

ステレオ画像を用いた書籍画像の形状歪み・陰影補正による文字認識 および読み上げ順序を考慮した音読システム

大石 将太 山下 淳 金子 透

静岡大学工学部 〒432-8561 静岡県浜松市中区城北 3-5-1

E-mail: {f0030009,tayamas,tmtkane}@ipc.shizuoka.ac.jp

あらまし 視覚障害者が活字から情報を得ることは困難である。従来、活字を音読することにより情報取得の手助けを目的とした製品が存在する。しかし、これらの製品はスキャナ装置によって書面の画像を読み取るため、書籍の綴じ目付近に形状歪み、陰影が生じる。この形状歪み、陰影は文字の誤認識の原因となるため、本研究では形状歪み、陰影を考慮し、正確な読み上げを行うシステムを構築する。

キーワード 文字認識、音声読み上げ、ステレオ計測、視覚障害者援助

1. 序論

書籍などの紙媒体の印刷物は我々の生活において重要な情報源の1つであるが、視覚障害者にとって紙媒体の印刷物から情報を得ることは容易なことではない。そこで従来から、活字を音読することにより視覚障害者の情報取得の手助けをする試みが行われている。視覚障害者のための既存製品としてはイメージスキャナとパソコンを組み合わせた音声読み上げ読書器などが製品化されている[1]。これらの製品は視覚障害者の情報取得の手段として有効である。しかし、文書画像をスキャナ装置によって読み取るため、書籍のような印刷物においては綴じ目部分がスキャナ面に密着しない場合がある。この場合綴じ目付近に文字の歪み・陰影が生じ、文字の誤認識につながるという問題点がある。さらに、1ページごとに書籍をスキャナ装置にセットすることは視覚障害者にとって煩雑な作業である。

一方、カメラにより書面画像を取得し、読み上げを行う研究、製品もあるが[2][3]、書面形状による文字の歪みが原因となり、文字の誤認識が問題となる。

そこで、本研究では書籍において正確な読み上げを可能にするため、書籍画像の形状歪み・陰影を補正し、文章を音読するシステムの構築を目的とする。

また書面内容を文字認識した結果に関して、読み上げる順序を考慮する必要がある。具体的には、書面内容がページ順序どおりに音読されない問題や、ヘッダー・フッターを本文と区別なく読み上げるなどの問題が生じ、書面内容全体として意味が通らなくなる場合がある。これらの問題の解決も、使用者に正確な情報を与える上で不可欠であるため併せて行う。

2. 処理手順

本システムの処理概略を図1に示す。システムの構築に必要な処理は以下の4段階に分けられる。

1. 書籍画像の取得を行い、計算機内に画像データを送る (2.1 節)
2. 計算機内で形状歪み・陰影の補正を行い、書籍画像を補正する (2.2 節)
3. 補正された画像より、文字認識を行い、文書情報を取り出す (2.3 節)
4. 読み上げ順序を考慮した音読を行う (2.4 節)
前提条件として対象は日本語書籍および英語書籍とし、図や挿絵のない本文が1段組のものを使用する。

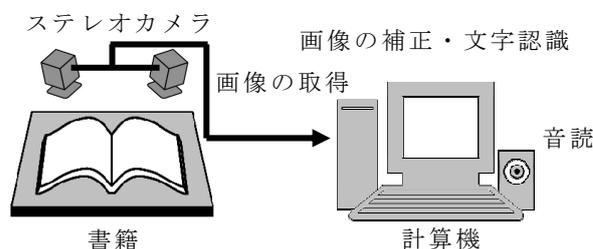


図1 処理の概略

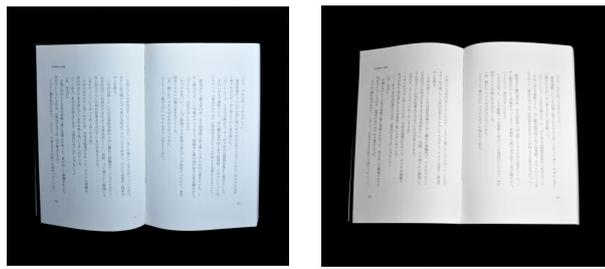
2.1 書籍画像の取得

書面が歪曲している見開き状態の書籍からでも画像が取得できるように、本研究ではカメラにより画像を取得する。撮影の際には、図2(a)(b)のように見開きの書面全体が写る様に書籍を設置し、2台のカメラによってステレオ撮影を行う。なお書籍を撮影する前にカメラキャリブレーションを行い、2台のカメラの位置・姿勢などの外部パラメータと内部パラメータを求め、以後の撮影ですべてのパラメータは変えないものとする。

取得した画像に関して、書籍以外のものが存在すると以後の処理が正常に行えない。そこで事前に2台のカメラで背景のみの画像を取得しておき、背景差分をとることにより書籍のみを自動抽出する。そして、次節の処理は平行ステレオモデルを想定しているため、取得画像を平行化する処理も一連の処理として自動で行う。

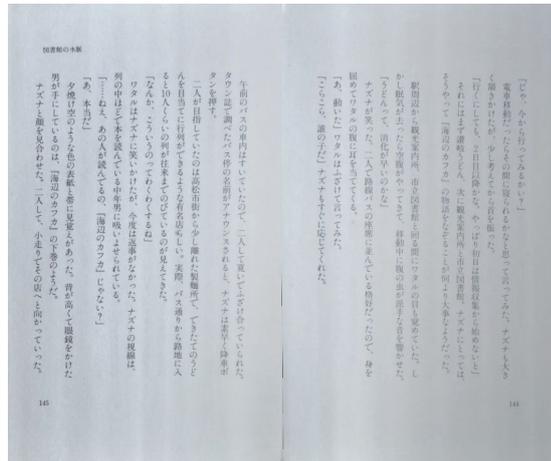
2.2 形状歪み・陰影の補正

カメラによる取得画像は書面形状の歪み・陰影により、そのままでは良い文字認識結果が得られない。そこで本研究では、ステレオカメラにより取得した書籍画像の形状歪み・陰影補正を行う手法[4]を用いる。手法[4]では、まず取得したステレオ画像により書面の形状を推定し、得られた形状を元に陰影の補正を行う。その後推定形状を引き伸ばすことにより形状歪みを補正し、最終的な補正画像を生成する。この補正結果を図2(c)に示す。



(a)左画像

(b)右画像



(c)補正結果

図2 形状歪み・陰影補正の結果

2.3 文字認識

ここまでの処理で得られた画像は文書としての情報を持っておらず、このままではコンピュータ内で文書として扱うことができない。そこで文字認識を行うことにより画像データから文書情報への変換を行う。本研究では市販の文字認識ライブラリを導入することにより文字認識を行う。

文字認識処理部では書籍の見開き2ページを1枚の画像として入力している。その場合、通常右ページと左ページにそれぞれ本文が記載されており、ページ綴じ目に伴う空白により本文が分割されている。また、章題やヘッダー・フッターなども空白により区切られ、読者が区別しやすいようになっている。これらの空白で分別され、ひとつのまとまりになっている複数の文字を本論文では文字群と呼ぶこととする。このような文字群の分別に関しては様々な研究が行われており[5-7]、また市販のソフトウ

エライブラリも存在する。本システムではソフトウェアライブラリを導入することにより文字群を分別する。分別された文字群の例を図3に示す。図3では分別された文字群単位を矩形で示している。その後、分別された文字群単位で書面画像は文字列に変換される。文字群分別のためのライブラリは紙面上の黒画素の分布状況、連続状態などから文字群の分別を行う。その際、画像中におけるそれぞれの文字群を囲う矩形の左上頂点の座標と幅、高さ、面積、及び文字群が縦書きと横書きのいずれかであるかなどのデータも取得しておき、2.4節の処理においてこれらの情報を用いる。

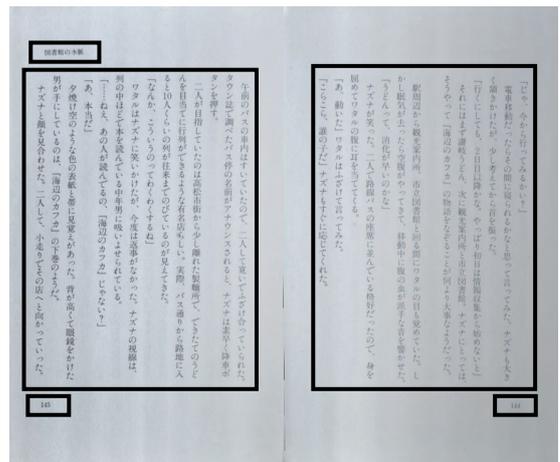


図3 文字群の分別例

2.4 読み上げ順序を考慮した音読

本研究では音声合成ライブラリをプログラム中に導入することにより、得られた文書情報を音声データへと変換し、スピーカーによる出力によって書籍の情報を使用者に伝える。しかし、2.3節で得られた文字列は出力順序が考慮されていないため、そのまま音読すると使用者に正確な情報を伝えることが出来ない。そこで、読み上げ順序を考慮した音読を行うための手法を提案する。

まず、ヘッダー・フッター部分が本文と混同しないような処理を行う。2.3節で得られた矩形の座標をもとに、分別された文字群全体がヘッダー・フッター領域内に存在するかどうかの判定を行う。ここでヘッダー・フッター領域とはそれぞれページの上・下からページ高さの一定割合内にあたる領域とする。この領域に存在すると判定された文字群は全てヘッダー・フッター候補として扱う。この際、本文をヘッダー・フッターとして認識してしまわないよう、文字群の面積が一定値より大きいものはヘッダー・フッター候補から外す。ヘッダー・フッター候補の中から最終的に残ったものをヘッダー・フッターとし、音読の際は出力しないこととする。

実際の認識例を図4に示す。図4において破線の矩形領域がヘッダー・フッター領域であり、実線の

矩形領域が文字群である。ヘッダー・フッター領域に存在する文字群の中でも、本来本文であるものはフッターとして認識しない。

次にヘッダー・フッターでない文字群を本文として扱い、読み上げ順序を決定する。2.3 節で得られた矩形の座標データをもとに、縦書き文書ではページの右側に存在する文字群から音読を行い、横書き文書では左ページの上側に存在する文字群から音読を行うことにより正しい読み上げ順序となるようにする。

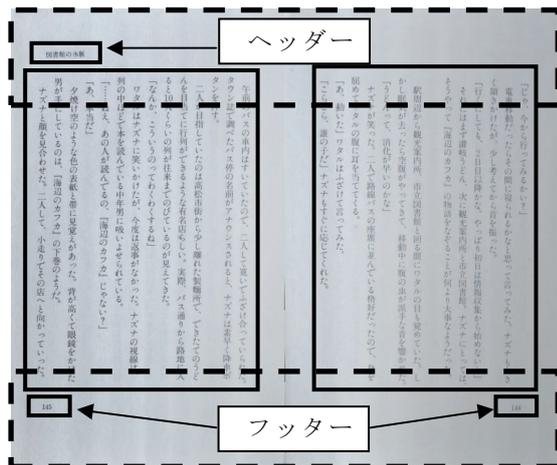


図4 ヘッダー・フッターの認識例

3. 実験

3.1 実験環境

実験は図5に示す環境で行った。テーブルに開いた書籍を置き、上方から2台のデジタル一眼レフカメラによって書面を撮影した。撮影した画像を計算機内に送って処理を行い、スピーカーにより音声を出した。カメラはNikon社製のD700を2台使用し、画像サイズは左右それぞれ4256×2832ピクセルである。処理に用いた計算機のCPUはCore 2 Duo 3.0GHz、メモリは3GBである。また、文字認識ライブラリはPanasonic社製の活字認識ライブラリ Ver.13を、日本語音声合成ライブラリには富士通社製のInspirium 音声合成ライブラリ V3.0を使用した。なお、英語音声の出力にはSoftFarm社製の簡読★ Textを使用した。



図5 実験環境

3.2 複数ページでの評価

本システムの1つ目の評価として日本語、英語書籍それぞれ1冊ずつに対し、連続した10ページ分の画像を取得し、読み上げを行った。その際、形状歪み・陰影の補正を行わなかった場合との文字認識率の比較結果を表1に示す。形状歪み・陰影の補正を行わない場合、10ページ中の平均文字認識率は日本語書籍では93%、英語書籍では90%となった。形状歪み・陰影の補正を行った場合には、日本語書籍では97%、英語書籍では97%となり、形状歪み・陰影の補正により文字認識率が向上したことがわかる。使用した10ページ中の総文字数は日本語書籍では4001字であり、英語書籍では6794字であった。形状歪み・陰影の補正を行った後にも文字の誤認識があり、日本語書籍では118、英語書籍では202文字であった。この原因については3.4節で考察する。また、ページをめくった後に単語がまたいでいる場合、単語は途切れて読み上げられる。

表1 文字認識率の比較結果

	ページ数	補正なし	補正あり
日本語書籍	10	93%	97%
英語書籍	10	90%	97%

3.3 読み上げ順序の評価

本システムの2つ目の評価として、構図の異なる日本語・英語書籍において1冊につき見開き2ページの画像を1枚ずつ処理した。また使用されているフォントは書籍によってそれぞれ異なる。2.4節での手法を日本語文庫本に適用して得られた読み上げ順序の例を図6に示す。この場合は右ページの本文を読んだ後、左ページの本文を読む順序となっており、正しい音読がなされることがわかる。

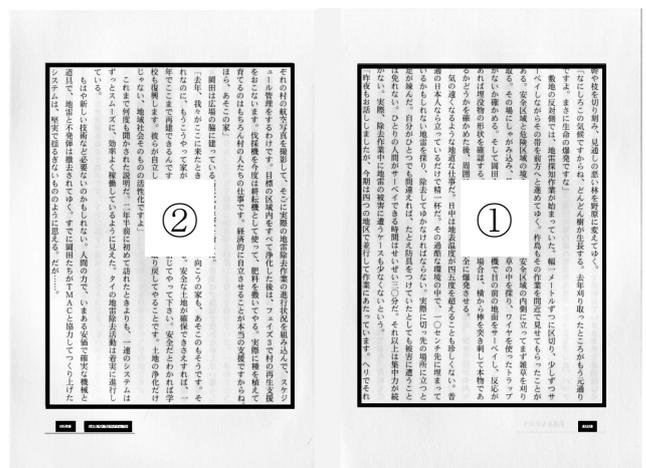


図6 日本語文庫本における読み上げ順序

同様に英語文庫本での読み上げ順序の決定結果を図7に示す。図7では左ページの本文を読み上げた後、右ページの本文を読み上げる順序となっており

正しい順序が得られている。

英語書籍画像 10 枚, 日本語書籍画像 10 枚で実験を行い, すべての画像において正しい読み上げ順序が得られた。また読み上げを行った 20 ページ中の平均文字認識率は日本語書籍では 96%, 英語書籍では 97% となり, 画像を入力してから音声が出力されるまでの 1 ページ分あたりの平均処理時間は日本語書籍では 13 秒, 英語書籍では 14 秒となった。これらの結果を表 2 に示す。

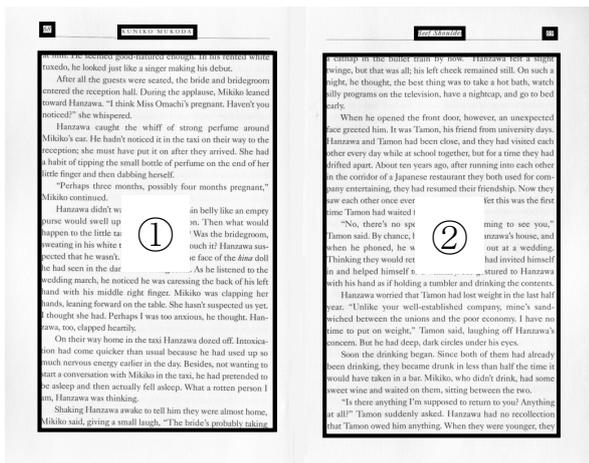


図 7 英語文庫本における読み上げ順序

表 2 実験結果

	ページ数	文字認識率	平均処理時間
日本語書籍	20	96%	13 秒/ページ
英語書籍	20	97%	14 秒/ページ

3.4 考察

2 章の処理手順で述べた前提条件の下, 実験を行った全ての書籍において読み上げ順序に関しては正確な読み上げを行うことができた。雑誌などに比べレイアウトが簡素である文庫本のような書籍に関して, 本研究での読み上げ順序の考慮は有効であることがわかった。

実験において実際に出力された音声を聞くと, 文字の誤認識により書籍の内容とは異なる文字を読み上げる箇所が見られた。使用者の常識でカバーできる程度の間違いはあるが, 一文字変わるだけで意味が大きく変わってしまう場合もありうる。

文字の誤認識に関しては, 英語書籍において数字の「1」とアルファベット小文字の「l」など似たような文字で見られた。日本語書籍においては画数の多い漢字で誤認識が多い傾向が見られた。これらの誤認識の原因は取得画像の解像度が低いためであると考えられる。従って解像度の高い画像を取得する手法が必要である。

4. 結論

本研究ではステレオカメラにより取得した書籍画

像の形状歪み・陰影補正を行う文字認識および読み上げ順序を考慮した音読システムの構築を行った。形状歪み・陰影の補正を行うことにより, 文字認識における誤認識を軽減させ, 使用者に伝える文字情報の正確さを向上させた。また, 読み上げ順序を考慮することにより, 書面に記載された内容を正確に使用者に音声情報として伝える手法を構築した。

実験結果より, 書籍の内容を正しい読み上げ順序で読み上げることができることを確認した。

今後の課題を以下に述べる。本システムではカメラを固定しているため, 見開き 2 ページを音読するためには書籍全体を写す必要がある。紙面からカメラまでの距離を大きくすると 1 文字あたりの解像度は低くなる。現状としては, 取得画像の解像度は文字認識ライブラリの推奨値の半分程度しか得ることができず, 複雑な漢字などの文字において誤認識が起こりやすい。

誤認識を減らすためには取得画像の解像度を高める必要がある。解像度を高めるため, 画面の一部分を拡大した画像を複数枚撮影し, それらモザイクングすることにより高解像度の画像を取得する方法が考えられる。自動的にモザイクングを行うことが出来る装置を構築し, 本システムに組み込む事により, 画像の解像度を高め, より正確な情報を視覚障害者に伝えるシステムの構築が今後の課題として挙げられる。

文 献

- [1] 韓星民: “文字情報へのアクセスとその支援”, 生存学研究センター報告 6 視覚障害学生支援技法, pp.31-68, 2009.
- [2] 大西淳児, 小野東: “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”, 筑波技術大学テクノレポート, Vol.16, pp.18-21, 2009.
- [3] Intel Reader <http://www.intel.com/about/companyinfo/healthcare/products/reader/index.htm>
- [4] 秋本廉太郎, 山下淳, 金子透: “ステレオカメラを用いた歪曲形状書籍画像補正の高速化”, 第 15 回画像センシングシンポジウム講演論文集, pp.IS1-14-1- IS1-14-8, 2009.
- [5] 辻正博, 米田正明, 長谷博行, 酒井充: “視覚障害者用文書画像認識システム”, 電子情報通信学会秋季大会講演論文集, Vol.1994, p.323, 1994.
- [6] 中島昇, 田中直哉, 山田敬嗣: “拡張スプリット検出法による文書構造・読み取り順序解析”, 電子情報通信学会総合大会講演論文集, Vol.1998, No.2, p.243, 1998.
- [7] Yasuto Ishitani: “Logical Structure Analysis of Document Images Based on Emergent Computation”, IEICE Transactions on Information and Systems, Vol.E88-D, No.8, pp.1831-1842, 2005.