), ICFHR Best Paper Award 各受賞。電子情報通信学会, IEEE, ACM, 情報処理学会各会員。

堀松 晃 (非会員) 2007年阪府大・工・情報卒業。2009



年同大大学院修士課程修了。在学中は文字認識の研究に従事。

丹羽 亮 (非会員) 2007年阪府大・工・情報卒業。2009



年同大大学院修士課程修了。在学中はカメラベースの文書画像処理の研究に従事。

黄瀬 浩一 (正員) 1986年阪大・工・通信卒業。1988年同大大学院博士前期課程修了。同年同大大学院博士後期課程入学。1990年阪府大・工・電気助手。1993年同情報助手, 1996年講師, 1999年助教授を経て, 2005年同大大学院工学研究科教授。現在に至る。博士(工学)。その間, 2000~2001年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。2006年電子情報通信学会論文賞, 2007年 IAPR/ICDAR Best Paper Award, 2010年 IAPR Nakano Award (最優秀論文賞), ICFHR Best Paper Award 各受賞。電子情報通信学会, 情報処理学会, 人工知能学会, IEEE, ACM などの会員。



内田 誠一 (非会員) 1990年九大・工・電子卒業。1992年同大大学院修士課程(情報)修了。セコム(株)勤務を経て, 現在, 同大システム情報科学研究院情報知能工学部門教授。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。2003年電子情報通信学会 PRMU 研究奨励賞, 2006年 MIRU 長尾賞, 2007年 IAPR/ICDAR Best Paper Award, 2008年電子情報通信学会論文賞, 2010年 ICFHR Best Paper Award 各受賞。IEEE, 情報処理学会各会員。



大町 真一郎 (正員) 1988年東北大・工・情報卒業。1993年同大大学院博士課程修了。同年同大情報処理教育センター助手。1996年同大工学部助手。1999年同大大学院工学研究科助教授。2009年同教授。博士(工学)。その間, 2000~2001年米国ブラウン大学客員准教授。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。2007年 MIRU 長尾賞, 2007年 IAPR/ICDAR Best Paper Award, 2010年 ICFHR Best Paper Award 各受賞。電子情報通信学会, IEEE, 情報処理学会, 人工知能学会等各会員。

