

花押を対象とした画像検索システムの構築

石村 隆博・村川 猛彦 (和歌山大学)

概要 : 花押とは、その人物を示す署名 (サイン) である。本研究では花押画像ファイルを入力にとり、花押彙纂データベースの 25,000 枚の花押画像の中で類似する画像を求めるシステムを構築した。類似度算出のための画像分析には、OpenCV を使用し、AKAZE に基づき画像の特徴点を抽出する。2 つの画像間の類似度は、特徴点の距離の合計値を特徴点の数で割ることで算出している。全体の処理は Python で記述し、画像検索のための Web アプリケーションには Flask を活用した。類似度計算と花押画像検索のそれぞれで評価実験を行った。

キーワード : 花押, 画像検索, OpenCV, Web アプリケーション

Construction of a Kaou Image Retrieval System

Takahiro Ishimura / Takehiko Murakawa (Wakayama University)

Abstract: Kaou is a signature of a person. In this study, we constructed a system that takes a kaou image file as input and searches for similar images among 25,000 kaou images in the Kaou Collection Database. We use OpenCV to analyze the images for the similarity calculation, where feature points are extracted from the images based on AKAZE. The system calculates the similarity between two images by dividing the total distance between the feature points by the number of feature points. The entire process of the system was written in Python, and Flask was used for an image retrieval Web application. Evaluation experiments were conducted for both the similarity computation and the image retrieval.

Keywords: kaou, image retrieval, OpenCV, Web application

1. はじめに

花押とは、その人物を表す署名 (サイン) である。空海や源頼朝、足利氏、徳川家康などの歴史上の人物だけでなく、江戸時代には一般庶民も花押を使用しており、現在でも閣議の署名に見られる。

現在、紙媒体の文書や重要文化財に指定されている文化的遺産など様々な歴史的文化的資料は虫食いや日焼けなどで劣化や損傷が進んでいる。この問題を解決するため、文化資源をデジタルアーカイブ化することが行われている。これにより電子的に文化的資料を閲覧することが可能となった。

本研究では花押の検索・閲覧を通じた活用に着目する。花押は、筆者特有のものとなっているが、社会的地位の変化に伴う形状の変更、血縁関係が反映されたデザイン、経年による筆跡の変化などの事象がある。そのため筆者の身分や血縁関係、年代の推定に利用されている[1][2]。

上述した花押画像の特徴から、花押画像の類似画像検索機能の実現は様々な利点をもたらす。まず、収集と管理の省力化である。花押を対象としたデータベースの作成と運用は膨大な作業を要することになりやすい。新たに発見された歴史史料から花押を見出して登録するとなったとき、手作業の労力は今後さらに増加することが見込まれ、労力の削減が不可欠となる。次に、花押画像

の相互関係の解明は、作業者の負担軽減と別の効果として期待される。前述したとおり、花押形状の解析を通じて、歴史上の人物の身分の変遷や他者との関係、文書の作成された年代が推定できる可能性が期待されている。このような分析は、従来人手で行われてきたが、コンピュータビジョンや人工知能のアルゴリズムおよび既存ソフトウェアを活用しながら、形状の類似した花押を検索し詳細を出力することも、求められる。

花押に関して電子化されている資料には、花押譜と花押彙纂データベースがある[3]。

花押譜は、花押が集められた書物であり、600名を超す人物の花押が掲載されている(多くは一人に対し1点の花押となっている)。花押譜は画像化されており、人文オープンデータ共同センターが IIF (International Image Interoperability Framework) の仕様に基づき公開している[4]。

花押彙纂データベースは、東京大学史料編纂所が管理しているもので、主に南北朝時代から織豊時代に活躍した人物の花押、約 25,000 件について、メタデータ(人名、和暦および西暦の年月日、文書名、位署書(花押を記した人物の位階、官職、姓名を一定の規格に従って書かれたもの)、管理番号など)および BMP 形式のモノクロ画像が整備されている。同一人物のレコードが複数存在することもある。図 1 に画像例を示す。なお、花押彙纂データベースに収録された足利尊氏の花押画像は 20 枚ある。

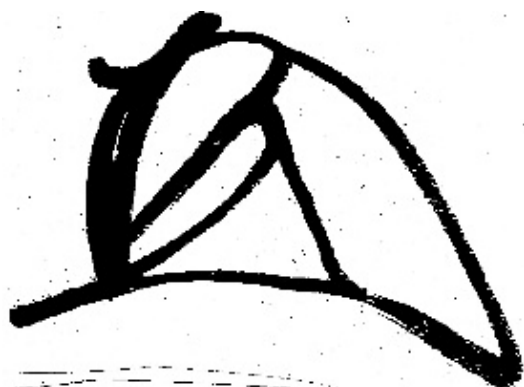


図1 花押彙纂データベースの画像例（足利尊氏）

2. 関連研究

鬼塚ら[5]は畳み込みオートエンコーダによる花押画像解析を行った。オートエンコーダは情報圧縮を行う符号化・復号化アルゴリズムである。通常、オートエンコーダは同サイズの入力層と出力層、およびそれらよりサイズの小さい中間層で構成される。花押の筆者をクラスとして、1,848のクラスを作成し、累積クラス一致率を求め、花押画像の解析を行い、得られた特徴量から類似度の高い花押画像を提示している。

筆者の研究室ではこれまで、花押情報の検索から閲覧までを行うことができ、花押画像を比較・検討ができる Web アプリケーションを構築してきた[6]。花押譜に掲載された画像と、花押彙纂データベースの花押画像を対照表示できる。その後、花押彙纂データベースの検索結果に花押のサムネイル画像を表示する機能を追加したが、検索の方法は人名などをキーワードで指定するものに限られていた。

花押譜では、どこにどの人物の花押があるのか分からない。一方、花押彙纂データベースのメタデータファイルについて、すべてを取り出すのでは件数が多く、表示や検索に時間がかかってしまう。また文献[6]のアプリケーションでは検索の手段が文字入力に限られることに加え、既出の花押にしか対応しておらず、新たに見つけた花押に対する検索支援は十分ではない。これらの課題点を踏まえ、本研究では画像検索を行う Web アプリケーションを新たに開発することとした。

3. OpenCV を用いた花押画像の類似度算出

画像処理に当たり OpenCV (Open Source Computer Vision Library)を採用した。オープンソース・マルチプラットフォームのライブラリで、

画像処理やコンピュータビジョンに関するアルゴリズムが多数含まれている[7]。C++、Python、Java のプログラミング言語から利用でき、本研究では Python で処理を記述した。

入力画像に対する比較画像は花押彙纂データベースの画像を使用した。画像のサイズを縦横ともに 200px のグレースケールに、また画像フォーマットは BMP から JPEG に変換している。

特徴点検出と特徴量記述を行う手法は AKAZE を使用した。入力された花押画像と花押遺産データベースの花押画像の特徴量を抽出し、その特徴量を cv2.BFMatcher (Brute Force Matching) オブジェクトの match 関数に入力し特徴点マッチングを行いマッチングした特徴点を用いて類似度を計算した。図1の画像に対して、特徴点を抽出した例を図2に示す。

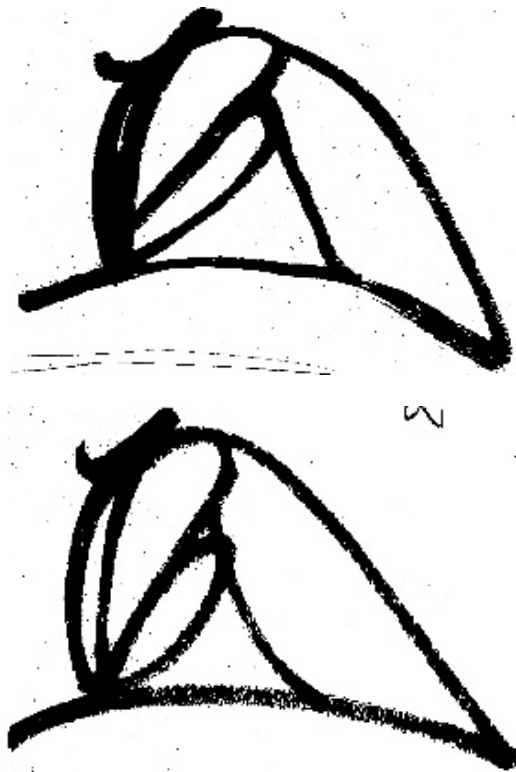


図2 特徴点検出の例（足利尊氏）

特徴点マッチングは、特徴点検出、特徴量記述、マッチングの順に行われる。特徴点検出では、画像の中から角や輝度の変化などの他と異なる特徴的な点の座標を検出する。図1の画像にはモノクロ化で生じた微細な黒の点が多くあるほか、左下には点線状のものも見られる。しかし、それらが特徴点として検出されることは非常に少なく、図2においては中央右に見える1点のみであり、特徴点は花押の筆画およびその周辺に集中している。特徴量記述では、検出した特徴点の固有性をベクトルやバイナリコード表現した値を特徴量として算出する。特徴量には特徴点の X 座標特徴点の Y 座標、特徴点の直径、特徴点の向きの角度、特徴点の強度などが保存されている。マッチングでは、2つの画像同士の特徴量を比較し、特徴量の距離が近いものを類似度が高いとし、特徴点同士を対応付けている。それらの特徴点を用いて、類似度を計算した。類似度の計算方法は以下の通りである。

$$\text{類似度} = \frac{\text{特徴点の距離の合計値}}{\text{特徴点の数}}$$

類似度は値が低いほど類似していることを示し、完全に一致する場合に0となる。花押彙纂データベースに収録された足利尊氏の2つの花押画像に対する類似度算出例を図3に示す。これらの画像間で、特徴点の距離の合計値は19078、特徴点の数は194で、類似度（小数第5位以下切り捨て、以下同じ）は98.3402となった。



類似度：98.3402

図3 花押画像の類似度算出例（足利尊氏）

花押譜から切り出した蘆名盛隆の花押画像（複数あるうちのの一つで、5節の評価実験では2枚を使用しているため、ここでは「蘆名盛隆1」とする）と、花押彙纂データベースの画像の一つ、およびそれらの類似度算出結果を、図4に示す。



(a) 花押譜画像



(b) 花押彙纂データベース画像

類似度：123.5454

図4 花押画像の類似度算出例（蘆名盛隆1）

ここまで述べた処理は、ノート PC（OS: Windows 11, CPU: Intel Core i7-1165G7 2.80GHz）で行い、1枚の画像の特徴点検出ならびに特徴量抽出、および2枚の画像の類似度計算は、いずれも瞬時（プログラム起動の時間を含めても0.1秒以内）であった。

4. 花押画像検索のための Web アプリケーション

本研究で構築した花押画像検索の全体像を図5に示す。まず花押画像を入力し、入力した花押画像の特徴量を抽出する。その後花押遺産データベースの花押画像の特徴量を用いて類似度を計算する。得られた類似度の値から対応する花押画像に付随する情報を取得し、花押画像とその付随情報を表示する。

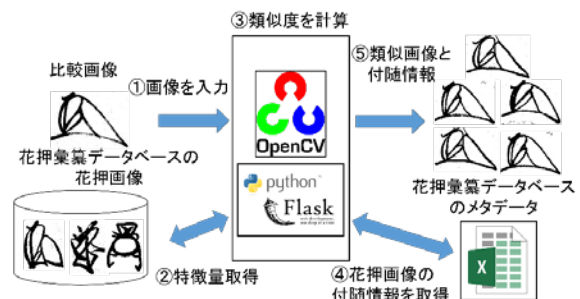


図5 花押画像検索システムの概要

前節で述べた類似度算出プログラムを用いて、検索を行うための Web アプリケーションを試作

した。画像ファイル（検索画像という）を入力にとり、花押彙纂データベースから抜粋した画像約25,000枚のそれぞれと類似度を求め、類似度の小さいものから5件を表として表示する。表に表示される情報は和暦年月日、西暦年月日、文書名、人名、資料群名、位階書、画像名（ファイル名）、花押画像（検索結果画像）、比較画像（検索画像）、類似度である。

図6に検索画面の初期状態を示す。



図6 検索システム初期画面

利用者は「ファイルを選択」ボタンを押して、検索画像のファイルを選択したあと、「POST」ボタンを押す。検索が終了すると、図7のように類似画像とその付随情報が表示される。ここでは図1に示した足利尊氏の花押画像を検索画像とした。

和暦年月日	西暦年月日	文書名	人名	資料群名	位階書	画像名	花押画像	比較画像	類似度
文和4年2月16日	13550020160.0	nan	足利尊氏	足利尊氏所屬文書	〔花押〕	00002535.jpg			4.484536082474227
観應元年7月28日	13500070280.0	nan	足利尊氏	足利尊氏所屬文書	〔花押〕	00002533.jpg			98.34020618556701
康永元年7月13日	13420070130.0	幕府御状	足利尊氏	足利尊氏所屬文書	〔花押〕	00000192.jpg			99.95154639175258
西暦5年12月27日	13490120270.0	nan	足利尊氏	足利尊氏所屬文書	〔花押〕	00002532.jpg			100.1340206185567
正徳元年8月3日	12570080220.0	nan	足利尊氏	足利尊氏所屬文書	〔花押〕	00001054.jpg			100.30412371134021

図7 検索システム結果画面例

図7の表の比較画像は、検索画像の縦横比を崩さずに幅200pxにしたものであり、行ごとに結果

を見比べられるよう、各行とも同じものを配置している。表頭を除いた1行目について、花押画像は、検索画像と同一であり、類似度が約4で、他より大幅に低い。花押彙纂データベースの他の花押画像をアップロードして検索した場合にも、同様の傾向が見られた。同一の画像であるのに0にならないのは、検索画像のサイズ変換が影響したと考えられる。表の他の列の情報を通じて、利用者は、その花押の人物名、記された年月日、文書の詳細などを知ることができる。

計算機内の動作にあたり Flask で Web サーバを起動し、Web アプリケーションとして動作させている。類似度の計算結果をもとに花押彙纂データベースのメタデータファイルから画像および付随情報を取得している。前節で述べたノート PC 上で動作し、1回の検索には約9秒を要する。

5. 評価実験

5.1 類似度計算

提案手法による花押画像検索の有用性を検証するために、評価実験を行った。ここでは類似度計算に関する実験を述べる。対象とする人物は、花押譜に1枚以上、花押彙纂データベースに6枚以上の花押画像がある者から選定した。蘆名盛隆については、花押譜に2枚の花押が掲載されていたので、「蘆名盛隆1」「蘆名盛隆2」として別々に画像を取り出し、合計9名を選んだ。それらの花押譜画像ごとに、花押彙纂データベースの同一人物の花押画像それぞれと類似度を算出した。別途、筆者の一人（石村）が目視で比較を行い、最も類似すると判断した、花押譜画像と花押彙纂データベースの画像の組と、類似度最小となった組とが一致したかどうかを調べた。結果を表1に示す。この表において「枚数」とは、花押彙纂データベースの該当人物の花押画像数をいう（表2も同様である）。「目視一致」の列には、目視で最も類似すると判断した組と、類似度最小の組が同一のときは○、そうでない場合には×を記した。

表1 類似度画像計算の結果

人物名	枚数	類似度 最小	類似度 最大	目視 一致
吉良治家	7	113.8333	127.7042	×
斯波義淳	8	127.7207	138.8649	○
武田信武	6	125.4414	139.3394	○
小笠原長基	6	109.4831	129.6250	○
尼子勝久	11	126.6666	134.9271	×
小山田信茂	16	114.3428	124.9048	○
蘆名盛隆1	58	123.5000	142.7051	×
蘆名盛隆2	58	123.7948	140.2901	×
上杉能憲	11	124.6796	131.4563	○

類似度最大と最小の差について、平均は13.3727, 最大は20.1419, 最小は6.7767となった。この差の大小と、目視一致の成否の間には、関係性が見られなかった。

蘆名盛隆2の画像について、花押譜から切り出した画像、花押彙纂データベースで類似度最小の画像、花押彙纂データベースで最も類似しているように見える画像を、図8に示す。



(a) 花押譜画像



(b) 花押彙纂データベースで類似度最小の画像 (類似度: 123.7948)



(c) 花押彙纂データベースで最も類似しているように見える画像 (類似度: 125.5256)
図8 目視一致の失敗例 (蘆名盛隆2)

蘆名盛隆について、花押彙纂データベースに登録された画像は58枚あり、似た形状をした花押画像が多く、判定が成功しなかったと考えられる。

5.2 花押画像検索

続いて、4節で述べた検索システムで類似する花押画像が出力されるか評価するため、花押譜からいくつかの花押画像を抽出しシステムに入力した。また花押画像の選定にあたり、10枚以上登録されている人物や、1枚しか登録されていない人物の花押画像など幅広く花押画像を選定した。

表2にその結果を示す。1から5までの列は、

検索結果画面の1位から5位までが、比較画像(花押譜の画像)の人物名と合致している場合に、丸を付けている。

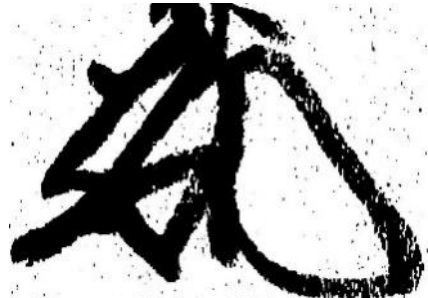
表2 花押画像検索の結果

人物名	1	2	3	4	5	枚数
吉良治家	○					7
吉良氏朝	○	○				3
今川範政						5
斯波義淳						8
斯波義雄	○					1
山名致豊					○	4
佐竹義篤					○	10
武田信武				○		6
小笠原長基	○	○				6
尼子勝久						11
蘆名盛隆1	○	○				58
蘆名盛隆2						58

計12枚の花押図の画像を入力に与え、1位に該当人物の花押画像が表示されたのは5/12, 上位5件までに表示されたのは8/12となった。斯波義雄や吉良氏朝といった5枚に満たない花押画像で上位5件に表示されたが、尼子勝久や蘆名盛隆2のように、数十枚登録されていても、上位5件に表示されないものもあった。図9に成功例を、図10に失敗例を示す。



(a) 花押譜画像



(b) 花押彙纂データベース画像
類似度 111.8333

図9 成功例 (吉良治家)



(a) 花押譜画像



(b) 花押彙纂データベース画像

類似度 120.1428

図 10 失敗例 (尼子勝久)

上位 5 件に表示されなかった理由として、次の 2 点が考えられる。一つは、花押譜の画像に対してグレースケール変換を行っているが、背景が白でないため、前後のページの裏写りした内容が影響を及ぼしている可能性である。もう一つは、花押彙纂データベースの花押画像には汚れがあるため、汚れが特徴点として抽出され、類似度がうまく計算されない可能性である。

6. おわりに

本研究では、花押画像における類似画像検索システムを開発し、類似度計算の妥当性と、画像検索に関する評価実験を行った。しかし実用性は十分と言いがたい。花押画像間の類似度計算をする際に縦横 200px に変換しているが、元の花押画像の縦横比を崩さずに類似度を計算することでより類似性のある花押画像が表示される可能性がある。

Web アプリケーションの利用に関して、現状では検索画像をファイルとしてアップロードする必要があるが、デスクトップ環境で保持・提供しているクリップボードからも獲得できれば利便性が向上する。また文献[6]で述べた検索システム

と統合し、多種多様な方法で花押を検索閲覧できるようにする。

検索時間の削減も行う必要がある。特徴量に基づく本手法のもとでは、例えば、花押彙纂データベースの花押画像に対してクラスタリングを行い、類似する画像ごとにファイルに画像を振り分けるといったアプローチが考えられる。それとは別に、近似最近傍探索ライブラリである Annoy への置き換えも候補となる。約 25,000 枚の花押画像から瞬時に結果を返せるよう改良する。

謝辞

花押彙纂データベースの画像およびメタデータは、東京大学史料編纂所から提供を受けたものです。ここに感謝申し上げます。本研究は JSPS 科研費 20H00022 の助成を受けたものです。

参考文献

- [1] 松尾剛次: 最上義光花押再考, 山形大学紀要 (人文科学), Vol.19, No.3, pp.1-22 (2020).
- [2] 浜野真由美: 和歌賛から見た島津家旧蔵「架鷹図屏風」の成立事情—近世初期の宮廷における寄合書の観点から, 日本研究, Vol.61, pp.7-43 (2020).
- [3] 林譲: 花押・筆跡データの網羅的収集と汎用的利用をめざして, 情報処理学会研究報告, Vol.2017-CH-115, No.10, pp.1-4 (2017).
- [4] 花押譜 | 日本古典籍データセット. <http://codh.rois.ac.jp/pmjt/book/200017283/> (2022-11-01 参照).
- [5] 鬼塚洋輔, 大山航, 山田太造, 井上聡, 内田誠一: 花押類似検索のための畳み込みオートエンコーダによる画像特徴抽出, 人文科学とコンピュータシンポジウム論文集 (じんもんこん 2018), pp.257-262 (2018).
- [6] 岡野元春, 村本俊哉, 村川猛彦: 花押を対象とした画像比較システムの構築, 人文科学とコンピュータシンポジウム論文集 (じんもんこん 2021), pp.212-217 (2021).
- [7] 小枝正直, 上田悦子, 中村恭之: OpenCV による画像処理入門 改訂第 2 版, 講談社 (2017).