

# 枠無しオンライン手書きパターンにおけるストローク特徴の解析

3N-03

木原伸次郎 福島貴弘 中川正樹

東京農工大学

## 1. はじめに

近年、筆記枠の制限を行うことなく文字を認識する“枠無し文字認識”へのニーズが高まっている。このような筆記枠の制限がない環境下では、文字だけでなく図形や数式なども同時に書かれる場合が数多く想定される。しかし、現状では文字、図形、数式に対する認識手法がそれぞれ異なるため、これらのパターンが混在したデータを処理するには、あらかじめ適用する処理ごとにパターンを分類する必要がある。本稿では、パターンの分類の自動化に対する試みとして、文字、図形、数式のそれぞれのパターンにおけるストロークの特徴を解析し、その結果を示す。

## 2. 解析に利用した手書きパターン

本稿では、我々が収集した筆記枠無しオンライン手書きパターンデータベース[1]に含まれる 5 人分のパターンセット(パターンセット 1-5)に対して解析を行う。各パターンセットは文字、数式、及び図形のパターンを含む。パターンセットごとに含まれるストロークの数は表 1 の通りである。個々のパターンセットは複数のページから構成され、各ページは筆記内容別に、①文字のみ、②数式のみ、③文字、図形が混在、④文字、図形、数式が混在、の 4 種類に分けられる。

表 1 解析したパターンセットのストローク数

パターンセット	1	2	3	4	5
文字	11056	9223	8636	10087	10895
数式	487	580	487	655	642
図形	257	355	258	333	281

Analysis of stroke features in on-line handwritten patterns without any writing box.

Shinjiro Kihara, Takahiro Fukushima, Masaki Nakagawa  
Tokyo Univ. of Agriculture and Technology, 2-24-16 Naka-cho,  
Koganei, Tokyo, 184-8588, Japan

## 3. ストローク個々の特徴解析

各パターンセットに対して、ストロークの長さ、外接矩形の長辺の長さ、及び運筆速度の分布を調べた。それぞれの分布に対する近似曲線を図 1~3 に示す。ストロークの長さや外接矩形の長辺の長さに関しては、文字、数式は短い方に偏りがあり、図形はほぼ均等に分布していることが分かった。また、数式は文字よりも図形のそれに近い分布を示した。これは、分数線や $\sqrt{\quad}$ 記号など、1 ストロークが長いパターンの影響を受けたためと考えられる。運筆速度についても同様に、図形ストロークの速度は他に比べて高くなる傾向が見られた。

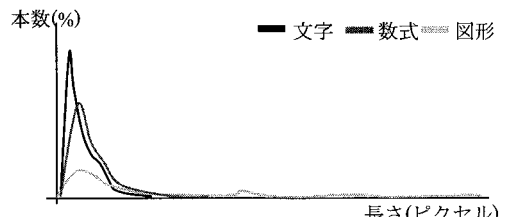


図 1 ストロークの長さ分布

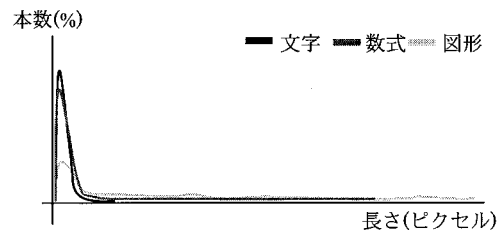


図 2 外接矩形の長辺の長さ分布

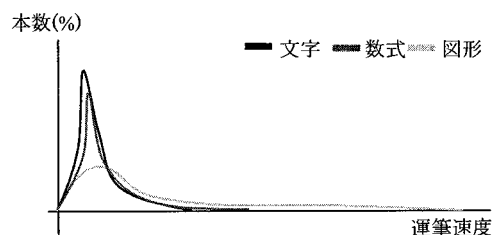


図 3 運筆速度の分布

次に、ストローク長の平均値と標準偏差について調べた。パターンセット1に対する結果を図4に示す。この結果から、文字や数式だけを書いた場合に比べて、図形も含まれている場合には平均値と標準偏差の関係に違いが見られた。これは、図形を構成するストローク長の分布のばらつきが大きくなるためと考えられる。他のパターンセットに対しても同様の傾向が見られた。

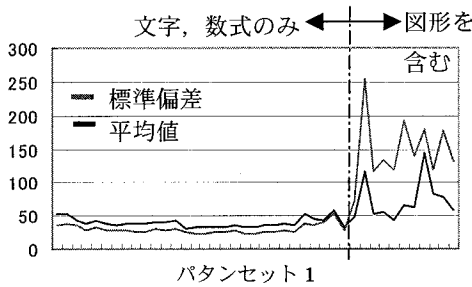


図4 ストローク長の平均値と標準偏差

#### 4. ストローク相互関係の解析

ここでは、各パターンセットにおける図形ストロークとの接触頻度を調査した。結果を表2に示す。その結果、文字、数式ストロークは図形ストロークと交わることは極めて少ないことが分かった。

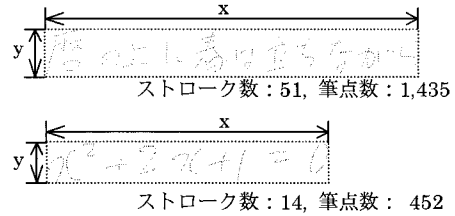
表1 図形ストロークと他ストロークの接触関係

パターンセット	1	2	3	4	5
文字	8	4	0	3	11
数式	0	2	0	0	0
図形	132	115	132	203	124

#### 5. ストローク列単位での解析

数式は一部の例外を除いて数字、アルファベット、及びギリシャ文字の組合わせで構成される。このため、個々のストロークの比較だけでは、文字列との区別が難しいと考えられる。そこで、ストローク列単位で見たときに、ストローク列外接矩形の大きさに対するストローク数、及び筆点数の含まれる頻度を調査した(図5)。結果を図6、7に示す。その結果、数式は文字列に比べて、ストローク列外接矩形内に含まれるストローク数、及

び筆点数が少ない傾向があることが分かった。



$$\text{ストローク数頻度} = \text{ストローク数} / (x \times y)$$

$$\text{筆点数頻度} = \text{筆点数} / (x \times y)$$

図5 文字及び数式のストローク列の外接矩形

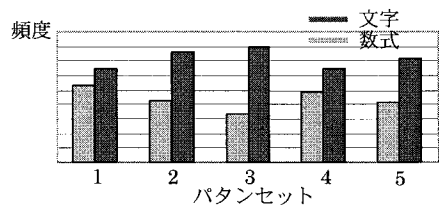


図6 文字列外接矩形内平均ストローク数頻度

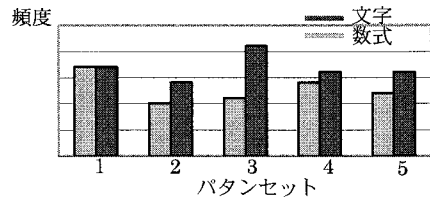


図7 文字列外接矩形内平均筆点数頻度

#### 6. おわりに

枠無しオンライン手書きパタンの文字、図形、数式におけるストローク特徴の解析を行った。

今後、これらの解析結果を基に効率的な分離手法や認識手法の考案をしていく必要がある。また、本報告とは異なった観点からの解析を行うことも重要である。

#### [参考文献]

[1]木原,中川,福島:"枠無しオンライン手書きパターンデータベース",情処 62 回全国大会 1M-6,2001,3.  
 [2]待井,福島,中川:"オンライン手書き図における文字と図形の分離",情処 45 回全国大会 4G-3,1992,10  
 [3]A.K.Jain and A.M.Namboodiri:"Structure in On-line Documents",Document Analysis and Recognition,844-8,Sep,2001