

小袖屏風を対象とした知的構造の抽出に関する研究

濱上 知樹¹ 澤田 和人²

概要：国立歴史民俗博物館蔵の小袖屏風（野村コレクション）のデジタルアーカイブデータ化と、機械学習・セマンティックデータ処理に基づく知的構造の抽出に関するプロジェクト研究について報告する。特に本稿では本プロジェクトの背景と取り組み全般について紹介するとともに、階層クラスタリングによるモチーフセグメントの抽出と、Deep Learning を用いた一般物体認識に基づく、モチーフセグメントの分類法について述べる。

1. はじめに

世界的な歴史資料のデジタルアーカイブ化が進む中 [1], [2], その莫大なデータの利活用の方法が模索されている。デジタルアーカイブは、デジタルデータとしての永続的な保存や二次的資料価値を有するが、加えてデータの意味や構造まで含めた知識抽出が可能になれば、新たな知見の発見や解釈の支援となる。すなわち、知的情報処理によるデジタルアーカイブの利活用は、今後の人文科学の新たな研究手段となりうる。

この研究の嚆矢として、平成 25 年より国立歴史民俗館所蔵の野村コレクション小袖屏風 [3] のデジタルアーカイブデータを用いた文理融合共同研究が開始された [4]。この研究では、歴史資料デジタルアーカイブを Big data とみなし、高度な画像処理・解析を初めとする特徴抽出と、これらの解析技術で培われた機械学習、セマンティックデータ処理を用いて、莫大なデジタルアーカイブからの知識発見をはかることをめざしている。

本共同研究は、(1) 高精細小袖画像からの特徴点と構造データの抽出、(2) メタデータベースの設計と高度知的検索システム、(3) 非破壊検査技術を利用した分析、(4) アーカイブデータを用いた知的展示・鑑賞方法など多岐にわたる内容を含んでいる。本稿では特に、小袖屏風画像の特徴量の抽出とその構造化、意味抽出の方法について述べる。

2. 小袖屏風デジタルアーカイブ

小袖とは、平安時代の貴族の下着が上着に変化し、現在

の着物の直接的な源流となった衣服である。桃山時代から江戸時代にかけて小袖の文様はその時代背景と強く結びつきながら変化し、様々な文様・配置・技法が時代とともに変遷してきた。

小袖屏風とは小袖を二曲一隻に仕立てた屏風である。国立歴史民俗博物館には、野村コレクションの一部である約 100 隻が収蔵されている。辻が花、縫箔、慶長小袖、寛文小袖、元禄小袖、友禅染などの小袖の諸相を網羅しており、芸術品としてはもちろん、染織史研究にとっても貴重な資料価値がある。

小袖の絵柄は様々なメタ情報を含んでいる。分類上の特質（文様、地色、地質、諸技法）のほかに、メタファー、背景にあるストーリー、マクロからミクロに至る構図上の特徴などの非言語的な情報も含んでいる。これらの小袖屏風のメタ情報が与える印象の類似性や、陽に現れない部分の共通性などが抽出できれば、小袖屏風に含まれる隠れたセマンティクスや新たな構造の発見につながる。また、小袖屏風の資料分析においては、小袖表面の柄や織、文様等の様々な特徴をもとに、他資料間との類似性や連続性に着目した分析が行われる。屏風上に貼られた小袖は、元来の様子のままとは限らず、断片からの再構成や創作となっている場合も多い。これらの制作過程までを含めた資料をまたぐ情報の構造が抽出できれば、これまでにない展示、検索等への応用が期待できる。

以上のことを背景に、本研究では、小袖屏風のデジタルアーカイブ化と知的データベース化を進めている。特に著者らは、アーカイブデータの高精細画像を機械学習により分析することで、アーカイブに含まれている意味や構造の抽出をはかる方法について検討を進めている。

¹ 横浜国立大学大学院工学研究院
Yokohama National University, Yokohama, 240-8501, Japan

² 国立歴史民俗博物館情報資料研究系
National Museum of Japanese History, Sakura, 285-8502, Japan

3. 特徴点抽出と構造分析

画像中の特徴抽出とその構造分析に関する研究は、画像検索 (IR: Image Retrieval) 技術として発展してきた [5]. IR には、以下の3つのカテゴリが知られている。

(1) Text Based Image Retrieval (TBIR)

画像に付与されたキーワードやアノテーションをもとに、これらのテキストの意味空間において画像特徴をマッピングし、検索を行う。精度の高い分類が可能であるが、キーワード・アノテーションの付与は人が行う必要があり、特徴の自動抽出には限界がある。

(2) Content Based (Visual) Image Retrieval (CBIR)

画像コンテンツの種々の特徴をもとに、その特徴空間中での比較に基づく検索を行う。画像から直接抽出可能な色や輪郭の統計的性質やパターン情報を用いるため、大量のデータを自動でマッピングすることが可能である。一方、物理的な画像特徴は、見た目の類似性しか評価できず、対象の意味的クラスとの乖離 (semantic gap) が生じやすい点が問題である。

(3) Semantic Based Image Retrieval (SBIR)

SBIR は、CBIR におけるセマンティックギャップの課題を解決することで、意味的空間へのマッピングによる検索を可能にする手法である。TBIR で得られた意味的特徴と CBIR で得られた物理的特徴を合わせることで、物理的特徴空間から意味的特徴空間へのマッピングを行う。精度の高い検索が可能であるが、事前の教師あり・教師なし学習が必要となり、これらの学習に用いるデータの質が重要となる。

抽象度の高いモチーフ検索のためには SBIR が有効である。しかし、小袖の柄のように人の創作や想像によってつくられた画像を対象とした場合、その学習データの獲得には以下のような問題が生じる。

(1) モチーフ範囲の決定が困難

様々なレベルのモチーフが複雑に組み合わされた画像中から、CBIR を行うモチーフ範囲を決定する方法が必要である。

(2) メタ情報を正確に表すテキストの付与が困難

抽象的な意匠を的確に表現する印象語の付与は個人差が大きい。数、精度の確保に問題がある。メタ情報に対応する特徴を自動的に付与する方法が必要である。

これらの問題に対し、階層的領域抽出と一般物体認識器を組み合わせた手法を提案する。階層的領域抽出では、視覚的に注視しやすい局所の特徴を要素とする階層的クラスタリングを行い、要素とクラスターの近接性に基づく構造を用いた領域抽出を行う。抽出された各領域に対し、一般物体認識器を用いて評価し、その印象を複数の物体存在確率の空間に射影する。システム全般の構成を図1に示す。

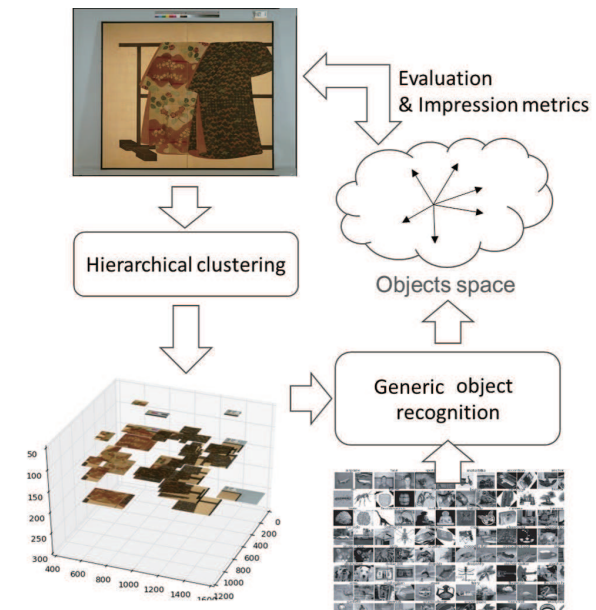


図1 階層的クラスタリングと一般物体認識器を用いたモチーフの印象評価

Fig. 1 The hierarchical clustering, and the impression evaluation by the generic object recognition system.

4. 高精細小袖画像からの特徴点と構造データの抽出

4.1 高精細画像の取得

8x10 ポジフィルムに撮影された96枚の画像を2000dpiでスキャニングする。サイズはおよそ19500×15500 pix 24bit である。これは小袖屏風自体の実寸に対する解像度に換算すると約200dpiに相当する。データベースの閲覧用には、オリジナルの画像を512×512pixのサイズの画像に階層的に分割する。およそ6レイヤーの階層画像集合が約1600枚得られる。

4.2 特徴点の抽出

画像中から注目領域を抽出する方法として、SIFT(Scale Invariant Feature) キーポイント [6] の密度を利用する。SIFT では、そのキーポイント検出に DoG(Differential of Gaussian) 処理を行い、各スケールの変化領域に含まれるエッジ情報の極値点を検出する。

各画像に対し、SIFT 特徴量を用いた類似画像領域の抽出を行う。SIFT 特徴量は画像の局所的な特徴を表し、スケールスペースを使った照明変化や回転、拡大縮小に不変な頑強な性質を持つ。ここでは、局所特徴ベクトルとして周辺領域を一边4ブロックの計16ブロックに分割したのちに、ブロックごとに8方向(45度ずつ)の勾配方向ヒストグラムを作成する。その結果、ヒストグラムの方向128次元が特徴量となる。これらの処理を経て、局所領域の特徴として有効な座標集合を得る。この集合の密度が高い領域を中心にモチーフの候補が存在すると仮定する。

4.3 階層化クラスタリングによるセグメントの抽出

得られた SIFT キーポイント集合の座標情報を用いた階層クラスタリングを Ward 法によって行う。Ward 法では、各要素をそれぞれ 1 つのクラスタとするところから開始し、その後、クラスタ内の分散が最小になるように統合を繰り返す。最終的にすべての要素が 1 つのクラスにまとまるまでこれを繰り返す。図 2 に Ward 法によって階層的にクラスタリングされたセグメントのデンドログラムを示す。この中で、200~400 までの距離に存在するセグメントをモチーフセグメントとして用いる。

4.4 Deep Learning による物体認識空間マッピング

Deep Learning は近年大きな進歩を遂げた多段ニューラルネットワーク学習法である [7]。特に画像認識の分野では、特徴量を事前に分析することなく、画像情報から直接教師なし学習によって特徴ニューロンを学習できることが示され、多くの分野にインパクトを与えている。

Deep Learning は一般物体認識としての利用法だけでなく、リファレンスモデルの中間層出力を用いて、高次の特徴空間を取り出すことが可能である。これにより汎用の特徴抽出器として利用することができる。一般物体認識で学習された Deep Learning の中間層出力とは、実世界における物体認識に必要な特徴空間とみなすことができる。すなわち、小袖の絵柄をこの特徴空間へ射影することにより、一般物認識としての意味抽出とモチーフの類似性評価が可能になる。

以上の考えに基づき、本研究では、Deep Learning のフレームワークに、CALTECH の BVLC が開発した Caffe[8]を用いた分析を行った。リファレンスモデルとして Caltech101 データベース [9] の 9145 画像から学習された結果を用いている。すなわち、101 次元での空間で、小袖屏風のモチーフ分類を試みた。

5. 小袖屏風画像を用いたモチーフ分析

小袖屏風画像 103 点に対して階層クラスタリングを行った結果、平均 24.3 階層のクラスが抽出された。各クラスに含まれる SIFT 座標データによってつくられる矩形領域のうち、抽出すべきモチーフが含まれる領域を、画像全体の面積の 5%以上、20%以下とし、かつその中に含まれる SIFT 特徴点が、総点数の 1%以上である場合を有効なモチーフ領域候補とした。図 3 に、現画像と SIFT キーポイントに基づく領域抽出結果の例を示す。なお、今回はこれらの候補選択のパラメータを経験的に決定しているが、分析対象の領域や解像度によって必要な領域の条件は変更する必要がある。103 点の画像から抽出されたセグメントは合計 3976 枚、1 屏風画像あたり平均 38.6 枚となった。

103 点の画像すべてを確認した結果、注目すべきモチーフ領域が候補として選択できている画像と、主要なモチー

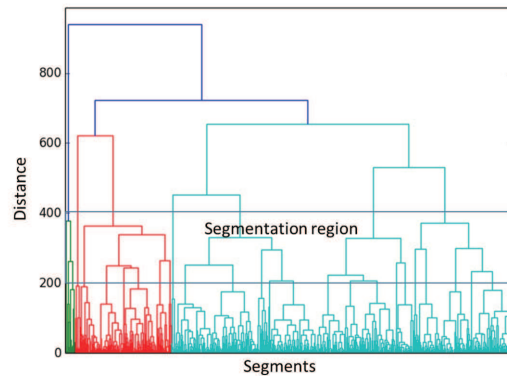
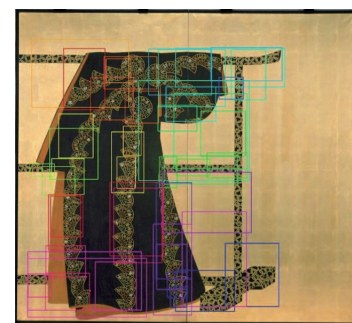
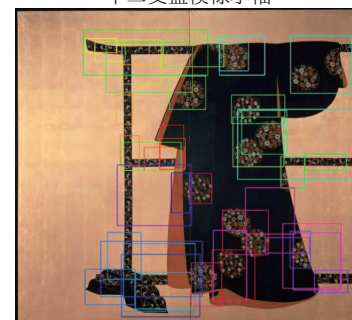


図 2 Ward 法による階層的クラスタリング

Fig. 2 The dendrogram for hierarchical clustering by Ward method.



十二支盤模様小袖



花入雪輪散し文様小袖

図 3 モチーフセグメンテーションの抽出例

Fig. 3 The examples of the motif segmentation.

フが分解されている画像がみられた。小紋のようにモチーフが局所的に集中・独立している場合には、抽出は容易であるが、大きな文字のように「空白」に意味のあるモチーフではモチーフが分解されてしまう傾向がある。これは、SIFT の近傍距離最小化を行う Ward 法では、「空白」を特徴とするモチーフの抽出が困難であることに起因する。

次に、抽出されたモチーフ領域を、Deep Neural Net を用いた一般物体認識によって評価した。101 次元のうち、上位 10 位までのオブジェクトの一致度を用いて距離を評価した。図 4 に、モチーフ領域候補に対応した一般物体空間上の候補例を示す。画像から直接受ける印象が、一般物体の持つ印象に対応して分析されていることがわかる。図 4(a),(b) のようにモチーフの画像分析から直接抽出するこ



図 4 モチーフの一般物体認識空間における評価例

Fig. 4 The examples of the generic recognition in each motif segment.

とが困難な意味を、一般物体認識空間におけるモノとしての見え方で表現することで定量的な評価が可能になった。

一方で、図 4(c) のように、モノに対する類似性が十分評価できていない領域に対しては、生地や単純な構造を持つモノの集合との関係しか見つからない場合もみられた。このような単純なモノは多くの領域に共通してみられる特徴であり、特徴空間の軸としての有効性は小さい。これは、Caltech105 のカテゴリが、小袖屏風画像のモチーフの意味を完全に網羅できていないことを意味する。

図 5 に、提案する類似性評価を用いたモチーフ間の類似性ネットワークを示す。いくつかの一般物体認識に対応したクラスが形成されていることが明らかになり、これらのクラスネットワークによる小袖屏風の分類や類似関係を可視化することが可能になった。

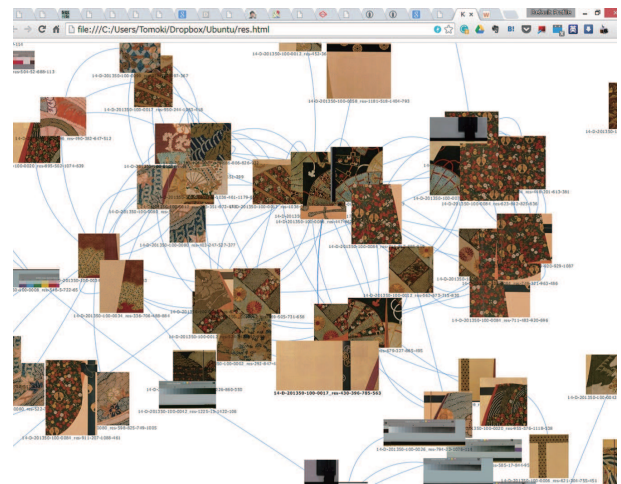


図 5 モチーフの一般物体認識空間上の距離ネットワーク表現

Fig. 5 The semantic network expression of motif segments in the generic object recognition.

6. おわりに

本稿では、特に歴史資料に含まれるモチーフ分析を目的とした、抽象画像の一般物体認識空間への射影を行う SBIR の方法を提案した。SIFT の密度分布に基づく階層的クラスタリングによるモチーフ領域候補の抽出と、Deep Learning による一般物体認識空間への射影により、モチーフの一般物体認識空間でのクラスタリングと距離に基づくネットワーク化が可能になった。

本手法により、モチーフの種別に寄らない一般物体認識空間上での意味抽出とそのネットワーク表現が可能になったものの、CALTECH101 データベースのカテゴリがモチーフの分析に最適であるとは限らない。今後は、あらかじめタグ付けされたモチーフキーワードに沿った画像をもとに同様の手法を試み、キーワードに準じた意味空間での表現を試みる。

参考文献

- [1] <http://www.emuseum.jp/>
- [2] <http://www.europeana.eu/>
- [3] 小袖屏風一野村コレクション, 国立歴史民俗博物館編 (2002)
- [4] 濱上知樹: 歴史資料デジタルアーカイブデータからの知的構造の抽出-小袖屏風を対象として-, 電気学会システム研究会, ST-13-124 (2013).
- [5] 柳井啓司: 一般物体認識の現状と今後, 情報処理学会論文誌 Vol.48, No.SIG 16, pp.1-24 (2007).
- [6] D. G. Low: Distinctive image features from scale-invariant keypoints, Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110 (2004).
- [7] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks", NIPS2012 (2012).
- [8] <https://github.com/BVLC/caffe>
- [9] http://www.vision.caltech.edu/Image_Datasets/Caltech101/