

## 色相情報に基づく正倉院文書の切り分け位置の検討

脇 正宏<sup>†</sup> 安達 文夫<sup>‡</sup> 徳永 幸生<sup>†</sup> 米村 俊一<sup>†</sup>芝浦工業大学<sup>†</sup> 国立歴史民俗博物館<sup>‡</sup>

## 1. はじめに

正倉院文書とは、奈良時代から東大寺正倉院に伝わる文書群を指す。本資料は、製作当時和紙が貴重であったため、表面には律令公文と呼ばれる地方から都への報告書、裏面には再利用された写経所文書がそれぞれ記載されている。しかし、明治時代の歴史研究者によって裁断、再構成が行われ、奈良時代に作られた時とは異なる並びとなっている[1]。

正倉院文書を元来の並びに復元するこれまでの方法は、主として文書の写真やその印刷物の手作業による並び替えであった。この方法では、対象の数が膨大となり、漏れなく並びを確認するには手間が掛かる。そこで文書をデジタル化した画像を用いて、表裏同時に表示し、効率的かつ正しく元の状態に復元する事ができれば、研究が大きく進展すると期待されている。

正倉院文書を並び替えて復元を行うには、まず資料を紙面ごとに切り分ける必要がある。そこで本研究では、資料画像を自動で切り分ける事を目指す。濃度情報に着目した切り分け位置の検出法については、既に報告した[2]。しかし、エッジ抽出の際に発生するノイズの影響により、適切な位置を検出できない例が存在する。本報告では、資料の色相情報に着目した切り分け位置の検出法について検討した。この手法を用いる事で、より適切な位置で正倉院文書を切り分けることができると期待される。

## 2. 資料画像の構成

現存する正倉院文書は、正集、続集、続々集など合計667巻の文書で構成されている。このうち、正集の複製をデジタル化した資料画像を用いて検出法を検討する。正倉院文書の画像例を図1に示す。図1画像で、本研究では破線で囲まれている領域を文書領域、破線外の領域を台紙領域と呼ぶ。対象とする資料の例を図2に示す。資料は文書領域の並び方に基づき、図2aに示す台紙領域を挟んで並ぶ「独立型」、図2bに示す台紙領域を挟まずに並ぶ「連続型」の2種類に分類できる。本報告では、既にデジタル化された正集45巻のみを対象としている。その中で、まずは対象資料の大半を構成している「独立型」を取りあげる。

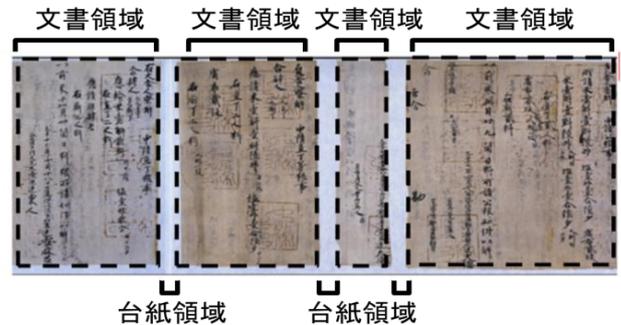
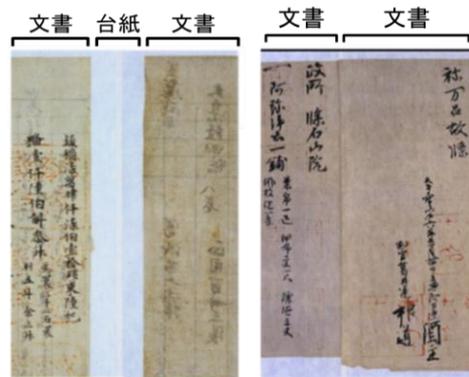


図1 文書領域と台紙領域

正倉院文書正集2巻複製 (国立歴史民俗博物館所蔵, 原品: 正倉院文書事務所蔵)



a: 独立型 b: 連続型

図2 文書領域の並びに着目した分類  
正倉院文書正集2巻複製 (同上)

## 3. 切り分け位置の検出法

## 3.1 切り分け位置の検出法

これまでに提案した従来の切り分け位置の検出法は以下の手順で行われていた。

まず、資料画像にエッジ抽出手法であるCanny法を適用し、エッジ抽出を行う[3]。抽出後、探索の開始地点を台紙領域内に手動で指定し、資料の縦方向(y)の各点から横方向(x)に向かって初めてエッジが検出されるまで探索を行う。探索終了後、エッジが検出された全ての位置データをプロジェクションする。

次に、プロジェクションしたデータを基にエッジとして検出されたx座標の平均値 $x_{avg}$ と標準偏差を求める。そして、算出した $x_{avg}$ を切り分け位置として文書領域を垂直に切り出す。しかし、この方法では資料の切れ目の形状が不均一なため、文書領域を切り落としてしまう可能性がある。そこで $x_{avg}$ に、算出した標準偏差 $\sigma$ の1.4倍分の領域を加えた検出範囲を設定す

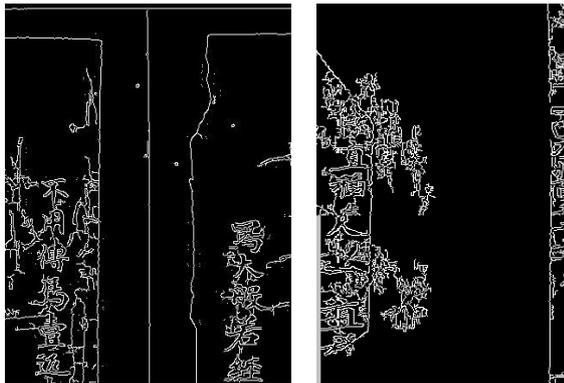
Cutting position detection based on hue for Shosoin monjo images.

<sup>†</sup>Masahiro WAKI(ma12119@shibaura-it.ac.jp)<sup>‡</sup>Fumio ADACHI(adachi@rekihak.ac.jp)<sup>†</sup>Yukio TOKUNAGA(tokunaga@sic.shibaura-it.ac.jp)<sup>†</sup>Shunichi YONEMURA<sup>†</sup>Shibaura Institute of Technology<sup>‡</sup>National Museum of Japanese Histories

る。更にその範囲内で最も外側にあるエッジを切り分け位置とする事により、文書領域を切り分けてしまう可能性を回避し、文書領域に近い位置で切り分ける事が可能となる。

3.2 濃度情報を用いた場合の問題点

Canny 法に濃度情報を用いたエッジ抽出例の拡大図を図3に示す。図3aのような台紙の繋ぎ目の存在や、図3bのような台紙領域上のノイズが誤った切り分け位置の検出する原因となっている。



a: 台紙の繋ぎ目                      b: ノイズ

図3 エッジ抽出例の拡大画像

4. 色相情報を用いた切り分け位置の検出法

より適切な切り分け位置を検出するため、画像の色相情報に基づいたアルゴリズムを検討した。

4.1 資料画像の色相

台紙の繋ぎ目とノイズは、台紙領域内で発生するという共通点がある。本報告では、検出法に用いる資料画像の台紙領域を単一の画素情報に置き換える事で、ノイズとなる画素を除外し、上記の問題点を解決する事を目指した。

まず、対象資料の色相情報分布について調査した。図4に色相情報の例を示す。資料画像の共通する特徴として、文書領域の画素は赤みがかかり、台紙領域の画素は青みがかっている点が挙げられる。そこで、正集45巻の表面全ての画像を対象に文書領域と台紙領域の色相を調査した。その結果、色相情報を円グラフで表した時、台紙領域は56度から260度の色相情報を持つ事がわかった。この結果を基に検出法に用いる画像の台紙領域を白色に置き換える事で、ノイズの発生を抑え、検出法の精度向上が可能となる。白色に置き換える前後の濃度ヒストグラムを図5に示す。また、色相情報を用いたエッジ抽出の結果例を図6に示す。

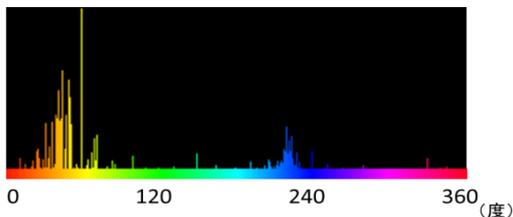


図4 色相のヒストグラム例(正集2巻)



図5 色相情報を用いたエッジ抽出結果

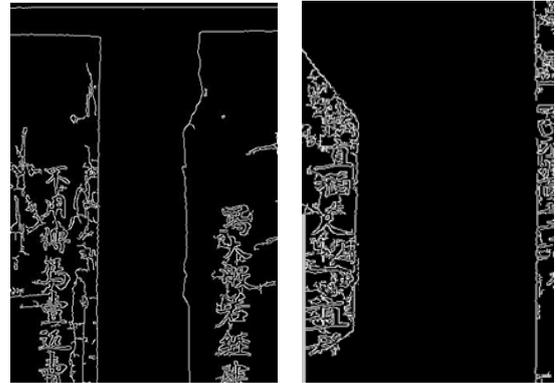


図6 色相情報を用いたエッジ抽出結果

4.2 切り分け位置の検出精度

検出法に用いる画像に、従来の濃度情報に着目した画像と色相情報に着目した画像を用いたアルゴリズムの精度を比較検討した。正集45巻表裏の切り分けが必要な606箇所を対象とし、実際に切り分けを行った。表1に切り分け位置の検出精度の結果を示す。ここでいう検出精度とは、切り分け位置が台紙領域内で検出された割合を表す。

表1 切り分け位置の検出精度

濃度	0.808
色相	0.912

その結果、濃度情報よりも色相情報を用いる手法の方が、より高い精度で切り分ける事が出来る事が明らかになった。また、色相情報を用いる事で濃度情報より、文書領域に近い位置で切り分け位置を検出する事が可能となった。

5. 結び

本報告は、文書領域と台紙領域の切り分け位置を色相情報に基づいて検出する方法を提案し、その有効性を明らかにした。今後は、切れ目の形状に沿った切り分けの方法について検討を行う。本報告をまとめるにあたり指導頂いた国立歴史民俗博物館の仁藤敦史教授に感謝する。

—— 参考文献 ——

[1] 丸山裕美子：正倉院文書の世界 よみがえる天平の時代、中公新書(2010)  
 [2] 脇正宏, 安達文夫, 徳永幸生, 杉山精：正倉院文書における画像処理を用いた切り分け位置の検出手法の検討, 画像電子ミュージアム学会(2012)  
 [3] Canny, J., A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714 (1986)