

## 既存フォントに対する不足文字の作成支援システム

大谷 紀子 研究室

2072081 三上 龍之介

### 1. 背景と目的

現在、インターネット上には大量のフォントが存在しており、株式会社モリサワが開発している「モリサワフォント」のみでも、1000種類以上のフォントを取り扱っている。また、橋本[1]により多くの感情が特定の文字形状と関係あることが示されており、多様なフォントはグラフィックデザインやプレゼンテーションの表現に幅を持たせることに繋がると考えられる。しかし、既存フォントには非常用漢字や旧字体などの文字が収録されていないもの、留学生の名前などに利用されている簡体字や繁体字が含まれていないものがあり、用いようとしていた肝心の文字が存在せず、フォントの使用自体を諦める問題が発生している。フォントを自作するにしても、誰もがデザインに精通しているわけではなく、個人が手作業でフォントの不足文字を作成することは難しい。別のフォントで代用したとしても、既存フォントが特徴的な文字形状である場合に統一感を持たせるのは困難である。また、既存フォント生成システムとして手書き文字からオリジナルフォントを自動生成するシステムが存在する。しかし、既存フォントに対する不足文字の作成支援を目的として作成されたシステムではない点やフォント製作会社向けフォント生成補助システムである点から気軽に利用できないという問題点が挙げられ個人的にフォントを利用したい者にとっては扱うことが難しい。

本研究では、既存フォントに対する不足文字作成支援システムの使用による利用可能なフォントの増大を目的としてシステムを構築する。

### 2. システム概要

本システムでは、複数の文字を分解した後、部首や筆画単位で変形・配置を行うことで文字を再構築する。文字の再構築に際して様々な部首や筆画が必要となることから文字の分解には OpenCV の `connectedComponents` 関数と `Density-based spatial clustering of applications with noise` (以下 DBSCAN) [2] を利用する。`connectedComponents` 関数は、OpenCV においてラベリング処理を行うことができる関数の1つであり、二値化された画像において、画素が繋がっている連結領域ごとにラベルを割り当てる関数である。したがって、二値化された文字画像を入力することで連結している部首や筆画を抜き出して出力する。DBSCAN は密度準拠クラスタリングのアルゴリズムであり、半径以内に存在するデータポイントの数で領域をクラスタとして判断し、近傍の密度がある閾値を超えている限り、クラスタを成長させ続ける。データポイントの最大距離の閾値である `eps` とクラスタと見なすために必要な最小のデータポイント数である `min_samples` を調整することで部首や筆画をクラスタとして抜き出すことができる。また、DBSCAN の長所として他のクラスタリング手法である `k-means` 法と違い、最初にクラスタ数を決めずともクラスタを作成することができるという点がある。例として「鬱」をラベリング処理した結果を図1に、クラスタリングした結果を図2に示す。また、ユーザは Google Slides を利用して抜き出された部首や筆画を変形・配置し文字を作成する。システムの処理手順を以下に記す。



図1「鬱」をラベリング処理した結果

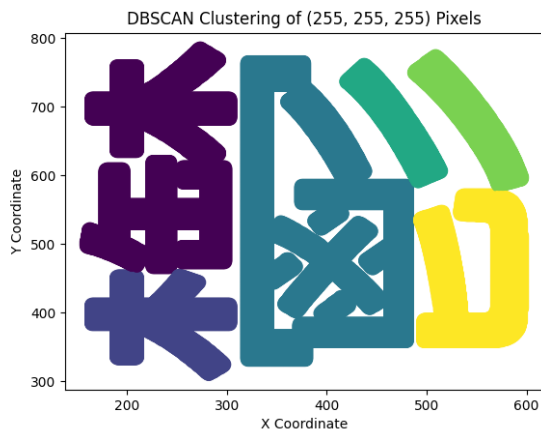


図2「鬱」をクラスタリングした結果

- ① 不足文字を構成する筆画や部首が含まれる文字を二値化する。
- ② 二値化された画像についてラベリング処理とクラスタリング処理を行い部首や筆画を出力する。
- ③ 出力された部首や筆画をユーザが組み合わせ文字を再構築する。

### 3. 評価実験

評価実験では、大学生6名を被験者として、任意の漢字が不足していると仮定し、本システムで文字を作成させた。システムの利用性についての項目としてインタフェイスや操作手順が直観的であったか、文字の分解についての項目としてアルゴリズムの出力結果に違いを感じたか、筆画や部首の抜き出しが適切であったかなどを1~5の5段階で評価させた。また、フォント内に不足文字

があった経験やシステムの改善点、感想などを自由記述で回答させた。

評価実験の結果、システム使用時に半数以上の参加者がストレスを感じたと回答した。文字の分解については、画素連結領域によるクラスタリングと DBSCAN によるクラスタリングで分解結果の違いを体感した被験者は少数に留まったが、分解が不十分であるという意見はなかった。また、自由記述欄には文字を配置する際の大きさを数値により指定したいという記述があった。

### 4. 考察

評価実験より、本システムを使用することでフォント内に収録されている文字の筆画から不足文字を作成することができたと考えられる。また、画素同士が連結している場合でも筆画の先端が細化した形状の場合、DBSCAN で分解した際に半径以内のデータポイント数が閾値を超えず、別々に分解される。したがって、文字形状が特殊なフリーフォントを分解する場合、筆画の先端が細化している可能性が十分にあることから、アルゴリズムの出力結果に多様性が生まれると考えられる。しかし、本システムには「れっか」や「さんずい」のような複数の連結領域からなる部首や、距離が極端に近い筆画同士を個別に抜き出すことが困難という問題がある。したがって、画素連結領域によるクラスタリングと DBSCAN によるクラスタリングとは異なる結果を出力するアルゴリズムを採用することで、文字の分解精度が改善されると考えられる。

### 参考文献

- [1] 橋本共生, “文字形状と感情の関連性に関する研究”, 武蔵工業大学環境情報学部情報メディア学科卒業論文, 2006.
- [2] M. Ester, H. Kriegel, J. Sander, X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise” Proceedings of KDD96, pp.226-231, 1996.