

文字認識結果の目視一括化方式

岩田泰明[†] 大峽光晴[†] 松本俊子[†] 小野山隆[†]

紙文書の電子化における文字認識結果の目視確認を効率化する目的として、目視一括化方式を提案する。本方式は、文字認識結果に対応する単語画像を、字体が類似するグループにクラスタリングし、クラスタ内の各画像を重ね合わせる合成処理を特徴とする。合成結果の画像により、同じ単語の画像のみが合成されているかどうかを判別できる。目視確認の担当者は、クラスタ内における合成画像と認識結果を比較することで、クラスタ内の全ての画像の文字認識結果の正誤の判断や修正を一括で行うことができ、作業工数を削減する。営業文書の文字認識結果を用いて評価を行ったところ、目視確認の回数を39%削減でき、本手法の有効性を確認できた。

Aggregating Visual Confirmation Method for Character Recognition Result

Yasuaki Iwata[†] Mitsuharu Oba[†]
Toshiko Matsumoto[†] Takashi Onoyama[†]

We propose an aggregating visual confirmation method for character recognition result by OCR to reduce the cost of document scanning and capture. Our method has two features. First, it has a clustering algorithm which is based on the recognition result and on the shape feature of the string image. Second, it creates the composite image by superimposing the images which are classified in the same cluster. The composite image enables operators to determine whether it consists of the same string image. Therefore, the operators can confirm and correct the recognition results in the same clusters at a time by comparing them. The experiment show that our method could cut down the number of visual confirmation by 39%.

1. はじめに

業務効率化、コンプライアンスの確保、情報活用などを目的として紙文書の電子化が進んでいる[1]。これに伴い、紙文書に記載された文字情報の文字コード化を効率化するためにOCRの利用が活発化している。特に、あらかじめユーザが指定したエリアの文字認識をする帳票OCR[2]は、金融、医療、及び官公庁や地方自治体などの業務に適用され、外部から受け取った紙文書に記載された文字情報をワークフローシステムや情報管理システムへ自動的に入力するために頻りに利用されている[3]。

これらの分野における紙文書は、事務処理に用いる情報が記載されるため、文字認識結果は全て正しいことが要求される。たとえば、自治体の住民税賦課業務において、法人から受け取る給与支払報告書は、自治体が作成したテンプレートに氏名、住所、および支払金額などが記載されている。これらの情報を用いて事務処理が行われるので、各項目の認識結果は全て正しくなくてはならない。しかし、OCRの文字認識は統計的手法[4]を用いることから認識精度は100%にならないため、OCR後に目視による文字認識結果の確認と誤認識の修正が必要となる。文字認識結果の確認と誤認識の修正を効率化する手法には、ユーザが修正した誤認識の情報を利用する手法[5]や認識結果の文字コードに対して形態素解析、N-gram、マルコフ連鎖モデルなどを用いた手法[6][7][8][9]がある。前者の手法[5]は、認識結果と文字の画像の特徴ベクトルを用いて、文字の画像をグループ化する。ユーザが文字認識結果を修正した場合、その文字の画像が属するグループを対象に、自動で同じ文字コードを割り当てる。この手法は、誤認識する字体が同一傾向であることを利用している。後者の手法[6][7][8][9]は、いずれも文書全体の文字列を対象とし、文書全体に散在するOCR誤認識を統計処理により自動検出する。しかし、両者の手法ともに、誤認識の検出精度は100%に至っていない。従って、文字認識結果が全て正しいことが求められる業務には、OCRの認識精度によらず、電子化を行った文書とその文字認識結果を一件ずつ照らし合わせて全件を目視しなくてはならない。また、帳票の受付期間を限定している住民税賦課業務など繁忙の差がある業務では、特定の期間に紙文書が集中することが多い。すると、期間内に大量件数の紙文書をOCRしなければならぬため、全件の目視確認をするコストが高くなってしまふ。このような場合、いかに認識結果の確認と誤認識の修正する作業工数を削減するかが課題となっている。

そこで本研究では、類似する文字の画像をクラスタリングし合成画像を計算することで文字認識結果を一括で確認できる方式を提案する。本方式により、確認作業の工数削減を目指す。

[†]株式会社日立ソリューションズ
Hitachi Solutions, Ltd.

2. 目視一括化方式

本研究では、文字認識結果の目視確認作業の工数が膨大になる課題を解決するため、認識結果についてまとめて確認できる合成画像を生成する目視一括化方式を提案する。目視一括化方式が生成する合成画像は、複数の認識結果に対応する画像において、各画像の画素について二種類の演算をとった画像である。ユーザは合成画像を確認するだけで全件の文字認識結果の確認を済ませることができ、目視確認作業の工数を削減できる。図1に目視一括化方式を適用した文字認識結果の目視確認作業の概要を示す。目視一括化方式によって生成された合成画像をユーザが確認することで、正しい認識結果を一括で確認済とすることができる。本章では、目視一括化方式の概要について説明する。

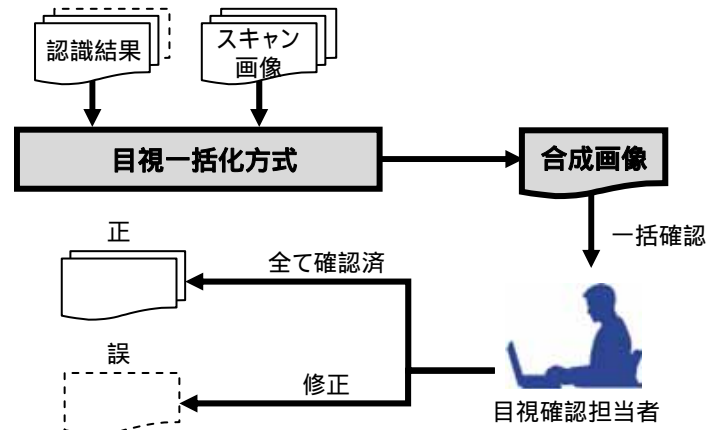


図1 目視一括化方式における文字認識結果の確認

2.1 目視一括化方式の概要

目視一括化方式は、同じ文字を重ね合わせたとき、字体が一致することに着目した画像合成をする方式である。本方式の概要を図2に示す。文字認識結果に対応する文字列部分の画像（以後、単語画像と呼ぶ）を切り出して、同じ文字列で認識した単語画像を合成処理する。合成処理には「AND演算」と「OR演算」があり、これらについて以下に示す。扱う画像の画素は二値であり、前景を黒、背景を白と表現する。

- AND演算：すべての画像における画素が黒の場合のみ黒、それ以外は白。
- OR演算：いずれかの画像で画素が黒の場合のみ黒、それ以外は白。

目視確認の担当者が二つの合成画像を比較すると、重ね合わせた単語画像は全て同じ文字列であることがわかり、合成した単語画像の認識結果について一括で確認できる。よって、目視する画像数が少なくなることから、目視確認の作業を削減できる。以下に合成画像についてより詳しく説明する。

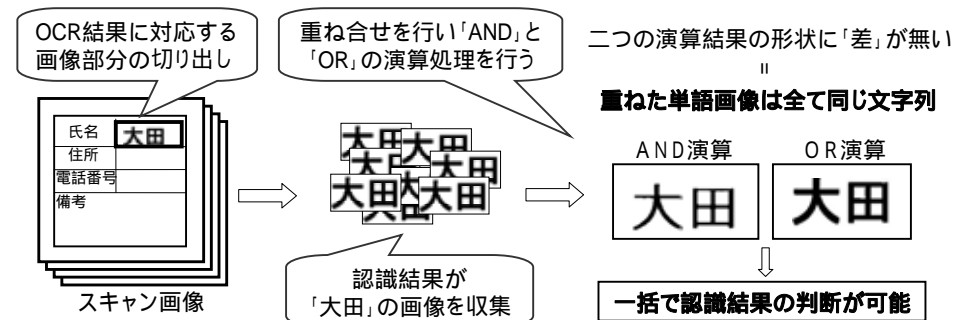


図2 目視一括化方式の概要図

2.2 重ね合わせについて

単語画像の重ね合わせの際に、字体が最も一致する位置に単語画像を平行移動させる。単語画像は、OCRにおける文字の外接矩形の座標情報を用いて切り出される。この座標情報は文字の切り出し処理における画像の特徴量解析[10]や文字の大きさの正規化[11]などで決定される。このことから単語画像によって外接矩形の位置が僅かに異なる可能性があるため、平行移動による位置合わせが必要になる。また、単語画像を平行移動させる際には、文字列の文字の間隔が単語画像によって異なる場合があるため、単語画像を文字ごとの画像（以後、文字画像と呼ぶ）に分割して各字体が最も一致する位置に平行移動する。字体が最も一致する位置を求める方法としてテンプレートマッチング[12]を用いる。テンプレートマッチングにおける各座標の類似度は正規化相互相関[13]により求め、最も高い類似度が算出された座標までの平行移動分を座標補正值とする。このように、単語画像を重ね合わせる際には、座標補正值の分だけ各文字画像を平行移動させる必要がある。

2.3 合成画像の確認について

ユーザが二つの合成結果を比較して、重ねた単語画像の認識結果について一括で確認する方法を以下に示す。

(1) 正しい認識結果の単語画像を重ね合わせた場合

「大田」と正しく認識した複数の単語画像を重ね合わせて合成処理をした画像例を図3に示す。二つの形状に差はなく、共に「大田」と読める。すなわち、合成した単語画像はいずれも同じ文字であり、かつ、それらの文字認識結果は「大田」であることから、ユーザは重ねた画像の認識結果が「大田」で正しいことを一括で判断できる。



図3 正しい認識結果の単語画像を合成した画像例

(2) 誤認識を含む単語画像を重ね合わせた場合

「大田」と認識した複数の単語画像の中に、「太田」の画像を「大田」と誤認識した画像が含まれていた場合の合成処理をした画像例を図4に示す。二つの形状に差があり、互いに異なる文字として読める。すなわち、重ね合わせた単語画像の中に異なる文字の単語画像が含まれているので、ユーザは重ねた画像の中に誤認識が含まれていることが認知できる。



図4 誤認識を含む単語画像を合成した画像例

3. 認識結果の一括確認可能な合成画像を生成するための課題

目視一括化方式を適用した目視確認作業において、月当たりの対象文書数や人件費などを考慮すると、本来の目視確認回数を25%以上削減できれば効果が見込める。このとき、合成画像から同じ単語画像のみを重ね合わせているかどうか判断できない場合（合成画像の字体が劣化するような場合）には重ねた画像を1件ずつに戻して目視し、誤認識した画像が含まれているかを確認しなければならない。よって、目視確認作業を削減するためには、重ねた画像を1件ずつ確認する作業をできるだけ発生させない、認識結果の一括確認可能な合成画像の生成が求められる。しかし、これを業務で扱われる帳票の画像を用いて実現しようとする場合、合成画像の字体が劣化する二つの課題がある。

3.1 異なる字体の合成による合成画像の劣化

帳票には発行元やフォーマットの指定元が同じであっても、項目の記載元が異なることがある。するとそれらの帳票から切り出した文字画像は、書体やサイズが異なるなど、同じ文字でも字体が異なる場合がある（図5）。部分的に線の太さや長さが異なるので、合成処理したときのAND演算は形状の不足部分が欠落し、OR演算は字体の余剰部分が膨張する。また、位置合わせにおいて二つの文字画像の字体が一致する配置は存在しない。つまり、文字画像によって座標補正值が異なってしまうため、字体が異なる数に比例して合成画像の字体の欠落と膨張は誇張される。このように、AND演算の結果は画素の数が少なく、また、OR演算の結果は画素の数が多くなるため、二つの演算の劣化した文字を比較したとき、同じ文字を重ねているか判断できない画像となってしまう。劣化した文字の合成画像とならないためには、字体に差が少ない類似した単語画像のみを合成する必要がある。

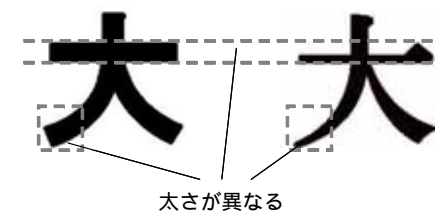
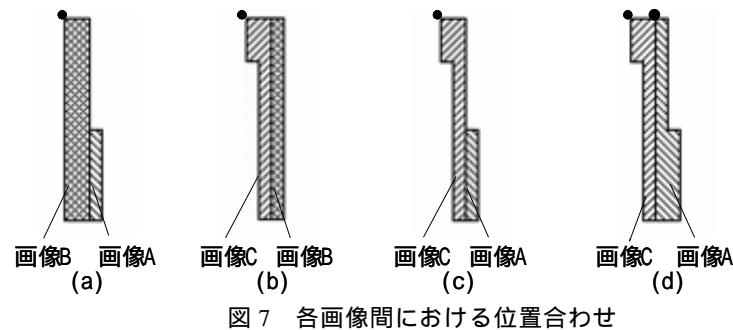
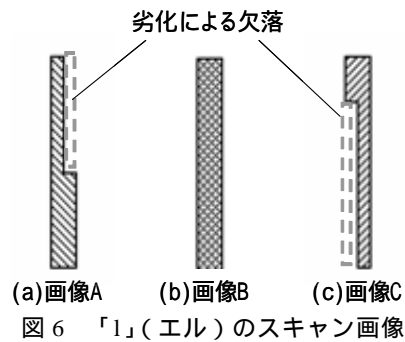


図5 部分的に形状が異なる同じ文字の例

3.2 部分劣化の字体による位置合わせの順序依存

単語画像を重ねる際の位置合わせにおいて、平行移動は順序によらないことから、いかなる重ねる組合せでも合成画像は同一となる。しかし、組合せによって合成画像が異なり、認識結果について一括で確認できない合成画像が求まる場合がある。たとえば、図6に示すようなゴシック体のアルファベット「1」(エル)の画像3つを重ね合わせる場合を考える。画像Aと画像Bはスキャン時の状態や二値化処理により一部が劣化した画像である。このような画像の位置合わせをすると、不足した画像部分の影響を受けて図7で示す配置の座標補正值が求まる場合がある。(d)の座標補正值を含む組み合わせを用いると、画像Aと画像Cが重なる部分がないため、AND演算の結果は黒の画素値はなくなってしまふ。同じ字体の文字画像を重ねているのにも関わらず、合成画像の文字が順序に依存することで劣化してしまう可能性がある。従って、最適な位置合わせを成す組合せを求めなければならない。



4. 単語画像のクラスタリングと位置合わせ条件の最適化

合成画像の生成における課題を解決するため、目視一括化方式では、字体が類似する単語画像のクラスタリングと最大全域木問題による位置合わせ条件の最適化を行う。本方式の処理フローを図8に示す。太線で囲んだ二つの処理が技術的課題に対する解決策に該当する。これらの処理について以下で詳しく説明する。

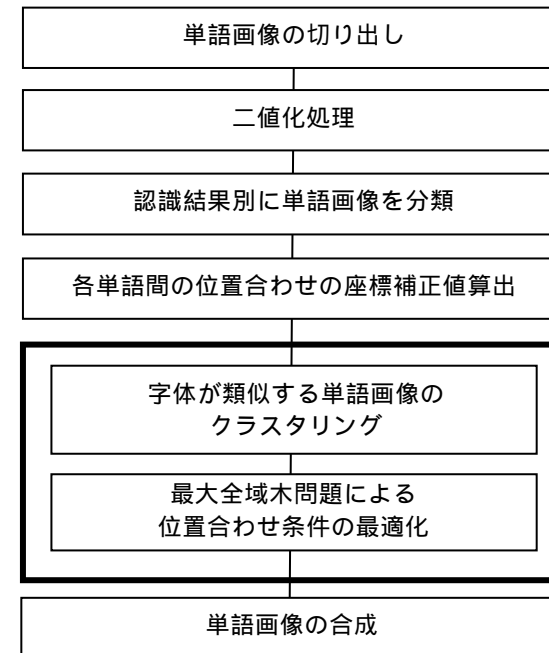


図8 一括化方式の処理フロー

4.1 字体が類似する単語画像のクラスタリング

一つ目の課題である異なる字体の合成により合成画像の字体が劣化することを解決するため、単語画像間の類似度を基準にした文字画像のクラスタリングを行う。クラスタリングにおける単語画像間の距離は、位置合わせの補正值を求めた際に算出し

た類似度の値に比例して距離が近くなる値を設定する。また、合成処理の結果は単語画像として取得するので、異なる文字が含まれているか確認できない文字が一つでもある場合、重ねた単語について1件ずつに戻して確認しなければならなくなる。低い類似度（長い距離）の文字を含めないようにするため、単語画像の距離は、各文字間で求めた類似度の中で最小のものとする。このようにして各単語画像間の距離を設定して、最短距離法により類似する単語画像のクラスタリングを行う。また、クラスタリングの際に、類似しているとする閾値以下の類似度であるクラスタ間は併合しないようにする。そして、クラスタ内の画像の合成処理を行うことで、字体の差が少ない画像の重ね合わせができる。

4.2 最大全域木問題による位置合わせ条件の最適化

2 つ目の課題である部分劣化した字体により位置合わせが順序に依存してしまうことを解決するため、前節で求めたクラスタを連結グラフと見なし、辺の重みの総和が最大となるスパニングツリー（最大全域木）を求める。連結グラフにおける辺は二つの単語画像を重ねることを示すため、スパニングツリーを構成する辺集合はクラスタが含む全ての単語画像を重ね合わせる組合せに対応する。そして、より高い重み（類似度）の辺を優先的に選択することで、字体がより一致する配置の組合せを求める。これは、最小全域木を求めるクラスカル法[14]のアルゴリズムを用いて最大全域木を求める。この処理により、位置合わせ条件の最適化が可能となる。

5. 評価

提案手法を評価するため、実際の業務で用いられる営業文書で評価実験を行った。OCR の目視確認を行う文字は、文書におけるタイトルの文字列を用いた。

5.1 実験方法

(1) 実験条件

実験条件を以下に示す。

対象文書：営業文書 1150 文書

目視確認項目：タイトル（1 文書に対してタイトルは 1 箇所）

タイトルの種類数：30

画像の解像度：200dpi

類似度の閾値：0.8

合成する最大画像数：50

(2) 測定方法

対象文書に対して、目視一括化方式のプロトタイプを適用し、削減率を測定する。なお、削減率の算出方法は以下の通りである。

目視回数の削減率： $1 - (\text{一括化後の目視回数} \div \text{全件の目視回数})$

（ただし、合成画像を目視した結果、認識結果について一括で判断できない場合は、合成画像を 1 件ずつに戻して目視回数を数える）

5.2 実験結果

上記の条件で実験を行った結果、一括化後の目視回数は 696 回となり、39% の削減率を達成できた。25% を上回ることから本手法の有効性が確認できた。以下に 3 章で述べた課題に対する提案方式の効果を確認する。

(1) 字体が類似する単語画像のクラスタリング

字体が類似する単語画像のクラスタリングにより得たクラスタの合成処理結果の例を図 9 に示す。各合成画像から同じ字体の文字を重ねていることがわかり、クラスタが含む画像は類似している。生成されたクラスタ 274 個の中で、このような文字認識結果を一括で確認できる合成画像を生成したクラスタは 257 個であった。また、文字認識結果を一括で確認できない 17 個のクラスタの中で、類似した画像のみを含むクラスタは 16 個であった。よって、適切に分類されたクラスタは、274 個中 273 個であり、実際の帳票における活字を字体が類似するグループにクラスタリングできることを確認できた。

画像数	AND 演算	OR 演算
40	御_見_積_書	御見積書
14	業務完了証明書	業務完了証明書
3	検収通知書	検収通知書
3	受領書	受領書

図 9 目視一括化方式で生成した合成画像の例

(2) 最大全域木問題による位置合わせ条件の最適化

適切に分類されたクラスタ 273 個において、合成画像から認識結果について一括で確認できるクラスタは 257 個であった。すなわち、これらのクラスタは位置合わせが最適なクラスタである。よって、最適な組合せを求めることを確認できた。

6. おわりに

紙文書電子化における OCR 認識結果の目視確認工数の削減を目的として、複数の認識結果の正誤をユーザが一括で判断可能な画像を生成する目視一括化方式を提案した。本方式は、単語画像を重ねて、各ピクセルについて論理演算をとる合成画像の生成を行う。また、実際の帳票画像に適用するために以下の二つの処理を特徴とする。

- 字体が類似する単語画像のクラスタリング。
- 最大全域木問題による位置合わせ条件の最適化。

本方式により生成した合成画像により、文字認識結果が正しいことをユーザが一括で確認できることを実現した。社内の営業文書を用いた実験により、従来の全件目視確認の工数を 39% 削減し、本手法の有効性を確認した。

参考文献

- 1) 社団法人 日本画像情報マネジメント協会: 統合文書情報マネジメント(ECM)関連市場動向調査 (2010).
- 2) 情報端末事業委員会, OCR 専門委員会: OCR 関連装置/ソフトの市場見通し (2010).
- 3) 社団法人 電子情報技術産業協会: OCR ガイドブック (2010).
- 4) 荒木哲郎, 池原悟, 小松康則: 日本語文の相互干渉誤り文字列の検出・訂正方法, 情報学基礎研究会報告, Vol.98, No.60, pp.9-15 (1998).
- 5) 宮原末治, 鈴木章: パターン間照合を利用した文字認識結果の確認修正法, 電子情報通信学会論文誌 Vol.J77-D-II, No.9, pp.1735-1743 (1994).
- 6) 荒木哲郎, 池原悟, 塚原信幸, 小松康則: マルコフモデルを用いた OCR からの誤り文字列の訂正効果, 自然言語処理, Vol.102, No.13, pp.97-104 (1994).
- 7) 竹内孔一, 松本裕治: 統計的言語モデルを用いた OCR 誤り訂正システムの構築, 情報処理学会論文誌, Vol.40, No.6, pp.2679-2689 (1999).
- 8) 竹内孔一, 松本裕治: 統計的形態素解析と文字 n-gram を利用した OCR 誤り訂正, IPSJ SIG Notes, Vol.99, No.22, pp.17-24 (1999).
- 9) Neubig, G., 森信介, 河原達也: 重み付き有限状態トランスデューサーを用いた文字誤り訂正, 言語処理学会第 15 回年次大会講演論文集, pp.332-335 (2009).
- 10) 本谷秀堅, 出口光一朗: マルチスケール解析に基づく文字領域抽出手法, Technical report of IEICE. PRMU, Vol.96, No.598, pp.1-8 (1997).
- 11) 川田貴俊, 長尾智晴: 大型ディスプレイを用いた手書き文字・図形の認識システム, 情報科学技術フォーラム一般講演論文集, Vol.5, No.3, pp.465-466 (2006).
- 12) Lewis, J. P.: Fast Template Matching, Proceeding of Vision Interface 95, pp.120-123 (1995).
- 13) Briechele, K. and Hanebeck, U. D.: Template Matching using Fast Normalized Cross Correlation, Proceeding of SPIE on Optical Pattern Recognition XII, Vol.4387, pp.95-102 (2001).
- 14) Kruskal, J. B.: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem, Proceeding of the American Mathematical Society, Vol.7, No.1, pp.48-50 (1956).