

汎用文字認識システムに関する研究

大町 真一郎 阿曾 弘具

東北大学大学院工学研究科

1 まえがき

汎用文字認識システムには、より多様な文書に対応できることが求められている。文書認識の立場で言えば、書式に依存しないあらゆるフォーマットの文書、図や表の混在する文書などに対応できることが求められ、文字認識の立場で言えば、どのようなフォントにでも対応できること、劣化の激しい文字にも対応できることなどが挙げられる。

本研究では、これらの要件のうち、劣化した文字画像を高精度に認識する手法について検討した。明瞭に印刷された文書と比較して、複写やファクシミリによる転送が行なわれたものは品質が落ち、認識誤りが多くなる。この原因の一つは、雑音の混入や量子化誤差により生じるつぶれ等の文字の劣化である。つぶれた文字画像は本来の文字画像とはかなり異なるものとなる。つぶれは偶発的に発生し予測できないものであり、しかも多種多様であるから、あらかじめ訓練データを集めて設計した従来の認識システムでは、訓練データとしてつぶれのある文字画像を大量に用いたとしても対応しきれず、認識率が低下する。実用的な文字認識システムにおいて、低品質文字を高精度に認識する手法の開発は不可欠である。

多少のパターンの変形に強い手法として、部分空間法 [1] や複合類似度法 [2, 3] が知られている。部分空間法は、訓練データ群から訓練データ群の分布を近似する正規直交基底を設計し、未知入力パターンのこの基底への射影成分を求めることで識別を行なう。複合類似度法は訓練データ群のすべてのデータを用いて類似度を計算すると同様の効果がある手法である。これらはいずれも訓練データから母集団の分布形状を推定し、分散の大きい成分の影響を軽減する効果がある。しかし、雑音等による入力画像の変形が非常に大きい場合、訓練データから推定された分布形状はこの変形に対応できず、入力画像の属すべきカテゴリーとの類似度がそれほど大きくならないため、誤認識の原因となる。複写機による複写やファクシミリで転送した文字画像を扱う場合、雑音や量子化誤差により生じるつぶれ等の劣化は大きく、しかもこれらは偶発的に発生するものであり、統計的手法を適用するための適切な訓練データの収集が困難である。また、そのような偶発的劣化は文字パターンの特徴とは独立なものであり、そのような訓練データを用いた手法は、高品質な文字パターンや訓練データとは異なる劣化をしている文字パターンの認識に悪影響を与える可能性がある (3.1の実験結果で確認される)。多様なつぶれに対応するため、入力画像からつぶれを検出し、つぶれに応じて動的に補正を加えることを考案した。これまでは雑音を動的に補正する手法についてはほとんど検討されていないと思われる。

低品質文字の認識に関しては、最近のデジタルカメラ等の普及から、カメラ画像に含まれる文字画像を対象とした研究が行なわれている [4, 5, 6]。[4] では、カメラを移動することで複数の画

像を得、カメラ画像の解像度の低さを補う手法を提案している。[5]では、ガウシアン・ラプラシアン処理による画像強調としきい値処理によって切り出しと認識を行なう手法を提案している。[6]では複数の参照パターンを用意し、切り替えることで認識精度が向上することが示されている。どのパターンを用いるかを文字列画像中の複数の文字画像から得られる劣化指数を用いて判定しており、さまざまな品質の文字に対応できるようになっている。しかし、いずれの手法も参照パターン（辞書）としては訓練データから作成したものをそのまま用いており、入力画像に応じて辞書パターンを積極的に補正するものではない。

本論文では低品質文字のうち特につぶれた文字を対象とし、低品質文字認識に適したアルゴリズムを提案する。つぶれた文字を高精度に認識することを目的とした研究としては文献[7]があり、つぶれた領域は特徴ベクトルの値が小さくなることを利用した重み付け法およびつぶれた文字を辞書に登録しておく手法を検討し、後者の方が有効であることが示されている。しかし、前述したように低品質の文字を辞書に加える手法では高品質な文字パターンや訓練データとは異なる劣化をしている文字パターンの認識に悪影響を与える可能性がある。また、文字画像をいくつかの小領域に分け、細線化を用いて領域ごとのつぶれを表す尺度（つぶれ度）を求め、重み付けを行なうことで低品質文字をある程度高精度に認識する手法が提案されている[8, 9]。距離尺度として[8]では重み付きユークリッド距離、[9]では単純類似度法が用いられているが、いずれもつぶれ度の大きい領域に対する距離値または類似度を小さくするような重み付けを行なっており、つぶれた領域の情報を用いないことで高精度化を達成するものである。本論文で提案する手法は、小領域に分割することなくつぶれた領域そのものを検出し、つぶれが生じた場合に特徴ベクトルの分布形状がどのように変化するかを考察し、つぶれに応じて部分空間法や複合類似度法で用いられる固有値・固有ベクトルを補正する方法である。補正によって特徴ベクトルの分布領域を近似する新たな正規直交基底および軸長を設計し、これらを認識に用いる。提案手法はつぶれのある文字画像を高精度に認識できるだけでなく、つぶれの全くない文字が認識対象となった場合は補正を行わないため、綺麗な文字画像に対する悪影響がないという特徴がある。また、実質的には、低品質文字の認識には自然言語等の知識を用いた後処理[10]が不可欠である。このような後処理と組み合わせる場合、未知入力文字画像に対して必ずしも常に候補を1個選出する必要はなく、読み取り不可能な場合には確実に棄却する機能も必要となる。これらを勘案した上で実用的な低品質文字認識手法を提案し、ファクシミリで転送して劣化した文字画像を用いた認識実験により手法の有効性を確認する。

2 文字認識アルゴリズムとつぶれの補正法

本章ではまず補正を加えない一般的な部分空間法および複合類似度法による文字認識手法について述べる。そして、文字画像につぶれが生じた場合の補正法を提案する。

2.1 部分空間法および複合類似度法による文字認識

ここでは、文字画像の変形の吸収と次元削減のため、前処理として以下の操作を行なう。まず文字画像を 64×64 の大きさに線形伸縮（正規化）する。黒画素を1、白画素を0とし、これにガウスフィルタをかけて 32×32 の多値画像とする。ここまでの処理の流れを図1に示す。更に定正規化操作[2]を行なった後、ノルムを1にすることで得られる1024次元の特徴ベクトルを用い

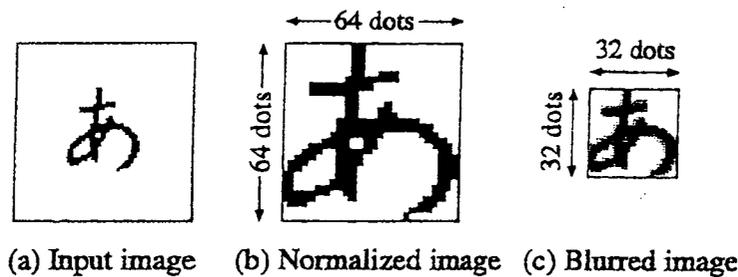


図 1: 前処理

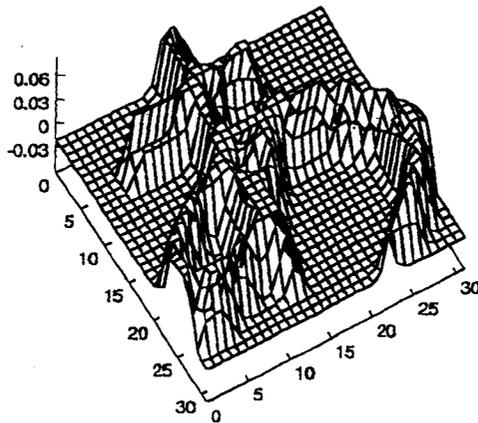


図 2: 特徴ベクトル

る。この特徴ベクトルを \mathbf{x} とすると、 $(\mathbf{x}, \mathbf{1}) = 0$ 、 $\|\mathbf{x}\|^2 = 1$ が成立している。但し、 $\mathbf{1}$ はすべての要素が 1 のベクトルを表す。図 1 の画像から求めた特徴ベクトルを 32×32 の画像として 3 次元表示したものを図 2 に示す。

部分空間法 [1] により認識を行なう場合、あらかじめ認識対象とする字種 (カテゴリー) ごとに訓練データを N 個用意し、上記 1024 次元の特徴ベクトルを求めておく。これらのベクトルから部分空間を表す正規直交基底を設計し、未知入力パターンのこの基底への射影成分を求めることで識別を行なう。通常、この基底を表す直交ベクトルとしては相関行列の固有ベクトルを用いる。すなわち、訓練データを $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ とおき、相関行列

$$R = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^t \quad (1)$$

を求め、 R の固有値を λ_j ($j = 1, 2, \dots, r$)、 λ_j に対応する固有ベクトルを ϕ_j とする。 r は識別に用いる部分空間の次元数であり、 n を特徴ベクトルの次元数 (= 1024) とすると、 $r < n$ である。固有ベクトルを用い、未知入力から求まる特徴ベクトル \mathbf{x} とすべての認識対象カテゴリーとの類似度を以下の式 (2) により計算し、最も値の大きいカテゴリーを認識結果とする。

$$S_s(\mathbf{x}) = \sum_{j=1}^r (\mathbf{x}, \phi_j)^2 \quad (2)$$

また、複合類似度法は式 (2) に固有値を用いた重み付けを行なった形の式で定義される。未知

入力から求まる特徴ベクトル \mathbf{x} の複合類似度 $S_m(\mathbf{x})$ は

$$S_m(\mathbf{x}) = \sum_{j=1}^r \frac{\lambda_j}{\lambda_1} (\mathbf{x}, \phi_j)^2 \quad (3)$$

で与えられる [3]. 部分空間法と同様に, 未知入力を与えられた場合, すべての認識対象カテゴリとの複合類似度を式 (3) により計算し, 最も値の大きいカテゴリを認識結果とする. $r=1$ のときは単純類似度法とよばれるが, このときの式 (3) は \mathbf{x} と ϕ_1 の正規化相互相関に他ならない.

ここで, 式 (2) と式 (3) に関する幾何学的意味を考える. まず, 式 (1) の R から決まる λ_j , ϕ_j は, カテゴリに属する特徴ベクトル (厳密には, 訓練データのベクトル) の集合を特徴付けるものである. すなわち, そのカテゴリの特徴ベクトルがある超楕円体状に分布しているとしたとき, 第 1 固有ベクトルは, その超楕円体状の分布領域が存在する方向を示しており, 第 2 以下の固有ベクトルは超楕円体の軸方向を表している. 各固有値の平方根はその軸の長さを表すものになっている. この分布領域に属するベクトルは, r 個の固有ベクトル ϕ_j を基底ベクトルとする r 次元の部分空間の元として表現できる. すなわち, ある係数の集合 $\{a_j\}$ を用いて

$$\hat{\mathbf{x}} = \sum_{j=1}^r a_j \phi_j \quad (4)$$

と表すことのできるベクトルは, その分布領域に属するとみなせる. 式 (2) と式 (3) の (\mathbf{x}, ϕ_j) は, この部分空間の基底ベクトルに関する \mathbf{x} の相対成分を与えており, $a_j = (\mathbf{x}, \phi_j)$ とおくと, 式 (4) で決まる $\hat{\mathbf{x}}$ は, \mathbf{x} の部分空間への射影, すなわち, 分布領域への射影ベクトルを与える. $\{\phi_j\}$ と \mathbf{x} とが同一カテゴリに属するなら, $\mathbf{x} = \hat{\mathbf{x}}$ が期待されている. 実際この等号が成立するとき, $S_s(\mathbf{x})$ は最大値 1 をとる. 以上から, $S_s(\mathbf{x})$ 及び $S_m(\mathbf{x})$ は未知ベクトル \mathbf{x} がカテゴリに対応する超楕円体状の分布領域に属する度合を与えており, これらの値で比較することは, 未知ベクトルがどのカテゴリの基底ベクトルの線形和としてもっとも近似できるかを判定しているといえる.

なお, 以下では, 訓練データから求まる固有値と固有ベクトルの組を辞書と呼ぶ. カテゴリの分布領域を考えると, 固有値, 固有ベクトルをそれぞれ, (平方) 軸長, 基底ベクトルと呼ぶ. また, 本論文では便宜上式 (2) および式 (3) の値のことを類似度と呼ぶ.

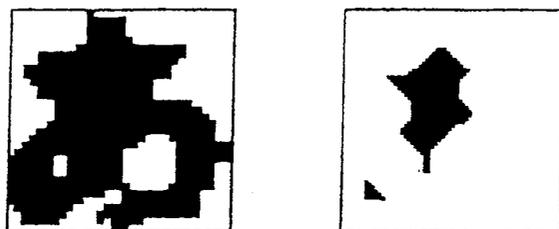
2.2 つぶれの検出と補正

文字画像のつぶれの検出には, 繰り返し回数を制限した細線化 [11] を基にした手法 [8] を用いる. 図 3(a) はつぶれのある文字画像を正規化したものであり, この画像より検出されたつぶれた領域を図 3(b) に示す.

次につぶれの補正法について述べる. まず, つぶれについて考える. 文字のつぶれは文字固有の性質ではなく, 外部からの影響で生じていると考えられる. 一方, 前節で, あるカテゴリの文字の特徴ベクトルはその分布領域に属し, 部分空間の基底ベクトルの線形和で表されることを述べた. すなわち,

$$\mathbf{x} \approx \sum_{j=1}^r a_j \phi_j \quad (5)$$

従って, 文字画像につぶれが生じて特徴ベクトル \mathbf{x} のある要素 x_k が変わってしまうのは式 (5) における基底ベクトル ϕ_j の対応する要素 ϕ_{jk} が変わってしまうためであると考えられる. 従



(a) Normalized image (b) Blurred area

図 3: つぶれの検出

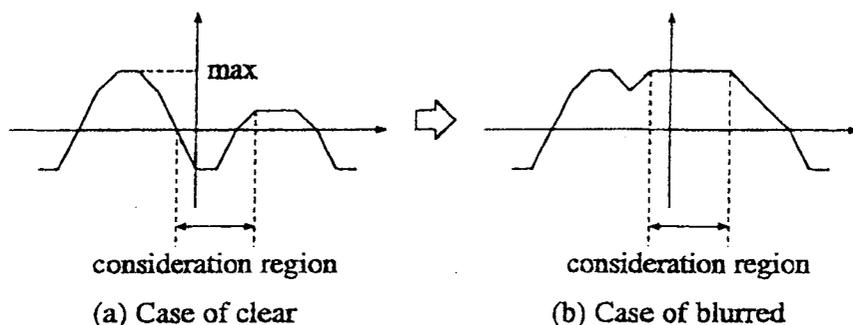
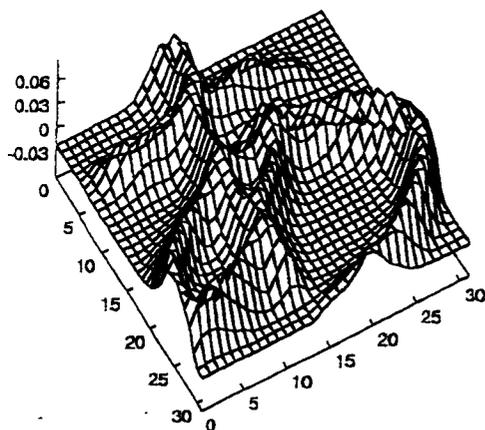


図 4: ベクトルの要素の変化

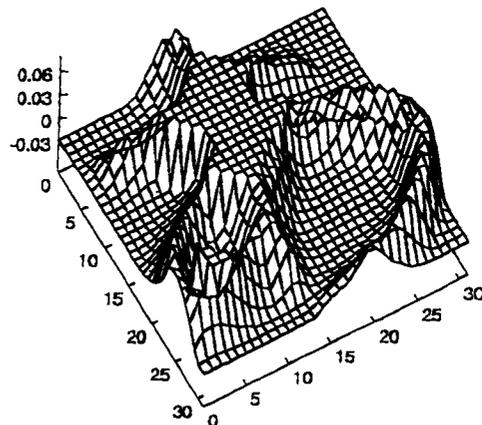
来手法をこの見方で説明すれば、つぶれの影響が係数 a_j の方にあるものとみなし、つぶれた画像の特徴ベクトルを準備してそれに対応する基底ベクトルを用意していたと考えられる。しかし、つぶれの偶発性を考えると、偶発的に基底ベクトルが変わったとみる見方に妥当性が感じられる。本論文はこの見方から、つぶれの補正法を考えている。

本節の最初に述べたように文字画像のつぶれた箇所は検出できる。その箇所に対応する基底ベクトルの要素の値をどう変化させるべきかを決める必要がある。入力画像のつぶれた領域はすべて黒画素になるが、これは、特徴ベクトルが濃度パターンとみなせることから、つぶれた領域に対応する要素の値がそのベクトルの要素の最大値に置き換わったものとみなせる。図 4 に、同一ベクトルのある区間の要素が、つぶれない場合からつぶれのある場合へ変化した例を示す。横軸が次元、縦軸がベクトルの要素の値を示す。図 4(a) がつぶれない画像から得られたベクトルであるとする、"consideration region" と示された要素に対応する領域につぶれが生じている場合、ベクトルは (b) のようにその領域における値が最大値になると考えられる。この見方を基底ベクトルの要素値の変化のさせ方に適用する。

図 5 にモデル図を示す。図に表すため 3 次元のみ表示してある。 e_1, e_2, e_3 が特徴ベクトルの元の座標軸であり、図中、網かけをした領域 R がつぶれない特徴ベクトルの分布領域であり、二つの固有ベクトル ϕ_1, ϕ_2 が基底ベクトルである。基底ベクトルは単位ベクトルであるが、図では、軸長に比例する大きさで表示している。ここでは、 e_1 に対応する領域につぶれが生じているものとし、 ϕ_1 の e_2 要素が最大であると仮定する。分布領域 R に存在する特徴ベクトルは、図 4 での考察から、 e_1 要素がその特徴ベクトルの要素の最大値と等しい値を持つように変化する。すなわち、特徴ベクトルの e_1 要素が e_2 要素と等しくなるため、領域 R は図 5 に示すような超平面 $e_1 = e_2$ 上の領域 R' に写像される。更に正規直交化することにより、単位超球上の領域 R'' に移される。こ



(a) Original dictionary



(b) Changed dictionary

図 6: 辞書の第 1 基底ベクトル

式 (10) より, ϕ_j は超平面 $e_1 = e_2$ 上にあることが分かる. この後, 式 (6), (7), (8) により正規直交化することになる.

分布領域の軸長の補正は次のようにする. 分布領域 R における軸長は固有値であり, 基底ベクトル (この場合は固有ベクトル) のノルムに関して基準化されているとみなせる. つぶれにより分布領域が R から R' に変化した場合, その基底ベクトルのノルムの変化に応じて軸長が変化するものとする. すなわち,

$$\lambda'_j = \lambda_j \frac{\|\phi_j''\|^2}{\|\phi_j\|^2} \quad (11)$$

ここで, 直交化後も基底ベクトルが表す軸方向が対応しているという近似を仮定した. 更に, 単位超球面への射影では, ϕ_j'' のノルムに比例する分の変化が全軸長におこるという近似を仮定することにした. すなわち, 最終的に補正された軸長を次のようにした.

$$\hat{\lambda}_j = \lambda'_j \frac{1}{\|\phi_j''\|^2} \quad (12)$$

図 6(a) を, 訓練データから求めた「あ」の辞書の第 1 基底ベクトル ϕ_1 とする. 図 3(a) の文字画像が入力された場合, 上述の手法により補正した後の第 1 基底ベクトル $\hat{\phi}_1$ を図 6(b) に示す. 図 3 のつぶれた領域を良く表現できていることが分かる.

2.3 低品質文字認識アルゴリズム

2.2 で提案した補正法を組み込んだ, 低品質文字の認識に適した実用的なアルゴリズムを提案する.

まず, 入力画像からつぶれた領域を検出する. そして, 式 (8) および式 (12) により $\hat{\phi}_j$, $\hat{\lambda}_j$ を求め, 式 (2) または式 (3) により類似度を計算する. これをすべてのカテゴリーについて行ない, 類似度が最大のカテゴリーを C_1 , その類似度を S_1 とおく. また, 類似度が 2 番目に大きいカテゴリーの類似度を S_2 とおく.

表 1: 訓練データの内訳

		明朝体	ゴシック体
標準辞書用	プリンタ	5, 6, 8, 10, 12, 14, 20, 25	6, 12, 25
	写植	6, 7, 8, 9, 10.5, 11, 12, 14, 20, 22	
追加用	プリンタ	8(f), 10(f), 12(f)	12(f)
	写植	6(c1), 7(c1), 8(c1), 6(c2), 7(c2), 8(c2)	

つぶれの補正を行なうことで誤認識となる場合にリジェクトを行なうことを考える。補正を行なうことは入力画像に合わせて辞書を変形することであるから、認識対象文字との辞書だけでなく類似文字の辞書との類似度も大きくなり、このときはカテゴリーごとの類似度の差が小さくなる。従って、 S_1 と S_2 の値を比較することでリジェクトの基準を設定することは妥当であると考えられる。ここでは、 $S_1/S_2 \geq \theta$ のときに C_1 を候補とし、それ以外のときにリジェクトする。但し、 θ は定数である。

3 実験

本手法の有効性を確認するため、認識実験を行なった。認識対象としたのは、JIS 第1水準の漢字 2965 字種と仮名 71 字種、計 3036 字種の印刷文字である¹。訓練データとしては 6 ポイントから 25 ポイントまでのマルチフォントの印刷文字および、これらをファクシミリにより転送したもの、複写機により複写したものを収集した。データの内訳を表 1 に示す。「プリンタ」はプリンタ出力により得られたもの、「写植」は写植文字を表す。論文誌等の認識を意図しており、明朝体中心でゴシック体を若干加えた形となっている。表中の数字はポイント数を表し、「(f)」はファクシミリにより転送したもの、「(c1)」は複写機の薄いモードにより複写したもの、「(c2)」は複写機の濃いモードにより複写したものであることを表す。訓練データから特徴ベクトルを求め、固有値・固有ベクトルを求めて辞書とするが、実験では、比較のため二種類の辞書を用意した。一つは「標準辞書」と呼び、表中で「標準辞書用」と表記したつぶれの少ないデータのみを用いて構成したものである。もう一つは「混合辞書」と呼び、表中の「標準辞書用」と「追加用」の両方のデータを用いて構成したものである。

評価用データとしては、低品質データと高品質データの二種類を用意した。低品質データとしては、訓練データとは別に出力した 6 ポイントの明朝体とゴシック体の 2 種類 (各 1 セットずつ) の印刷文字をファクシミリで転送したものをを用いた。また、高品質データとしては、訓練データとは別に出力した 12 ポイントの明朝体とゴシック体の 2 種類 (各 1 セットずつ) の印刷文字を用

¹データベース ETL9B [12] に含まれている 3036 字種と同じ字種である。

あ	い	う	え	お	か	が	き	ぎ	く
あ	い	う	え	お	か	が	き	ぎ	く

亜	唾	娃	阿	哀	愛	挨	始	逢	葵
亜	唾	娃	阿	哀	愛	挨	始	逢	葵

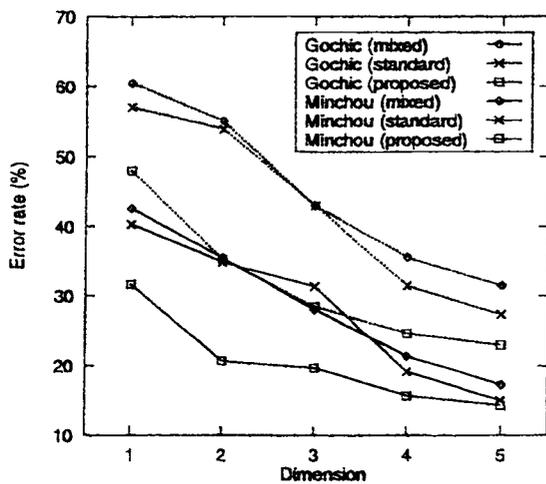
(a) 6 point Minchou

あ	い	う	え	お	か	が	き	ぎ	く
あ	い	う	え	お	か	が	き	ぎ	く

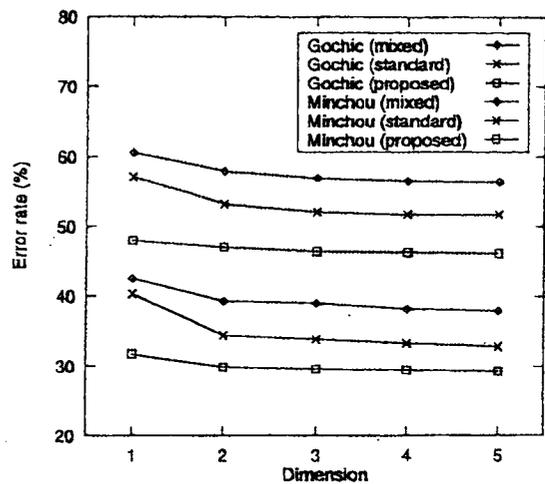
亜	唾	娃	阿	哀	愛	挨	始	逢	葵
亜	唾	娃	阿	哀	愛	挨	始	逢	葵

(b) 6 point Gothic

図 7: 性能評価用データの一部



(a) Subspace method



(b) Multiple similarity

図 8: 実験結果 (6 ポイント)

いた。低品質データの一部（仮名と漢字、それぞれ最初の10字種）を図7に示す。非常につぶれが多く品質の悪いデータであり、特に漢字は人間でも判断が難しいことがわかる。

3.1 つぶれの検出・補正法の評価実験

まず、2.2で提案したつぶれの検出による部分空間の補正法の有効性を確認するため、リジェクトを行わず常に候補を出力した場合の認識性能を調べた。

評価用データからつぶれを検出し、標準辞書に対して補正を行なった後の基底ベクトルを用いて部分空間法(式(2))および複合類似度法(式(3))により認識を行なう。比較のため、標準辞書または混合辞書をそのまま用い、評価用データを認識する実験も行なった。認識に使用する次

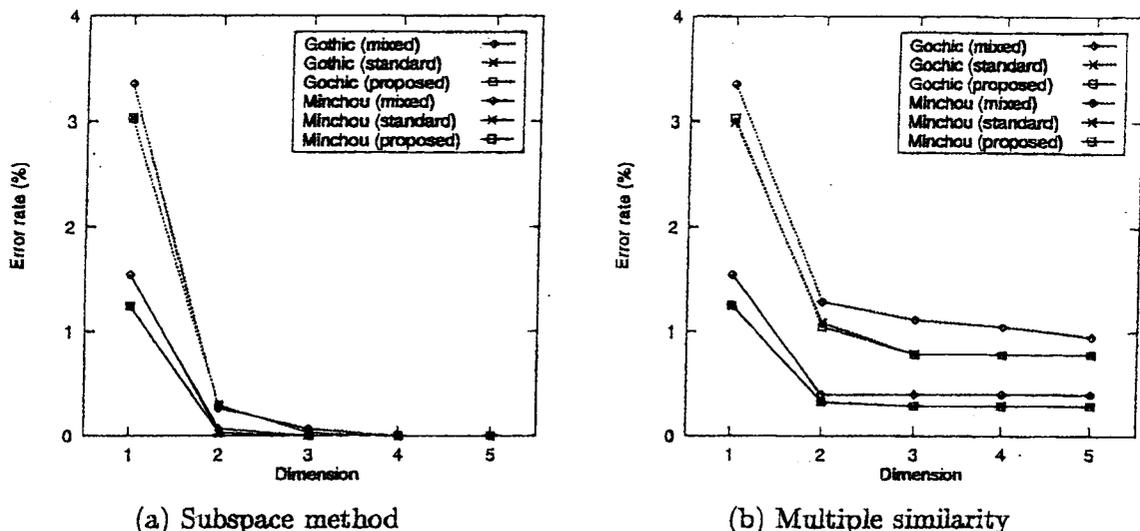


図 9: 実験結果 (12 ポイント)

元数 r は 1 から 5 まで変えた。6 ポイントの評価用データを用いた認識結果を図 8 に、12 ポイントの評価用データを用いた認識結果を図 9 に示す。いずれも (a) が部分空間法によるもの、(b) が複合類似度法によるものであり、横軸は用いた次元数、縦軸は誤り率を表す。また、“proposed” と表記したものは提案手法、“standard” は標準辞書を用いた手法、“mixed” は混合辞書を用いた手法である。図 8 より、低品質文字の認識においては部分空間法、複合類似度法とも、すべての次元において本手法で最も低い誤り率が得られていることが分かり、特に用いる次元数が小さい場合、認識性能改善の効果が大きく、つぶれによる補正の有効性が確認された。混合辞書を用いた手法で標準辞書のみによる手法よりも誤り率が高くなってしまったことは、つぶれ方が多種多様であり、辞書につぶれた画像を加えても入力画像がそれらと異なる劣化をしている場合にはむしろ悪影響が生じることを示すものである。また、高品質データを用いた実験 (図 9) では本手法と標準辞書を用いた手法でほぼ同様の認識率が得られていることが分かり、高品質データを用いた場合には認識性能の面で本手法による悪影響がないことが分かる。混合辞書を用いた場合は誤り率が増加しており、高品質データの認識には低品質データを用いることが悪影響を及ぼすことも確認された。

つぶれの補正の影響を詳しく分析するため、次元数 5 の部分空間法を用いた場合、6 ポイントゴシック体で、補正を加えない場合には誤りであったが補正を加えることによって正しく認識されるようになった例を表 2 に示す。認識対象となった文字画像と、つぶれによる補正を行なった場合の候補 (正解) の字種およびその字種の辞書との類似度、補正を行なわない場合の候補 (誤り) とその類似度が示されている。表より、補正なしの場合は、正解の文字と概形が同じでより線密度の高い複雑な字種が候補として得られていることがわかる。これは、つぶれによって本来白画素が配置されるべき領域も黒画素になってしまい、結果として黒画素が多く線密度の高い字種との類似度が大きくなったことが原因と考えられる。つぶれによる補正を行なうことで正しく認識され、また類似度の値そのものも大きくなっていることがわかる。

つぶれによる補正は、つぶれた領域にのみ作用し、つぶれない部分の特徴はそのまま生かすという特徴がある。すなわち、つぶれない領域の情報を最大限に利用し、高精度な認識を達成するものである。更に、つぶれの全くない文字が認識対象となった場合は全く補正を行なわない

表 2: 補正を加えることによって正しく認識された例

	(a)	(b)	(c)	(d)	(e)
画像	娃	始	扱	裕	庵
補正あり (類似度)	娃 (0.71)	始 (0.69)	扱 (0.59)	裕 (0.70)	庵 (0.62)
補正なし (類似度)	娘 (0.61)	鎗 (0.59)	緩 (0.57)	輪 (0.62)	塵 (0.60)

表 3: 処理時間および使用メモリ量 (部分空間法)

次元数		1	2	3	4	5	
処理時間 (秒/文字)	6 ポイント	類似度計算	0.09	0.18	0.25	0.34	0.41
		辞書変換	0.70	1.35	2.26	3.71	5.34
		合計	0.79	1.53	2.51	4.05	5.75
	12 ポイント	類似度計算	0.09	0.17	0.25	0.34	0.41
		辞書変換	0.26	0.51	0.86	1.39	2.02
		合計	0.35	0.68	1.11	1.73	2.43
使用メモリ量 (Mbyte)		11.9	23.7	35.6	47.4	59.3	

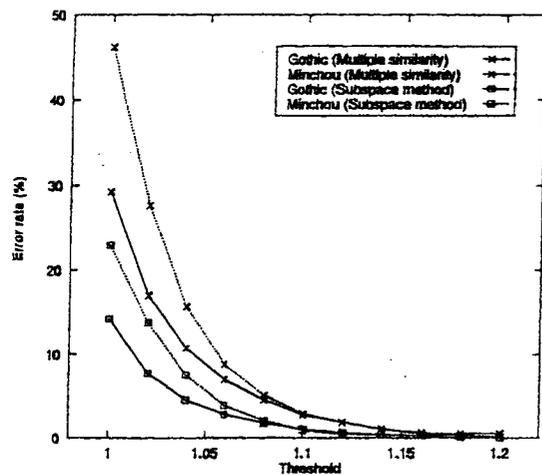
ため、綺麗な文字画像に対する悪影響は皆無である。

但し、逆に補正を加えない場合には正解であったが補正を加えることによって誤りに転じた文字も存在した。これは、つぶれによる補正を加えた場合は、どのような辞書も未知入力画像に合わせて変更されるため、類似文字の辞書との類似度が大きくなって区別できなくなったことが原因であると考えられる。

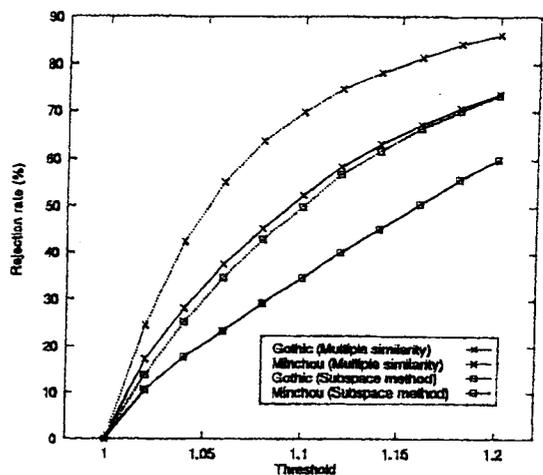
表 3 に、部分空間法を用いた場合の処理時間と辞書格納に要するメモリ量を示す。複合類似度法についても部分空間法と同じ結果が得られた。表中、6 ポイントとあるのは 6 ポイント明朝体と 6 ポイントゴシック体の平均、12 ポイントとあるのは 12 ポイント明朝体と 12 ポイントゴシック体の平均である。「類似度計算」は式 (2) の計算に要する時間であり、従来法の処理時間に等しい。「辞書変換」とはつぶれの検出と基底ベクトルの変換に要する時間の合計であるが、つぶれの検出は基底ベクトルの変換に要する時間と比べて無視できるほど小さく、0.001 秒/文字程度であった。6 ポイントのデータは 6072 文字中 6068 文字でつぶれが検出されており、表中の辞書変換の時間はつぶれが検出された場合に辞書変換に要する時間にほぼ等しい。また、本手法の使用メモリ量は従来法と同じであり、次元数に比例した値となる。

3.2 総合評価

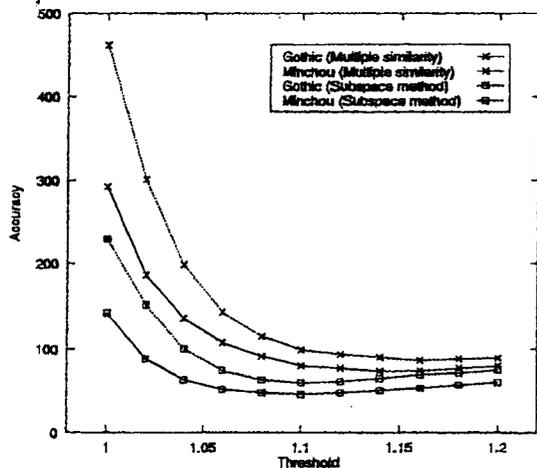
2.3 で提案した文字認識アルゴリズムの有効性を確認するため、認識実験を行なった。評価用データとしては、6 ポイントの低品質データのみを用いた。認識結果を図 10 に示す。図の横軸はすべて 2.3 で定義したしきい値 θ の値であり、(a) は誤り率 E 、(b) はリジェクト率 R 、(c) は認識精度を表す。認識精度 S は $S = 10E + R$ で定義され [13]、値が小さいほど良い認識系であると



(a) Error rate



(b) Rejection rate



(c) Recognition accuracy

図 10: 実験結果 (リジェクトあり)

言える。

しきい値 θ を大きくすることで誤り率は小さくなり、リジェクト率は大きくなる。図 10(c) より、認識精度は部分空間法では $\theta = 1.10$ 、複合類似度法では $\theta = 1.16$ 程度するとき最小となることがわかる。部分空間法を用いた場合、 $\theta = 1.10$ とするとリジェクト率は 6 ポイント明朝体で 35%、6 ポイントゴシック体で 50%と高いが、誤り率はそれぞれ 1.05%、0.96%と非常に低い値となった²。図 7 に示されるような人間でも判断の難しい低品質文字画像でありながら、本手法を用いることで認識できない文字を確実にリジェクトし³、認識できる文字に対しては誤り率 1%程度の高精度で候補を選出することが可能であることが示された。

²類似度の大きい順に候補を 3 個選出すれば、誤り率 (3 個とも誤りである率) はいずれも 0.4%となった。

³例えば図 7(b) の「重」は本手法によりリジェクトされた。

4 むすび

本論文では、つぶれのある低品質文字を高精度に認識することを目的とし、つぶれを検出・補正する認識手法を提案した。文字画像からつぶれた領域を検出し、つぶれが生じた場合にデータの分布がどのように変化するかを考察し、つぶれに応じて部分空間法や複合類似度法で用いられる基底ベクトル・軸長を補正することで特徴ベクトルの分布領域を近似する新たな正規直交基底および軸長を設計する手法を提案した。更に、低品質文字の認識に適したアルゴリズムを提案し、ファクシミリで転送して劣化した文字を用いた認識実験を行ない、手法の有効性を確認した。

従来の認識手法は、主に訓練データのみから特徴量の分布形状を推定し、認識を行なうものであった。しかし、複写機による複写やファクシミリで転送した文字画像を扱う場合、雑音や量子化誤差により生じるつぶれ等の劣化は大きく、しかもこれらは偶発的に発生するものであり、訓練データを集めて分布形状を推定しただけでは対処しきれないと思われる。本手法は、未知入力画像から求まるつぶれに応じて部分空間法や複合類似度法で用いられる基底ベクトル・軸長を動的に補正するものであり、従来の統計的な手法の欠点を補うものである。しかも、つぶれの全くない文字が認識対象となった場合は補正を行なわないため、綺麗な文字画像に対する悪影響は皆無であるという特徴がある。

本手法では、低品質の文字が入力された場合には基底ベクトルの変換を必要とするために処理時間が増加する。しかし、メモリ使用量は従来法と変わらず、また、つぶれの全くない高品質の文字が入力された場合にはつぶれの検出に要するわずかの時間が増加するのみで処理時間・認識率に対する影響がほとんどない。実際の認識システムを構築する場合、メモリ量・処理時間の制限のもとで最大精度を得るためには、従来法と本手法を組み合わせる必要もあるが、その検討は今後の課題である。本手法を自然言語等の知識を用いた後処理と組み合わせてより実用的な認識システムを構築することも今後の課題である。

謝辞

本研究に対し、石田(實)記念財団から御援助を頂きました。ここに深く感謝致します。

参考文献

- [1] E. Oja, "Subspace methods of pattern recognition," Research Studies Press, 1983.
- [2] 飯島泰蔵, "パターン認識理論," 森北出版, 1989.
- [3] 橋本新一郎, "文字認識概論," 電気通信協会, 1982.
- [4] 小佐位 潤, 星野雄一, 岡本正義, 加藤邦人, 山本和彦, "低解像度画像からの文字認識手法について," 信学技報, PRMU97-221, Feb. 1998.
- [5] 澤 和弘, 若林哲史, 鶴岡信治, 木村文隆, 三宅康二, "FAにおける低品質文字列の認識," 信学技報, PRMU97-214, Jan. 1998.
- [6] 澤木美奈子, 村瀬 洋, 萩田紀博, "変形推定に基づいた参照パターンの自動選択による文字認識," 信学技報, PRMU97-187, Dec. 1997.

- [7] 矢田勝啓, 鶴岡信治, 木村文隆, 三宅康二, “加重方向指数ヒストグラム法をつぶれ文字への対応,” 信学技報, no.PRU90-128, pp.21-26, Feb. 1991.
- [8] 大町真一郎, 阿曾弘具, “つぶれを考慮した低品質印刷文字の高精度認識,” 信学論 (D-II), vol.J79-D-II, no.9, pp.1534-1542, Sept. 1996.
- [9] 大町真一郎, 阿曾弘具, “品質を考慮した印刷文字認識手法,” 情処学論, vol.38, no.12, pp.2490-2498, Dec. 1997.
- [10] H. Mori, H. Aso, and S. Makino, “Robust n -gram model of Japanese character and its application to document recognition,” IEICE Trans. Inf. & Syst., vol.E79-D, no.5, pp.471-476, May 1996.
- [11] 阿曾弘具, “並列処理向き細線化アルゴリズム,” 信学論 (D-II), vol.J76-D-II, no.9, pp.2148-2150, Sept. 1993.
- [12] 斉藤泰一, 山田博三, 山本和彦, “JIS 第 1 水準手書漢字データベース ETL9 とその解析,” 信学論 (D), vol.J68-D, no.4, pp.757-764, April 1985.
- [13] 川又文男, 若原 徹, 松井敏弘, 能見 正, 山下郁生, 堤田敏夫, “手書き漢字宛名認識に関する第 3 回文字認識技術コンテストの実施,” 信学'94 秋大, 情報・システム, D-321, Sept. 1994.