

## 印字を対象とした書体名称スクリーニングアプリケーションの開発 Development of the Font Name Screening Application for Printed Characters

根本 訓央<sup>†</sup>  
Norio Nemoto

### 1. はじめに

印刷物の印字の書体(フォント)は、ワープロソフトの設定、ソフトウェア付属フォントの利用、印刷時のプリンタフォントによる置き換え等、コンピュータ及び周辺機器の環境並びに設定によって変化するため、印字のフォント名称を識別することで印刷物を作成した機器等を絞り込むことができる。これは科学捜査分野において用いられる手法であるが、印字からフォント名称を識別するアプリケーションはほとんど無いため、目視にてフォント形状を比較、識別する手法が一般的である。一方、同検査手法では対照するフォント数が数百種以上に上ることから、識別には多くの時間を要する問題点がある。よって、印字からフォント名称の候補をスクリーニング及びソート処理するアプリケーションを開発し、符合する確率の高いものから順に比較することで識別の簡略化を実現する。なお、この研究では検査対象となる印字について、日本語フォントのひらがな文字に限定して実験、開発を行った。

### 2. スクリーニング手法

スクリーニングには、文字に外接する四角形を用いる。常に文字と外接するよう四角形を回転させると同四角形の面積が傾きにより増減し、また、その増減の幅はフォントの種類及び文字種ごとに異なる。この現象を利用して印字に外接する四角形を $\theta$  ( $\theta=0,1,2,\dots,88,89$ )度回転させ、その面積に対する印字の画素数の割合を $y$ 軸、回転角度を $x$ 軸とするグラフにプロットし、最小二乗法により連続する10点間の割合データの回帰直線の傾き値を求め、前もって計測した各フォントにおける傾き値データと繰り返し比較することで候補の絞り込みを行う手法を提案した[1](以下、前手法と呼称する)。しかし、割合値のみを比較する同手法では印字と各フォントデータ採取時の文字サイズの差やジャギー(印字曲線部のギザ)、印字精度の影響を受け識別率が低下することが分かった。よって、本研究では画素数の比率の比較手法を再検討すると共に、印字の重心、及び、外接四角形の辺の比、印字の画線の太さを新たなスクリーニングの指標として用い、識別率の向上を図った。

### 3. スクリーニングの流れ

#### 3.1 画像の閾値の決定

まず、検査すべき印字画像に対してグレースケール化処理を行い、印字及び印字周囲の濃淡値を計測する。続いて、文字と背景を分ける濃淡値の分散が最大となる閾値を判別分析法で求める。その後、同閾値を基準として画像中から印字部分を決定し、印字の上下左右部の端点に接する四角形を描き、画素数、及び、四角形の横辺に対する縦辺の比、面積、印字の重心位置の計測を行う。

#### 3.2 外接四角形の回転、計測

続いて外接四角形を $\theta$ 度回転させ、前項と同様に回転後の四角形の辺の比及び面積を計測する。回転手法は以下の通りである。傾き $0$ 度の外接四角形の中心点 $(x_0, y_0)$ を基準として、四角形の各頂点 $(x, y)$ から時計回りに $\theta$ 度回転した座標 $(x', y')$ について、次の座標変換の式①から求める。

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} \cdots \textcircled{1}$$

ここで、回転後の座標同士を結ぶ各線分が印字の端点に接する式 $y=ax+b$ を求めた後、線分同士が直交する座標を求め、同座標が $\theta$ 度回転後の頂点となるよう座標同士を線分で結ぶ。これにより、印字に対して $\theta$ 度回転した外接四角形が描かれる。

#### 3.3 外接四角形面積に対する印字面積割合の算出

外接四角形の回転及び計測後、外接四角形面積に対する印字面積の割合を算出する。外接四角形の傾きに関わらず印字部の面積は常に一定であるため、外接四角形の面積で印字画素数を除算することで外接四角形面積に対する印字部面積の割合値(以下、検査データと呼称する)が求まる。

#### 3.4 フォント名称のスクリーニング

##### 3.4.1 印字部面積割合の比較

外接四角形面積に対する印字部面積の割合を $y$ 軸、外接四角形の回転角度を $x$ 軸とするグラフは、同一種類のフォント及び文字種においてほぼ同じ軌跡を描く。また、印字の大小により生じることのある割合値の差は小さく、また差の値は角度に依らずおおむね一定である。よって、前もって計測した各フォントの印字面積割合値(以下、フォントデータと呼称する)から検査した印字の角度 $0$ 度の割合値を減算した値について、回転角度 $1\sim 89$ 度のフォントデータから減算し、各回転角度の検査データとフォントデータのユークリッド距離が最も小さいものから順に点数付けする。また検査データとフォントデータの割合差が一定値内に属するものについて加点をを行う。

##### 3.4.2 印字の重心位置の比較

印字の形状は電子写真方式やインクジェット方式等のプリンタの種類により若干の誤差は生じるものの、ほぼ一定であることから重心位置についてもおおむね一定であると考えられる。よって、印字を成す画素の位置 $(x_i, y_i)$  ( $i=0,1,2,\dots,n-1$ )の平均値(重心)を以下の式②から求め、各フォントの重心位置と比較し、重心位置が近いものから順に点数付けを行う。

$$\left( \sum_{i=0}^{n-1} x_i / n, \sum_{i=0}^{n-1} y_i / n \right) \cdots \textcircled{2}$$

<sup>†</sup> 茨城県警察本部科学捜査研究所  
Forensic Sci. Lab., Ibaraki Pref., Police H.O.

### 3.4.3 外接四角形の辺の比の比較

外接四角形の横辺に対する縦辺の比の値はフォントの種類及び文字種間でそれぞれ差を生じ、また外接四角形の回転角度により値が変化する。よって角度 $\theta$ 度時の外接四角形の辺の比の値について、フォント種別毎の比の値と検査すべき印字の比の値のユークリッド距離を求め、その値が小さいものから順に点数付けを行う。

以上の比較結果の累計点数が高いものから順にソートし、形状が一致する確立の高いフォントをスクリーニングする。

## 4. アプリケーションの開発

今回開発したアプリケーションでは、印刷物画像中の任意の印字を枠線で囲み、フォント名称をスクリーニングする。よって、複数の文字の印字ある印刷物において1文字ずつ印字画像を撮影する、あるいは、切り出す必要がなく、印刷物全体画像からの連続した検査が可能である。なお、スクリーニング結果はテキスト及び文字画像で表示される。また、スクリーニングの所要時間は約2秒(比較対照数350フォント時)である。

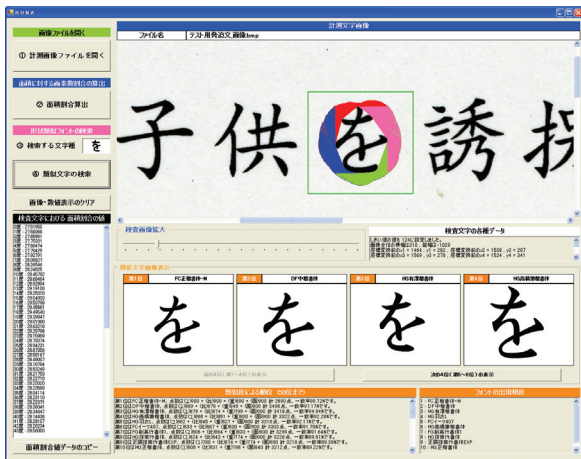


図1. 書体名称スクリーニングアプリケーション外観

## 5. 対照用データの収集

対照用データの収集には図2に示すアプリケーションを使用する。同アプリケーションは濁音、半濁音を除くひらがな文字46文字種について、3.1項、3.2項と同一の処理の流れを用いて自動的にデータ収集を行う。なお、データ収集するフォントは、リスト化したコンピュータ中のフォントから任意に選択する方式とした。また、データ収集にかかる時間は、1フォントにつき約2分である。

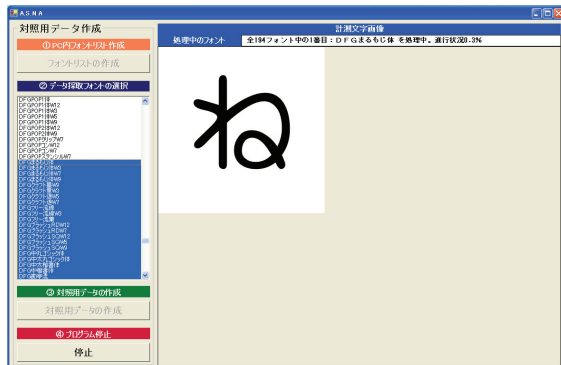


図2. フォントデータ自動収集アプリケーション外観

## 6. 性能評価実験

20～50歳代の当県科学捜査研究所職員男女10名に対してフォント名称の識別実験を行った。実験では従来の手法である目視によるフォント形状の比較検査、並びに、アプリケーションを使用した検査を実施し、所要時間、労力、識別の難易度について比較した。

### 6.1 実験要領

実験は80書体で印字された「お」字(文字サイズ18pt)の印字ある用紙を被験者に配布し、「お前の子供を誘拐した」等の印字ある模擬脅迫文(文字サイズ22pt)中に使用されているフォントを上記用紙から識別する方法とした。なお、アプリケーションによる検査では、あらかじめ模擬脅迫文を解像度600dpiで撮影した画像を使用して識別を行った。

### 6.2 実験結果

実験の結果、識別に掛かった検査時間について、目視による比較では5～10分程度の時間が最も多かったが、アプリケーションによる比較では全員が3分以内に識別できていたことが分かった。また、識別の難易度、労力についても10名全員に改善が見られた。なお、目視による手法では「形状の類似するフォントが見分けづらい」、「目が疲れる」等の意見があったが、アプリケーションによる比較では「客観的で非常によい」、「使いやすく簡単な操作で識別できた」等の意見が得られた。これより同アプリケーションによるスクリーニングはフォント名称識別検査において一定の効果が見られたものと判断された。

なお、本手法では前手法で影響のあったジャギー及び印字と対照用データの文字サイズ差の影響による識別率の低下は見られず、スクリーニング結果は常に上位に入った。

## 7. まとめ

本研究では、フォント名称の識別検査に掛かる時間及び労力を減少すべく、以前提案した前手法の指標を見直したアプリケーションの開発を行った。

識別実験を行った結果、今回開発したアプリケーションによる識別は、目視による検査に比べて所要時間、労力、検査難易度、及び、前手法と比較した識別率に改善が見られた。よって、アプリケーションによるフォント名称のスクリーニングがフォント名称識別において有効な手法となるものと判断された。今後は、印字の傾きを補正する手法の開発、及び、対照用データ収集の短時間化を行い、実用化に向け改良する。

### 謝辞

本研究は、独立行政法人日本学術振興会科学研究費補助金奨励研究(課題番号23919017)の助成を受け行った。

### 参考文献

- [1] 根本 訓央, 古川 猛, 川本 英雄, “回転する外接四角形を利用したオフライン文字識別”, 日本法科学技術学会第16回学術集要旨集, pp. 157, 2010.
- [2] 小林 一彦, 菅原 茂, 赤尾 佳則, 関 陽子, “文字認識を併用したフォント検索システムの検討”, 日本鑑識技術学会第5回学術集要旨集, pp. 141, 1999.
- [3] 常盤 公德, 鈴木 崇弘, “文字の数値化に関する研究”, 日本鑑識技術学会第9回学術集要旨集, pp. 180, 2003.