

制限付文字列読み取りの一検討

STUDY OF READING LIMITED CHARACTER LINES

佐瀬慎治・辻善丈・津雲淳

Shinji SASE, Yoshitake TSUJI, Jun TSUKUMO

日本電気(株)

NEC Corporation

あらまし 少数の単語よりなる自由に書かれた手書き文字列を読み取る方式を提案する。これらの文字列の中には文字間の接続などが含まれている場合もある。このような文字列を読み取るために文字切り出しと文字認識と単語照合を統合した文字列認識方式を考案した。この方式では、文字列より特定な文字を取り出すところにその特徴がある。本報告では文字列認識の構成とこの文字抽出の方法について述べる。実験により文字間が接続しているなど、従来技術では文字切り出しが困難な文字列に対しても読み取りが可能であることを確認した。

ABSTRACT In this paper, we propose a method to read character lines of a small number of words which were freely written and occasionally have some connections between characters. To read such lines, system of reading character lines is constructed by combining segmentation, character recognition and word recognition. In this structure a specific character image in a line should be able to be extracted. In addition to system structure, a method of this extraction is also shown as a character recognition. On the experiment applied to reading lines of hand-written Japanese addresses it is shown that lines which are thought to be difficult by usual segmentation method can be read.

1. はじめに

従来、多くの光学的文字符号読み取りは、読み取り対象から読みとるべき1文字の文字符号を切り出し、抽出された文字符号を文字符号カテゴリに判定する方法を採用していた。しかし、このような現状の方法では自由に書かれた手書き文書を正しく読みとることは困難であり、読み取り対象に制限があることが報告されている。^{1), 2)}特に、文字符号接続がある場合には、現状の文字符号切り出しの能力が対象制限の主因となっている。

また、1文字の文字符号に対する文字符号認識においても、認識対象を丁寧に書かれた文字符号に限定しているものが多い。³⁾光学的文字符号読み取り装置が走り書きのメモなどを読みとる人間の能力に近づくためには、この他にも多くの問題を解決しなければならない。

本報告では、認識対象として、郵便物の宛名のように、記載内容に制限はあるが文字符号接続・筆書き文字符号などを含む自由に書かれた文字符号

を用いる。このような文字符号の読みとりを試みることにより、従来では困難とされている認識対象に対する文字符号認識方式を提案し、実験評価を行なう。

本文字符号認識の基本的考えは、文字符号の先頭から順に文字符号を切り出さずに認識すること、および文法を十分かつ優先的に利用することにある。

この方法を用いて、実際の郵便物に記載された宛名に対する評価実験を行なったところ比較的良好な結果が得られた。

本報告では、2章で問題を定義し、3章でその解決方法について概説する。この方法のキーポイントは文字符号認識と全体のシステム構成にあるのでこれらについてそれぞれ4章と5章で述べる。6章では、郵便物の宛名読み取りに適用したシミュレーション結果を示す。

2. 解決すべき問題点



図2.1 変動が大きい。

5)出現文字種・出現単語種に大きな限定ができる。

上記特徴のうち1)-4)は、新たに解決すべき問題、5)は1)-4)を解決するために利用できる大きな情報となる。

文書読み取りは、情報伝達の立場から考えると、記載内容を読み出せばよい。出現単語およびその遷移の情報が厳格なこの場合、各単語を読みとることが主要なテーマとなる。

従来このような問題を1点1点限定された状況で解決しようとする種々の優れた試みがなされている。¹⁾⁻⁷⁾しかし、個々の問題点の設定の違いにより、これらの方法を組合せ的に使用してもこの問題を解決するには至らないと思われる。

3. 文字列読み取りの概要

2章で述べられた問題点を解決するために文字列認識方式の概要について述べる。

従来多く用いられてきている、文字切り出しと文字認識と単語照合をカスケードに並べた方法では、現状における文字切り出し性能と文字認識性能のため上記問題を解決することは相当に厳しいと思われる。その原因の1つとして、読まないと正しく切り出せない文字切り出しと、切り出さないと正しく読めない文字認識の間の

本報告では認識対象として郵便物の全面画像から切り出された文字列(文字行)画像を用いる。認識対象の1例を図2.1に示す。

解決すべき問題点を明示するためには、帳票読み取りなどの文字読み取り装置の認識対象と比較した、郵便物記載の宛名の特徴を以下に列記する。

1)文字記載のガイドラインがないため、文字間の接続があり、文字の大きさの変動が大きい。

2)認識対象が漢字であるため分離文字を含んでいる。

3)筆記具に限定がなく、文字線幅の変動および文字形の変動が大きい。

4)自由手書きであるため文字形の変動が大きい。

5)出現文字種・出現単語種に大きな限定ができる。

ジレンマがあげられる。この種の問題は、音声認識においても指摘されている問題である。⁸⁾

そこで、文字切り出し、文字認識、単語照合を統合することによりこの問題の解決をはかる。統合の流れとしてトップダウン的な流れを用いることにした。理由は、品質の低い文字画像からボトムアップ的に安定した特徴を得ることが困難なのに比較して、単語の情報は安定していてかつ文字間の遷移の情報として大きな情報であることと、人間がこのような文字列を読みとるときにはトップダウンの要素が大きく働いているのではないかと考えたことによる。人間がトップダウン的な読み取りを行なう状況としては、文書を高速に読む場合に読む前に次の内容を予測しているらしいこと、一文字では読めない画像でも文章中では比較的容易に読みとってしまうことなどがあげられる。

具体的には、以下のようアプローチを用いる。

文字列の読み取りは単語単位にすすめられる。文字列内に出現する単語はその遷移情報とともに単語群として準備する。

最初に、与えられた単語群の中から、単語間の遷移の情報をもとに、読みとるべき単語の候補を決定する。読み始めは、最も出現内容が限定される位置から始める方が望ましい。ここでは、文字列の先頭から読み始めることにする。

次に、読みとるべき単語の各候補に対して、文字列画像の先頭付近における類似度を求める。単語の類似度としては、単語を構成する各文字の参照パターンと文字列画像中の対応する画像との類似度の和とする。このために、単語を文字単位に分解する。

各単語の類似度を求めるために、決められた単語内の決められた文字を、文字列画像の先頭付近から読み出す。文字認識は、文字を読みとる前に文字列より正しく文字を切り出すのは困難であるとして、文字を切り出さないまま特定の文字が文字列の先頭にあるか否かを「検索」する形式とする。さらに、該当する文字があった場合、「抽出」という形式で文字列よりその文字に該当していると考えられる部分を切り取る。切り取られた文字の直後が、残りの文字列の先頭となる。類似度は、抽出時にマッチングの距離として得られる。従来文字認識で重視さ

れている、文字画像を多くの候補カテゴリの中の1つに特定する'識別'はここでは行なわない。

入力画像に対して、検索・抽出・切取りの操作を単語を構成する文字数回繰り返すことにより単語毎の類似度が得られる。

各単語の類似度を比較することにより、文字列の先頭の単語を判定する。この結果をもとに文字列画像より対応する単語画像を切り落とし、次に読みとるべき単語の候補を決定する。同様の処理を繰り返すことにより文字列認識が行なわれる。

上述の方法を実現するためには、文字列の先頭付近より特定の文字画像を抽出しその類似度（抽出信頼度）を求めることと、文字切り出し、文字認識および単語照合を統合したシステムを構成することがキーポイントである。前者を文字認識と称して4章で、後者を切り出し・文字認識・単語照合の統合法として5章で述べる。

4. 文字認識

文字列の先頭付近より文字を切り出すことなく特定の画像を抽出し、その類似度を求めることが可能にする文字認識の方法について述べる。従来、文字認識は正しく切り出された文字画像を対象としているのでこのような機能を実現することは困難である。

2章の考察と3章のアプローチの方法を考慮にいれると、文字認識には上記の抽出機能に加えて以下のような能力を持つことが必要となる。

- 文字サイズの変動に対応できること（文字単位のサイズの正規化は不可）

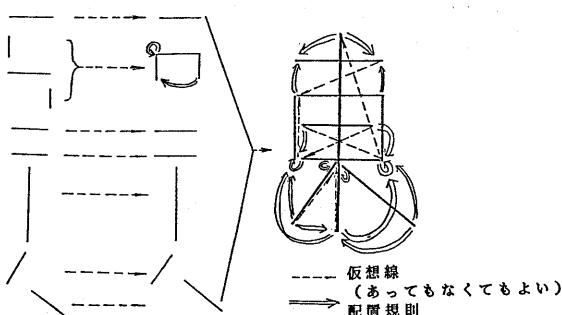


図4.1 文字モデルの例

- 続け字などにもある程度対応できること（自由手書きを含む）

- 文字の識別よりも抽出を重視すること（識別は単語により行なう）

このような機能実現のために、文字認識処理を画像抽出部と類似度演算部より構成する。

画像抽出部では、広く線幅や形状の変動を吸収する文字モデルを用いて、文字列の先頭よりもその文字と似ていると考えられる部分を検索し抽出する。類似度演算部は、モデルにより求められた抽出画像と抽出に用いられたモデルとの形状比較を背景を重視するマッチングにより行い、類似度を求める。

まず、画像抽出部について説明する。この方法は手書き数字認識に対して提案されたRubber String Matching（以下RSMと呼ぶ）に類似している。⁹⁾まず、画像抽出において最も重要な意味を持つ文字モデルについて述べる。

文字モデルは、文字を筆記順の線分系列で表現したものをもとにしている。線分系列を構成する線分には、文字のストローク部分を表現する実線と実線ではないが実線間の筆点の動きを表現する仮想線よりなっている。¹⁰⁾筆記順線分系列を実線分間の終点始点の相対位置関係をもとに簡単な図形よりなる構成単位にグループ分けする。線分系列よりなる各構成単位と構成単位間の相対位置関係で文字を階層的に記述したのが文字モデルである。

図4.1に文字モデルの例を示す。図の左側は文



図4.2 モデルによる続け字の表現

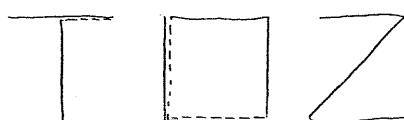


図4.3 構成単位の例

字'東'を構成する各線分を筆記順に上から並べたものである。この実線分の始点・終点の相対的な位置関係をもとに線分を構成単位にグループ分けしたのが中央に示された図である。中央の図でも筆記順の規則は保たれている。右側の図は、中央の構成単位を2次元上に配置し相対位置関係を曲線の矢印で筆記順の情報である仮想線を点線で付加し文字モデルを画像的に表現したものである。

ここで、最初に筆記順の線分系列を用いたのは、楷書と筆記順に書かれた続け字を1つのモデルにより表現するためであり（図4.2）、簡単な図形を構成単位としたのは、文字は（特に複雑な漢字は）線分を配置していると言うよりは、むしろ線分により構成される簡単な図形を配置するという階層性を有していると仮定したからである。図4.3に、簡単な図形の例を示す。図4.1-4.3において、実線は実際に描く線を表わし、点線は筆点の移動のみを意味する仮想線を表わしている。

本モデルにおける線分は、有向線分としての方向（8方向、図4.4）と、長さの変動許容値（上限・下限の2種）、他の線分との始点間の相対位置関係（上下左右の4種）、他の線分との中点間の相対位置関係（同じく4種）の11のパラメータにより表現される。相対位置関係のパラメータは参照する線分番号として各方向に代表的なものが1つだけ定義される。例えばその線分より上にある他の線分の線分番号を相対位置関係の'上'の部分に記述する。実線と仮想線の区別は長さの変動許容値の上限側の符号を用いている。簡単な図形の表現は、文字全体の線分系列に句切り目を与えることによる。

文字モデルをF、その構成要素を e_i ($i=1, \dots, n$)、構成要素における線分を s_j ($j=1, \dots$)

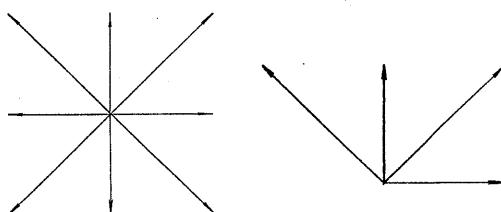


図4.4 モデルの8方向

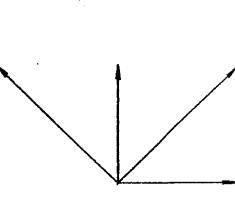


図4.5 画像の4方向

, m_i) で定義する。このとき

$$F = (e_1, e_2, \dots, e_n)$$

$$e_i = (s_{i1}, s_{i2}, \dots, s_{im})$$

で表わすことにする。この時、各線分 s の長さを定義された範囲内で様々に変化させると多様な文字形・文字サイズの画像を生成することができる。

文字の抽出は、上記文字モデルの各構成要素を文字列画像の先頭付近より求めることにより始まる。この部分が前述のRSMに類似している。詳しい説明は参考文献にゆずるとして、RSMと異なる点のうち主要なものは3点ある。

第1に、RSMでは多値画像に対して処理を行なっていたのに対して、ここでは4方向の方向画像（図4.5）に対して処理を行なっている点があげられる。これは、処理速度の向上と線分対応付けの安定を目的としている。またRSMでは各文字線の長さの変動許容値をもとに形状変動を制御していたが、ここでは文字サイズの変動に対応するため文字線の長さの変動許容を相当に緩め、線分間の相対位置関係により予めウインドウをかけることにより形状変動を制御しているというのが2点目である。実際上、複雑な形状に対しては相対位置関係を用いた方が形状は保たれる。オンラインの実験でも相対位置関係により漢字が十分表現できることが報告されている。¹¹⁾ところで、この相対位置関係は構成要素抽出の段階においては、各構成要素を構成する線分間の相対位置関係のみが参照され、残りの相対位置関係は、後の段階で用いられる。また、構成要素抽出においては、1つだけが抽出されるのではなく複数の図形が文字の構成要素の候補として残される。さらに、3点目として、RSMでは直線対応しか許していなかったのに

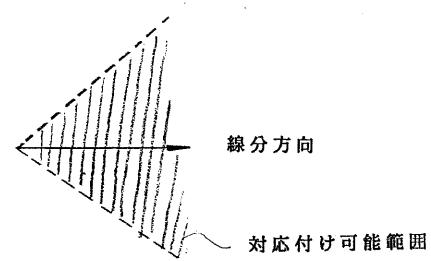


図4.6 線分対応範囲

対して、ここでは所定の方向を中心に 90° の範囲内(図4.6)での対応を認めている。これは、人間が1本と感じる直線はモデル上でも1本で表現することにより文字モデルの作成を容易にするためと、方向変動の激しい自由手書きに対応するためである。

構成要素 e を画像とし、文字列画像 I の座標を (x,y) で表わす。この時、1つの i 番目の構成要素より生成される様々な画像($j=1, \dots$)をモデルの終点の位置を用いて $e_{i,j}(x,y)$ で表わし、モデルの生成画像と文字列画像間の類似度を s とすれば構成要素抽出 $e_i(x,y)$ は

$$e_i(x,y) = \operatorname{argmax}_j \{ s(e_{i,j}(x,y), I) \}$$

for all x, y, i

となる。ここで、類似度 s としては、RSMと同等の類似度を用いている。こうして、文字列画像上の各座標に対して各構成要素の抽出候補を求める。ただし、上式の値がある閾値より小さいときは抽出できなかったものとする。

これらの各構成要素の抽出候補のあらゆる組合せから相対位置関係をチェックすることにより、すべての相対位置関係を満足するものが、

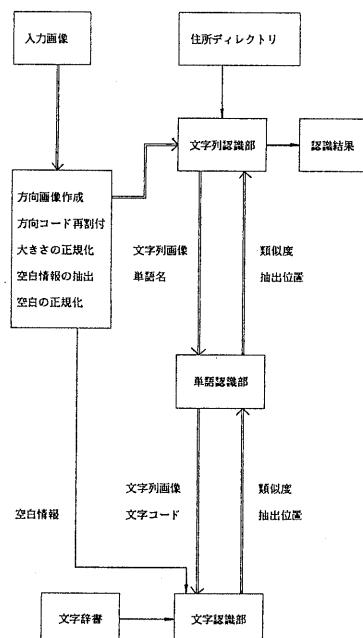


図5.1 文字列読み取りのシステム構成

文字抽出候補として残される。文字抽出候補が複数個ある場合は各図形の一致度の和が最も大きいものを抽出图形として選択する。

すべての構成要素間の相対位置関係の条件を \mathcal{C} 、構成要素の抽出候補の組合せを $p_k = \{ e_1, e_2, \dots, e_n \}$ とすれば最終結果は

$$\operatorname{argmax}_k \sum_{e_i \in p_k} e_i(x, y)$$

for all k by which p_k satisfies \mathcal{C}

で与えられる。

類似度演算部では、抽出した文字图形を太めて線幅の変動による影響を小さくし、仮想線部をも実線部と同等に扱うことにより掛け字を意識する。こうして作成された参照パターンと抽出された領域の画像とで背景のテンプレートマッチングを行ない類似度を求める。⁸⁾

このようにして、文字列画像の先頭付近より、特定の文字を抽出してその類似度を求めることが可能となる。

5. 文字切り出し・文字認識・単語認識の統合

本章では、上記文字認識を文字認識部に持ち、文字切り出しおよび単語照合と統合した文字列認識について述べる。文字列認識の構成図は図5.1のごとくなる。

図をもとに、処理の流れを説明する。入力画

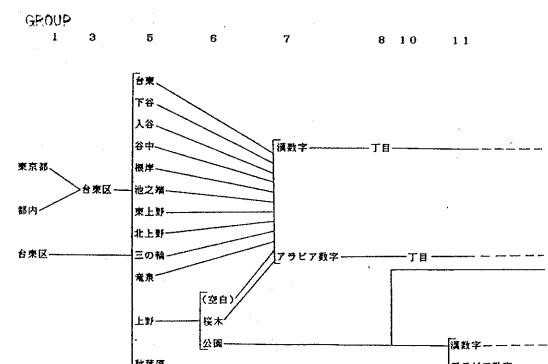


図5.2 単語テーブルの構成図

像は2値の文字列画像である。（図2.1）

この入力画像に対して、読み取り方向の決定
・2度の方向コード割付・文字列単位のサイズの正規化・空白情報の抽出・空白の大きさの正規化などの各処理が前処理としてなされる。2度の方向コード割付とは、文字認識の精度をあげるために最初固定方向に対して線分の方向を求め、次に右上がりの文字などに対応するために先に求めた方向コードの分布などをもとに方向コードを再割付する事を指している。

これらの前処理が施された文字列画像が文字列認識部に送られると、文字列認識部では、読みとるべき単語を選択し、単語名を付与して文字列画像の先頭部分を単語認識部に送る。

文字列認識部における読みとるべき単語の選択は、図5.2に示されるようなテーブルを用いている。このテーブルにおいて、出現すべき各単語はその単語が属するグループ、この単語が判定された場合次に読むべき単語のグループ、単語を構成する文字数およびその文字コードの各パラメータからなっている。最初はグループ1から出発し、判定結果にもとづいて次々と読むべき単語のグループが選択されていく。このグル



図5.3 処理における文字列の遷移

ープの遷移と共に、認識対象となる文字列画像も短くなって行き、最終的に先頭から順に文字列がすべて読みとられることになる。

単語認識部においては、文字列認識部より受け取った単語名を文字に分解し、単語の第1文字と文字列画像（図5.3(a)）を文字認識部に送る。

文字認識部では、与えられた文字コードに対応した文字モデルを用いて、文字列画像の先頭部分から該当すると思われる文字イメージを抽出する。その結果、抽出位置と類似度を単語認識部に返す。ただし、該当する文字が抽出できなかった場合にはリジェクトコードを返す。

抽出位置は、抽出したモデルの各線分の位置が記憶されているので、これをもとに抽出図形の外接枠を求め、外接枠の対角の2頂点で表現している。

単語認識部は、文字認識部より受け取った認識結果をもとに抽出した文字を切り取る。切り取られた後の残りの画像（図5.3(b)）と単語の第2文字目の文字コードを再度文字認識部に送る。このような処理を単語の文字数回繰り返すことにより、単語の類似度と切り出し位置が得られる。単語の類似度・切り出し位置の表現は文字のそれと同じである。

文字列認識部は、同一グループ内のすべての単語に対して抽出位置と類似度（あるいはリジェクト）を求める。この結果、最も類似度の高い単語を文字列の先頭にある単語と判定する。判定結果に対応する抽出位置部分を文字列より切り取り、次に読みとるべき単語グループを選択して、文字列の続きを同様の処理により読み取る。

このような処理を文字列画像がなくなるまで繰り返すことにより文字列認識がなされる。

東京都、都内
台東区
三ノ輪、下谷、慈泉
根岸、入谷、台東
東上野、北上野、上野
谷中、池之端、秋葉原
上野桜木、上野公園

図6.1 読み取り対象文字

6. 評価実験

4章、5章で述べた方法を用いて、郵便物の宛名画像に対して評価実験を行なった。対象は特定の郵便番号に対応するある地域の住所とした。この場合、出現単語は図6.1に示す17種類である。出現する文字数は図6.1の文字に、漢数字、アラビア数字、記号を加えた52字種である。

住所ディレクトリは図5.2のごとく作成した。文字辞書は1カテゴリ当り1モデルで、電総研のデータベースETL6およびETL8を参考に作成した。

実験データは実際に配達された郵便物をもとに準備した。これらの郵便物より住所行のみのイメージを54通分収集した。

実験結果は以下のようであった。

町名まで	41/54(75.9%)
丁目まで	25/47(53.2%)
番まで	12/48(26.1%)
号まで	8/42(19.0%)

さらに町名までの誤読は0であった。

読みとることのできた例を図6.2に示す。図に

おいて左が認識に用いられた実際の画像で正規化済みの2値画像である。中央の細線が抽出图形でありその右が判定結果である。各実験結果に對して左端から若干の説明を加える。

- (a)は分離文字の多い漢数字の部分の文字切り出しが正しくできていることを示している。
- (b)は筆によるつぶれ文字の例である。若干抽出を失敗している。(c)は文字の大きさの変動をある程度許容している。特に、野と二の大きさには9:1-16:1程度の大きさの違いがある。(d)は、草書の例であるが、このような文字も正しく判定している。しかし抽出图形を見てみると、この当りが文字変形吸収の限界であるともいえる。
- (e)は、漢字の部分が接続している例である。

本方法の効果は、文字を読みとった後に文字を切り出す操作と判定を単語単位で行なうという、システム構成にあると考えている。そして、このシステム実現の上で最も大きく貢献したのが文字認識である。

文字認識の性能は線分より構成される基本構成图形の抽出とその相対位置関係の評価による

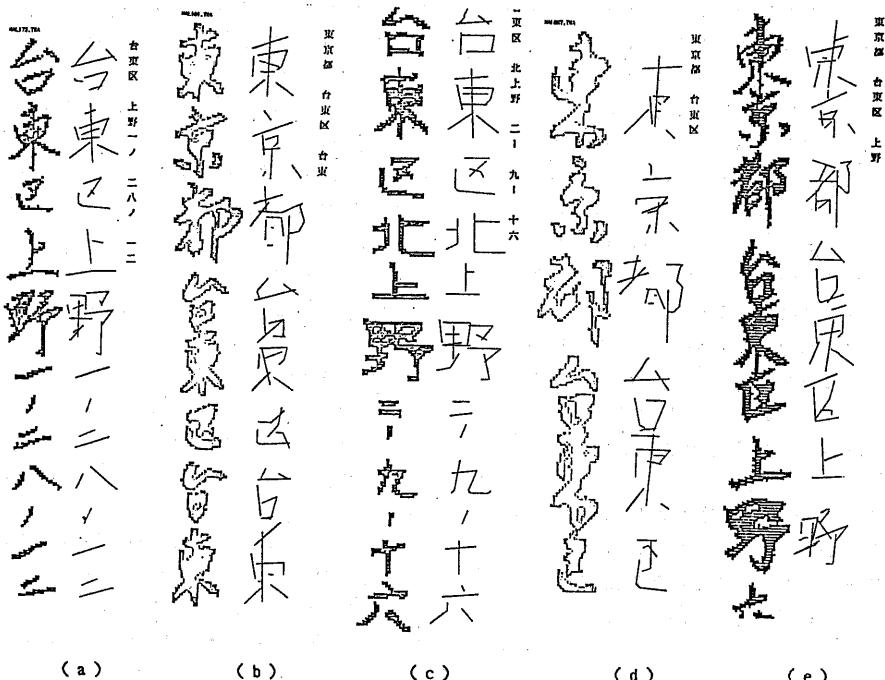
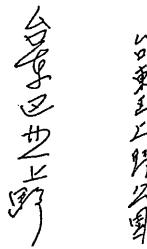


図6.2 読み取り結果の例



三 ↔ 五
六 ↔ 八

図7.1 読取不能(1) 図7.2 読取不能(2)

文字画像抽出にあると考えている。

7.まとめ

文字が接続していたり、草書などの筆文字が含まれる文字列の読み取りを試みた。ここでは、簡単のために記載内容に大きな限定がある郵便物の住所を対象に開発・評価を行なった。

今回の文字列読み取り方式のキーポイントは、文字切り出し・文字認識・単語照合のトップダウン的な統合とこれを可能にした文字列から文字を抽出する形態の文字認識にあった。

評価実験の結果、本方式の効果が確かめられた。特に、文字間に接続がある場合、文字列内で文字のサイズが著しく異なる場合、漢数字の並びで文字の境目を特定するのが困難な場合などでも、本方式では読み取る能力を備えていることが示された。

今後、文字読み取りの高精度化・前処理の修正・文字辞書の自動作成そして限定単語から通常の文書読み取りへの拡張を考えて行きたい。

図7.3 対して読みとることができないものがいくつかある。これらを示すことにより今後の課題としたい。

図7.1は文字モデルが文字変形に対応できなかった文字の例である。図7.2は、識別できない漢数字の例である。図7.3は文字列の正規化により画像の下部が潰れてしまう例である。

(謝辞)

本研究は筆者がC & C情報研究所・パタン認識研究部においてすすめてきたものである。

研究の機会と協力を戴いた郵政省郵務局システム企画室殿、いろいろ御指導を戴いたパタン認識研究部の浅井前部長、天満部長および産業オートメーション事業部・認識応用システム技術部の久保田部長、石川課長、西嶋主任に感謝を表したい。また有益な助言・討論および評価プログラムの作成の援助を戴いたパタン認識研究部の方々およびデータ収集をして戴いた認識応用システム技術部の方々にも謝意を表したい。

(参考文献)

- 1)馬場口、塚本、相原、"認識処理の導入による手書き文字切り出しの一改良"、信学論、Vol.J69-D, No.11, pp1774-1782 (1986)
- 2)依田、松浦、前田、南部、"手書き日本語文書からの文字切り出し方式"、信学技報PRU86-28, pp63-72(1986)
- 3)津雲、浅井、"文字認識技術の最近の動向"、信学技報IE88-5, pp31-38(1988)
- 4)Kovalevski, "Image Pattern Recognition", Springer-Verlag, New York(1980)(Tr. from "Metody Optimal'nyh Resenii v Raspoznavani i Izobrazhenii", Nauka, Moscow(1977))
- 5)Casy and Nagy, "Recursive segmentation and classification of composite character patterns", Proc. of 6th ICPR, pp1023-1026(1982)
- 6)藤沢、道野、"接触した手書き文字の自動分離を行なう文字切り出し方式"、昭59信学総全大、1588(1984)
- 7)福島、"視覚パターン認識における選択的注意機構の神経回路モデル"、信学技報MBE85-107, pp225-234(1985)
- 8)Sakoe, "Two-Level DP-Matching: A Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-27, No.6, pp588-595(1979)
- 9)迫江、"Rubber String Matching法による手書き文字認識"、信学資PRL74-20 (1974)
- 10)佐瀬、辻、"文字切り出し能力を伴う文字認識の一手法"、昭62信学総全大、1506(1987)
- 11)石井、"ストローク代表点による漢字の構造分析について"、信学論、Vol.J65-D, No.7, pp558-568(1986)