

# 自己相似性特徴を用いたオフライン手書き文字認識の提案

篠 沢 佳 久<sup>†</sup>

本論文においては文字の局所的な構造を解析し、認識に役立てる手法として、文字画像の持つ自己相似性という性質に焦点を当てる。文字画像の中にはその一部が他の一部と相似関係になっているものもあり、これは書き手が異なっても、構造上の制約になっており、その関係もしくは制約を文字固有の特徴としてとらえることができる場合もある。そこで図形の形状や構造上の特徴において自己相似性特徴を表現するのに有効なフラクタルの技術を用いて文字画像の局所的な領域間の相似関係の特徴として表現し、認識に利用できないかと考えた。本論文においては上述した文字画像内の自己相似性特徴をフラクタル幾何学におけるレンジブロックとドメインブロックの相似関係を利用することによって抽出する手法を提案し、文字認識において有効な特徴となりうるかどうかを調べる。そしてこの自己相似性特徴を利用した認識手法を提案し、産業技術総合研究所提供の ETL8B を用いた認識実験を通してその可能性について検証する。

## The Offline Handwritten Character Recognition by Self-similarities

YOSHIHISA SHINOZAWA<sup>†</sup>

We focus on the self-similarities of the character patterns. A character pattern has the self-similarities that a part of it, such as a stroke or a line, is similar to the other. The self-similarities may represent structural features and restrictions of a character pattern. We use the self-similarities of character patterns as useful features in recognizing them. The fractal geometry is useful technique that can extract the self-similarities from the images. We propose how to extract the self-similarities from offline handwritten character patterns by the fractal geometry and recognize character patterns by using the self-similarities. We examine effectiveness of our approach throughout the experiments with ETL8B.

### 1. ま え が き

手書き文字認識の研究は古くから行われており、今日に至るまでさまざまな認識方法が考案されている。字体が書き手によって異なり、筆順情報を利用できないオフライン手書き文字認識の研究はパターン認識の分野において最も困難な課題の1つである。日本で使用されている文字、特に漢字は偏や旁などの組合せにより構成されているため、文字画像から1画1画(これをストロークと呼ぶ)を取り出し、ストロークどうしのつながり、その構成や位置情報などストロークによる構造上の特徴を抽出し、それらを認識に役立てようとする研究(ストローク構造解析法)がさかんに行われてきた。特にオンライン文字認識においては、ストロークや筆順情報がオフライン文字認識と比較して容易に抽出できることからストローク構造解析法が主流となっている<sup>1)</sup>。しかし一方でオフライン文字認識

においては文字画像からストロークを1画1画厳密に抽出することは困難であるため、文字画像全体から文字固有の特徴を抽出し、その特徴を認識に利用する研究(パターン整合法)が主流となっている<sup>2)</sup>。

本論文では文字の局所的な構造を解析し、認識に役立てる手法として、文字画像中における自己相似性という性質に焦点を当てる。たとえば「米」という文字は「一」という線をさまざまな角度で回転させ、それらを組み合わせることによって構成されている。あるいは「回」という文字は大きさの異なる2つの「口」という文字から構成されている。このように文字の中には、その文字画像中のある一部分を組み合わせることによって構成されているものもある。また文字を学習していく過程においては、次の1画を書く際、その1画と他の部分との相対的な位置、形状を考えながら、文字を覚えていくという一面もある。したがって上記の例のように、文字の中にはその一部が他の一部と相似関係になっており、その関係がその文字固有の特徴として利用できるものもある。すなわち文字を構成する一部分や領域、局所的な線(ストロークとは限らな

<sup>†</sup> 慶應義塾大学理工学部管理工学科

Faculty of Science and Technology, Keio University

い)は拡大,回転などの操作(アフィン変換)を行うことによって他の一部分と一致する場合もある。複雑な漢字でも局所的な線に限定すれば,線と線との間に相似性を見つけだすこともでき,その相似関係(これを自己相似性と呼ぶ)は文字固有の特徴である場合もある。本論文においてはこの文字の自己相似性という特徴に焦点を当て,文字画像から自己相似性特徴を抽出する方法およびこれを認識に利用する方法を提案する。

このように画像内の自己相似性に焦点を当てた既存の技術としてフラクタル幾何学があげられる。フラクタル幾何学において,画像内の微小部分が全体と相似関係にある画像のことをフラクタル画像と呼び, Mandelbrot によって名付けられた。この技術の応用としては画像圧縮の分野で利用されているフラクタル画像圧縮があげられる<sup>5)</sup>。フラクタル画像圧縮とは,画像内に含まれている自己相似性を利用した画像符号化方式の総称で,画像の構造的な特徴を利用する新しいタイプの符号化方式であり,性能に関しても従来の手法以上の圧縮効率が得られている。

近年このフラクタルの技術を画像認識に利用する試みが行われている<sup>6)~11)</sup>。まず形状の複雑さ,形状的な特徴を表す指標としてフラクタル次元を導入し,分類や認識に利用している研究があげられる<sup>6)~8)</sup>。これらの研究においては画像中における自己相似性をフラクタル次元によって定量的に数値化し,対象画像の特徴として認識に利用している。またフラクタル幾何学の性質(フラクタル不変定理)を用いて,画像中から自己相似性特徴を抽出した後に,これを利用し,画像の復元処理をすることによって認識を行う方法が提案されている<sup>9)</sup>。この手法は前述したフラクタル画像圧縮と非常に似通った手法であり,画像内の局所的な自己相似性を基礎に認識を行うという方法である。そして画像の構造的な類似性特徴をフラクタル画像圧縮によって得られた符号情報を用いて表現し,画像分類,検索に利用することも提案されている<sup>10)~11)</sup>。特に画像認識においてフラクタル幾何学を利用した場合の利点として,縮小拡大や回転に強いことがあげられている<sup>9)</sup>。

以上のように,文字画像中には局所的に相似関係にある領域間もあり,これは書き手が異なっても,構造上の制約になっており,文字固有の特徴としてとらえることができる場合もある。そのため図形の形状や構造上の特徴において自己相似性特徴を表現するのに有効なフラクタルの技術を用いて文字画像の局所的な領域間の相似関係を特徴として抽出し,認識に利用できないかと考えた。そこで,本論文においては文字画像

から自己相似性特徴をフラクタル幾何学の局所的な領域間の相似関係を利用することによって抽出する手法を提案し,文字画像の認識において有効な特徴となりうるかどうかを調べる。そしてこの特徴を利用した認識手法を提案し,産業技術総合研究所提供の ETL8B を用いた認識実験を通してその可能性について検証する。

## 2. フラクタル幾何学

### 2.1 フラクタル変換

フラクタル幾何学においては,画像全体がそれ自身の局所的な一部の画像と相似関係にあるという考えから成り立っており,その相似関係を反復関数系(IFS: Iterated Function System)を用いて記述する<sup>3)~5)</sup>。しかし画像全体の縮小コピーを画像内部に見つけるのは困難なことであり,そこで Jacquin は画像を互いに重なり合わない領域に分割し,各領域において局所的な IFS を見つけるという分割反復関数系(PIFS: Partitioned Iterated Function System)を提案した。また Barnsley は PIFS をフラクタル変換とも呼んでいるため,本論文でも PIFS をフラクタル変換と呼ぶ。

フラクタル幾何学においては図1に示すようにまず画素数  $X \times Y$  の入力画像  $I$  を互いに重ならないように1辺の大きさが  $r$  の正方形  $R_i (i = 0, 1, \dots, M)$  に  $M$  個に分割する ( $M = X/r \times Y/r$ )。この正方形  $R_i$  をレンジブロックと呼ぶ。次にそれぞれのレンジブロックと最も相似関係にある図形を入力画像  $I$  の中から検索する。すなわちレンジブロック1つ1つに対して  $P$  倍の拡大および  $\theta$  度の回転(アフィン変換)を施した図形  $R'_i$  を考える。そしてこの図形  $R'_i$  と最も類似している(最も距離の短い)ブロック  $D_i$  を入力画像  $I$  の中から探し出す。このブロック  $D_i$  をドメインブロックと呼ぶ。図1の場合,レンジブロック

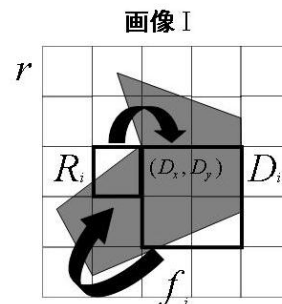


図1 フラクタル変換

Fig. 1 Fractal transformation.

$R_i$  を 2 倍に拡大し、時計方向に 180 度の回転させた図形が、ドメインブロック  $D_i$  とほぼ一致している。したがってドメインブロック  $D_i$  から任意のレンジブロック  $R_i$  に対する変換  $f_i$

- 拡大 (縮小) 率  $P'$
- 回転角  $\theta'$
- ドメインブロックの左上の座標  $(D_x, D_y)$

をそれぞれ  $M$  個求める。通常、拡大率  $P'$  は一意に固定し、回転角  $\theta'$  は 45 度ごとの回転および対称性のみを考慮するといったように制限を設ける。したがって、見つかったドメインブロックからレンジブロックへの変換  $f_i$  は対応するブロックの位置座標を含んだアフィン変換によって記述される。

$$R_i = f_i(D_i) \quad (1)$$

このアフィン変換の集合をフラクタル変換  $F = (f_1, f_2, \dots, f_M)$  と呼ぶ。本論文では任意の画像から抽出したフラクタル変換を自己相似性特徴として利用する。すなわち抽出されたこの自己相似性特徴は、その画像を構成する局所的な領域間の相似性を表現すると同時に、構造上における制約ではないかと考え、これをその画像固有の特徴として利用する。

そしてフラクタル変換を適用する場合、入力画像  $I$  から上記の方法で求めたアフィン変換の集合  $F$  を直接用いる。画像  $G$  (画素数は  $X \times Y$ ) へのフラクタル変換の適用は以下のように記述される。

$$F(G) = \bigcup_{i=1}^M f_i(D_i) \quad (2)$$

すなわち画像  $G$  においても 1 辺が  $r$  の大きさの領域に重ならないように分割し、フラクタル変換中に記述されているパラメータを用いて、指定されているドメインブロックから  $M$  個のレンジブロックを復元し、合成すると 1 つの画像を得る。またフラクタル変換を繰り返し適用することを、 $F^n = F(F^{n-1}(G))$  と記述する。そして  $n \rightarrow \infty$  としたとき、

$$A = \lim_{n \rightarrow \infty} F^n(G) \quad (3)$$

というように  $A$  に収束する。この  $A$  をフラクタル変換  $F$  のアトラクタと呼ぶ。

## 2.2 フラクタル不変定理

フラクタル画像圧縮は上述したフラクタル変換の応用例の 1 つである。このフラクタル変換をパターン認識に応用するために Neil らのフラクタル不変定理 (Fractal Invariance Theorem) を利用する<sup>9)</sup>。フラクタル不変定理とは「フラクタル変換  $F$  が  $A$  をアトラクタとするような変換であるとき、 $F(A) = A$

は成立する。かつこれはフラクタル変換  $F$  と  $A$  のみに成り立つ性質である」というものである。すなわち任意の画像  $A$  から抽出したフラクタル変換  $F_A$  を用いて、その画像自身に変換を行うと元の画像  $A$  に復元される。

$$F_A(A) = A \quad (4)$$

一方別の画像  $B$  に対してこの変換  $F_A$  を用いてフラクタル変換を行っても元の画像に復元することはいけないという性質である。

$$F_A(B) \neq B \quad (5)$$

また Neil らはこの定理を拡張し、類似画像に対してもこの性質が成立することを示している。

$$A' \approx A \rightarrow F_A(A') \approx A' \quad (6)$$

この性質を文字画像の認識において利用する。

## 3. 文字認識への応用

### 3.1 文字認識への応用

2 章で述べたフラクタル幾何学における局所的な領域間の相似関係、すなわちレンジブロックとドメインブロックの関係を利用したオフライン文字画像からの自己相似性特徴の抽出および認識方法は以下のとおりである<sup>15),16)</sup>。

#### (1) 自己相似性特徴の抽出

文字種  $i$  ( $i = 1, 2, \dots, N$ ) の画像から自己相似性特徴を抽出する。画素数が  $X \times Y$  の二値画像に対して、1 辺の大きさが  $r$  の正方形の領域に重ならないように区切る。すなわちレンジブロックの大きさを  $r \times r$  と固定する。各レンジブロックに対してアフィン変換のパラメータ (回転角, 拡大率) に制約を設けたうえで、最も類似しているドメインブロックを検索し、フラクタル変換 (以後、自己相似性特徴と呼ぶ)  $F_i$  を抽出する (図 2)。特徴数は  $M = (X/r) \times (Y/r)$  個である。レンジブロックとドメインブロックの類似性についてはユークリッド距離を用いて求める。認識対象 ( $N$  個) となるすべての文字画像についてそれぞれ自己相似性特徴  $F_i$  を求め、表 1 のようにデータベースに保存する。

#### (2) 認識対象となる文字画像の復元

未知の文字画像  $G$  (画素数は  $X \times Y$ ) を認識する場合、データベースに保存してある自己相似性特徴  $F_i$  ( $i = 1, 2, \dots, N$ ) を用いて式 (2) によるフラクタル変換を行って復元し、画像  $G_i = F_i(G)$  を得る。すなわち対応するドメインブロック  $D_i$  から自己相似性特徴中にパラメータ化されている値 (表 1) を用いて式 (1) のアフィン変換によってそれぞれのレンジブロッ

ク  $R_i$  を復元する．そして  $M$  個のレンジブロックを復元した後で，1つの画像として合成する．復元は文字種ごとに  $N$  個の自己相似性特徴を用いて別々に行い， $N$  個の復元画像を得る（図3）．

(3) 文字画像の認識

未知の文字画像  $G$  と  $N$  個の変換後の画像  $G_i$  を比較する．すなわちフラクタル不変定理より任意の文字画像に対して同字種のフラクタル変換を適用した場合は，その変形は少なく，一方で異なった文字種のフラクタル変換を適用した場合は，変形が大きくなるはずである．したがって画像  $G$  が文字種  $k$  に属する場合， $F_k(G) \approx G$  が成り立つが，画像  $G$  が文字種  $k$  に属さない場合  $F_k(G) \neq G$  となる．そこで式 (7) より文字画像  $G$  と変換後の画像  $G_i$  との距離  $d_i$  を求め（図3），最も変形の少ない（距離の小さい）画像  $G_k$

を検索し，文字種  $k$  を認識結果とする．

$$d_i = \|G - G_i\| \tag{7}$$

もしくはオフライン手書き文字認識で利用されている一般的な特徴を変形後の画像  $G_i$  から抽出し，認識を行う．

オフライン手書き文字画像から正確にストロークを切り出し，ストローク間での相似性を求めることは困難であるため，PIFSの求め方に従って画像領域を一定区間ごとに区切り，それをレンジブロックとした．またフラクタル不変定理という性質は，その画像もしくは類似した画像のみにあてはまる性質である．そのため活字文字のように比較的変形の少ない文字であれば，同字種の文字からその文字固有の自己相似性特徴を抽出することができ，この性質は十分にあてはまり，認識も比較的容易であると考えられる．しかし手書き文字のように書き手によって変形が激しい文字においては，同字種の文字からもその文字固有の自己相似性特徴を抽出することは難しいため，この性質はあてはまらない．そこで認識率の向上のためにいくつかの改良が必要である．

3.2 問題点とその改良

オフライン手書き文字画像に3.1節で述べた認識方法を適用した場合，以下のような問題点が生じる<sup>17)</sup>．

(1) 同字種内での特徴抽出における問題

手書き文字は書き手によって異なるため，書き手が異なった場合，ずれや変形などのために，同字種においてもある特定のレンジブロックに対して異なったドメインブロックが対応づけられてしまう（アフィン変換のパラメータに差が生じてしまう）．すなわち図4に示すように，同じ文字種においても特定のレンジブロックと対応づけられるドメインブロックが一致しないため，任意の文字から自己相似性特徴を一意に特定することが困難であり，特徴として一意に定めることが難しい．

(2) 文字固有の特徴の抽出としての問題

一方で異なった字種の文字において，あるレンジブロックと対応するドメインブロックが一致してしまうこともある（図5）．この場合，画像復元時において，正解でない文字の自己相似性特徴を利用して也正しくド

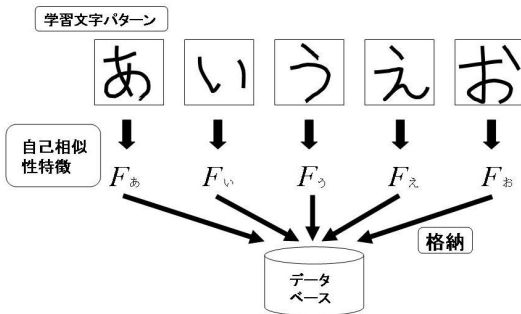


図2 自己相似性特徴の抽出

Fig. 2 Extraction of the self-similarities from character patterns.

表1 自己相似性特徴

Table 1 Elements in the self-similarities.

	$R_i$ の座標	$D_i$ の座標	拡大率	回転角(度)
$f_1$	(0, 0)	(3, 43)	0.5	45
$f_2$	(4, 0)	(24, 31)	0.5	180
⋮	⋮	⋮	⋮	⋮
$f_M$	(72, 72)	(11, 65)	0.5	90

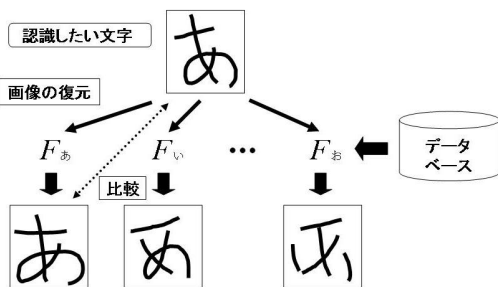


図3 画像の復元と認識

Fig. 3 Recognition by using the self-similarities.

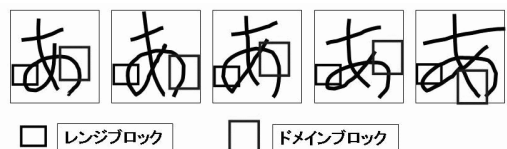


図4 自己相似性抽出時の問題その1

Fig. 4 The trouble with extracting the self-similarities.

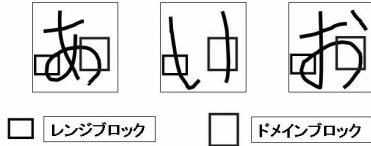


図 5 自己相似性抽出時の問題その 2

Fig. 5 The other trouble with extracting the self-similarities.

メインブロックからレンジブロックに復元されてしまう。すなわちこのような場合、その文字固有の特徴として利用することが難しい。

### (3) 画像復元の際の問題

画像を復元する際、認識対象の文字画像と自己相似性特徴を抽出するために利用した文字画像（学習文字パターン）の間にも問題点(1)と同様に、同字種であってもずれや変形がある。そのため正解である自己相似性特徴を利用して認識対象の文字画像において正確にドメインブロックからレンジブロックを復元することが難しい。

上述した問題点(1)(2)は自己相似性特徴の抽出時におけるものであり、問題点(3)は認識の際の復元時におけるものである。これらの問題はレンジブロックを固定しているため、書き手によって、そのレンジブロックの領域に書かれる文字の構成部品にずれや変形が生じてしまうからである。これらの問題点について下記の改善方法によって改良を試みる。

#### (1) 自己相似性特徴の抽出方法の改善

問題点(1)(2)を解決するためには、任意のレンジブロックに対応するドメインブロックを抽出する場合、同字種とは相似性が高く、異なった文字種とは相似性が低いブロックをドメインブロックとして一意に抽出する必要がある。そこで問題点(1)については、同字種において複数個( $P$ 個)の文字画像を利用し、同字種中の画像内における各領域間の相似性の傾向を調べる。文字種  $i$  において任意のレンジブロック  $R_{ij}$  と対応するドメインブロックを決定する際、レンジブロック  $R_{ij}$  にアフィン変換を施した図形  $R'_{ij}$  と画像中の座標  $(x, y)$  から  $R'_{ij}$  と同じ大きさのブロック  $B_{i,(x,y)}$  を切り出し、式(8)より2つの図形間の距離  $d_{ij,(x,y)}$  を求め、最小距離となる図形をドメインブロック  $D_{ij}$  として決定する。距離は全座標  $(1 \leq x \leq X - r, 1 \leq y \leq Y - r)$  において求める必要がある。

$$d_{ij,(x,y)} = \|R'_{ij} - B_{i,(x,y)}\| \quad (8)$$

手書き文字のように変形の激しいパターンでは、1字

種を1パターンで代表させることは困難なため、学習用の文字として  $P$  個利用し、統計的に相似性の傾向を調べる。すなわちレンジブロック  $R_{ij}$  にアフィン変換を施した図形  $R'_{ij}$  と切り出したブロック  $B_{i,(x,y)}$  との距離を求める場合、 $P$  個の学習用の文字パターンを用いてそれぞれ同様に距離  $d_{ijk,(x,y)}$  ( $k = 1, 2, \dots, P$ ) を求めた後、合計する(式(9))。

$$\begin{aligned} d'_{ij,(x,y)} &= \sum_{k=1}^P d_{ijk,(x,y)} \\ &= \sum_{k=1}^P \|R'_{ijk} - B_{ik,(x,y)}\| \quad (9) \end{aligned}$$

したがって距離の合計  $d'_{ij,(x,y)}$  を最小にするブロックが文字種  $i$  のレンジブロック  $R_{ij}$  と最も相似性の高い図形であり、0に近いほど、文字の構造上において制約が強いと考えられるので、そのブロックをドメインブロックとして一意に決定する。

次に問題点(2)については、レンジブロック  $R_{ij}$  と相似性のあるブロックを検索する際、他の文字種  $l$  ( $l \neq i$ ) においてはその変換によって復元されないものとしなければならない。そこでこの場合、文字種  $l$  においても式(9)より距離の合計  $d'_{lj,(x,y)}$  を求める。そして文字種  $i$  のレンジブロック  $R_{ij}$  と対応するドメインブロックを決める際、文字種  $l$  においては相似性が低いと考えられるブロック、すなわち距離の合計値  $d'_{lj,(x,y)}$  が大きいブロックを選ぶようにする。

まとめるとレンジブロックに対応するドメインブロックを決める際には、他の文字種とは相似性が低く、対象の文字種とは相似性の高いブロックをドメインブロックとして一意に決定する。そこで式(10)を最小にするブロックをレンジブロックとする。

$$\begin{aligned} d''_{ij,(x,y)} &= \alpha \times N \times \sum_{k=1}^P \|R'_{ijk} - B_{ik,(x,y)}\| \\ &\quad - \sum_{l=1, l \neq i}^N \sum_{k=1}^P \|R'_{ljk} - B_{lk,(x,y)}\| \quad (10) \end{aligned}$$

式(10)中の  $\alpha$  は定数であり、第1項目は式(9)と同様に文字種  $i$  のレンジブロック  $R_{ij}$  と相似性の高いブロックを選択することを意味する。第2項目は他の文字種  $l$  とは相似性の低いブロックを選択することを意味する。すなわち第2項目によって文字種  $i$  以外のすべての字種についての距離を求め合計し、最も距離の大きい(相似性が低いと考えられる)ブロックを選び出す。そしてこれら2つの式をまたまた式(10)を用いて、レンジブロックに対応するドメインブロックを一意に抽出し、これを自己相似性特徴として利用

する。

## (2) 画像復元方法の改善

未知の文字画像を復元する場合、自己相似性特徴を抽出するために利用した文字画像の間には変形やずれといった雑音があることも考慮しなければならない。そこで、各レンジブロックをドメインブロックから復元する場合、ずれを考慮し、自己相似性特徴中で指定されたドメインブロックの左上の座標(表1)を中心に、その近傍付近に復元すべきブロックが存在するものとする。そのためその周辺の領域の中で、最もレンジブロックと類似しているブロックをドメインブロックとして復元を行う。

### 3.3 自己相似性特徴の抽出と認識

改善点(1),(2)を考慮し、3.1節で述べた文字画像からの自己相似性特徴の抽出と認識手法を改良した提案手法を下記に示す。

#### (1) 自己相似性特徴の抽出

文字種  $i (i = 1 \dots N)$ 、画素数が  $X \times Y$  の二値画像においてレンジブロックの大きさを  $r \times r$  と固定する。任意のレンジブロック  $R_{ij}$  に対応するドメインブロックを求める場合、式(10)を用いて、全ブロックとの距離を求め、最小となるものをドメインブロックとする。抽出される自己相似性特徴  $F_i$  の個数は  $M = (X/r) \times (Y/r)$  個である。認識対象 ( $N$  個) となるすべての文字画像についてそれぞれ求め、データベースに保存する(図2)。

#### (2) 認識対象となる文字画像の復元

未知の文字画像  $G$  を  $N$  個の自己相似性特徴  $F_i$  を利用して別々に復元し(図3)、 $N$  個の画像  $G_i = F_i(G)$  を得る。復元する際には自己相似性特徴のパラメータ中に、復元先のドメインブロックの座標  $(D_x, D_y)$  は一意に指定されているが、その近傍  $((D_x - c, D_y - c), \dots, (D_x + c, D_y + c))$  を含めて検索し、復元すべきレンジブロックと最も相似性の高いブロックをドメインブロックとして復元する。

#### (3) 文字画像の認識

未知の文字画像  $G$  と  $N$  個の変換後の画像  $G_i$  との距離を求め、最も距離の短い  $G_k$  を検索し、文字種  $k$  を認識結果とする(図3)。

以上、上記の提案手法を用いてオフライン手書き文字画像の認識を行う。

## 4. 認識実験

### 4.1 実験条件

提案手法の実現性を検証するため、認識実験を行う。実験に利用した文字画像は産業技術総合研究所提供の

表2 抽出された自己相似性特徴

Table 2 The self-similarities of a character pattern.

	$R_i$ の座標	$D_i$ の座標
$f_1$	(0, 0)	(45, 12)
$f_2$	(4, 0)	(9, 67)
⋮	⋮	⋮
$f_M$	(72, 72)	(34, 22)

文字画像データベース ETL8B, 956 文字を対象とした。1 字種につき最初の 120 文字を学習文字パターンとして利用し、自己相似性特徴を抽出し、残りの 40 文字を未学習文字パターンとして認識に利用した。実験条件は下記のとおりである。

- 前処理として、画素数  $64 \times 63$  の二値画像に太線化処理および非線形正規化処理<sup>12)</sup>を施し、 $72 \times 72$  の大きさにした。これは文字画素の密度を一定にして、文字線をできるだけ均一にするためである。
- 自己相似性特徴を抽出する際、ドメインブロックの大きさはレンジブロックの 2 倍と固定し、レンジブロックの大きさを変化させ特徴を抽出した。またアフィン変換において回転角は考慮せず、恒等写像のみとした。これは二値画像を対象としているため、回転角を考慮しなくても、対応するドメインブロックを見つけることが困難ではなかったからである。したがって自己相似性特徴の各要素は、各レンジブロック  $R_i$  に対応するドメインブロック  $D_i$  の左上の座標のみをパラメータとして持つだけである(表2)。
- 式(10)において、予備実験より  $\alpha = 1.5$  とした。
- 自己相似性特徴を用いて画像を復元した後、元の画像と復元画像との比較にはユークリッド距離によるテンプレートマッチング(重ね合わせ法)を行った。これは提案する自己相似性特徴の有効性のみを正確に評価するためには、3.1 節で述べたようにオフライン手書き文字認識で一般的に使われている特徴を復元後の画像からさらに抽出し、認識を行うべきではないと判断したからである。実験は認識率に直接影響を及ぼすと考えられる以下の要因について検証を行う。
- 提案する自己相似性特徴の抽出方法の有効性について評価する。
- レンジブロックの大きさ(特徴数)を変化させ、自己相似性特徴を抽出したときの認識率の違いに

$\alpha$  の値を 0.5 から 2.0 まで 0.1 刻みで変えて特徴を抽出し、4.2 節で述べる実験条件で認識を行ったところ、 $\alpha = 1.5$  のとき、最も高い認識率が得られた。

ついて調べる．

- 認識時において，レンジブロックを復元する際に，対応するドメインブロックの左上の座標付近  $((D_x - c, D_y - c), \dots, (D_x + c, D_y + c))$  も候補として，復元する．その際パラメータ  $c$  が認識率に与える影響について調べる．

4.2 自己相似性特徴の抽出

自己相似性特徴の抽出方法についての評価を行う．式 (9) を用いて同字種内での相似性のみを考慮した抽出方法と式 (10) を用いて他の文字種との相似性も考慮した抽出方法とを比較する．レンジブロックの大きさは  $6 \times 6$  (したがってドメインブロックの大きさは  $12 \times 12$ ) と固定した．抽出する特徴数は 144 である．その結果を表 3 に示す．

表 3 には 2 つの抽出方法による第 1 位から第 3 位までの認識率の結果を示す．表 3 より，式 (9) を用いて同字種内での相似性のみを考慮した場合は認識率は約 81%，式 (10) を用いて他の文字種との相似性も考慮した場合，認識率は約 92% という結果が得られた．

一方で 1 字種につき 120 文字の学習文字パターンから 1 つのテンプレートとなる文字画像 (120 パターンの平均値) を作成し，ユークリッド距離によるテンプレートマッチング (重ね合わせ法) による認識を行った場合，その認識率は 86.71% であった．今回の認識方法は，通常のテンプレートマッチングの手法と比較すると，認識したい未知画像に対して自己相似性特徴による画像復元の操作を前処理として行い，未知画像を変形させた後にテンプレートマッチングによって認識を行っていることになる．そこでテンプレートマッチングの結果 (86.71%) と比較すると，式 (9) を用いて抽出した自己相似性特徴を前処理的な操作として利用し，認識を行っても，テンプレートマッチングによる認識結果に及ばないものの，一方で式 (10) による自己相似性特徴の抽出方法では，テンプレートマッチングによる認識結果よりも認識率の向上が図れることが確認できる．すなわち式 (10) による抽出方法は式 (9) による抽出方法と比較して，認識率が約 11%，テンプレートマッチングの結果と比較して認識率が約 5.3% 向上していることから，同字種内だけでなく他の文字種との相似性も考慮し，自己相似性特徴を抽出した方がよいことが分かる．以上のことから提案する自

表 3 自己相似性特徴の抽出方法の結果

Table 3 The results of extraction of self-similarities.

認識率	第一位	第二位	第三位
式 (9) による抽出	81.17%	88.00%	90.53%
式 (10) による抽出	92.07%	95.79%	97.01%

己相似性特徴を用いての文字画像の認識が可能であることが確認できる．

4.3 レンジブロックの大きさ

次にレンジブロックの大きさを変化させ自己相似性特徴を抽出し，認識を行った．レンジブロックの大きさを  $6 \times 6, 3 \times 6, 6 \times 3, 4 \times 4, 3 \times 3$  と変えて行った (ドメインブロックはそれぞれのレンジブロックの大きさの 2 倍である)．その結果を表 4 に示す．

表 4 には，それぞれの大きさのレンジブロックによって抽出したときの特徴数，第 1 位から第 3 位までの認識率の結果を示す．表 4 より，レンジブロックの大きさを細かくする，すなわち特徴数を増やすほど，認識率が向上する傾向が分かる．そして図 6 にレンジブロックの大きさを変化させ，復元したときの画像の一例を示す．図 6 中の「あ」と「話」という 2 つの文字画像は，レンジブロックの大きさが  $6 \times 6$  のときには，誤認識してしまったが，レンジブロックの大きさを細かくした場合，正しく認識できた例である．

図 6 より，レンジブロックの大きさを細かくすると，より正確に復元できる様子が分かる．またレンジブロックの大きさが  $6 \times 6$  のとき，認識率は 92.07%，特徴数を 2 倍にして  $6 \times 3$  のとき，93.02% と約 1.0% 認識率が向上できたのと比べ，さらに特徴数を 2 倍にして  $3 \times 3$  のとき，93.55% と約 0.5% 程度しか認識率は向上せず，特徴数の増加にともない認識率の向上の差も小さくなっていくことも分かる (さらにレンジブロックの大きさを  $2 \times 2$  (特徴数 1296) として，認識

表 4 レンジブロックの大きさを変化させたときの結果

Table 4 The results by changing the size of range blocks.

	特徴数	第一位	第二位	第三位
$6 \times 6$	144	92.07%	95.79%	97.01%
$6 \times 3$	288	93.02%	96.44%	97.40%
$3 \times 6$	288	93.06%	96.33%	97.27%
$4 \times 4$	324	93.45%	96.60%	97.58%
$3 \times 3$	576	93.55%	96.60%	97.52%

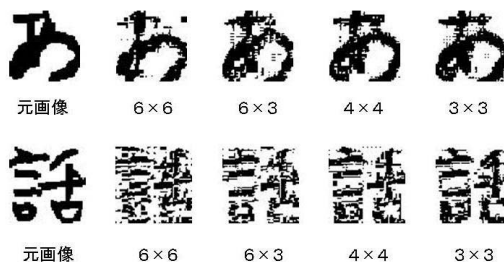


図 6 レンジブロックの大きさによる画像の復元

Fig. 6 The restored images by changing the size of range blocks.

表 5 特徴抽出および認識にかかる時間

Table 5 The time for extraction and recognition.

	特徴数	特徴抽出時間	認識文字数
6 × 6	144	150 (秒/字種)	80 (文字/分)
6 × 3	288	165	73
4 × 4	324	180	70
3 × 3	576	200	68

を行っても認識率は 93.49% とほとんど変わらない)。

レンジブロックの大きさを細かくし、特徴数を増やすと、

- より局所的な領域間での相似性を調べることができる、
- 一方で大局的な領域間での相似性の情報が欠落する傾向がある、

ということが考えられる。すなわち、レンジブロックの大きさが大きい(粗い)と、おおまかに復元を行うため、細かい局所的な領域での正確さには欠けるが、一方で、レンジブロックの大きさが細かいと、局所的な領域で正確に復元できるものの、文字画像が大きく変動した場合は、大局的な相似性の情報が欠落するため、全体的にずれが生じ、おおまかな復元ができなくなってしまふからだと考えられる。また特徴数の増加に比例して特徴抽出や認識に必要な計算量も増加する。実験を行った計算機環境において、レンジブロックの大きさを変えたときの処理時間を表 5 に示す。表 5 には学習文字パターン 120 文字を利用して 1 字種あたり自己相似性特徴を抽出するのに必要な平均時間および 1 分間あたりの平均認識文字数を示す。

表 5 より特徴数が増えるにつれ、特徴抽出、認識処理に必要な時間も増加していくことが分かる。すなわち特徴数の増加にとともに、認識率はある程度まで向上していくものの、必要な計算量も増加していく。特に認識処理においては、特徴抽出処理とは異なり、即座に認識結果を求める必要があるため、計算量の増加を改良する措置が必要である。そのため特徴数の少ないもので候補をある程度絞り込み(大分類)、次に特徴数の多いもので候補を 1 つに絞り込む(詳細認識)という大分類・詳細認識の手法を導入し、計算量の増加に対処しながら、認識率の向上を図るといった改善方法が考えられる。

#### 4.4 画像の復元

画像の復元時において、各レンジブロックを復元する際、自己相似性特徴中に指定されているドメイ

表 6 近傍を候補として復元したときの結果

Table 6 The results by using neighboring blocks.

認識率	第一位	第二位	第三位
$c = 0$	92.07%	95.79%	97.01%
$c = 1$	92.74%	96.19%	97.24%
$c = 2$	93.10%	96.42%	97.40%

表 7 レンジブロックの大きさを変え、近傍を候補として復元したときの結果

Table 7 The results by changing the size of range blocks and using neighboring blocks.

	特徴数	$c = 0$	$c = 1$
6 × 6	144	92.07%	92.74%
6 × 3	288	93.02%	93.63%
3 × 6	288	93.06%	93.60%
4 × 4	324	93.45%	93.90%
3 × 3	576	93.55%	93.97%

ンブロックの左上の座標(表 2)を中心に、その近傍( $(D_x - c, D_y - c), \dots, (D_x + c, D_y + c)$ )付近で最も相似性の高いブロックをドメインブロックとして用いて復元する。その際パラメータ  $c$  の値を変え、認識を行った。レンジブロックの大きさは  $6 \times 6$  とした。その結果を表 6 に示す。

表 6 には、 $c = 0, 1, 2$  と変化させたときの第 1 位から第 3 位までの認識率の結果を示す。表 6 より、 $c = 1$  のときは 92.7% と  $c = 0$  と比較して約 0.7%、 $c = 2$  のときは 93.1% と約 1.0% 認識率が向上している。自己相似性特徴中に一意に指定されている座標付近を中心に  $c = 0, 1, 2$  と近傍付近を含めて相似性の高いブロックを用いて復元した方が認識率の向上を図れることが確認できる。しかし一方で  $c = 0$  を基準に復元に必要な計算量を考えると、 $c = 1$  の場合は 9 倍、 $c = 2$  の場合は 25 倍必要となる。

次にレンジブロックの大きさの異なる特徴に対して  $c = 1$  と固定し ( $c = 2$  では必要な計算量が大きいため)、同様に認識を行った。その結果を表 7 に示す。

表 7 にはレンジブロックの大きさ、特徴数およびそれをういて  $c = 0$  もしくは  $c = 1$  としてレンジブロックを復元したときの第 1 位の認識率を示す。表 7 より、いずれの大きさのレンジブロックにおいても、 $c = 1$  のときの方が認識率の向上を図れることが確認できる。特にレンジブロックの大きさが  $3 \times 3$  で  $c = 1$  として復元した場合、93.97% の認識率を得ることができた。

以上の結果から認識対象となる未知画像と自己相似性特徴を抽出した学習文字パターンとのずれを考慮し、近傍付近を含めてレンジブロックを復元した方が、認識率の向上を図れることが確認できた。一方で復元に必要な計算量を含めて考慮すると、復元すべきドメイ



表 8 平仮名を対象としたときの認識率

Table 8 The recognition rate of hiragana characters.

	特徴数	$c = 0$	$c = 1$
$6 \times 6$	144	96.40%	96.50%
$4 \times 4$	324	96.75%	96.80%
$3 \times 3$	576	96.60%	96.70%

ンブロックの左上の座標を中心として遠くの範囲までの検索は困難でもある。

## 5. 考 察

実験結果より、提案する自己相似性特徴を用いた文字画像の認識は可能であることが確認できた。本論文においては文字画像中における自己相似性という特徴に焦点を当て、議論をすすめてきたが、本研究の関連研究として、同字種の 2 つの文字画像間において、一方を他方の文字画像に変形させマッピングすることによって認識を行う手法<sup>13),14)</sup>があげられる。本研究においては自己相似性特徴として、画像内の局所的な一領域に対してその画像内から類似する領域を検索し、アフィン変換によってマッピングし、これを実現していることになる。そこで文献 13) と同様に、ETL8B の中の平仮名 46 字種のみを対象に認識を行ってみた。その結果を表 8 に示す。

表 8 には、特徴数、近傍検索時におけるパラメータ  $c$  を変えていったときの第 1 位の認識率を示す。表 8 より、レンジブロックの大きさが  $4 \times 4$ 、 $c = 1$  の近傍検索を行ったとき、認識率は 96.8% であり、本研究においても文献 13) とほぼ同じ認識率 (97.0%) が得られた。またレンジブロックの大きさが  $3 \times 3$  の場合、 $4 \times 4$  のときよりも若干認識率は低くなっていることも分かる。平仮名は漢字と比較して曲線からなるストロークから構成され、その数も少なく、白画素部の面積も広いので、大局的な相似性の特徴の方が認識しやすいのではないかと考えられる。

さらに ETL8B 全字種を対象とした認識結果 (表 3) を分析すると、文字種によっては認識がしにくいものもあった。特に前述した平仮名は漢字に比べて認識率が低い傾向があった。平仮名は漢字と比べると曲線から構成されるものが多いので、前述した理由のほかに、レンジブロックの形状を四角形とすると、相似関係のある領域を求めにくかったためであると考えられる。特に「ぼ」や「ぽ」などの濁点、半濁点を持つ文字

表 9 複数のレンジブロックの大きさの特徴を利用した結果

Table 9 The result by using two kinds of self-similarities.

	特徴数	第一位	第二位	第三位
$6 \times 6, 4 \times 4$	468	93.60%	96.68%	97.55%
$6 \times 3, 3 \times 6$	576	93.72%	96.75%	97.65%
$4 \times 4, 3 \times 3$	900	93.93%	96.91%	97.80%
$3 \times 3$	576	93.55%	96.60%	97.58%

の区別は、正確に濁点、半濁点の領域をレンジブロックとして切り出さないとその正確な判別は困難である。

認識率をさらに向上させるためにはいくつかの工夫が必要となる。文字認識においては、認識率の向上のため、特徴を複数併用して使用し、認識を行うという試みも行われている。そこで本研究においても複数の特徴を組み合わせて認識を行うことを試みてみた。表 9 には、2 つの異なるレンジブロックの大きさの自己相似性特徴を用いて、未知画像を別々に復元し、その後で 2 つの変換後の画像に対して論理和によって 1 つの画像に合成し、認識を行った場合の結果を示す。

表 9 には 2 つの大きさの異なるレンジブロックの組合せによる特徴数および第 1 位から第 3 位までの認識率を示す。表 9 より、レンジブロックの大きさが  $3 \times 3$  (特徴数 576) の場合よりも、レンジブロックの大きさが  $6 \times 6$  と  $4 \times 4$  を組み合わせた場合 (特徴数 468)、もしくは  $6 \times 3$  と  $3 \times 6$  (特徴数 576) を組み合わせた場合の方が、同じ特徴数もしくは少ないにもかかわらず、わずかではあるが高い認識率を得られていることが確認できる。さらにまたレンジブロックの大きさが  $4 \times 4$  と  $3 \times 3$  (特徴数 900) を組み合わせた場合も、認識率の向上を図ることができた。以上のように、特徴を複数併用して認識を行うことによって認識率を向上させることも確認できた。

しかしこれまで実験結果で述べてきたように、特徴数を増やす、レンジブロックの復元の際、自己相似性特徴中に記述されているドメインブロックの左上の座標の近傍付近も考慮して行うといったことで、認識率の向上を図ることもできるが、それにともない計算量も増加する。そこで、計算量の増加に対処しながら、認識率の向上を図る工夫として 4 章で述べた大分類・詳細認識の手法を導入する。すなわち特徴数の少ない自己相似性特徴 (レンジブロックの大きさが  $6 \times 6$ ) を用いて候補を  $P$  個に絞り込み、次に特徴数の多い  $3 \times 3$  の自己相似性特徴および近傍検索  $c = 1$  によって候補を 1 つに絞り込む。その結果を表 10 に示す。

表 10 には大分類で第  $P$  位 ( $P = 2, 3, 4, 5$ ) まで候補を絞り込んだ場合の認識率を示す。表 10 より、 $P = 4$  のとき 93.9% の認識率が得られ、大分類を行わ

本研究においては 1 字種につき最初の 3 セット 120 文字を学習文字パターンとして利用している。

表 3 の式 (10) による認識結果の内訳は平仮名 75 文字の認識率は 90.93%、漢字 881 文字の認識率は 92.17% であった。

表 10 大分類・詳細認識の結果

Table 10 The result by using two kinds of self-similarities.

大分類での候補 $P$	認識率
第 2 位	93.22%
第 3 位	93.76%
第 4 位	93.90%
第 5 位	93.88%

ず、はじめからレンジブロックの大きさが  $3 \times 3$  の自己相似性特徴および  $c = 1$  として復元することによって候補を 1 つに絞り込む場合 (93.97%) とほぼ同じ認識率を得ることができた。この場合の計算量は、レンジブロックの大きさが  $6 \times 6$  で認識を行うときとほぼ同じで済む。このように大分類・詳細認識の手法を導入することによって、計算量を削減しながら、認識率の向上を図れることが分かる。

しかし、本論文で提案した認識方法では、抽出した自己相似性特徴を前処理的な操作として利用し、未知画像を変換した後で、認識処理 (今回はテンプレートマッチングで行った) を行わなければならないために、認識時に必要とする計算量が大きい。そこでさらなる計算量の削減のためにも、認識対象である未知の文字画像からも自己相似性特徴を抽出し、データベース内に保存されている学習文字パターンの特徴と直接比較することによって、認識を行うことができる手法を考案することが必要である。さらに本論文で提案した手法はフラクタル不変定理の性質に基づいているため、類似画像のみに適用し、縮小や回転といった大きく変動のある文字画像に適用することは難しい。そのため学習文字パターンと認識対象文字から自己相似性特徴を抽出し、2 つの特徴間での相互関係を調べ、認識に直接利用することによって、フラクタル幾何学を利用した認識方法の利点である拡大縮小および回転といった変動に強い認識を行うことも考えられる。

## 6. む す び

本論文においては文字画像の局所的な領域間の相似性を、文字固有の特徴としてとらえ、これをフラクタル幾何学におけるレンジブロックとドメインブロックの関係を利用して抽出し、自己相似性特徴として認識に利用する方法を提案した。そして産業技術総合研究所提供の ETL8B による認識実験を通して、提案した自己相似性特徴を用いた文字画像の認識が可能であることを示した。今後は認識対象の文字画像からも自己相似性特徴を抽出し、これを直接用いて認識が可能手法を考案し、計算量の削減を図ると同時に、拡大

縮小および回転といった変動に強い認識方法を考え、認識率のさらなる向上を図る予定である。

謝辞 本研究を行う際に、貴重な手書き文字データベースを提供してくださいました産業技術総合研究所の皆様へ感謝いたします。また研究開始当初、さまざまなお助言をくださいました大日本印刷の眞瀬智子さんには深く感謝の意を表します。

## 参 考 文 献

- 1) 森 健一：パターン認識，電子情報通信学会 (1988).
- 2) 藤澤浩道：企業研究者から見た文字認識の進歩と課題と将来，電子情報通信学会技術研究報告，PRMU2003-247, pp.93-100 (2004).
- 3) Jacquin, A.E.: Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations, *IEEE Trans. on IP*, Vol.IP-1, No.1, pp.18-30 (1992).
- 4) Barnsley, M.F. and Hurd L.P.: マルチメディア・フラクタル画像圧縮，トッパン (1995).
- 5) Nerson, M. and Gailly, J-L.: データ圧縮ハンドブック，トッパン (1996).
- 6) 上條, 中沢, 清水：フラクタル特徴変数による複雑な形状を持つ物体の同定 — LSI ウェーハ上の異物のカテゴリー分類，電子情報通信学会論文誌，Vol.J78-D-II, No.5, pp.794-802 (1995).
- 7) 本田, 小谷：フラクタル分析による顔画像の特徴抽出，電子情報通信学会技術研究報告，IE96-104, pp.43-48 (1996).
- 8) 井上, 村山, 武田, 岡田：フラクタル幾何に基づく古文書画像の解析・蓄積・呈示，電子情報通信学会技術研究報告，PRMU96-168, pp.61-68 (1997).
- 9) Neil, G. and Curtis, K.M.: Shape Recognition Using Fractal Geometry, *Pattern Recognition*, Vol.30, No.12, pp.1957-1969 (1997).
- 10) 横山, 渡辺, 菅原, 上野：フラクタル符号に基づく構造的な類似性特徴抽出の提案，電子情報通信学会技術研究報告，PRMU2001-285, pp.105-112 (2002).
- 11) 横山, 菅原, 渡辺：フラクタル符号の写像情報に基づいた圧縮領域における類似度，電子情報通信学会技術研究報告，PRMU2003-177, pp.1-6 (2003).
- 12) 山田, 斉藤, 山本：非線形正規化法の改良，電子情報通信学会総合大会，D-439 (1988).
- 13) 内田, 迫江：単調連続 2 次元ワーブ法によるオフライン手書きひらがな認識，電子情報通信学会技術研究報告，PRMU98-163, pp.103-110 (1998).
- 14) 内田, 迫江：単調連続 2 次元ワーブ法によるオフライン手書き文字認識実験，電子情報通信学会論文誌，Vol.J83-D-II, No.4, pp.1198-1200 (2000).
- 15) 眞瀬智子：フラクタル変換を用いたオフライン

手書き文字認識，慶應義塾大学理工学研究科修士論文 (2000).

- 16) 篠沢佳久，大駒誠一：フラクタル幾何学を用いたオフライン手書き文字認識，情報処理学会第 62 回全国大会，Vol.2-245 (2001).
- 17) 篠沢佳久：自己相似性を利用したオフライン手書き文字認識，情報処理学会第 66 回全国大会，Vol.2, pp.105-106 (2004).

(平成 16 年 6 月 21 日受付)

(平成 17 年 1 月 7 日採録)



篠沢 佳久 (正会員)

1994 年慶應義塾大学理工学部管理工学科卒業．1996 年同大学大学院理工学研究科管理工学専攻前期博士課程修了．1999 年同大学院後期博士課程修了．博士 (工学)．現在慶應義塾大学理工学部管理工学科助手．機械学習，パターン認識に興味を持つ．現在オフライン手書き文字認識，ニューラルネットワークによる言語獲得の研究に従事．電子情報通信学会会員．