

## 手書き風フォントの自動生成について

5S-1

安本 護\*\*\* 池田尚志\*\* 豊倉完治\*

\*三洋電機株式会社 \*\*岐阜大学工学部

## 1 はじめに

マルチメディア環境の進展に伴い、今後は従来のような画一的なフォントだけではなく、ユーザごとの個性を表現できる文字デザインが要求されると考えられる。しかしながら、膨大な数の日本語フォントをユーザごとに入力することは事実上不可能である。そこで、我々は、ヘンやつくりなどの部分字形をユーザの手書き文字から抽出し、それを組合せることによって手書き風フォントを自動生成することを考えた。

## 2 字形データベース

漢字は部首等の部品を特定のパターンにしたがって並べることにより構成されたものと考えられる。ここでは、漢字を構成する部品を部分字形、部分字形の並べ方を構成パターンと呼ぶことにする。構成パターンには、漢字を横方向に分割するもの、縦方向に分割するもの、カマエやニョウのように文字を囲む周辺部分とその内部に分割するもの、さらに以上の組合せによるものがある。

我々は、JIS X0208に含まれる漢字 6355 文字の部分字形と構成パターンを分析し、階層的な字形データベースを構築した。すなわち、漢字は部分字形を組合せたもの、また、部分字形もさらに下位部分字形を組合せたものとして字形データベースに登録した。したがって、ある漢字に使われている部分字形が他の漢字にも使用される場合は、下位部分字形に分解可能でも分解せずにそのまま登録し、部分字形登録の際に下位部分字形へ

の分解を行なった。この結果、字形データベース作成にかかる作業量を削減するとともに、人手による分析結果の揺らぎをなくすことができた。

## 3 手書き文字データの収集

我々が生成対象としている手書き風文字は、ボールペンで書いたように文字の骨格だけで表されるものである。そこで、入力デバイスにはペン入力のディジタイザを使用し、ペンの軌跡を時系列の座標データとして収集した。システムの構成は、UNIX ワークステーション<S4-330>にディジタイザを接続したものであり、X-Window 上にデータ収集用のプログラムを作成した。図1にデータ収集中の画面を示す。なお、ディジタイザのサンプリング速度は20点/秒、文字は25mm角の矩形枠内に記入した。手書き文字の座標データと字形

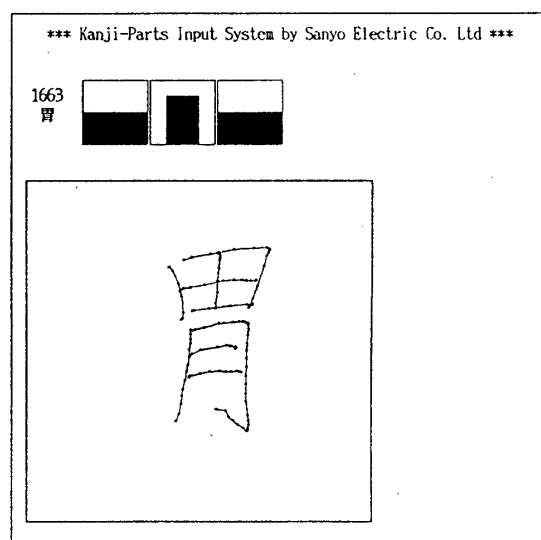


図1: 手書き文字データ収集システム

## A Generation of Kanji Script

Mamoru Yasumoto\*\*\*, Takashi Ikeda\*\*,  
Kanji Toyokura\*

\*Sanyo Electric Co., Ltd.,

\*\*Faculty of Engineering, Gifu University

データベースの対応づけを手書き文字データの入力後に行なうことは非常に困難である。そこで、データ入力時にこの対応づけを行なうこととし、図1に示すようにシステムの入力要求している部

分字形が構成パターン中のどの箇所であるかを反転表示する方法を取った。以上のシステムを使用し、常用漢字 1945 文字の手書き文字データを収集した。

## 4 標準部分字形の抽出

標準部分字形とは、文字生成に用いる部分字形の形状データであり、部分字形コードによって指定される。我々の作成した字形データベースは、同一概念を持つ部分字形には原則として1つの部分字形コードしか与えていない。しかし、部分字形「禾」のようにヘンの位置に来る場合「秒」とカンムリの位置に来る場合「委」で、形状の変化するものもある。そこで、このような形状変化に対応するために、部分字形の出現位置に基づく以下の4種類の属性を付加した。

- P 属性 横方向分割の最も左(ヘン)。
- C 属性 縦方向分割の最も上(カンムリ)。
- F 属性 周辺と内部に分割されるとき周辺の。
- E 属性 その他。

部分字形を単独に入力して標準部分字形を作成するよりも、漢字として入力されたものから部分字形を抽出し、標準部分字形を作成するほうがより自然な形状のデータを得られると考えた。さらに、標準部分字形は、できるだけ少ない文字から抽出できたほうが文字入力の負担を軽くできて好ましい。そこで、以下のアルゴリズムによって標準部分字形抽出に用いる漢字を絞り込んだ。

### <アルゴリズム>

- (i) 文字ごとに含まれる相異なる部分字形を数えあげ、その数を  $N$  とする
- (ii)  $N$  が最大の文字  $C$  をサーチする
- (iii) *if*  $N \neq 0$   
*then* 文字  $C$  を標準部分字形抽出の対象に加える  
*else* 終了
- (iv)  $N > 0$  である全ての文字について、文字  $C$  に含まれる部分字形を取り除く
- (v) *goto* (i)

このアルゴリズムにしたがって、標準部分字形抽出に必要な漢字を求めた結果を表1に示す。このようにして、決定された漢字の手書きデータから外接四角形の面積が最大となる部分字形を抽出し、標準部分字形とした。

表 1: 部分字形抽出に必要な文字数

生成対象文字	出現属性なし	出現属性あり
常用漢字	259(13.3)	411(21.1)
JIS 第1水準	285(9.6)	481(16.2)
JIS 第2水準	349(5.5)	687(10.8)

カッコ内は、生成対象文字集合全体に対する割合(%)

## 5 手書き風文字の生成

手書き文字データを構成する部分字形の外接四角形に合わせて標準部分字形をスケールし、配置することにより、手書き風文字の生成を行った。図2は手書き風文字の生成例である。上から順に出現属性なしの場合、出現属性付きの場合、オリジナルを示している。部分字形相互の位置ずれが発生しているため、オリジナルの文字品質には及ばないが、出現属性を利用することで、かなり自然な文字生成結果の得られることがわかる。

ニ 洋 電 機 研 究 秒 季  
 ニ 洋 電 機 研 究 秒 季  
 ニ 洋 電 機 研 究 秒 季

図 2: 文字生成の例

## 6 まとめ

漢字字形データベースと手書き文字データから抽出した部分字形データを用いて手書き風文字を自動生成する手法を提案し、常用漢字 1945 文字の生成実験を行なった。今後の課題としては、生成文字品質の向上ならびに標準部分字形抽出に必要な文字数を削減することがあげられる。