コンピュータビジョン　52—16
（1988. 1. 22）

# 文庫本点字訳のためのパソコンによる印刷文字認識

# Printed KANJI Recognition for Braille Translation of Novel Books using A Personal Computer

島田恭宏　Yasuhiro SHIMADA
塩野充　Mitsuru SHIONO

岡山理科大学　工学部　電子工学科
Okayama University of Science

あらまし　視覚障害者の文字情報獲得のための点字訳作業は，ボランティアによるものが主であり，おのずとその作業量には限界があり機械化が望まれる．英語文化圏の文字は英数字だけであり，その文字種はきわめて少ないのでＯＣＲシステムの構築は比較的容易である．アメリカにおいてはKurzweil読書器などが実際に使われている．しかし日本においては，文字種はＪＩＳ第２水準漢字まで含めると７０００種を越え，このようなシステムを構築した場合，非常に複雑なものとなる．強力な大型計算機によりシステムを構築する場合は，物理的制約が少ないため比較的問題点は少ない．しかし本研究では，ホームユース等のためにできるだけローコストなシステムの実現を目指すためにパソコンをベースとした構築を行った．その結果，文庫本小説を対象として９９．７％の認識率を得ることができた．

Abstract　The amount of character information acquisition of a blind person is extremely few as compared with that of a healthy eyed person. Because, the voluntary work is the mainstream of Braille translation and the amount of the work has the limitation. A book reading machine with OCR is already realized in U.S.A.. But Japanese character, especially KANJI, is very difficult to recognize with OCR. A big main frame computer will make it easy to recognize KANJI characters, but our purpose is to make a low cost system as possible. So that a personal computer is used for the host computer, and the recognition rate of 99.7% was obtained using a paperbacked book.

## 1 まえがき

視覚障害者の文字情報獲得は，聴覚，触覚等の残存感覚で代行させている．しかし，これらの感覚による文字情報の獲得量は晴眼者のそれに比べると極めて少ない．これは日々出版される書物の量が膨大であるにもかかわらず，これらを点字に翻訳したり，朗読し録音する作業はボランティアによるものが主流であり，その作業量にはおのずと限界があるからである．従ってこれらの作業の機械化は大きな意義を持つ．海外においてはoptakonやKurzweil読書器等[1][2]が実際に使用されているが，日本語の場合，漢字ＯＣＲの困難性（文字種が多く，複雑な文字を多く含む等）があり，現在積極的に研究されている課題である．筆者らはこのシステムの性格上，個々の視覚障害者の多種多様な読書要望にきめ細かく対処しうるためにも，大型計算機による集中型システムではなく家庭や小規模の図書館でも使用できるようなパソコンレベルのシステムの実現が必要と考えた．そこで第１段階としてイメージスキャナとパーソナルコンピュータ（以後スキャナ，パソコンと略す）を用いて文庫本点字訳のためのパソコンによる印刷文字認識の実験を行った．[3][4][5][6][7]入力データには文庫本を使用しているがその理由は，文庫本は種類が豊富で安価であること，小

説類の書物の中では文字が小さく印刷状態が悪く，これを対象にシステム構築を試みることで他の大型本の書物にも対応できるものと考えたからである．

## ２．システム構成

図１に本システムのブロック図を示す．まず文庫本をページ単位で入力し，この画像に対して傾き補正等の文書画像処理を行う．次に印刷文字認識を行い，その後，言語処理を施し最終的な出力として音声出力や点字出力を計画している．本稿では破線内の文字認識までの報告を行う．本システムのジェネラルフローチャートを図２に示す．文書画像処理部において画像の入力は，本実験では読み取り線密度の比較的高いスキャナを用いたためパソコンのグラフィック画面２枚で文庫本の１ページが表示できる大きさとなる．すなわち，スキャナの線密度は300dpi(dot per inch)でこれは１２本／mmの解像度である．一方，文庫本の１ページの大きさは縦約１２０mm，横約８０mmゆえ，１ページのドット数は１２０×１２＝１４４０，８０×１２＝９６０，すなわち縦１４４０ドット，横９６０ドットとなる．使用したパソコン（ＰＣ９８ＸＡ）のグラフィック画面は横１１２０ドット，縦７５０ドットゆえにパソコンのグラフィック画面１枚では文庫本の１ページの上半分または，下半分が表示できる．そこで全体の処理を１ページ単位で行うには，スキャナから送られてくる生のデータを，一時ファイルとして一旦ディスクに格納し，画像の傾き検出用に必要な画像は，１／４に縮小して画面に全体を表示する．パソコンには１枚のグラフィック画面をそのまま記憶できるＶＲＡＭ（ビデオＲＡＭ）が複数枚内蔵されているので，画像の傾き補正以後は，ディスクに格納してある画像データを読み出して，パソコンのＶＲＡＭ１枚に文庫本のページ上半分を，他のＶＲＡＭ１枚にページ下半分を記憶させ，原情報本来の大きさで処理を行う．処理単位を文庫本１ページ毎とするには，小説類は縦書きゆえページ上半分の１行の処理が終わるとＶＲＡＭのページ切り換えを行いその行の続きの処理を行う．なお，各行のつながりは，スクリーン座標の位置関係から判断してそのつながりを保障している．文字認識処理部において距離及び類似度の比較にソートを用いている

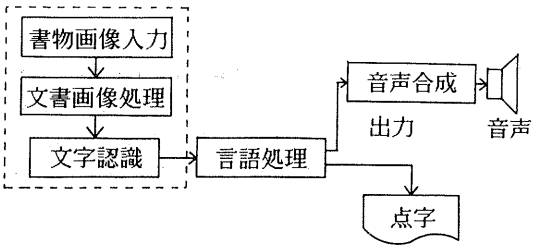


図１．本システムのブロック図

Fig.1 A block diagram of the proposed system.

が，これは実際のシステムとは違い，研究のため中間的なデータをとるため，１つのデータを比較しながら得るよりも処理も簡単で短時間で実行できると考えソートを用いている．なお，本システムの使用形態としては晴眼者が操作を行い，視覚障害者にサービスを行うことを想定している．

## ３．文書画像処理

文書画像を入力し認識処理を行うには，入力された画像の傾き補正，文字パターンの個々の切り出し，パターンの位置の正規化が必要である．これらについて以下に述べる．

### ３－１．傾き補正

画像の傾きはＬＰＰ(local projection profile)法[8]を用いて検出する．この方法は，傾き角度の検出可能範囲が狭く数度程度であるが，本実験で用いたスキャナは原稿固定式のため，注意深く入力すると大きな傾きを生じず，よってマクロ的な補正を行った後，ミクロ的な補正を行わず，ＬＰＰ法のみで原稿の傾きを求める．図３にＬＰＰ法の例を示す．まず第１に，画像を複数個の並行する帯状領域に分割し，それぞれの帯状領域内の文字並びの方向に沿って（今の場合，縦方向）周辺分布を求める．第２に，互いに隣接する帯状領域の周辺分布の相関値を計算し，位相のずれから傾きを検出する．位相のずれ$\alpha_k$を求めるには，式（１）を用いる．

$$\max_{-\beta \leq x \leq \beta} \{\sum_j P_k(j)\, P_{k+1}(j-x)\} = \sum_j P_k(j)\, P_{k+1}(j-\alpha_k) \quad (k = 1,2,3,4,5) \qquad (1)$$

ただし，$P_k$：第ｋ帯状領域の周辺分布値

$\alpha_k$：第ｋ帯状領域の位相のずれ

$\beta$ ：位相のずれを求めるための相関処理範囲

各帯状領域について求められた位相のずれ$\alpha_k$の平均値$\alpha_m$を算出し，式（２）で画像の傾き角度$\theta$を求める．

$$\theta=\tan^{-1}\left(\frac{\alpha_m}{S_h}\right) \qquad (2)$$

ただし，$S_h$：帯状領域の幅

この傾き角度$\theta$を用いて，アフィン変換の回転で補正座標を求め，変換座標での濃度値を求め補正を行う．なお，濃度補正は最近傍法を使う．

### ３－２．文字の切り出し

文書画像からの文字の切り出しは，文字列抽出と，各文字ごとの切り出しからなる．文字列抽出は，まず文書画像を垂直方向に投影し周辺分布を求め，基準となる文字列の平均ピッチを算出する．そして，各文字列に対する幅（文字列の周辺分布の幅）をこの平均ピッチと比較し，これと近いものはこれを文字列として切り出す．また，これよりかなり小さいものは，振り仮名の文字列と判断して切り出さない．この他，この平均ピッチよりもかなり大きい場合，振り仮名列と本文列の癒着と考えられので，その周辺分布の下の部分をある程度切捨てて周辺分布を各列毎に切り離して，先の２つの処理を繰り返して列の切り出しを行う（図４参照）．各文字の切り出しは，各文字列での周辺分布を幅方向にそれぞれ取り，基準となる線をその列の上部より最初に周辺分布が出現した点に引く．このとき，他の列のその位置よりも半文字分下にあれば，その距離を上に持ち上げて基準線を引く．個々の文字は，ほぼ正方形に近い矩形で囲むことができるという性質を利用し[3]，その文字列の幅で候補点をあげていく．この候補点は，周辺分布より判断し，周辺分布が無い位置，また，その周辺において，その値が最小の位置を仮切り出し位置としてあげる（図５参照）．一列の切り出しが終わった後，文字の平均切り出しピッチを計算し，この値を用いて先の切り出し法を再度行い，切り出し位置の決定を行う．全画面の切り出し終了後，切り出せなかった部分について，マウスを用いて手動で補正を行う．

### ３－３．文字パターンの正規化

本報告で用いた文字パターンの画面サイズは，

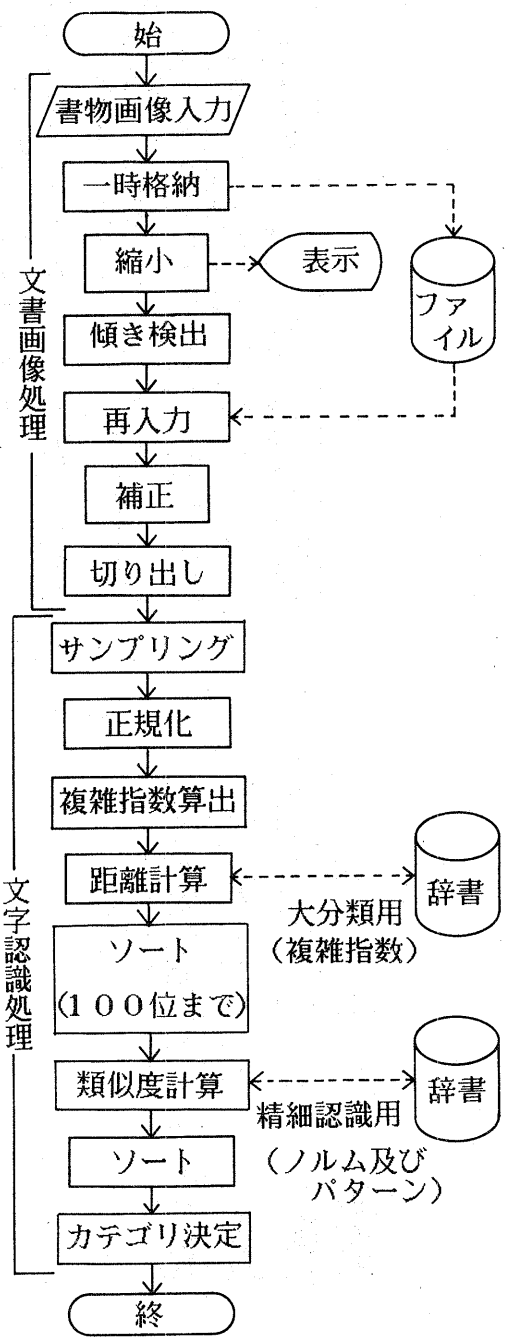


図２．本システムのゼネラルフローチャート

Fig.2 General flowchart of the proposed system.

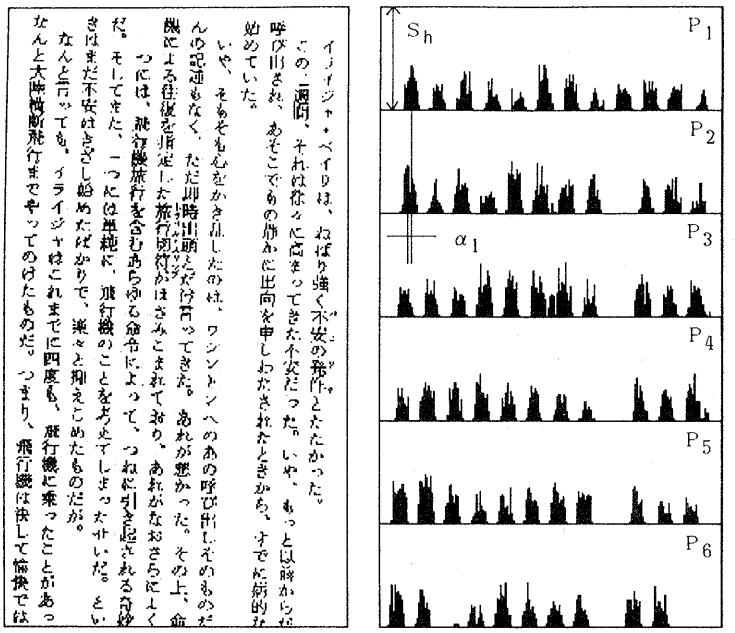


図３．紙面の部分的な周辺分布
上に生ずる位相のずれ

Fig.3 Phase shift on Line Projection Profiles.

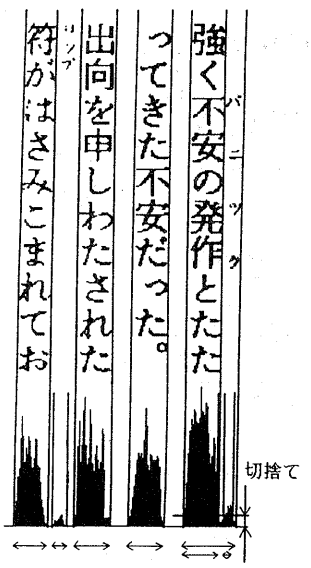


図４．文字列抽出

Fig.4 Extraction of text lines.

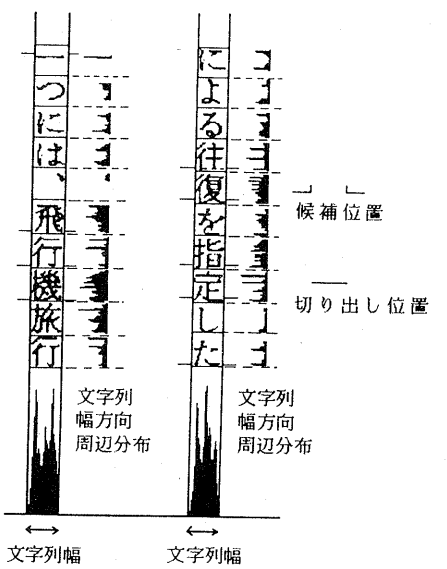


図５．各文字の切り出し

Fig.5 Segmentation of each character.

白枠の部分を含め４０×４０画素である．文字パターンの正規化は，文字パターンの重心を求め，それを画面の中心に平行移動することによって重心位置の正規化を行う．また，マッチングにより認識を行う場合，文字パターンを画面の枠一杯に拡大した方がよい結果が得られることが報告されている．しかし，文書画像中の文字を認識する場合「っ」，「ッ」，「。」等の小さい文字が含まれるため，本研究では，大きさの正規化は行わないこととした．

## ４．認識アルゴリズム

入力された文字を効率よく認識処理するために本稿では，大分類（分類処理）と精細認識（識別処理）で構成する．大分類に複雑指数を用い，入力された未知パターンと辞書パターンの複雑指数の距離（非類似度）を求め，１００位までを候補として分類する．精細認識には，単純類似度法を使用する．以下に各処理の詳細を述べる．

### ４－１．複雑指数(9)(10)

漢字パターンの図形としての複雑さを示す指標として複雑指数がある．この複雑指数は，式（３）により定義される．

$$C_x = \frac{\ell_y}{\sigma_x} \quad , \quad C_y = \frac{\ell_x}{\sigma_y} \qquad (3)$$

ここで$\ell_x$および$\ell_y$は，それぞれ横方向および縦方向の文字線の長さの和であり，次の方法で近似的に求める（$\sigma_x$，$\sigma_y$は後述）．図６でＮ及びＢはそれぞれ白地の内点，黒字の内点であり，Ｈ，Ｖ，Ｌ，Ｔは白と黒の境界点である．そしてＨ及びＶはこの点の近傍での原パターンの輪郭線の方向が水平または垂直の直線に近い場合，Ｌは斜線に近い場合，そしてＴは細い斜線の両縁の場合に得られる量子化パターンである．原パターンの輪郭線のどの部分もＨ，Ｖ，Ｌ，Ｔのいずれかに対応するので，輪郭線の長さはＨ，Ｖ，Ｌ，Ｔに適当な長さを割り当てることによって近似的に求めることができる．Ｈ，Ｖに単位１の長さを割り当てれば，Ｌ，Ｔに相当する長さはそれぞれ$1/\sqrt{2}$，$\sqrt{2}$と考えてよい．図６の太線は各パターンに割り当てられた長さを示している．従って原パターンの輪郭線の近似値は，パターン全面についてＨ，Ｖ，Ｌ，Ｔの和$\Sigma H$，$\Sigma V$，$\Sigma L$，$\Sigma T$を計数

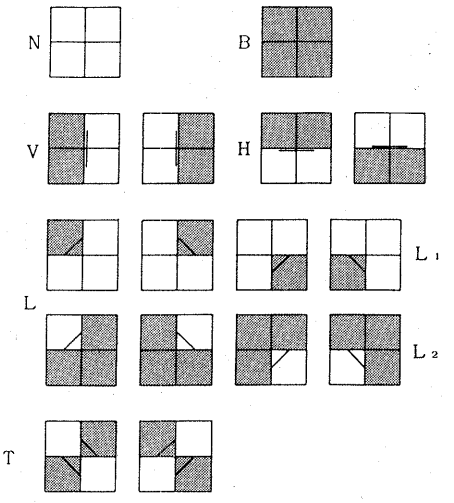


図６．線長及び平均線幅を計算するための2×2メッシュ・パターン

Fig.6 2×2 mesh patterns to calculate line length and mean line width.

しそれぞれに１，１，$1/\sqrt{2}$，$\sqrt{2}$の重みをかけて加え合わせることによって求めることができる．すなわち輪郭線の長さをλとすると，

$$\lambda = \Sigma H + \Sigma V + \frac{\Sigma L}{\sqrt{2}} + \Sigma T \qquad (4)$$

である．そして輪郭線の長さλは文字線分の長さの総和$\ell$の２倍に近似的に等しいので，

$$\ell = \frac{1}{2}\left( \Sigma H + \Sigma V + \frac{\Sigma L}{\sqrt{2}} + \Sigma T \right) \qquad (5)$$

となる．さらにＬに割り当てられた長さの横方向の成分と縦方向の成分を考えると，それぞれ１／２，１／２であり，Ｔの場合はそれぞれ１，１である．従って輪郭線の横方向成分$\ell_x$と縦方向成分$\ell_y$は式（６）で与えられる．

$$\left.\begin{aligned} \ell_x &= \frac{1}{2}(\Sigma H + \Sigma L/2 + \Sigma T) \\ \ell_y &= \frac{1}{2}(\Sigma V + \Sigma L/2 + \Sigma T) \end{aligned}\right\} \qquad (6)$$

次に$\sigma_x$および$\sigma_y$は漢字パターンの横方向，縦方向への拡がり量を示す．この２つの量は，漢字パターンの重心回りの２次モーメントから与えられ，式（７）により求めることができる．

$$\sigma_x = \sqrt{M_{20}^*} \quad , \quad \sigma_y = \sqrt{M_{02}^*} \qquad (7)$$

ただし，$M_{20}^*$，$M_{02}^*$ の星印は，$M_{20}$，$M_{02}$ を濃度について正規化したものである．$M_{20}^*$，$M_{02}^*$ は式（８）で与えられる．

$$\left.\begin{aligned} M_{20}^* &= \frac{1}{m_{00}} (m_{20} - m_{10}^2 / m_{00}) \\ M_{02}^* &= \frac{1}{m_{00}} (m_{02} - m_{01}^2 / m_{00}) \end{aligned}\right\} (8)$$

$\ell_x$，$\ell_y$ および $\sigma_x$，$\sigma_y$ は，いずれも横方向と縦方向に関して独立な分布をすること，位置と濃度に関する不変量であることが知られている．$M_{20}^*$，$M_{02}^*$ は［長さ×長さ］，$\ell$ は［長さ］であり漢字パターンの大きさに直接関係する量である．漢字パターンの大きさは，文字の本質的な情報ではないので，大分類のための量としては無次元量であることが望ましい．そこで，文字線分の長さの和である $\ell_x$（$\ell_y$）を文字の拡がりの平方根 $\sqrt{M_{20}^*}$（$\sqrt{M_{02}^*}$）で除すれば無次元の量を得ることができる．この複雑指数は，単位当りの拡がりに対してどれだけの線分が含まれているかという線分の密度を表している．大分類にはこれらの量を未知パターン，辞書パターンについて求め，式（９）の標準ユークリッド距離[11]を用い，距離の最小の候補から１００位までを取り分類処理を行う．

$$\left.\begin{aligned} D_{QP} &= \left(\sum_{i=1}^{p} (x_{Qi} - x_{Pi})^2 / S_i^2\right)^{1/2} \\ S_i^2 &= \sum_{Q=1}^{L} (x_{Qi} - \bar{x}_i)^2 / (L-1) \\ \bar{x}_i &= \sum_{Q=1}^{L} x_{Qi} / L \end{aligned}\right\} (9)$$

ここで，ｐは変数の個数（$C_x$，$C_y$ の２つ），Ｌは個体数（全カテゴリ数）である．

### ４－２．単純類似度法

入力された未知パターンＰと辞書パターンＱに対して類似度 $S^{\ell}(P)$（Ｐ）は，式（１０）により与えられる．

$$S^{\ell}(P) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} Q_{i,j}^{\ell} \cdot P_{i,j}}{\|Q^{\ell}\| \cdot \|P\|} \qquad (10)$$

ここで $\|Q^{\ell}\|$ は辞書パターン $Q_{i,j}^{\ell}$ のノルムであり，

$$\|Q^{\ell}\| = \left(\sum\sum (Q_{i,j}^{\ell})^2\right)^{\frac{1}{2}} \qquad (11)$$

となる．$\|P\|$ についても同様である．この単純類似度の式をパターン空間で考えると，ＱとＰのなす角度の余弦に相当しており２つのベクトルの向きの離れ具合いを表している．認識は類似度 $S^{\ell}(P)$ が最大となるカテゴリを未知パターンのカテゴリとすることで行われる．スキャナより入力した画像は２値画像ゆえ，サンプリングしたパターンは０，１で表現される．よって，類似度の計算は整数値の０，１ではなく，パターンの画素を横方向に見て，８画素を１バイト分のビットパターンに変換して計算する．これにより，計算の高速化，辞書の小容量化が可能となる．

## ５．辞書

認識に用いる辞書は，複雑指数，文字パターン，ノルム，パターン識別用ＪＩＳコードからなり，カテゴリ数は本来ならば，小説一編は異なる字種が約２４００種程度で構成されている[12]ので，この程度のカテゴリ数が必要ではあるが，今回はとりあえず１１６４カテゴリで構成している．辞書作成には１カテゴリに対し出現順に５枚までを４０×４０画素でサンプリングする．大分類辞書は，５枚のそれぞれのパターンの複雑指数を算出し，この平均を辞書値とした．精細認識に用いるパターン辞書は，この５枚を重ね合わせ，しきい値を１以上として２値化を行い，先に述べたようにビットパターンに変換して格納している．辞書作成に用いたデータは，認識手法として単純類似度法を用いているために，同じ字形を用いる方が良いと考え，入力データに使用するデータと同じものからサンプリングすることとした．

## ６．処理構成及びシミュレーション

装置として，パソコンはＰＣ－９８ＸＡ（ＣＰＵは８Ｍｚ相当の８０２８６（８０２８７付加））で，プログラム記述にはＣ言語を用いた．スキャナは主，副走査線密度ともに３００ｄｐｉのＰＩＳ－３０（Ｉ・Ｏ データ機器社製）を用いた．入力データは，文庫本（早川書店，ア

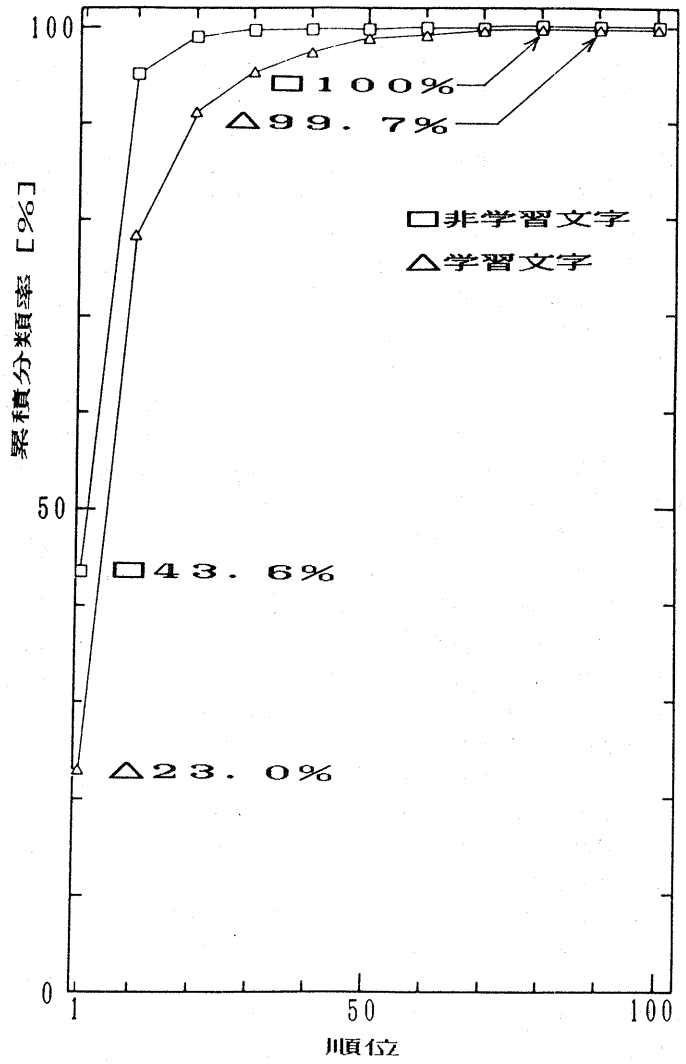


図7．累積分類率

Fig.7 Accumulative classification rate.

表1．認識率

| | 認識率 |
|---|---|
| 学習文字 | 97．7 % |
| 非学習文字 | 99．7 % |

表2．大分類及び精細認識における失敗文字

| 大分類 | 飛，　頑，　境 |
|---|---|
| 精細認識<br>（正→誤） | ベ（カタカナ）→べ（ひらがな）<br>ば→ば，び→び，き→ぎ，未→末<br>聞→閣，自→目，立→五，青→景<br>ペ→べ，ラ→フ，ぼ→ぼ，子→千<br>で→て，ぶ→ぶ，互→耳，宮→官<br>カ→ヵ（カタカナ），壁→璧 |

イザック・アシモフ著，冬川亘訳『はだかの太陽』９～１３ページ，総文字数３０８４文字，このうち学習文字１０５８字，非学習文字２０２６字）を用いた．実験結果として，図７に学習文字，非学習文字の累積分類率を，認識率を表１に示す．また大分類失敗文字と精細認識失敗文字を表２に示す．なお本報告では５ページ分の認識結果について示した．処理時間は，サンプリングからカテゴリ決定までの間を計測し，平均約５．７秒／字を得た．また，文書画像処理における各文字の切り出し成功率は，９９．３％を得た．

## ７．むすび

文庫本点字訳のためのパソコンによる印刷文字認識の実験を行い，非学習文字の認識率９９．７％，学習文字の認識率９７．７％，一文字当りの平均認識時間５．７秒を得た．今回の実験の結果において，学習文字よりも非学習文字の認識率が良い結果となっているのは，認識実験を行った文字数が少ないためと考える．複雑指数は文字パターンの雑音に直接影響される．今回の実験の結果では，複雑な文字に対して分類を失敗している．これは複雑な文字ほど雑音が大きく影響し，安定した文字パターンが得られず分類処理を失敗していると考える．単純類似度法による識別処理は，時間的な効率を上げることには成功した．しかし識別能力については不十分であり，ことに類似文字の識別に失敗している．これは入力パターンの雑音からくる文字の曖昧さが影響していると考えられ，ことに濁点と半濁点の識別は失敗率が高い．人間でも濁点か半濁点か不明瞭な文字が単体で提示された場合，識別は難しい．しかしこれが文章としてならば，その前後関係を利用して識別は容易である．ゆえに今後は文脈情報を利用して文字認識を進めていくことが必要と考える．処理時間を各処理についてみた場合，識別処理部分が最も時間を費やしている．従って，分類処理における候補数をより絞り込むことで識別処理にかかる負担を減らし処理時間を短縮することが考えられる．これには，距離計算において特徴量の付加，他の分類手法の併用を，また演算を行う場合，実際に計算を行うのではなく，計算結果を記した表を作成しておき，その表を用いるテーブル参照方式をとることで処理速度の高速化を図ることを考えている．今後，付加する機能としては，読み取り棄却機能があげられる．これらの課題をふまえながら言語処理部の研究も進める計画である．

最後に本実験に関し種々御助言いただいた本学小畑正貴講師，大学院生の大倉　充氏，実験に協力頂いた卒研生の大林一雄君，武中裕司君に感謝する．

［参考文献］

（１）海保ほか：”漢字を科学する”，有斐閣，1984.
（２）篠原正美：”視覚障害者用読書器について”，システムと制御，vol.29,no.1,pp.23-29,1985-01.
（３）島田，塩野：”パソコンによる文庫本の認識システム”，情報処理学会中国四国支部研究会，昭62-01.
（４）島田ほか：”パソコンによる文庫本認識システムについて”，昭62電気四学会中国支部連合大会，092107,1987-10.
（５）島田，塩野：”パソコンによる小説認識システムの一検討”，昭62電子情報通信学会情報システム部門全国大会，79,1987-11.
（６）島田ほか：”パソコンによる文庫本点字翻訳システムについて”，昭62電気関係学会関西支部連合大会，G8-35,1987-11.
（７）島田，塩野：”パソコンによる小説書物の認識システム”，第１８回画像工学コンファレンス，4-4,1987-12
（８）秋山，増田：”書式指定情報によらない紙面構成要素抽出法」，信学論(D)，vol.J66-D,no.1,pp111-118,1983-01.
（９）坂井，森：”２０００文字種を１００文字／秒で読む印刷漢字ＯＣＲの開発”，日経エレクトロニクス，no.172,102-128,1977.
（10）坂井，森：”漢字パターンの大分類”，信学技報，PRL73-17,1973-05.
（11）田中ほか：”パソコン統計解析ハンドブックⅡ多量解析編”,共立出版，1984.
（12）林ほか：”図説日本語”，角川書店，1982.

盛光印刷所