

多種の特徴パラメータによる 花押の画像データベースシステム

高崎浩二[§]柳沼良知[†]林譲[‡]坂内正夫[†][§]中央大学大学院[†]東京大学生産技術研究所[‡]東京大学史料編纂所

1 はじめに

人文系の書物、資料を電子化して保存したり、何か有効利用しようということが考えられる。本論文では鎌倉、室町時代の花押をコンテンツの1つとして取り上げ、花押の画像データベースを構築する。花押画像として多種の特徴パラメータをとり、データベース化することにより、類似検索を試みる。さらにある特定の人物の花押の時系列での形体変化を調べる。これにより、年代不明瞭なものや真偽などを推測することができる。

花押は文字を崩したもののや、蛇のような形体をしたものなどがあり、一様に文字認識をするのは難しい。文字認識としての関連研究は文字のループ、交点、端点などの特徴をとっている。また、文字ではなく意匠、商標を対象にしたもの[2] 画像検索システムがあるが、これらは領域としての特徴を取っており、画像を回転しても対応できる。本論文では花押を文字としての特徴、さらに文様としての特徴も取り入れ、検索システムを構築する。

2 本システムの概要

本システムを構築する上で、次のような方針で行う。

- (1) 花押の画像データをスキャナーで入力する。取り込んだ画像にフィルタを通す。
- (2) 画像の特徴をとり、データベースを構築し、検索システムを作成する。

- (3) ユーザーインターフェイスを設計する。

まず、1では花押カード、花押かがみ[3] から画像データをスキャナーなどで入力する。画像によってはかすれやノイズがあるのでそれらをフィルタを通して正規化し、ノイズを除去している。特徴量として花押画像の大きさなど取るため、同じ解像度で取っている。

2では入力した画像をデータベースに蓄積するのであるが、データベースに格納するのは元の画像を圧縮したものと各画像のそれぞれの特徴量である。特徴量は元の画像と大きさを正規化したものからそれぞれ取っている。検索の方法として、花押の画像を例示してその画像と似たものを取り出すといった類似検索が可能となっている。

3ではインターフェイスの設計であるが、ナビゲート型の検索ができるようになっている。インターネットを介して試験的に実行できるようにしている。

3 特徴パラメータの抽出

花押の形体を認識するためには、画像の大きさ、形状(線、空間)、位置などの指標が必要となってくる。

花押画像は線要素と面要素が重要になるため、それらの特徴パラメータを利用する。花押画像の特徴や類似性を記述するため、特徴パラメータとして以下のものを取り、データベースに格納する。

- 濃度分布：画像をそれぞれ縦、横8分割にして各領域の画素数をあらわす。
- 細線化によって概略形状：花押の内部の形状、その外部の形状を抽出する。部分構成要素である数、形状、大きさ、位置を抽出する。

細線化した画像(図1)から、穴の数、その面積、分岐している枝のそれぞれの数、分岐していない線の長さの合計を抽出している。

Image database system of Kaou using many feature parameters

[§] Graduate School of Science and Technology, Chuo University

[†] Institute of Industrial Science, University of Tokyo,

[‡] The Historiographical Institute, University of Tokyo

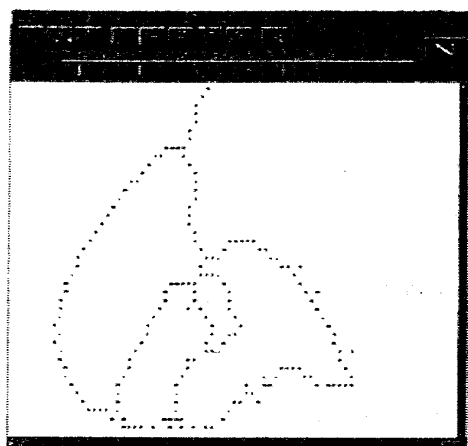


図 1: 細線化画像

本システムでは例示画による類似画像検索が可能になる。ここでの類似度の尺度として各画像のそれぞれの特徴パラメータを例示画像と比較して距離を計算し、例示画との距離が近い順番から候補を出している。結果を図 1 に示す。プロトタイプを本システムでは、画面上にある花押を指定すると 1159 個の花押で実験した検索結果を図 2 に示す。

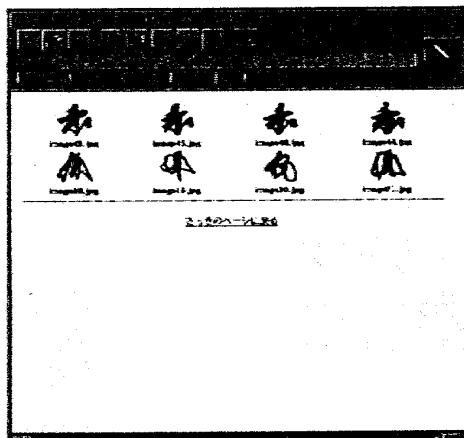


図 2: 検索結果

図 3 に各画像の適合率と再現率を示す。横軸をしきい値にとり、適合率と再現率をす。本システムではしきい値として 0.20 から 0.22 を選ぶことによって検索が可能となっているが、類似度はユーザーによってそれぞれ違うため、ナビゲーションすることによってよりユーザーに近い指標によって、検索することが可能となっている。

また、時系列変化に対応するためにもそれらの特徴を取る。例として足利尊氏の花押をとり、各特徴パラメータ（縦・横比など）の時系列変化ををとり変化を分析している。

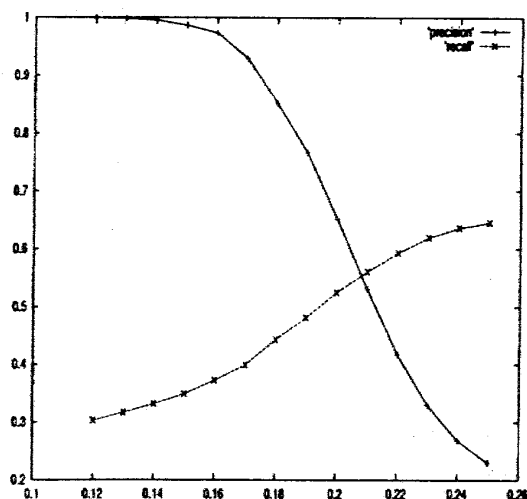


図 3: 適合率と再現率

類似検索は大まかな特徴パラメータを取って重みづけをしたが、尊氏の場合は細かく特徴パラメータを取り、変化の多いものに重みづけをしている。

4 おわりに

花押画像から細線化などの処理をすることにより、多種の特徴量を得ることができ、類似検索が可能になった。また、尊氏の花押を例に、時間的な推移による形体変化を追うことができた。

ただ古文書にある花押はかすれが多く、また、他の文字と重なっていることが多いため、パラメータを取る際に不要な情報が入ることがあるのが課題となる。

参考文献

- [1] 高崎 浩二, 柳沼 良知, 坂内 正夫, 林 譲: 花押画像の電子計算機処理について, 第三十回 日本古文書学会大会研究発表要旨, (1997)
- [2] 加藤, 下垣: マルチメディア商標・意匠データベース TRADEMARK, 電子情報通信学会技術報告, PRU88-9, pp.31-38 (1988)
- [3] 東京大学史料編纂所: 花押かがみ 四 鎌倉時代三 (1985)