

合法パターン頻度分布に基づく9か国文字の類別†

美濃導彦^{††} 松瀬哲朗^{††} 坂井利之^{††}

文書画像は、本質的に濃淡値をとる写真のような領域と二値である線図形領域に大きく分けられる。線図形領域はさらに文字領域のように行という構造をもつ領域と図のように、特定の構造をもたない領域に分割できる。文書画像を処理・蓄積・伝送するためには、まずこれらの情報を大きく類別することが必要である。本論文では、大きく類別された後の文字領域を対象とした特徴抽出とそれによる9か国文字パターンの類別について述べる。線図形が十分細かくサンプリングされているときは、それを3×3の単位メッシュで観測すれば、そのうち起こりうるパターンはほとんどが合法パターン66種である。合法パターンに対してローカルな方向性を定義すれば、それに基づいて傾斜方向・曲線性などの線特徴パラメータが定義できる。これを利用すれば、文字領域を対象とする画像に対して合法パターンの統計をとるという簡単な処理により、含まれている文字の種類を類別することが可能となる。実験の結果、これらのパラメータは、フォントの違い・画質の低下に対しては変化が激しいが、文字パターンの大きさの変動に対してはほぼ不変であることが確認できた。すなわち、これらのパラメータは文字領域の特徴をよく表現しているため、文書画像の処理・蓄積・検索に役立つと考えられる。

1. まえがき

文書画像処理の需要の増大に伴い、多品質・多種類の文書画像を処理する必要性が高まってきている。文書画像は一般に、線図形領域（文字領域と図表領域）、写真領域および空白領域から構成されており、性質が本質的に異なる情報や品質の異なったものが混在したものである。したがって、文書画像を処理・蓄積・伝送する場合は品質を判定し雑音除去を行う¹⁾と同時に、情報を大きく類別しそれぞれに応じた処理を施すことが重要である。情報の大類別は入力時に実時間で行える可能性がある²⁾ので、本論文では大類別後の文字領域を対象にし、その処理、とくにその特徴抽出処理について述べる。

わが国で使用されている文字は、漢字・ひらがな・カタカナおよびアルファベット等であるが、多くの民族が共存する国では多種類の文字が使用されているので、これらを類別することは今後重要な意味をもつと考えられる。そこで、文字として世界の代表的な文字系から9種類を選定し、これらの文字の特徴抽出を行う。ここでは、2値の線図形に関するMOLD理論(Mesh Oriented Line Drawings' Theory)^{3),4)}で定義された66種類の合法パターンの出現頻度分布に基づいて各種の線特徴パラメータ(表2参照)を定義し、これをもとにして文字を類別する⁵⁾。

† Classification of 9 Kinds of Characters Based on the Statistics of Legal Patterns by MICHIIHIKO MINOH, TETSUO MATSUSE and TOSHIYUKI SAKAI (Department of Information Science, Faculty of Engineering, Kyoto University).

†† 京都大学工学部情報工学科

2. 9か国文字パターンの視覚的特徴

本論文で対象とする文字は、現在世界で使用されているものを識字人口(=人口×識字率)によって大分類⁶⁾し、各カテゴリから選択したハングル文字、チベット文字(北インド系)、ギリシャ文字、タイ文字(南インド系)、アラビア文字(セム系)、漢字、ロシア文字、ローマ文字、日本文字の9種類の文字である。

図1は各文字パターンの例である。ギリシャ文字、ロシア文字、ローマ文字の3種類に対しては類似性を感じられるが、他の6種類の文字からはそれぞれ異なった印象を受ける。そこで、各文字パターンの構成単位(字素)を表1にまとめた^{6),7)}。漢字を除けば、文字パターンはそれぞれ数十個程度の字素(音節文字あるいは音素文字)から構成されている。

したがって、この字素の集合に注目することにより各文字パターンについて次のような視覚的特徴が明らかになる。

ハングル文字：ほとんどの母音文字・子音文字は鉛直と水平方向のストロークの組合せから成っている。

チベット文字：子音文字の大多数は上部に水平方向のストロークをもつ。また右下方向への独特なストロークが存在する。

ギリシャ文字：全体に右上方向に傾斜し、丸みのあるアルファベットである。

タイ文字：子音文字のほとんどは鉛直方向に長いストロークをもち、その端に小さな円が存在する。

アラビア文字：直線ストロークがほとんど存在しない。

<p>체으로 인구문제 어라는 사정을 도 계속 경제적 의 문제는 인구:</p> <p>ハングル文字 チベット文字 キリシヤ文字</p>	<p>ཕ་རབ་ལོན་ཤིང་ དཀའ་ལས་སུ་ འགྲུབ་པར་སྒོམ་ནས་ མི་ཁྲི་འདི་</p> <p>チベット文字</p>	<p>γάρ ἄν ἐκεῖ 20 ὁ γὰρ π αὐτῶ ὁ αὐτ. ἔργα, ἵνα ὁ ἐγείρει τοὺς οὖς θέλει ζῆ</p> <p>ギリシヤ文字</p>
<p>พิจารณา อักษรกราฟิก \$39 แทนตัวเลข เป็นคํ่าไป และ</p> <p>タイ文字 アラビア文字 漢字</p>	<p>ي حيناً ما بها ، والساطور رة إلى الجزع!</p> <p>アラビア文字</p>	<p>在鸳鸯树下相 我，一心想害 息也没有用啦 样地难受，便</p> <p>漢字</p>
<p>струкций (не ми их семан практический сю граммати автоматическ водить любые</p> <p>ロシア文字</p>	<p>dom than in th small in compa years, mergers o significant reduc Thus, in F General Electric</p> <p>ローマ文字</p>	<p>に、が多い。それにも 非日常的なカナ 本章の最初の 相通じる場合があ</p> <p>日本語</p>

図1 9種類の文字で書かれた文書画像

Fig. 1 Document images written in 9 kinds of character sets.

表1 各文字の構成要素

Table 1 Constitutive elements of each character sets.

単位	文字	字素の例	
音素文字	ハングル文字	母音文字	ㅏ ㅑ ㅓ ㅕ ㅡ ㅗ
		子音文字	ㄱ ㅋ ㆁ ㆅ ㅊ ㅄ
	チベット文字	母音記号	~ ̣ ~ ~
		子音文字	ཀ ཁ ང ། ས ཁ ། ཉ
	タイ文字	母音文字	๓ ๔ ๕ ๖ ๗ ๘ ๙
		子音文字	ก ๓ ๔ ๕ ๖ ๗ ๘
ローマ文字	a b c d e f g ...		
	α β γ δ ε ζ η ...		
	а б в г д е е ...		
かな文字	かな文字	あ い う え お か き ...	
	アラビア文字	母音記号	، ى ، ٴ ، ٲ ، ٲ
子音文字		، ٲ ٲ ٲ ٲ ٲ	
漢語文字	漢字	亞 委 宇 江 御 化 蓋 區 毛 子 差 賓 州 瀾 棠 一	

漢字：水平、鉛直方向のストロークが多く、文字パターンが複雑である。

ロシア文字：直線ストロークが比較的多いアルファベットである。

ローマ文字：アルファベットがもつ丸みはギリシヤ文字とロシア文字の中間程度である。

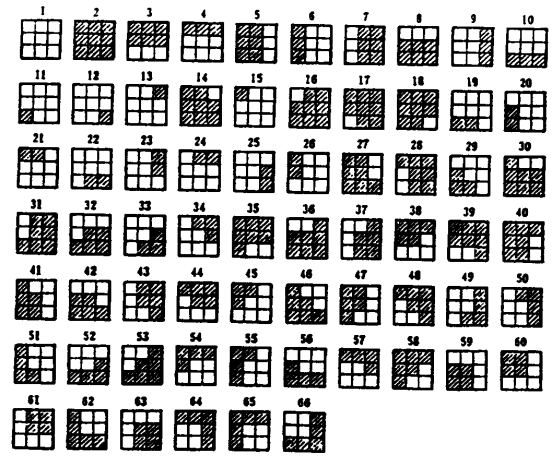


図2 合法パターン

Fig. 2 Legal patterns.

日本語文字：カナ文字が混じっている分だけ漢字より曲線ストロークの割合が多い。

3. 合法パターンの形状に基づく線特徴パラメータ

3.1 合法パターン

前章で述べたような文字パターンの視覚的特徴を表す線特徴パラメータを定義するために、まず基本となる合法パターンについて述べる。合法パターンは、2値の線図形の生成時にもつ強い拘束条件を単位メッシュにおいて観測・表現することを考えた MOLD 理論^{3),4)}で定義されている。

ファクシミリ等から入力された2値の線図形を構成する線は一般に幅をもっている。サンプリング間隔が十分に細かく、線幅が単位メッシュの一辺(3画素)より広いという仮定が満たされると、単位メッシュ内には白画素領域と黒画素領域がたかだか一つずつしか現れない。この制限によって、理論的に出現しうる単位メッシュパターンの種類は、2⁹=512個のうち66個に限られ、これらを合法パターンと呼ぶ。図2にこの合法パターンを示す。

交点や尖点が存在する場合は、たとえ線幅の仮定が満たされても、合法パターン以外の単位メッシュパターンの出現が理論的に考えられる。しかし、CCITTのファクシミリテスト原稿について調べた結果、合法パターンのみの累積率は99%以上となり、最も頻度の高いオール白のパターンを除外しても95%以上となった。このことから線図形の大局的な傾向を把握するだけならば、合法パターンに注目するだけで十分であると考えられる。

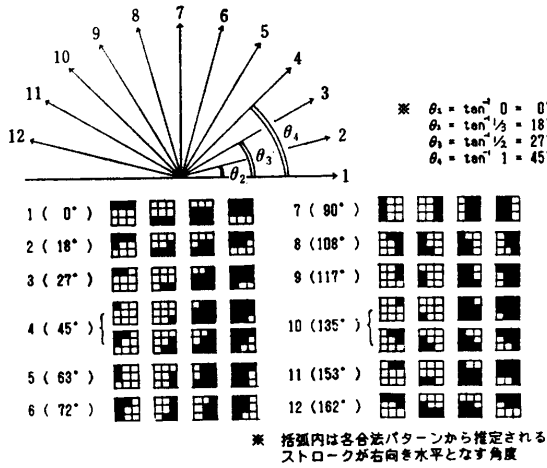


図3 合法パターンの方向性による分類
 Fig. 3 Classification of legal patterns by local direction characteristics.

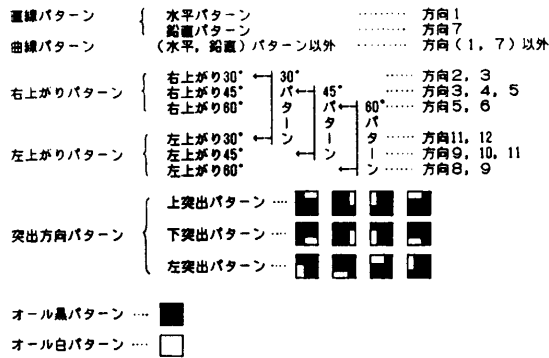


図4 方向性特徴パターンの定義
 Fig. 4 Definition of directional characteristic patterns.

※ 数字は方向性分類の数字と一致する

各合法パターンがもつ局所的な方向性は、そのもつ黒画素領域が線の境界の一部であると考えられることにより定義することができる。図3はこの方向性に注目した合法パターンの分類である。各方向が水平方向となす角度は、方向1： $\theta_1=0$ 度、方向2： $\theta_2=\tan^{-1}(1/3)=18$ 度、方向3： $\theta_3=\tan^{-1}(1/2)=27$ 度、方向4： $\theta_4=45$ 度であり、他の方向も同様に計算できる。図4はこの分類を用いて定義した合法パターンの集合(方向性特徴パターンと総称する)を示している。

3.2 線特徴パラメータ

文字パターンは前節で述べたように国語の種類ごとに形状が異なり、ストロークの方向性や丸みに特徴がある。そこで線特徴パラメータとして、線の方向性に関する傾斜方向、傾斜角度、縦横比および曲線性を考える。また、文字パターンの特殊な形状を表すと考

表2 線特徴パラメータの定義

Table 2 Definition of line characteristic parameters.

傾斜方向	1 左上りパターン/右上りパターン 2 左上り30パターン/右上り30パターン 3 左上り45パターン/右上り45パターン 4 左上り60パターン/右上り60パターン
傾斜角度	5 60パターン/30パターン 6 右上り60パターン/右上り30パターン 7 左上り60パターン/左上り30パターン
曲線性	8 曲線パターン/直線パターン
縦横比	9 鉛直パターン/水平パターン
突出方向	10 下突出パターン/上突出パターン 11 上突出パターン/左突出パターン
合法パターン比率	12 オール白パターンを除く全パターンのうち合法パターンが占める割合

表3 特徴パラメータが示す性質

Table 3 Distinctive features derived from the line characteristic parameters.

傾斜方向	線の方向は左上りと右上りのどちらの傾向が強いか
傾斜角度	線の傾きは大きいものと小さいものどちらが多いか
曲線性	線は曲線性が強いのか直線性が強いのか
縦横比	線の鉛直成分と水平成分の大きさの比
突出方向	交点あるいは屈折点から出る線の方向性
合法パターン比率	線図形が複雑なものであるかどうか

られる突出方向性と線の構成の複雑さの指標となる合法パターン比率も線特徴パラメータとして採用する。表2は各線特徴パラメータの定義を示したものである。線特徴パラメータはすべて対象画像に含まれている線の絶対量に依存しないように方向性特徴パターンの出現数の相対的な比率によって定義している。たとえば、傾斜方向は左上りパターンと右上りパターンの出現数の比で定義され、ストロークの方向性を示す。また、曲線性の指標には曲線パターン*と直線パターンの出現数の比を用い、鉛直成分と水平成分が少ないほど値が増大する。表3に各線特徴パラメータが表現していると考えられる線の特徴をまとめておく。

処理対象とする画像の合法パターン頻度分布は、3×3画素の大きさの窓を1画素単位で走査し、統計を取ることによって求める。したがって、任意の位置

* 厳密には局所的に水平、垂直でないパターンであるが、ここでは便宜上この名称を用いる。

적으로 인구문제에
더하는 사정을 인정
도 계속 정책적 배
의 문제는 인구조
소장관의 기능과 경
클래스0

적으로 인구
더하는 사정
도 계속 정
클래스2

적으로 인구문제
더하는 사정을
도 계속 정책적
의 문제는 인구
클래스1

적으로 인
더하는 사
도 계속 정
클래스3

図5 4通りの大きさの文字パターン
Fig. 5 Character patterns with 4 kinds of size.

にある一つの画素は窓がかかる位置の違いによって、
頻度分布には9回計数されることになる。

4. 文字の特徴抽出

4.1 文字特徴抽出実験

図1に示した9種類の画像は1回の処理対象の1例
である。対象画像は小さいほうが望ましいが、統計に
基づく処理をするため、ある程度の大きさが必要であ
る。今回、大きさは255×255画素(約3.2×3.3cm)
とした。この画像には3×3画素の単位メッシュパ
ターンが計64,009個含まれており、統計をとるには
十分であると考えられる。

各文字ごとにフォントは1種類で品質のよいものを

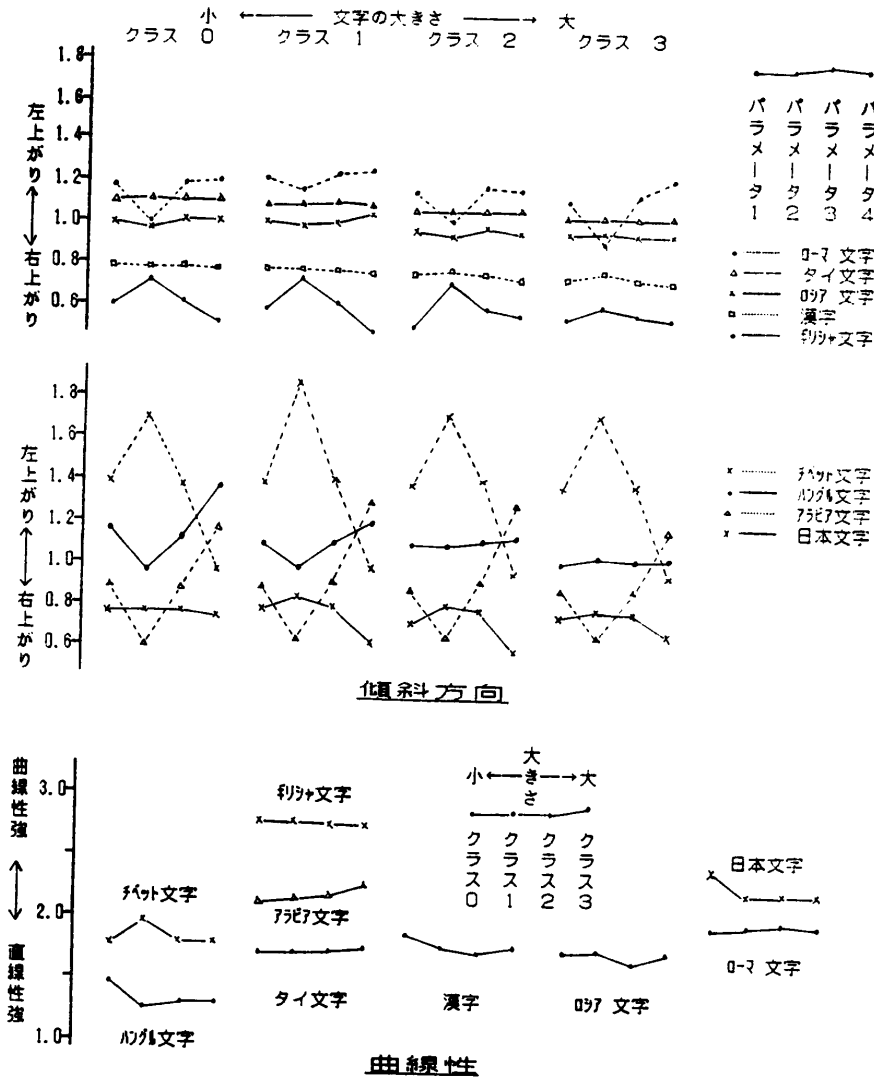


図6A 線特徴パラメータの平均値(1)
Fig. 6A Averages of the line characteristic parameters(1).

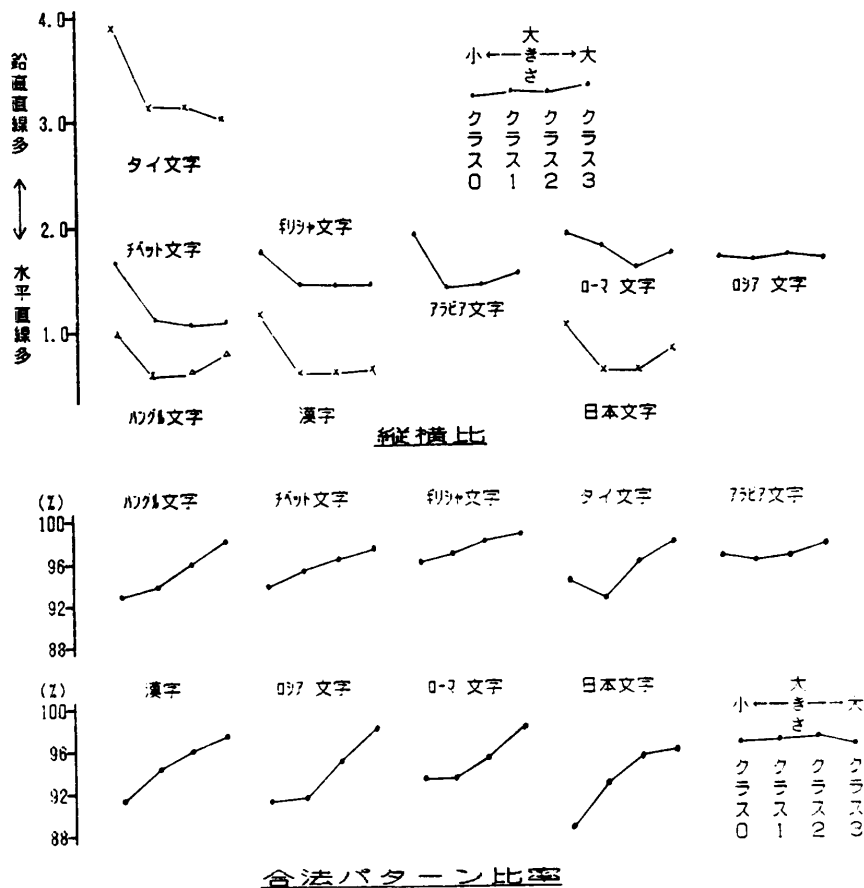


図 6B 線特徴パラメータの平均値(2)
Fig. 6B Averages of the line characteristic parameters(2).

選んだが、線特徴パラメータの値は文字パターンの大きさに依存する可能性があるので、文字ごとに4通りの大きさのデータを作成した。この文字パターンを大きさの順にクラス分けし、最小のものから順にクラス0~3とする。クラス0とクラス3の大きさの比は約1対2とした。また、9種類の文字間では、同一クラスの文字パターンの大きさがほぼ同程度に揃うようにした。図5はクラス0からクラス3までの四つの画像例(ハンゲル文字)である。255×255画素の処理範囲内に含まれる文字数は、クラス3(大きさ最大)のときでも10文字程度は存在するので、データの違いによって各パターンの出現傾向が変化することはあまり考えられず、線特徴パラメータの値の分散は比較的小さいと思われる。

実験ではまず、対象画像から合法パターン頻度分布を求め、それをもとに線特徴パラメータの値を表2の定義により計算する。多数のデータについて線特徴パラメータごとに統計をとり、平均・分散を算出するこ

とにより文字の特徴を抽出する。

処理対象とした画像の枚数は、4通りの大きさについてそれぞれ40枚、各文字ごとに計160枚である。入力装置としては解像度8本/mmのファクシミリを使用した。計算機はサイクルタイム650nsのPANAFACOM U-300を使用し、プログラムはFORTRANで作成した。また、画像は、ディスク上にあり、これから255×255画素の画像を切り出し、すべての線特徴パラメータの値を算出するのに要する時間は約35秒であった。

4.2 実験の結果と考察

実験の結果、得られた線特徴パラメータの平均値のグラフの一部を図6に示す。個々の線特徴パラメータの値と文字の特徴との対応関係について以下に考察する。

・傾斜方向について

強い左上りの傾向を呈しているのがチベット文字である。これはチベット文字に特有の右下へのストロー

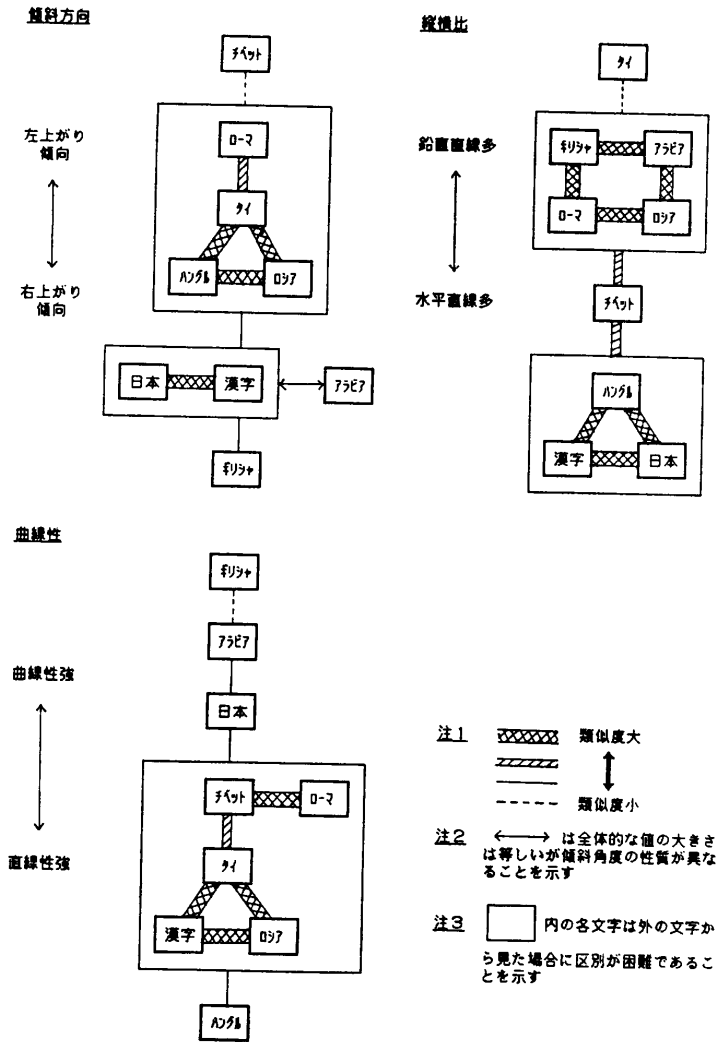


図7 各パラメータからみた文字の類似度
Fig. 7 Similarity of character patterns based on the line characteristic patterns.

クが寄与しているためである。逆に右上りの傾向が強いのが、ギリシャ文字、漢字、日本文字である。アラビア文字は全体として右上り傾向であるが、60度方向のパラメータは左上り傾向を示している。この原因は左上り60度方向に長いストローク（\表1参照）をもつ子音文字が存在するからである。

・曲線性について

ギリシャ文字、アラビア文字、日本文字は比較的強い曲線性を示し、ハンゲル文字は、直線性を示している。ギリシャ文字とアラビア文字の曲線性は鉛直、水平方向のストロークが少ないことが原因であり、また日本文字はカナ文字を含んでいる分だけ漢字の場合より大きな強い曲線性を示している。ハンゲル文字は構

成字素のほとんどが鉛直あるいは水平ストロークから成っているため、強い直線性を示す。

・縦横比について

タイ文字が縦横比で非常に大きな値（鉛直ストロークが多いこと）を示している。これはタイ文字の子音文字の多くが水平方向と比較して鉛直方向に長い直線をもっている（表1参照）ことが原因である。一方、ハンゲル文字、漢字、日本文字はいずれも鉛直方向のストロークが少ないという結果が現れている。漢字と日本文字のパラメータ値がほぼ一致しているのは、二つの文字間の差異であるところのカナ文字にはこのパラメータに影響を及ぼす鉛直、水平ストロークが少ないためである。

・合法パターン比率について

文字パターンの大きさが小さくなるに従って相対的にストロークが混みあってくるため、合法パターン比率は単調に減少する。

タイ文字の場合、パラメータ値が単調に変化しないのは、ほとんどの構成字素がもつ小さな円が原因であると思われる。すなわち、文字パターンが小さくなると円の内部の白画素領域もしいに小さくなり、合法パターン以外のパターンが増加するが、ある程度小さくなると円がつぶれてすべて黒画素領域となり、再び合法パターンで表現される領域となるからである。

漢字や日本文字はストローク数が多く、文字そのものが複雑であるので、合法パターン比率はあまり大きな値にならない。

以上、おもな線特徴パラメータについて、実験結果から得られた文字の特徴について述べた。合法パターン比率を除けば、各文字とも線特徴パラメータの値はほとんど文字パターンの大きさに依存せず、ほぼ一定の傾向を示すことがこの実験から確かめられた。とくに、ストロークの方向性は大きさの変動に対して安定であるから、各文字に固有の性質を示すことになる。この実験で得たことをまとめる意味で、図7に線特徴パラメータからみた各文字間の大まかな類似度を示した。

5. 文字の類別

5.1 文字の類別実験

前章までの基本的考察をもとに、入力画像に含まれている文字の種類を類別する実験を行った。

まず、各文字に対して線特徴パラメータの各文字に対する標準値を四つの大きさの文字パターンが示した値の平均としてそれぞれ一つだけ定める。これは線特徴パラメータの値が文字パターンの大きさにほとんど依存しなかったことによる。

傾斜方向・傾斜角度・突出方向を表すものとして、注目する方向性の違いによってそれぞれ4個、3個、2個の線特徴パラメータを定義しているが、これらは互いに独立に扱うよりもひと固まりの類別パラメータと考えるほうが方向性の特徴をより正確に把握することができる。そこで、各グループの線特徴パラメータを一つの絶対値とパラメータ間の相対値で表現する類別特徴パラメータを考える。たとえば、傾斜方向ならば表2に示した線特徴パラメータ1の値と、他の三つの線特徴パラメータ(2~4)との相対値で表現する。傾斜角度、突出方向についても同様である。この相対値をそのまま表現するか、あるいは表4のコード表によってコード化して表現するかの違いによって2通りの実験を行った。なお、他の線特徴パラメータはそのまま類別パラメータとして用いており、類別のために合計12個のパラメータを用いる。

未知画像中の文字パターンがあるクラスに属する可能性を示すスコアは未知画像から算出される類別パラメータ値 $i=(i_1, i_2, \dots, i_{12})$ とその文字の標準値 $h_j=(h_1, h_2, \dots, h_{12})$ ($j=1, 2, \dots, 9$) との差を用いる。すなわち、 $w=(w_1, w_2, \dots, w_{12})$ として

$$\min_j |w \cdot (i - h_j)|$$

なる j をその未知画像が属する国語のクラスとする。ただし、コード化する場合は、コード化したパラメータについては差として街区画距離を用いる。また、 w はパラメータの信頼性を表す重みであり、傾斜方向のように各文字の標準値が分散していて、類別に対して有効なパラメータほど大きく設定する。

5.2 類別実験の結果および考察

良質文書を用いて9種類の文字の類別実

験を行った。各文字とも、データには特徴抽出の際に使用した文書を用いたが、ファクシミリから再入力するので処理対象範囲は異なる。データ数は文字の大きさの四つのクラスについて各20枚、各文字ごとに計80枚である。

表5に類別結果を示す。この表は各文字の類別画像の入力に対して、最高スコアを得たクラスを出力として判定し結果をまとめたものである。この判定では全体として約86%の正答率であるが、実際の処理では上位2クラスのスコアが接近したときはリジェクトしており、約93%の正答率を得ている。

表5の中で目につくことは漢字と日本文字の誤答が非常に多いことである。類別パラメータの標準値を比較してみると、二つの文字間にはほとんど差異が見られず、カナ文字の丸みが影響を及ぼす曲線性のパラメータ値に少々差異が存在している程度であることからこの結果が理解できる。

表5の左欄(相対値をそのまま用いた実験)と右欄(相対値をコード化した実験)の数字を比較すると、ハングル文字、ロシア文字を漢字と判定する誤答が右欄より左欄のほうが多い。この原因は、相対値を用いれば傾斜方向等の類別パラメータの相対値と標準値との差がそのままスコアに現れるので、その変動が大きくなり、標準値に近い文字と誤りやすいためであると考え

表4 相対値に与えたコード

Table 4 Code given to the relative values.

相 対 値	斜 傾 方 向			傾 斜 角 度			
	-0.05~	+0.05 ~-0.05	~+0.05	-0.04~	-0.05 ~-0.04	+0.05 ~-0.05	~+0.05
コ ー ド	0	1	2	0	1	2	3

表5 自動類別の実験結果

Table 5 Experimental results of automatic classification.

入力 出力	1	2	3	4	5	6	7	8	9
1 ハングル文字	75	78				1	5	1	
2 チベット文字		78	80				1	1	
3 キリシヤ文字			80	80					
4 タイ文字	1			72	69	1	1	3	4
5 アラビア文字	2				71	72	2	2	1
6 漢字		2	1		1	2	57	51	3
7 ロシア文字		1		2	3	1	16	5	59
8 ローマ文字	1	3		2		3	2	2	9
9 日本文字			2			2	5	14	18

右欄：相対値を用いた実験，左欄：コード化した実験

られる。これに対して、コード化すれば多少のパラメータ値の変動は同一コードに吸収されるので、このような誤りは少ない。しかし、逆に、得られた類別パラメータ値と標準値との間に大きな差が存在する場合でも、コード表現ではその差が縮小されてしまい、相対値を用いれば生じにくい誤りが発生することがある。アラビア文字とローマ文字相互間の誤答はこの典型例である。

この実験結果からわかるように、二つの類別方法にはそれぞれ一長一短がある。より高い正当率を得るためには、入力画像から得られた類別パラメータ値と標準値とのユークリッド距離の大きさによって二つの方法を使い分けることが考えられる。

異なるフォントの文字パターンや雑音文書中の文字パターンについて、標準値を変更せずに類別実験を試みたが類別は困難であった。これは線特徴パラメータが合法パターンのローカルな方向性を利用しているため、線の境界部分の性質が保持されないパターンの変化に対しては類別パラメータの値がかなり変化してしまうからである。たとえば、イタリック体はゴシック体より曲線性が強く、右上りの傾向がある。また、汚染文書では直線パターンの出現数が減少するために曲線性がきわめて強くなる。かすれ文書ではファクシミリ

の性質上、走査線方向に雑音がのりやすく、水平方向のパターンが相対的に減少するので縦横比が大きくなる。したがって、ローマ文字とロシア文字のように類別パラメータの標準値が接近している文字間の類別を行うためには他からの知識が必要である。

6. む す び

本論文では、合法パターンの頻度分布を利用して、文字の特徴抽出およびその特徴を利用した9か国文字の類別について述べた。

この手法は入力画像を1回走査して合法パターンの統計をとるという簡単な処理によって、対象領域に含まれる文字の類別がその大きさに関係なくできるのが大きな特徴である。ここで用いた類別パラメータはおもに線の方向を示す線特徴パラメータに基づいているので、文字に対してだけではなく図表領域においても

有効である。また、合法パターンが通信に対する符号化にも適することからこの処理は通信と併用することが可能である。すなわち、送受信時の符号化・復号化と同時に統計をとり文字の類別を行うというある種の情報フィルタ的な機能ももちうる。

この手法はフォントの変化に対しては弱いのが、同一の国家・社会で利用されている文字の種類がおもなものでふつう数種類であることを考えれば、フォントの異なる文字パターンが存在しても、それを別の文字として扱えば、実用的に十分利用できると思われる。

今後は、文字以外の文書画像領域(グラフ・表ならびに写真)の特徴抽出および文書画像の蓄積・検索について研究を進める予定である。

謝辞 文字パターンのデータを提供していただいた、国立民族学博物館助教授の杉田繁治氏に感謝します。

参 考 文 献

- 1) 美濃, 坂井: 二値線図形に対する画質判定と雑音除去法, 情報処理学会論文誌, Vol. 24, No. 2, pp. 182-190 (1983).
- 2) 王, 美濃, 坂井: 3×3 単位メッシュ内の濃淡パターンに基づく文書画像領域分割方式, 信学技法, PRL 83-62 (1984).
- 3) 美濃, 坂井: MOLD 理論—二値線図形の一表現法—, 電子通信学会論文誌, Vol. J65-D, No. 7, pp. 928-935 (1982).
- 4) Minoh, M.: Theoretical and Experimental Studies on Basic Relations between Real World Pictorial Patterns and Their Generating Constraints, Doctorial Thesis, Kyoto University (1982).
- 5) 和田祐一: 現代世界の文字—グラフ, *energy*, Vol. 6, No. 2, pp. 26-27 (1969).
- 6) 日下部文夫: 現代世界の文字—比較文字論, *energy*, Vol. 6, No. 2, pp. 19-21 (1969).
- 7) 西田龍雄(編): 世界の文字, 大修館書店, 東京 (1981).
- 8) 松瀬, 美濃, 坂井: 合法パターン頻度分布に基づく9か国文字の類別, 信学技法, PRL 83-9 (1983).

(昭和59年4月9日受付)

(昭和59年9月20日採録)