ヒューマンインタフェース 82−10
（1999. 1. 29）

視覚障害者用ＯＣＲシステムの開発

菅原一秀
日本アイビーエム株式会社東京基礎研究所
E-mail: sugawara@jp.ibm.com

視覚障害者が独力で印刷文書を読むことができるシステムを紹介する。従来のこのようなシステムは印刷された文章の音声への置き換え機能を提供するだけなので、音声の一次元性により、ユーザが読みたいところにすばやくアクセスすることができなかった。この問題に対して、われわれは印刷文書の論理構造を抽出し、それを利用してユーザが求める情報へすばやくたどりつくことができる新しい方式を提案した。また、そのために木構造を渡り歩くためのコマンドを提供した。階層化された音声メニューやヘルプも提供されている。われわれはこの方式を雑誌を読む課題に適用した。表紙による雑誌の判別や、目次構造の利用、ページ番号の取得、構造化された文書の管理などの機能も含め、雑誌からの総合的な情報取得のためのシステムを作り上げた。

Development of an OCR System for the Visually Disabled

Kazuhide Sugawara
Tokyo Research Laboratory, IBM Japan Ltd.

In this paper, we present a system for reading printed documents, which can be used by the visually disabled. Our system uses speech to transmit the printed information to the users. To cope with the slowness of the information transmission speed caused by the one-dimensionality of speech, the logical structure of the scanned documents is extracted and used to accelerate the access to the information. A tree-traversing command is provided for navigating through the logical tree generated by the system. A layered speech menu/help is used to guide users. The scanned and recognized pages can be linked on the basis of the page numbers detected, and can be stored in a file system for later use.

1 はじめに

最近まで視覚障害者が印刷文書を読むことは、晴眼者によって点訳された一部のものを除いて、不可能であった。近年、一部の文字認識システムでは視覚障害者が独力で印刷文書を読むことができるようになっているものもある。しかし、これらのシステムは印刷された文章の音声への置換え機能を提供するだけであり、音声を時間系列に沿って追いかける必要があるので、ユーザが読みたいところにすばやくアクセスすることができない。この問題に対して、われわれは印刷文書の論理構造を抽出し、それを利用してユーザの求める情報へ素早くたどりつくことができる新しい方式を提案した。［5］われわれはこの方式を雑誌を読む課題に適用した。そこでは単に本文のOCR結果を読み上げるのではなく、表紙による雑誌の判別や、目次構造の利用、ページ番号の取得などの機能も含めた、雑誌からの総合的な情報取得のためのシステムとして、光学的メディアリーダ（OMR）と名づけ、試作を行った。

システムはスキャナから入力された文書のページから論理構造をツリーの形で抽出する。ユーザはシステムの提供するツリーを辿る機能を利用して文書を読むことができる。ツリーを辿る基本機能としては、下位ノード、同位ノード間、上位ノードへの移動／読み上げ、深さ優先でのノード自動遷移／読み上げがある。これらの移動コマンドはユーザが容易に操作できるようキーボード上の数値入力キーパッドに割り当てられている。さらに、ページに関する物理的な情報；ヘッダやフッタは文書本体とは別に取り扱われ、本体の論理構造を辿って読んでいるときにも、それとは独立にアクセスすることが可能となっている。現在光学的メディアリーダはユーザによる評価を行っている。

以下では、第2節でシステムの概要を説明し、第3節で主な機能と論理構造抽出の手法について記述する。第4節では複数のページにわたる文書の管理について説明する。最後に第5節でまとめと今後の課題について述べる。

2 システムの概要

システムへの入力は文書をスキャンすることによって得られる。OMRの主な機能は雑誌の表紙の判別、及び目次ページ、本文に対する論理構造抽出と読み上げ、ページ番号処理である。

雑誌の判別の際の入力はカラーで行われる。入力された画像は、あらかじめ保管されていた雑誌の表紙のテンプレートに対しマッチングが行われる。

目次や本文を読ませる場合の入力はモノクロである。スキャンされた画像は文書の傾きを検出し修正する前処理を経たあと、テキス

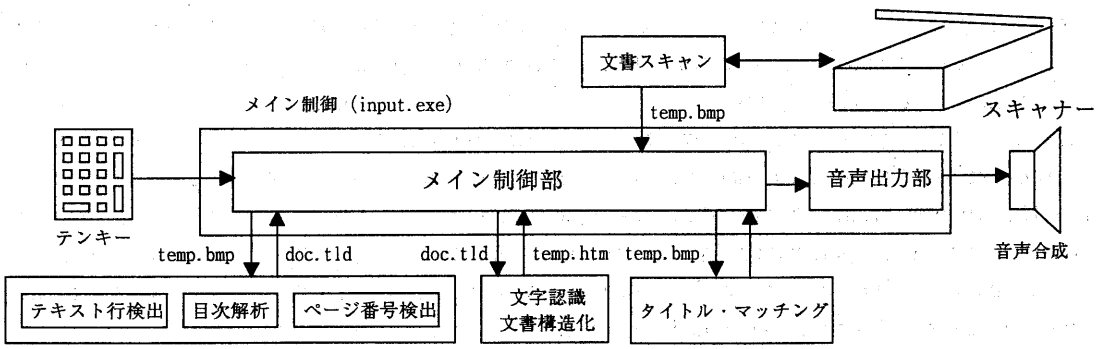


図1．システムの構成

ト行の検出が行われる。検出されたテキスト行の画像はＯＣＲエンジンに送られ、文字コードに変換される。 その後、レイアウト解析が行われ、論理構造が抽出される。テキストはその論理構造を示すタグとともにファイルに格納され、合成音声によりユーザに提示される。

ユーザは論理構造を表す木のどこに移動するか、合成音声の速さ、ピッチなどの制御、文書管理などのコマンドは数値キーパッドから入力する。点字出力機器もシステムに接続することができる。システム構成を図１に示す。

3 主な機能

3.2 タイトルマッチング

雑誌の名前を判別するためにタイトルマッチング機能を作成した。多くの場合、雑誌の表紙の雑誌名の部分は色使いを除いて一定している。この領域が雑誌の特定に使用される。一定の色を持つ領域の輪郭が検出され、あらかじめ作成されていた表紙のテンプレートとマッチングを取られる。 最良のスコアを与えるテンプレートに対応する雑誌名がスコア付きで読み上げられる。

3.2 傾き検出と修正

テキスト行を検出するため、ＯＭＲは水平および垂直方向への投影操作を行う。そのためスキャンした文書画像の傾きを検出し、修正することが重要となる。われわれは文書画像注の黒画素連結領域の中心のハフ変換を利用した。[1, 2, 3] 検出範囲は±１０度、分解能は 0.1 度である。文書画像の傾きは縦行及び横行に対し同時に計算され、その結果に基づき正規の位置まで回転される。

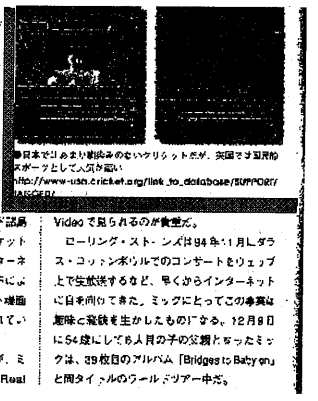
3.3 テキスト行検出

多くの日本語の雑誌は同じページ上に縦行および横行が現れる。そこで、行検出は縦、横両方向を対象にする必要がある。われわれのテキスト行検出方法は２段階で行われる。まず、テキスト領域を決め、次に、それぞれのテキスト領域の中の連結領域の配置の具合により、テキスト行の方向を決定する。

テキスト領域の決定には黒画素連結領域の高さ、幅、黒画素の平均ランレングスなどの統計的特徴を利用する。

図や写真が一つの段の幅の倍数でないときにはテキストがその横や縦に流し込まれて、凹状のテキスト領域が形成されることがある。これが従来の縦・横への投射に基づくテキスト領域検出方式，再帰的ｘ－ｙ 切断方式 [4] による誤りの原因の多くを占めていた。

われわれは再帰的ｘ－ｙ切断方式を拡張した。従来法では、テキスト領域、画像領域を分離するのに利用できるのが縦又は横方向への切断だけだったものを、新しい方式、Ｌ－字型分離法ではＬ字型の切断も許すよう拡張した。(図２)


ローリング・ストーンズのミック・ジャガーは、クリケットの熱狂的なファンとしても知られるが、昨年秋、メジャーなクリケットの試合をウェッブで放送する新会社「ジャグド・インターネットワークス！」を設立。本格的にインターネット事業に進出した。その第一弾は『アカイ・シンガー・チャンピオンズトロフィ』。PM年12月11日から19日までUAE（アラブ首長国連邦）のシャルジャで行われ、英国、インド、パキスタン、西インド諸島が参加したトーナメント戦だ。組合クリケットサイト「Cricinfo」内のジャグド・インターネットワークスのページで、全7試合の音声による生中継とRealVideoによるハイライト場面の映像が放送された。過去の試合も視聴されているので必見。

写真、記録等のデータも充実しているが、ミック・ジャガー自身の新築コメントがReal Videoで見られるのが貴重だ。

ローリング・ストーンズは94年11月にダラス・コットン・ボウルでのコンサートをウェッブ上で生放送するなど、早くからインターネットに目を向けてきた。ミックにとってこの事業は趣味と実益を生かしたものになる。12月8日に54歳にして6人目の子の父親となったミックは、39枚目のアルバム「Bridges to Baby on」と同タイトルのワールドツアー中だ。

日本ではあまり馴染みのないクリケットだが、英国では国民的スポーツとして人気が高い

http://www-usa.cricket.org/link_to_database/SUPPORT/RASCEB/

図２ Ｌ字型分離領域

ある領域に対し、最良の分離を求めるために、その領域内の黒画素連結領域の外接矩形に対し、縦・横方向の投影の４つの組み合わせを計算する。例えば「┌」の形の分離をするのには、上向き及び、左向きの投影を行う。その時内部に残された最大の矩形がこの

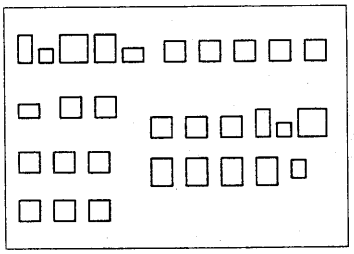


図3 黒画素連結領域の外接矩形

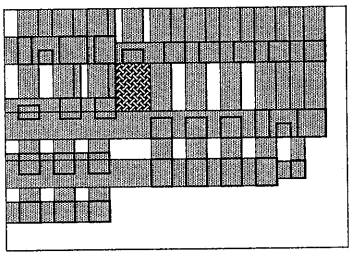


図4 上、左への射影と内部最大矩形

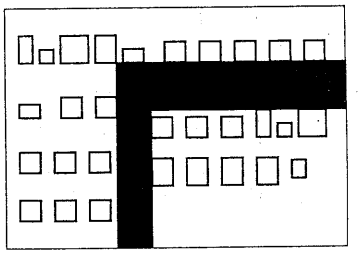


図5 検出されたL字型領域

形の分離の内で最良のものを与えるL字型の角になる。(図 3, 4, 5) 同様な操作を残りの3つの投影の方向の組み合わせに対して行う。すべての中で最良の分離を与えるものを選び、領域の分離を行う。この操作は再帰的に繰り返される。

テキスト領域の中でテキスト行の向きを定めるために、われわれは縦、横2つの方向に対してもっともらしさを計算する。

あるテキスト領域に対し、縦及び横方向のテキスト行候補を独立に生成する。テキスト行は連結領域を縦および横方向に投射して生成する。

Tを横行の候補としよう。WをTの幅とし、W'をT中の連結領域の幅の和とする。高さについても同様にHとH'を定める。nをT中の連結領域の個数とする。テキスト行は横方向であるという仮定から、$R_{hzH}(T)=H'/(n \times H)$ 及び、$R_{hzW}(T)=W'/W$ はどちらも 1 に近いことが期待される。同様にもしTが縦行であれば、$R_{hzH}(T)=H'/H$ 及び、$R_{hzW}(T)=W'/(n \times W)$ は1に近いことが期待される。

われわれはテキスト行の方向の判別をするためにこれらの比の対数の二乗和を計算する。

T_{hz} 及び T_{vt} をテキスト領域の横及び縦方向のテキスト行の集合とする。この領域に対する横及び縦方向の指数 S_{hz} 及び S_{vt} を次のように定義する。

$$S_{hz}=\sum_{T\in T_{hz}}((logR_{hzH}(T))^2+(logR_{hzW}(T))^2)$$

$$S_{vt}=\sum_{T\in T_{vt}}((logR_{vtH}(T))^2+(logR_{vtW}(T))^2)$$

この指数が小さければ小さいほどそのテキスト領域の行の方向はその指数に対応する可能性が高い。このようにしてテキスト領域内の行の方向及び行候補が求まる。

3.4 ページの向きの検出

われわれの使用しているOCRエンジンは文字の向きを決定できない。それは文字の画像が上下正しいことを仮定し、最良の認識スコアを与える文字のリストを返す。もし、上下逆さの文字の画像を与えると、OCRの返す文字をならべたものは理解できないものになってしまう。われわれのOMRはこの問題をページの向きを少量のテキスト行候補の認識スコアを調べることで解決した。テキスト全体を使わないのは文字認識に要するコスト(時間)が比較的大きいためである。検査に使うテキスト行は次の基準で選ぶ。

1. テキスト長がページ内の平均より長い
2. 行の高さがページ内の文字の高さの平均と同程度
3. ページ内の分散した場所から選ばれている

われわれの予備的な実験によればテキスト行5つ位でたいていの場合に正しいページの

向きを判定することができた。

3.5 テキストブロック生成とパラグラフ接続

スキャンされたページへの素早いアクセスを提供するために、システムはページの論理構造を抽出し、その構造に沿ってＯＣＲした結果を配置する。最初に文書の物理的配置を再帰的 x-y 切断を用いて抽出する。テキストブロックはこの段階で得られる。次に論理構造が、テキスト行の配置に基づいて推定される。

テキスト行の画像はＯＣＲエンジンに送られ、認識結果がスコアとともに返される。認識結果はテキスト行の座標と共にテキストブロックの生成に使用される。テキストブロックは行間の大きさ、インデントの有無、行末の位置などいくつかの要素により決定される。

もし、テキストブロックが少数の行からなっており、その連結領域のサイズがページの平均より大きければそれは見出しと判定される。そうでなければパラグラフとみなされる。パラグラフはしばしば画像領域によって分断されたり、別の段に続いたりする。このようなパラグラフの断片を接続するためにわれわれは以下の条件を調べる。

1. 文字サイズの違い
2. パラグラフ断片中の行数
3. 先行パラグラフの最終行のパラグラフ断片領域に対する相対位置
4. 後続パラグラフのインデントの深さ

目次ページは別方式で扱う。このような特別なレイアウトを持つページの処理にはレイアウトモデルを利用する。現在われわれのシステムで扱えるのは横行で構成された目次ページにかぎられている。

3.6 認識結果のユーザへの提示

テキストブロックの認識結果は抽出された論理構造とともにタグ付きの文書として保管される。タグはテキストブロックの論理構造及び物理的な情報を表す事ができるように定められる。現在、見出し、パラグラフ、ページ番号、座標値などを使用している。

認識されたテキストは合成音声で読み上げられる。ユーザへのインターフェースとして認識結果をナビゲートするための論理構造木を辿るコマンド及び、階層化された音声メニュー／ガイドが提供される。すべてのコマンドは数値キーパッド領域のキーに割り当てられている。システムは「文書読み上げ」、「入力機器制御」、「合成音声制御」などのいくつかの操作モードを持っている。それぞれのモードは数値キーへの独自のキーマップを持ち、操作モードつまりキーマップ自体の切り替えは特別に割り当てられたキーを押すことで実行する。

それぞれのモードでは基本的な操作がツリー状に配列されているので木構造をたどることで操作の選択／実行ができる。例えば親子関係を辿るのは「６」と「４」、兄弟関係を辿るのは「２」と「８」である。文書読み上げモードでは文書自体がツリー状に構成されているので自然に操作できる。特殊な操作、例えば現在位置から木構造を深さ優先ですべて読み上げるとか、一時的に本文を離れ、ページの物理情報を読み上げるとかなどは未使用のキーに割り当てられている。

文書を読み込んだ場合の最初の読み上げ位置はそのタイトル（又はファイル名）で、次の層は見出し群からなる。ユーザは見出しを走査して興味のあるところを見つけ、木構造を降下するキーを使って詳細を聞くことができる。この構成によってユーザは興味のある個所に素早く辿り着くことができる。

4 構造化文書管理

雑誌のページはＯＭＲを利用してスキャンされ、認識され、読み上げられ、必要によっては保管される。われわれは雑誌一冊を一つの文書としてまとめて扱えるよう設計を行っ

た。構造化文書の管理を容易にするためにページ番号に基づいて、それがない場合はスキャンされた時刻の順序に基づいて、スキャンされたページ同士のリンクを自動的に作成する機能を開発した。目次ページはこの場合重要な役割をになう。以下では目次ページの存在を仮定する。（スキャンされた目次ページがない場合は仮想的に作成する）

1. 目次ページにおいてページ番号と記事のタイトルとの対応を抽出
2. 各本文ページからページ番号を抽出
3. 目次ページのページ番号とタイトルそれぞれに対応する本文ページがあればリンクを作成
4. 本文ページのページ番号(又は作成時刻)をソートし、本文ページを整列
5. 本文ページに目次ページへのリンク及び前後ページへのリンクを作成

図６にリンク結果の模式図を表す。

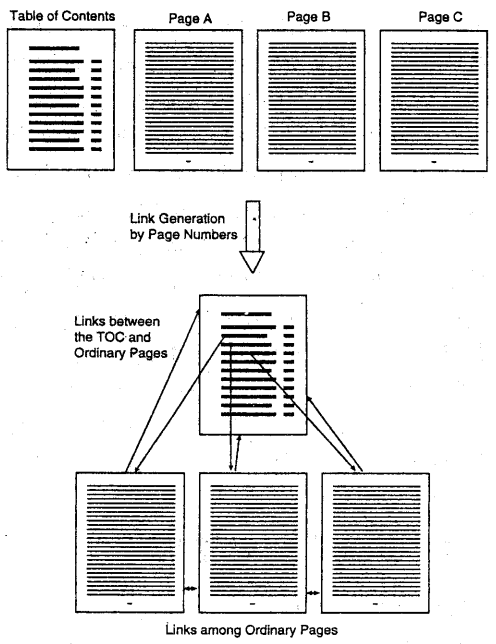


図６ リンク結果の模式図

5 結び

われわれは視覚障害者のための文書読み上げシステムを作成した。合成音声を使うことによるアクセス速度の遅さに対処するために、スキャンされた文書から論理構造を抽出して利用した。システムの効率は論理構造の抽出精度に左右されるので、より精度の高い抽出手法の開発が必要である。文書管理については目次ページを中心としたページ相互のリンクを作成することで、ユーザが効率的にアクセスできる組織的な文書とすることができた。今後はＯＭＲのユーザ評価を行い、実用レベルまで高めていく計画である。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] Y. Nakano, Y. Shima, H. Fujisawa, J. Higashino, and M. Fujinawa, "An Algorithm for the Skew Normalization of Document Images," in Proc. ICPR '90, pp. 8-11, 1990.

[2] S. C. Hinds, J. L. Fisher, and D. P. D'Amato, "A Document Skew Detection Method Using Run-Length Encoding and the Hough Transform," in Proc. ICPR '90, pp. 464-468, 1990.

[3] Kazuhide Sugawara, "Weighted Hough Transform on a Gridded Image Plane," in Proc. ICDAR '97, pp. 701-704, 1997.

[4] Jaekyu Ha, Robert M. Haralick, Ihsin T. Phillips, "Document Page Decomposition by the Bounding Box Projection Technique," in Proc. ICDAR '95, pp. 1119-1122, 1995.

[5]菅原一秀：" 視覚障害者用ＯＣＲシステムの設計"、情報処理学会第５７回全国大会講演論文集、1S－4, Oct, 1998