

汎用ウェブカメラを用いた墨字文書自動読み取りソフトウェアの試作

筑波技術大学保健科学部情報システム学科

大西淳児 小野 東

要旨：この研究の目的は、パーソナルコンピュータと安価な汎用ウェブカメラを利用して自動的に墨字文書を読み上げるソフトウェアを構築することである。ここでは、既存のOCR技術を巧みに利用することによって、汎用ウェブカメラから文書画像から文字を自動認識し、スクリーンリーダーの技術を活用して、自動読み上げを行う。現在、ソフトウェアの試作段階であるため、この報告では、汎用ウェブカメラを用いて、墨字文書画像からの文字認識の可能性について報告する。

キーワード：視覚障害、情報補償、文字認識、画像処理

1. はじめに

日本には、約30万人の視覚に障害のある人がおり、そのうちの約69.7%の人がインターネットを利用していると言われている[1]。視覚に障害のある人が日常生活を送る中で数々の困難なことがある。ほとんどの情報は、印字された媒体で映像として情報を得るものであり、これらの人々にとってこのような情報をアクセスするのは容易ではない。近年では、情報技術の高度化によって、このような情報アクセスの障壁を取り除くための様々な方法が生み出されている。特に、コンピュータを利用した文書読み上げ技術やコミュニケーション技術もしくは、インターネットでの情報検索などによって、幅広い情報にアクセスすることができるようになってきている。

一方で、このような技術は、視覚に障害のある人がテキストをベースとした情報によりアクセスすることが可能となる効果をもたらした。たとえば、拡大読書機などの工学的に映像を拡大する装置や点字などへの変換、テキスト情報を音声へ変換するなどの技術など実に様々な方法が生み出された。墨字文書へのアクセス手段としては、音声出力装置とコンピュータおよびイメージスキャナで構成された装置に、テキスト音声変換技術と光学文字認識技術を融合したソフトウェアを導入する方法などが既にある。しかしながら、これらの方法は非常に効果的であるものの、装置が大きかったり、価格が非常に高価であるなどの問題もある。また、操作性の面においても、ペーパーベースの媒体をスキャナ装置にセットしなければならず、操作が少々煩雑になる欠点がある。そのため、これらの問題を解決するための様々な研究がなされている。カメラを取り付けた携帯端末による文字認識システムでは、アルファベット文字の認識を可能にしている[2]。また、鈴木らは、環境内の文字を認識するシステムを構築し、その有効性を報告して

いる[3]。

そこで、この研究では、より容易に操作ができ、また、非常に安価なコストで墨字文書へアクセスするため、広く一般に市販されている安価な汎用ウェブカメラを用いて、墨字文書を取り込み、文字認識とテキスト音声変換の既存技術を活用した墨字文書の自動読み上げソフトウェアの開発を試みる。現在の研究進捗状況は、ソフトウェアを開発しつつ、墨字文書の読み取り能力向上のための様々な試験を実施している段階である。そのため、この報告では、汎用ウェブカメラから得られる画像から文字認識およびテキスト情報化の可能性について中間報告としてまとめるものとする。まず、第2節では、ソフトウェアの開発環境の概要について述べる。第3節では、ソフトウェアの処理概要について解説する。第4節では、実際に汎用ウェブカメラから得られた画像から文字認識を行った結果について述べる。最後に5節では、今後の予定と課題についてまとめる。

2. 試作ソフトウェアの動作と開発環境の概要

まず、試作ソフトウェアが動作するシステム構成について述べる。このソフトウェアは、Windowsを搭載したパーソナルコンピュータで動作し、ウェブカメラは、Logicool社のQCAMPRO9000を用いた。このカメラは、量販店で約1万円弱程度で購入できる汎用のウェブカメラである。また、文字認識した結果を音声で伝えるために、パーソナルコンピュータには、ワイヤレスのヘッドホンを接続した。

ソフトウェアの開発においては、マイクロソフト社のVisual Studio 2008を用いた。

図1は、開発中のシステムの外観を示したものである。このシステムでは、ディスプレイを2台接続し、向かって右側のディスプレイをサブ画面として、ディスプレイの上部に文書画像を取り込むためのウェブカメラを接続してい



図1 開発環境

る。ディスプレイの下には、読み取り対象の文書を置くためのボードを設置し、視覚に障害のある人が文書をどこに置くかわかりやすいようにしておく。

3. 自動読み取りソフトの処理概要

開発するソフトウェアでは、図2に示すように、3つの処理モジュールから構成される。まず、カメラから画像を取り込むモジュールでここで得られた画像はソフトウェアのキャプチャーウィンドウ内で確認できるようになっている。次に、所定の場所に読み取り対象の文章を置き、画像を取り込む、このとき、画像は、ウェブカメラの持つ最大の解像度で取り込む。この研究で用いたウェブカメラでは、800万画素の解像度で取り込むことができる。この取り込んだ画像を、自動文字認識エンジンのモジュールで文字領域の抽出および文字認識を行い、最終的にテキストデータとして出力する。このテキストデータを音声変換モジュールで音声として、パソコンに取り付けられたヘッドホンより出力する仕組みになっている。この音声処理については、既存のスクリーンリーダーの技術を活用することにした。

図3に開発中のソフトウェアのインターフェースを示しておく。

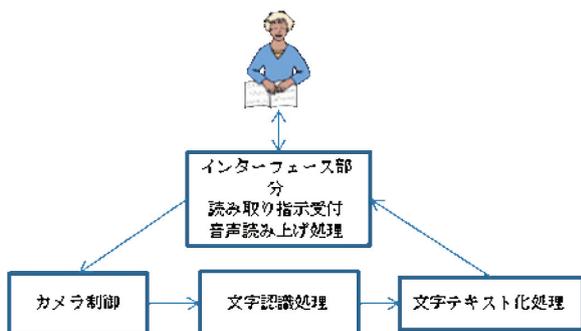


図2 ソフトウェアの処理概要

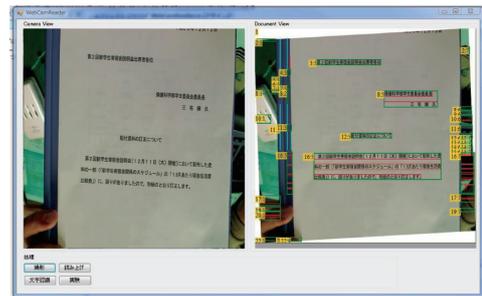


図3 インターフェース画面

4. 文字認識実験

この研究では、ウェブカメラから取り込んだ画像を対象として文字認識を行う。このとき、スキャナを使う場合に比べて文字認識を困難にする要因としては、まず、対象文書以外の複雑な背景画像が含まれることがあげられる。つまり、読み取りたい文書が画像のどの部分に存在するのか認識をすることが困難になる。このような複雑画像から特定の領域を抽出と文字領域のクラスタリングに関する手法も数多く存在する [4][5]。

一方、ウェブカメラによって得られる画像は、スキャナを用いる場合に比べて、きわめて低品質の画像になり、ノイズ、つぶれ、かすれなどの画質劣化してしまう。このような状況に対応するための文字認識手法はすでに様々な報告がされている [6][7]。画質が悪い状態で文字認識するには、一般的に、つぶれた領域を検出して識別関数を補正するなどの工夫がされる。文字の線幅は本来それほど変動しないことから、文字画像の骨格化の処理を応用することによりつぶれた領域を検出ができる。骨格化は、文字画像を輪郭から順に一画素ずつ削っていく細め処理を繰り返すことで行われる。ここで、削る回数に上限をもうけておけば、綺麗な文字は全領域が完全に骨格化される。ところが、低品質文字はつぶれた部分の線幅が1にならずに残る。このことから、黒画素として残った領域を検出することでつぶれた領域を検出が可能である。つぶれによって特徴の分布がどのように変化するかをあらかじめ調べておき、検出されたつぶれによって識別関数を変換することで高精度な認識を実現するなどの方法がある。

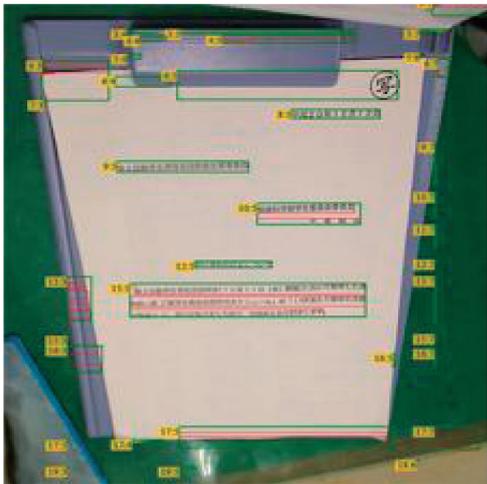
それでは、ウェブカメラから取り込まれた画像から文字を認識する事前実験を行ってみよう。

まず、2節で解説した開発環境を使って、ウェブカメラから文書画像を取り込む。取り込んだ画像を試作ソフトウェアに組み込む文字認識モジュール処理を用いて、文字領域の抽出およびテキスト出力を行ってみる。この結果を図4に示そう。

この結果を見ると文字の領域の認識において、文字以外の部分も誤検出が多少存在していることがわかる。実験では、対象とする墨字文書を原稿台におき、直接画像を取り込み処理を行って、文字領域の検出を試みてみた。このため、複雑な背景画像が要因となって、既存の領域認識のソ



(a) 原稿画像



(b) 領域認識結果



(c) テキスト化処理結果

図4 実験結果

フトウェアモジュールでは正確な検出が困難になっていることがわかる。一方、ウェブカメラは動画として画像を取り込むこともできるため、文書画像を設置する前の状況を記録しておき、これを元に認識対象文書の領域を特定するなどの工夫をすることである程度、改善できる可能性がある。

図4(c)は、テキスト化した結果を示している。この結果にも背景画像による不正な文字多少含まれているものの、ある程度の文字の認識には成功している。しかしながら、認識に失敗した文字が多く見受けられ、読み上げ処理にデータを引き渡すレベルには至っていない。この失敗の原因は、カメラが対象画像を斜め方向から撮影しているため、ウェブカメラを文章の真上方向から撮影できるように、原稿台にカメラ設置のスタンドをとりつけて、撮影条件を改善するなどの工夫を加えることで文字検出率を高めることも可能となると考えている。

5. おわりに

この報告では、汎用ウェブカメラを用いて墨字文書を自動読み上げるソフトウェアの試作の中間的な報告と完成させるまでに解決すべき課題について述べた。4節の実験結果から、自動読み上げを行うには、文字領域と文字認識の過程において、認識率を高めるための工夫が必要となることがわかった。これらの解決には、ウェブカメラが静止画像ではなく、動画としてデータを取り込める特性を活用して、文字認識を行う前に適切な前処理を施すことで解決するなどの方法を検討するなど、さまざまな試験を実施することによって問題を解決することが重要となる。今後は、これらの課題の解決に取り組み、ソフトウェアを完成させる予定である。

謝 辞

本研究は、平成18年度筑波技術大学教育研究等高度化推進事業（競争的教育研究プロジェクト事業）の助成を受けて実施したものである。ここに記して深く感謝申し上げます。

文 献

- [1] 総務省 情報通信政策研究所：「障害のある方々のインターネット等の利用に関する調査報告書」, 2003
- [2] V. Gaudissart et al., 「SYPOLE : mobile reading assistant. for blind people」, In International Conference on Speech and Computer, pp. 538-544, 2004.
- [3] 鈴木・大西他：「視覚障害者のための環境内の文字

- 情報システム」電気学会論文誌 (C), Vol.124-C, no.6, pp.1280-1287, 2004.
- [4] V. Wu, R. Manmatha : 「Textfinder : an automatic system to detect and recognize text in images」, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.21, no.11, pp. 1224-1229, 1999.
- [5] A.C. Bovik, M. Clark, W. Geisler : 「Multichannel texture analysis using localized spatial filters」, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.12, pp. 55-73, 1990.
- [6] Shin'ichiro Omachi, Fang Sun and Hirotomo Aso : 「A Noise-Adaptive Discriminant Function and Its Application to Blurred Machine-Printed Kanji Recognition」, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.3, pp.314-319, March 2000.
- [7] 大町真一郎・阿曾弘具 : 「低品質文字認識におけるつぶれを動的に補正する部分空間法」, 電子情報通信学会論文誌, D-II, vol.J82-D-II, no.11, pp.1930-1939, November 1999.

Development of Reading Assistant Software by Using Web Camera for Blind People

ONISHI Junji, ONO Tsukasa

Faculty of Health Science, Tsukuba University of Technology

Abstract: The main aim of this study is to build an automatic text reading assistant using existing hardware associated with innovative algorithms for blind or visually impaired people. A web camera was chosen for its low-cost price. Three key technologies are necessary: text detection, optical character recognition and speech synthesis. The software is under development now. Thus, this report describes an evaluation of character detection from the images captured by a web camera.

Keyword: Visually Impaired, Communication Support, Image Processing, Character Recognition