

視覚障害者用OCRシステムの設計

1 S - 4

菅原一秀

日本アイ・ビー・エム株式会社 東京基礎研究所

1 はじめに

視覚障害者が簡単に操作でき、日常的な印刷情報にアクセスすることのできる文字認識システム、光学的メディアリーダー（以下、OMRと呼ぶ）の試作を行なった。[1][2] 出力は合成音声によるので、文書の内容に効率良くアクセスするには文書の論理構造に基づいた構造化が必要となる。また、表題や、ページ番号などヘッダ、フッタなどに記載されている情報については、これらの主に物理的配置に基づいた構造の抽出を行ない、本文とは別にアクセスできるようにした。

2 システムの概要

OMRは、スキャナーから雑誌などの日常の文書を読み込み、音声で音声で出力するシステムである。システムは図1に示すように、スキャン、画像前処理、文字認識、文書構造化、および、それらを制御し、音声出力を行う部分からなる。以下に、主要要素技術について述べる。

3 ページ単位の読みとりに関する主な機能

3.1 タイトルマッチング

タイトルマッチングとは雑誌の表紙をスキャンし、あらかじめ登録しておいた雑誌のうちからもっとも適合するものを選びだし、音声で雑誌名を伝える機能である。

3.2 文書画像の回転角の検出及び補正

文字領域及び画像領域の検出では水平及び垂直方向への射影に基づく手法を使っているため、スキャンした画像の回転（傾き）は検出結果に重大な影響を及ぼす。光学的メディアリーダーでは画像前処理として、この画像の回転角の検出及びその値に基づく画像の補正を行なっている。

画像の回転角の検出には黒画素の連結成分の中心をサンプル点とするハフ変換を利用している。[3]

3.3 文書の上下方向のチェック

通常は文書の下部分をスキャナーの手前側に置いて操作を行なう。もし、上下を逆にしたまま操作を続けると、文字認識の段階で全く異なった文字と認識され、意味不明の出力を受けとることになる。このことを防ぐため、本システムでは文書の向き（上下方向）のチェックを行ない、文書の向きがどちらでも正しい方向の画像が認識エンジンに送られて処理されるようにした。

文書の上下方向のチェックは、文書のテキスト行の一部を認識エンジンに送り、そのスコアを調べることにより行なわれる。予備的な実験の結果、ある程度の長さを持ち、平均的な高さを持つ5行程度をサンプルとして調べることにより上下の判定について信頼性のある結果を得ることができていることが分かった。

3.4 パラグラフ接続

テキスト行は一行毎にその座標を認識エンジンに送られ、そこで文字認識が行なわれ、結果が返される。認識結果は座標と共に処理されて、インデントの有無や、行末の位置によって、見出しやパラグラフの元となるブロック単位に編成し直される。また、この際、縦書き、横書き、段組を考慮して各ブロックの読みの順序を決定する。具体的な手順は次の通りである。

1. 縦書き、横書きが共にある時はそれぞれを別に扱う
2. 横書き部分について、x-y 切断に基づく方式によってブロックを並べる。
3. 縦書き部分についても同様に、x-y 切断に基づく方式によってブロックを並べる。
4. 縦書き、横書きが共にある時は、分量の多い方が主要部であるのでそれを先に並べる。

ブロックがパラグラフの場合は段組、または他の領域（画像など）を越えて接続可能かどうかを、次の要素を考慮して決定する。

- 先行ブロックと後続ブロックの文字高の一致度
- 先行ブロック及び後続ブロックの行数が2以上

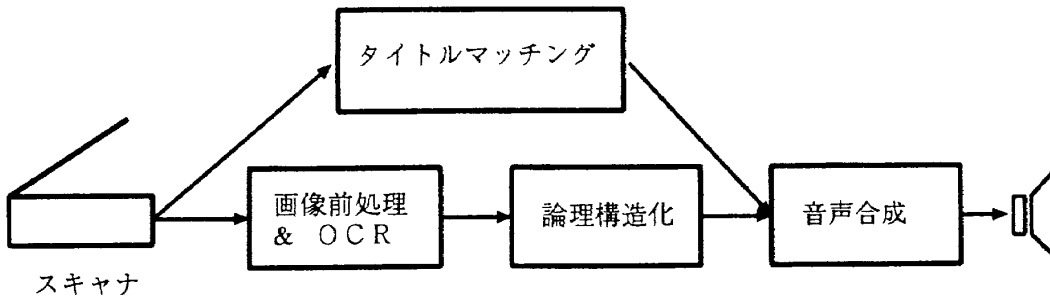
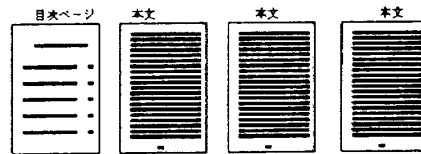


図 1: 光学的メディアリーダシステム

- 先行ブロックの最終行の行末の位置がブロックの端に近い。
- 後続ブロックのインデントが浅い



ページ番号による
リンク作成

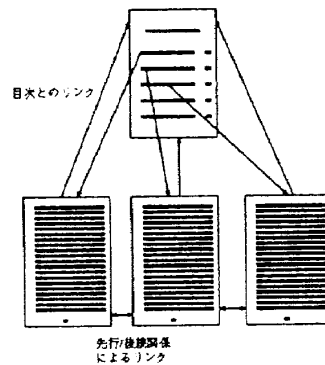


図 2: 複数文書の結合

3.5 文字認識結果の出力

文字認識の結果は、行位置の情報に基づいて、論理的な構造化が行なわれ、その情報を表すタグ付きのテキストという形式でファイルに出力される。これは最終的には音声/点字を通じてユーザに伝達される。

4 論理構造化文書の管理

目次情報から、ページ番号を介して対応する本文のページを読み込んだファイルに対するリンクを張る機能を開発した。

4.1 ファイル管理

認識をさせたファイルを保管する場合はある作業ディレクトリの下にファイルを作成する。目次ページの場合には特別なファイル名をつける。

4.2 ページ番号によるリンク機能

あるディレクトリの下でいくつかのページ番号を振られた、構造化された文書がある時、それをページ番号順にならべ、前後へのリンクを作成する。また、目次ページがある場合、そのページ番号と、本文のページ番号の対応を取り、目次のタイトルから、対応するページへのリンクを作成する。

図 2 に目次ページを含む複数文書の結合の概念図を示す。

無理にブロックを接続してしまった場合、その部分の意味が非常に分かりにくくなってしまいますので、接続のためのパラメータは条件がきつめに設定されている。

5 まとめ

視覚障害者用文字認識システムを試作した。文書の論理構造に基づいた構造化を行ない、内容に効率良くアクセスできるように設計した。今後はユーザによる評価を行ない、より使いやすいシステムへと改良していく予定である。なお、本研究は医療福祉機器技術研究開発制度の一環として、新エネルギー・産業技術総合開発機構 (NEDO) からの委託により実施したものである。

参考文献

- [1] 菅原一秀: “視覚障害者用文字認識システム”, 情報処理学会第 51 回全国大会講演論文集, 4T-7, Sept., 1995
- [2] 菅原一秀: “視覚障害者用文字認識システムにおけるレイアウト/論理情報抽出法”, 情報処理学会第 53 回全国大会講演論文集, 1Q-8, Sept., 1996
- [3] Kazuhide Sugawara: “Weighted Hough Transform on a Gridded Image Plane”, 4th ICDAR, pp. 701-704, Aug., 1997