

文字の部分的な特徴量による筆者の自動判定

情報科学科 田中 秀

指導教員：何 立風

1 はじめに

遺言書などの筆者を識別する際に重要な証拠として筆跡は注目を浴びることが多いが、目視による識別には証明力に限界がある[1]。また、同じ人が同じ文字を書いた場合でも、全く同じ文字になるという事はほぼなく、どこかに異なる部分が生じてしまう。これを変動性といい、筆者識別の難しさの原因の一つである。

そこで本研究では、筆者識別の精度向上のため、文字の部分的な特徴量に着目した判定方法を提案する。

2 提案手法

提案手法の流れを以下に示す。

1. 文字のデータ化
2. 二値化
3. 細線化
4. 特徴部分ごとの分解
5. 特徴量の抽出
6. 特徴量の記録・更新
7. 筆者の判定

細線化とは、画像中の図形から線幅 1 の中心線を抽出する処理であり、文字認識やパターン認識の分野で多く用いられている[2]。細線化処理においてその図形に求められる特徴としては、線幅が 1 であること、中心線が元の図形の長さを維持していること、交点が歪まないこと、連結性を維持していることなどが挙げられる。細線化の代表的な方法としてヒルディッチの逐次型細線化処理などがある[3]。

特徴部分とは、「とめ」、「はね」、「はらい」などの文字を構成している特徴的な部分を指し、本研究ではこれらの特徴部分を抽出した後、各特徴部分で特徴量を抽出する。図 1 に細線化した文字画像の例と抽出した特徴部分を示す。

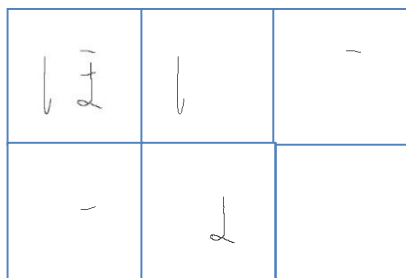


図 1 細線化画像と特徴部分

本研究における特徴量とは、各特徴量の高さ、幅、長さ、線の傾き、はねの角度、丸み、輪の大きさを指す。筆者が判明している文字の特徴量を記録・更新する場合は、筆者ごとの各特徴部分の各特徴量に対してそれぞれ最大値と最小値を記録・更新する。筆者が不明の文字が、ある筆者が書いた文字であるかを判定する場合は、その文字から抽出した特徴量と、あらかじめ記録してある筆者の特徴量の最大値、最小値のデータを比較し、最大値より小さく最小値より大きければ 1、そうでない場合は 0 とする。1 文字ごとに全ての特徴量を比較し、その結果を合計していき、

最終的に合計値から比較した特徴量の総数で除算し、100 倍して小数第 1 位で四捨五入したものを、その文字が比較した筆者と同一である確率(%)とする。

3 実験

複数の人物によって書かれた文字に対して本研究の提案手法を実行し筆者の判定を行うことで、研究目的である文字の特徴部分に着目した筆者判定の精度の向上を図る。被験者 8 人にボールペンで 3 種類の文字をそれぞれ 12 回ずつ書いてもらい、10 回分を筆者が判明している文字として特徴量を記録し、残りの 2 文字を筆者が不明の文字(一方を a、残りを b とする)として、筆者ごとの特徴量のデータと比較し判定するために使用する。表 1 に被験者 A の、筆者が不明の文字を用いた筆者判定の結果を示す。

表 1 被験者 A の文字の判定結果

	被験者 A	被験者 B	被験者 C	被験者 D
て a	100	45	55	100
て b	91	27	36	55

	被験者 E	被験者 F	被験者 G	被験者 H
て a	91	36	82	27
て b	55	27	55	27

筆者が不明の文字が正しい筆者と同一であると判定されたもの(この場合は最も高い確率となった筆者を判定の結果とする)が、被験者 A が 6 文字中 4 文字、被験者 B が 6 文字、被験者 C が 3 文字、被験者 D が 5 文字、被験者 E が 6 文字、被験者 F が 4 文字、被験者 G が 3 文字、被験者 H が 3 文字となった。単純な計算をすれば、全ての被験者において筆者判定の精度は 50%を超えるものとなった。しかし、全ての文字が正しく判定された被験者もいるが、半分しか正しく判定できなかった被験者も存在するため、安定した結果を得られなかった。

原因としては、被験者同士で似た文字を書く傾向が見られる場合や、データ量が少ないことなどが挙げられる。

4 まとめ

実験を通して、いくつかの問題点があることがわかったが、文字画像に対して提案手法が特徴量を抽出でき、一定の精度で筆者が正しく判定されるという結果が見られた。提案手法の問題点を解決し、より筆者の精度を向上させるには、抽出する特徴量の追加、複数の文字での判定、細線化の性能の向上、判定方法の追加が挙げられる。

参考文献

- [1] 吉村ミツ, 吉村功: "筆者認識研究の現段階と今後の動向", 電子通信学会技術研究報告 PRMU96(6), pp. 81-90, 1996-06-28
- [2] 高井信勝: "技術報告 2 値画像の細線化における連結数の役割の再検討", 北海学園大学工学部研究報告, 2011(38), pp. 155-172, 2011-02-14
- [3] 和田潔, 井宮淳: "Hilditch の細線化の並列処理", 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM), 1990(6), pp. 33-40, 1990-01-25