

百認一取：色情報を用いた実画像からの取札領域抽出

伊藤壮真¹ 長嶺和紀² 吉川唯人¹ 角田唯隼¹ 佐藤礼一郎¹ 岡本学¹ 筒口拳¹

概要：我々は、コンピュータと人間による現実の取札を用いた百人一首の対戦を可能にするシステムの構築をめざしている。配置された取札を認識するには、撮影した取札配置画像から各取札の領域を抽出し、各取札の歌番号を付与する必要がある。これまでに、シミュレーション画像からは正しく抽出できたが、実画像では照明変動や取札の重なり等により個々の取札を正しく抽出できないという課題があった。本研究では、安定した各取札の領域抽出を目的とし、色情報およびモルフォロジー変換を用いた手法を試みた。また、本手法をシミュレーション画像および実画像に適用し、その効果を確認した。

キーワード：画像処理、色空間、幾何補正、OCR

Hyakunin-Isshu Project: Extraction of Fuda Areas using Color Information from Real Images

Soma Itou^{†1} Kazuki Nagamine^{†2} Yuito Yoshikawa^{†1} Yuito Kakuta^{†1}
Reiichirou Satou^{†1} Manabu Okamoto^{†1} Ken Tsutsuguchi^{†1}

Abstract: Our goal is to construct a system that enables a computer and a human to play a game of Hyakunin Isshu (100 poems) using realistic cards. In order to recognize the cards, it is necessary to extract the area of each card from a captured image of the cards and assign a song number to each card. In the past, it has been possible to correctly extract the cards from simulation images, but in real images, individual cards could not be extracted correctly due to illumination fluctuation and overlapping of cards. In this study, we attempted a method using color information and morphological transformation for stable extraction of each trumpet area. The method was applied to simulated and real images to confirm its effectiveness.

Keywords: Image processing, Color space, Geometric Correction, OCR

1. はじめに

本研究は、リアルな場で人間とコンピュータが百人一首の対戦が可能となるシステムの構築をめざす「百認一取プロジェクト」の一環である。「リアルな場」とは、実物の取札を畳の上などの現実世界に配置し、読み上げられた歌に該当する取札を物理的に取る、または指定することを意味する。人間は視覚で取札の位置を確認し、聴覚で歌を識別し、手で取札を取得するのに対し、コンピュータは「カメラ入力された画像の認識による取札位置の識別」、「音声認識による歌の識別」、「プロジェクション等による取札の指定」を行うことを想定している(図1)。これまでに我々は、取札の合成画像(プレゼンテーションツールを用いて電子的に生成した画像)を競技かるたの配置に沿ってランダムな順序で配置したシミュレーション画像に対し、取札領域を抽出して文字認識とラベリングによる取札領域への歌番号付与を可能とするシステムを実装した[1][2]。しかし、実画像では照明変動や取札の重なり等により取札領域を正しく抽出できないことがある、という課題があった。本研究では、取札の枠の色に着目して実画像から(下の句の書か

れた領域ではなく)枠領域を抽出することで、安定した取札領域抽出を試みた。また、実画像においては撮影距離や光の反射等により取札の文字部分と背景(地)部分の輝度差が低下し、文字認識精度が低下することも考えられるため、輝度差による文字読み取り精度についても検討を行った。

以下、第2章で先行研究とその課題を述べ、第3章で色情報を用いた改善手法について提案する。第4章で実験について述べ、第5章で実験結果について考察し、第6章でまとめる。

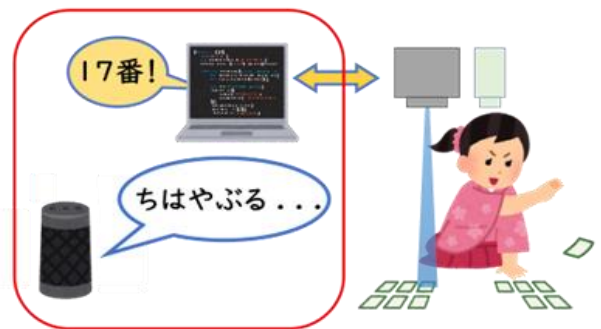


図1 百認一取プロジェクトの概要

¹ 崇城大学 情報学部
Faculty of Computer & Information Sciences, Sojo University
² 崇城大学大学院 工学研究科
Graduate School of Engineering, Sojo University

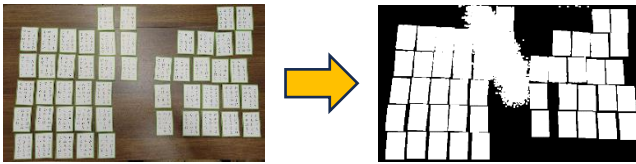


図2 実画像の二値化の失敗例

2. 先行研究とその課題

本研究においては、文字認識に Tesseract OCR[5]を用いる。Tesseract は縦書の日本語にも対応しており、百人一首の取札も認識できる。認識した取札がどの歌に該当するかは、認識した文字列を事前に準備した 100 首の下の句の文字列と比較し、最も類似する歌を探索することで行う。先行研究において、取札領域の画素数が横 44 画素、縦 62 画素程度であっても Tesseract OCR は高い認識精度を保つことが確認されている[2]。従って、取札の下の句が書かれている矩形領域を正しく抽出することができれば、取札領域に歌番号をラベル値として付与することができる。文献[1]では取札配置画像（合成画像）を 2 値化したあと、Morphology 変換と Watershed により、画像を前傾と背景に分離することで取札領域を良好に抽出できていた。しかし、この手法はその成否が 2 値化処理に依存しており、光の反射等がある実画像の場合は、正確に取札領域を抽出できないという課題があった。図 2 にその例を示す。2 値化処理そのものを改善することも考えられるが、さまざまな状況下で撮影された画像への適用を考慮すると、光の反射の影響を受けにくい取札の枠の色に着目し、枠部分を抽出して札領域の境界とすることとした。

3. 提案手法

3.1 前提条件

本研究では、取札を配置した「競技場」を撮影する画像のサイズとして HD (1920×1080 画素) を想定する。競技かるたのルールによれば取札 1 枚のサイズは横 52mm、縦 73mm であり、取札を配置する競技場は横 87cm と定められており、縦方向の最小値は 50.8cm である。従って、競技場 1mm あたりの横方向の画素数は最大で $1920/870=2.207$ 、縦方向の画素数の最大値は 2.126 となる。

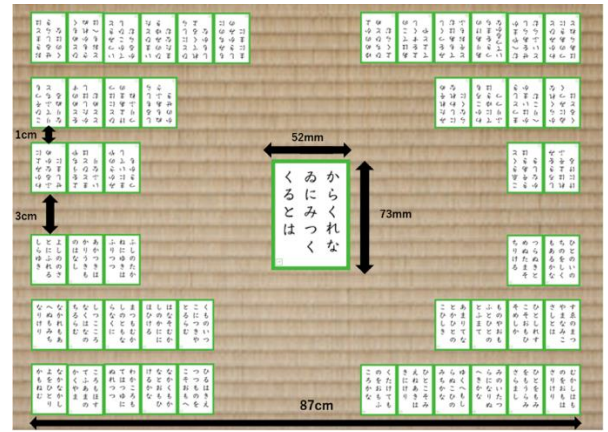


図3 競技かるた場のサイズ

このことから、取札 1 枚の最大画素数は約 110×155 画素となるが、実際は競技場が HD サイズよりも小さく撮影されると想定し、本研究では取札の基準画素数をおよそ 104×146 画素と想定する。実画像においては射影変換されることもあるが、以降、このサイズを基準として用いることとする。

3.2 取札領域の抽出

競技用の標準かるたは現在 1 種類のみであり、取札の枠は薄緑色である。そこで、撮影した取札配置画像を RGB 色空間から HSV 色空間に変換し、緑を抽出したマスク画像を作成する。RGB 色空間から HSV 色空間への変換後、HSV 空間における緑色の成分は、
 下限 (H, S, V) = (30,10,0)
 上限 (H, S, V) = (90,255,255)
 となる。

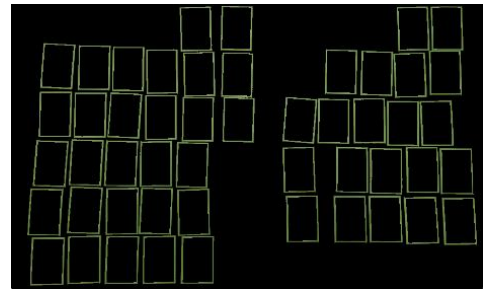


図3 HSV による緑色の抽出

抽出した枠領域に対し Morphology 変換で膨張・収縮を行った後に、枠領域を境界（背景）として白単色の画像を区切り、分離することで暫定的な取札領域を得る。しかし、取札を配置する畳は一般的に緑色に近いため、取札ではない領域がノイズとして混ざる。そこで、抽出する緑の閾値を以下のように変更した：

下限 (H, S, V) = (30,10,0)

上限 (H, S, V) = (90,255,255)

次に、分離された領域を連結成分とみなして、前述の取札領域のサイズをもとに領域面積（画素数）が極端に大きい、もしくは小さい領域を除去することで不要な領域を削除する。基準となる取札領域の画素数はおよそ 14,000 程度

としている。次に、除去できなかったノイズを除去するため、残った取札領域から領域面積の中間値を求め、これより大きく異なる画素数の領域はノイズとして除去する。最後に、Morphology 変換を適用後、改めてラベリング処理を施して取札の領域を取得する。

4. 実験

本研究の目的は枠部分の抽出による取札領域の抽出であるが、実画像においては札領域が正しく抽出できても、光の反射による文字領域と背景領域の輝度差の低下により、正しく文字が認識できない可能性がある。このため、枠領域の抽出に先立って、まず輝度値の差による OCR での読取精度の評価実験を行い、その後、実画像による枠領域の抽出実験を行った。

4.1 取札領域の文字認識

光の反射で文字が薄くなることによる文字認識への影響を調査するため、シミュレーションの取札画像 (104×146pix) を用いた実験を行った。取札画像を二値化し、文字部分に該当する黒領域の画素値を 0 から段階的に増加させたものに文字認識を行い、その精度を計測した。画素値は 0~250 までは 25 ずつ増加させ、250~255 では 1 ずつ増加させた(図 5)。

プログラムは Python を用いて実装し、前述の通り取札の文字認識には Tesseract OCR [5]を用いた。読み取った文字列と下の句との比較には、difflib の SequenceMatcher 関数を用いたシーケンスの類似度の計測で行った。その結果を図 9 に示す。文字部分の画素値が 0~252 までは高い認識精度を維持しているが、253 で大きく低下し、それ以降はほとんど認識できない結果となった。しかし、画素値 253 以降の取札は目視でも文字は確認できいため、かなり良好な精度であると考えられる。この実験結果から、取札の文字が書かれている領域が抽出できれば、書かれてある文字も高い確率で認識できることが確認できた。

4.2 取札領域の抽出

提案手法を実装したプログラムによる取札の抽出結果を以下に示す。文献[1]の手法、および本項で提案した手法ともに、シミュレーション画像および光の反射の少ない実画像では適切な取札領域が抽出できた。しかし、光の反射のある画像では、文献[1]の手法が不適切な結果となっているのに対して、提案手法では適切な取札領域が抽出できている。抽出結果に対し、結果がわかるよう色付けした結果を図 8 に示す。



図 5 画素値による文字の変化 (左 : 0, 右 : 240)

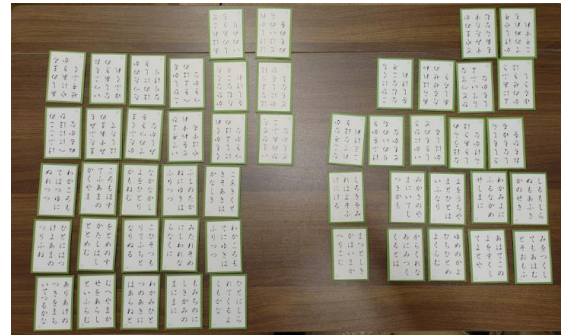


図 6 光の反射のある実画像例



図 7 光の反射のある実画像に対する旧手法の結果例

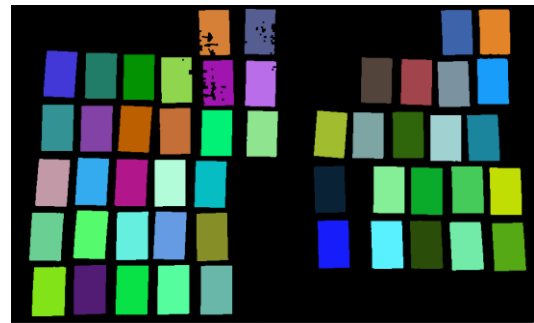


図 8 光の反射のある実画像に対する提案手法の結果例

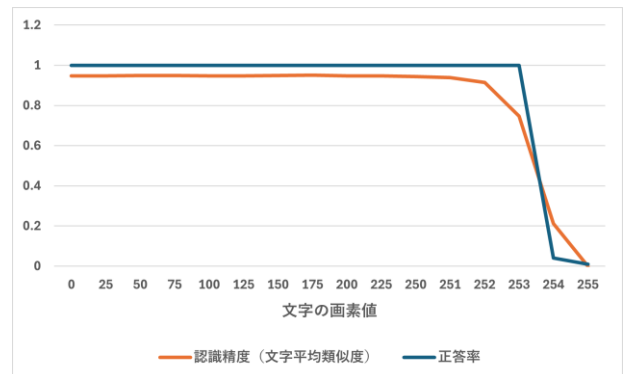


図 9 文字の画素値による文字認識の精度

5. まとめ

本研究では、光の反射等の影響がある実画像においても安定して取札領域を中種できる手法を提案した。撮影画像中の取札1枚あたりの画像サイズを見積ることでノイズ領域を良好に除外でき、また、枠である緑色の部分をHSV空間で抽出することにより、札領域を良好に抽出できた。

また、文字部分と背景部分との輝度差による文字認識は、合成画像ではあるものの輝度差2程度でも認識できることがわかり、取札領域を正しく抽出できれば、かなりの精度で歌を特定できることが明らかとなった。

今後の課題として、真上からの撮影は競技者の身体により取札が隠蔽される可能性があるため、補助的に斜めからの画像も撮影することが考えられるが、文字認識が射影変換に耐えうるかどうかの検討が必要である。また、本プロジェクトの音声認識システム[3][4]との統合が挙げられる。

参考文献

- [1] 長嶺和紀, 角田唯隼, 仲菜夏, 中内真希, 岡本学, 筒口拳: 百認一取(1): 取札配置画像に対する歌番号付与システム, 電子情報通信学会九州支部学生会講演会・講演論文集, D-26, 2023.
- [2] 吉川唯杜, 佐藤礼一郎, 安慶直哉, 長嶺和紀, 岡本学, 筒口拳: 百認一取(2): 取札画像のサイズに対する文字読み取り精度の評価, 電子情報通信学会九州支部学生会講演会, D-27, 2023.
- [3] 濱武右京, 木村翔真, 黄思韵, 柴田美桜, 筒口拳, 岡本学: 百認一取(3): 歌読み上げのための音声認識システムの検討, 電子情報通信学会九州支部学生会講演会, D-28, 2023.
- [4] 木村翔真, 濱武右京, 黄思韵, 柴田美桜, 筒口拳, 岡本学: 百認一取: 歌読み上げのための音声認識システムの評価, 火の国情報シンポジウム 2024.
- [5] tesseract-ocr ,
[https://github.com/tesseract-ocr/tessdoc\(2024/02/09](https://github.com/tesseract-ocr/tessdoc(2024/02/09) アクセス).