

博士学位論文

統合的手法を用いた  
手書き文字切り出しと認識の研究

2000年3月

西村 広光



# 要 旨

現在までに様々なオフライン手書き文字認識の研究が行われ、実用に供されている。しかし既存の手法は、接触した手書き漢字は認識できない。また一方では、相補性を有する複数の認識手法を組み合わせることが有効なことが知られたため、新しい認識手法が望まれている。本論文では、より実用的かつ高い性能を持つ文字認識系を実現するため、統合的手法を用いた手書き文字の切り出し法・認識法を提案した。

接触文字を含む単語を認識する場合には、単語から一文字を切り出して認識する手法、単語全体を一つのパターンとして認識する手法がある。後者は、認識する単語の語彙数が多い場合、対象パターンと照合されるパターン数が膨大化するため、一般的な認識手法として適当ではない。前者は英数字のように比較的各文字の構造が単純なものに対しては成果が報告されているが、手書き漢字のような複雑なパターンに対する切り出し手法は知られていなかった。また、接触文字切断の場合、最終的には文字認識・単語認識と協調して確定すべきであるが、文字切り出し単独での能力も評価することが望ましい。本論文では、接触を含む漢字パターンを対象とした文字切り出し手法を提案し、文字切り出し能力を単独で評価する手法を新たに開発して評価した。

提案手法を実現し、文字接触が含まれるパターンを対象に性能評価実験を行い、人間が適当と考える切断箇所での切断線を検出できたかどうかで評価した。検証の結果、接触パターンの64%で適当な文字の切り出しに成功した。また、適当な箇所での切断線の検出はできなかったが、検出した切断線により分割したパターンが認識可能と判断できるものも成功と評価した場合、接触パターンの約75%で文字の切り出しに成功した。

次に新しい文字認識手法の開発を行った。現在までに様々なオフライン文字認識の手法が研究されているが、一つの手法では完全でなく複数の手法を組み合わせることが有効であることが知られ、既存手法と相補的な特性を有する手法の開発が望まれている。最近 Hidden Markov Model (HMM) を利用した認識系は、HMM の表現能力が高いこと、音声認識等で高い性能が実証されていること等から、より高い性能を発揮する認識系の可能性が期待されている。現在までに、音声認識等で性能が確認されている1次元 HMM を利用した手法、2次元情報を利用できるように改良した2次元 HMM を利用した手法が報告されているが、前者は認識率が不十分であること、後者はモデルのパラメータ数が膨大化すること等の問題が生じている。モデルのパラメータ数膨大化は、学習時に必要とするサンプル数の増大、学習時の計算時間の長大化という問題を生じるため、大量のデータを収集することが実際上困難であるオフライン文字認識においては、実用上大きな問題となる。

ここまであげた理由により、既存の1次元・2次元の HMM を利用した文字認識手法は、ともに実用には適していない。そこで、本論文では HMM を利用した認識系をより効果的に実装する手法を提案した。一つは、多方向から特徴抽出を行い、パラメータ数の少ない1次元 HMM 複数を組み合わせ高い認識性能を実現する手法、もう一つは Bagging 法を HMM 認識系に適用し、より高い認識性能を実現する手法である。

前者を実現し認識結果を検証したところ、2次元の HMM よりもはるかに少ないパラメータで同等の認識性能を発揮する認識系の構築に成功した。後者においては、厳密な Bagging 法の適用以外に、HMM の特性を反映した Voting 法を提案し、実験により提案手法による認識性能の向上を実証した。

以上から本論文で提案した手法は、オフライン文字認識において、従来の手法と比較して実用レベルで高い認識性能を実現可能にするものと考えられる。

# 目 次

第1章	序論.....	4
1.1	オフライン文字認識の課題.....	4
1.2	本研究の方針.....	6
1.3	論文の構成.....	7
第2章	手書き文字切り出し.....	9
2.1	本研究の目的.....	9
2.2	接触文字切り出しによる認識可能性.....	9
2.3	従来の研究.....	10
2.3.1	Oversegmentation.....	12
2.3.2	接触がないときの日本語住所の文字切り出し.....	14
2.4	本研究の方針.....	16
2.5	統合的手法による手書き文字切り出し法.....	17
2.5.1	線幅の推定と正規化.....	18
2.5.2	直線成分の検出.....	20
2.5.3	簡略化パターンの生成.....	22
2.5.4	パターン分割候補領域の推定.....	24
2.5.5	切断線の確定.....	26
2.5.6	検出切断線の絞り込み.....	28
2.5.7	文字の切り出しの評価法.....	29

第 3 章	Hidden Markov Model を利用した文字認識.....	3 0
3.1	従来の研究.....	3 0
3.2	1 次元 HMM の解説.....	3 1
3.2.1	Hidden Markov Model の解説.....	3 1
3.2.2	1 次元 HMM の定義.....	3 2
3.2.3	1 次元 HMM の確率計算と最適パスの算出.....	3 4
3.2.4	1 次元 HMM のパラメータ推定.....	3 7
3.3	1 次元 HMM を利用したオフライン文字認識.....	4 0
3.3.1	オフライン文字認識への適用.....	4 0
3.3.2	1 次元 HMM を利用した文字認識の問題点.....	4 3
3.4	2 次元 HMM を利用した文字認識.....	4 4
3.4.1	MRF-HMM の解説.....	4 4
3.4.2	MRF の解説.....	4 5
3.4.3	MRF-HMM の定義.....	4 7
3.4.4	MRF-HMM の確率計算と最適パスの算出.....	4 9
3.4.5	MRF-HMM のパラメータ推定.....	5 2
3.4.6	2 次元 HMM を利用した文字認識の問題点.....	5 3
3.5	本研究の方針.....	5 4
3.6	1 次元 HMM の特徴の改良.....	5 5
3.7	多方向特徴抽出 HMM.....	5 7
3.8	Bagging 法を利用した Voting 法の適用.....	6 0
3.8.1	Bagging 法の解説.....	6 0
3.8.2	Bagging 法の HMM への適用.....	6 1
3.8.3	HMM の特性にあわせた Voting 法の拡張.....	6 2



第 4 章	評価実験.....	6 5
4.1	文字切り出し実験.....	6 6
4.1.1	文字切り出し実験対象.....	6 6
4.1.2	文字切り出し結果の評価システム.....	6 7
4.1.3	文字切り出し実験結果.....	6 8
4.1.4	文字切り出し手法の評価.....	7 0
4.2	HMM を利用した文字認識実験.....	7 1
4.2.1	各手法の学習効率の比較.....	7 1
4.2.2	文字認識実験対象.....	7 2
4.2.3	1次元 HMM の特徴改良の認識実験結果.....	7 3
4.2.4	多方向特徴抽出 HMM の認識実験結果.....	7 4
4.2.5	Bagging 法と応用した Voting 法の適用実験結果.....	7 5
4.2.6	提案認識手法の評価.....	7 6
第 5 章	結論.....	7 8
	謝辞.....	8 0
	参考文献.....	8 1
	研究業績.....	8 4

# 第1章 序章

## 1. 1 オフライン文字認識の課題

文字認識は、パターン認識の中で最初に実用化された対象であり、産業的にも大きな成功を収めた[27]。このように歴史的には1950年代に初めて実用化され、それ以後着実に進歩しているにもかかわらず、手書き文字認識においては未だに人間の認識能力には遠く及ばない。本研究は機械の手書き文字認識能力を高めようとする努力の一つである。

手書き文字認識は、文字筆記時に認識を行い、筆記情報（ペンの動きの情報）を認識で利用できるオンライン文字認識と、紙に筆記されているパターンのみを利用するオフライン文字認識とに大別される。

前者のオンライン文字認識の研究は未だに完成されたとは言えず、現在もさらなる精度向上にむけた研究はされているものの、ペン入力の手帳やワードプロセッサ等さまざまな用途での利用が始まっている。後者のオフライン文字認識の研究はオンライン文字認識よりも早く実用レベルに到達し、広く実用化されているが、未だ問題点は多い。本論文で扱うのは、後者のオフライン文字認識である。

現在までにオフライン手書き文字認識で実用に至っているのは、各種帳票の認識（手書き OCR）、郵便番号の読み取りシステム、英字の住所読み取りシステム、日本語住所読み取りシステムなどである。しかし既存の手法は、続け書きされた文字や、変形の大きな文字などに対しての認識性能は十分とはいえないため、筆記箇所の制限や筆記する単語の限定などの拘束を課して実用化されることが多い。

OCR（光学式文字読取装置）は1950年代に実用化されたが、当初はラインプリンタ出力の数字を読むものであった。1960年代末期に手書き数字を読む OCR が出現したが、切り出し認識性能が低いため、赤インクで印刷した



記入枠（赤枠）の中に丁寧に書いた文字に限定されていた[27]。その後認識対象は手書き漢字にまで拡張されたが、丁寧に書かれた文字でないと認識率は低く、数字以外の接触文字はほとんど認識できていない。

同じく1960年代末期に実用化された郵便番号の読み取りシステムでは、読み取る文字は数字10クラスに限定され、筆記箇所の制限により続け書きが防止される。続け書きされている場合でも記入欄の位置情報から一文字に切り出すことは容易である。さらに、リジェクトされた書状は人間が処理できるという逃げ道があった。

米欧で実用化されている（あるいは実用化に近い）英字住所の読み取りシステムでは、筆記されている文字は英数字記号を含めても高々100クラスで、文字クラス数はあまり多くないが、筆記体なので文字切り出しが難しい。しかし、筆記されている単語が住所に限定されているため、文字認識性能が不十分であっても知識処理によって結果を改善することができる。さらに地名とZIPコード（郵便番号）の照合なども行われる。

最近日本で実用化された手書き住所読み取り装置でも、個々の文字が接触しておらず字形の崩れが大きい場合は、ある程度の文字認識性能が得られているので、英数字と同様に知識処理で改善することができる。

これらの応用に限定されず、オフライン文字認識を様々な分野で実用化するためには、オフライン文字認識の認識率向上はもちろん、続け書きされた文字、あるいは接触した文字をより高い精度で認識すること、認識する文字・単語の種類が多い場合でも高い認識性能を発揮すること等が求められている。すなわち、文字切り出しおよび文字認識自体の性能向上は、文字認識普及に不可欠な要素といえる。

現在までに有効であると報告されている認識手法においては、これ以上の性能向上が困難なものが多い。そのような状況下で、最近注目されているのが、複数の手法を組み合わせることで認識率を高める方法である[1]。そのためには、新しい認識手法の提案が重要であり、Hidden Markov Model（隠れマルコフモデル、以下HMMと記す）やニューラルネットワークが有望と考える研究者も多い。

しかし、ニューラルネットワークを利用した文字認識系や HMM を利用した文字認識系等、学習によって構築したモデルを利用する文字認識系では、高い認識性能を実現するため、膨大な数の学習データを必要とする。しかし、大量のデータを収集することは困難である。そのためモデルの学習効率を高め、より実用的な利用方法を考案する必要がある。

ここまでにあげた課題の他にも、変形の大きいパターンに対して十分な認識性能が得られていないこと、日本語以外の中国簡体字、繁体字、ハングルなどのクラス数が膨大な文字認識で十分な成果が得られていないことなど、オフライン文字認識を様々な分野でさらに広く活用するために残されている課題は多い。

このような技術課題の中で、手書き日本語認識における接触文字の自動分離と、新しい認識手法の提案は、文字認識における大きな技術課題であると言えるよう。

## 1. 2 本研究の方針

本論文では 1.1 節であげたオフライン文字認識のさらなる実用化に向けた課題のうち、「接触文字の認識性能の問題」と「学習によってモデルを構築する HMM を利用した文字認識系の学習効率と認識性能の問題」を取り上げ、その解決方法を提案し、検証を行った。

接触文字、あるいは連続筆記体単語を認識する手法は、大きく分けて二つの手法が今までに提案されている。一つは、続け書きされた一つの単語を接触点と思われる複数の箇所を切り離し、各パターンを切り出した後に、文字認識を行い、文字認識結果が得られる（リジェクトされない）箇所が真の接触点として、文字の切り出しと認識を同時に行う手法である。この方法は接触した数字の認識[2]や英単語認識[3]に採用されており、oversegmentation として知られている。なお英単語の場合は認識部のみで切り出しの良否を判断するのは困難なため、単語辞書との照合を行って切断箇所を推定している。



もう一つは、単語を文字単位で認識するのではなく、単語全体のかたまりを一つのパターンとして認識を行う手法である。この方法は **holistic approach** と呼ばれる[4]。この方法は、認識する単語の種類が膨大な場合、対象パターンと照合されるパターン数が膨大となるため、一般的な文字認識システムとして実用化するのに適していない。

前者の手法は、漢字のように文字の構造が複雑なパターンの場合であっても、文字が接触していなければ適用できる。すなわち、複数の部分パターンを統合することにより多重の文字切り出し仮説を作り、認識、知識処理で確定すれば良い。しかし、接触文字が存在する場合は未解決であった。

以上の考察に基づいて、本研究の一つの目的を、郵便書状住所として書かれた文字のうち、接触した漢字パターンを切り出すこととした。最終的には文字認識、知識処理との協調が必要であるが、切り出し部のみの性能評価をするため、文字認識結果なしで文字単位で切り出す手法を提案し、検証を行った。

また、HMM を利用したオフライン文字認識の研究は、HMM の表現能力の高さ、音声認識・オンライン文字認識での高い性能等の理由から現在も様々な研究が進められ、高い認識性能の可能性が期待されている。そこで本論文では、学習によってモデルを構築する文字認識系としてHMMを利用した文字認識系をとりあげ、学習効率・認識性能向上のための手法を提案し、検証を行った。

### 1. 3 論文の構成

本論文は、1.2 節の方針に従い、オフライン手書き文字認識の高度化に向けて、接触日本語文字パターンからの文字切り出し手法の提案、ならびに HMM を利用した文字認識システムの学習効率・認識性能の向上を実現する手法の提案を行い、それぞれについて検証を行ったもので、5 章により構成される。

第2章では、漢字を対象として接触文字も含む単語パターンからの文字切り出しを実現する手法を提案する。

第3章では、HMM を利用した文字認識系の既存研究の紹介と、認識率向上

に向けた手法の提案、学習効率・認識率の向上を実現するための HMM 認識系構成方法の提案、学習効率・認識率向上のために HMM に適用した Bagging 法を基とした Voting 手法を提案する。

第 4 章では、第 2 章・第 3 章で提案した手法を検証するために行った幾つかの実験結果を示し、その結果を定量的評価とともに示し、考察する。

最後に第 5 章では論文のまとめと今後の課題について述べる。



## 第2章 手書き文字切り出し

### 2.1 本研究の目的

本章で述べる研究内容は、(株)日立製作所中央研究所との共同研究で進められたもので、同所で開発中であった郵便物宛名読取区分機に実装することを目的としている。

最終的な目標は、文字切り出し、認識、知識処理の統合であるが、このシステムは中央研究所で開発が進行中であったため、連日のようにシステムが変化しており、その中で作業するのは開発速度を鈍らせる恐れがあった。

そのため、単独で開発できる接触文字の切り出し（接触候補点の発見とパターン分離）を当大学で研究し、手法が完成した後実システムに組み込む方針を立てた。

### 2.2 接触文字切り出しによる認識可能性

多くの文字認識系は、一文字単位の認識のみを行うため、複数文字が接触したパターンはそのままでは認識系でほとんど認識できないが、接触パターンを接触箇所で切断し、切り出されたパターンを文字認識系に送れば、それまで認識不能であったパターンも認識可能になることもある。

しかし、適切な接触箇所での切断を行っても、切断された接触パターン全てが認識可能になるとは限らない。切断により切り出したパターンの変形が大きい場合に誤りが発生する可能性があるためである。

このように認識系の不完全性のため、文字切り出しの成功率と文字認識率の向上率は完全に一致することはない。しかし、正しい箇所での切り出しができれば、切り出されたパターンが正しく認識される可能性は高いと考えられ、認識率が向上する可能性は高いと予想される。

## 2. 3 従来の研究

接触を含む手書き文字パターンを認識するための文字切り出しの研究は、英数字に対してはいくつかの研究成果が報告されている[3][5][7][8]。それらは以下のような手法である。

[A] 周辺分布を利用して切断箇所を探索する手法

[B] 文字の大きさを推定し、その大きさを利用して切断箇所を探索する手法

[C] 輪郭線の変化から切断箇所を探索する手法

手法 [A] は、図 1の例のように、パターン全体を筆記方向の軸に黒画素投影した黒画素周辺分布を利用して、極小値から切断箇所を探索する手法であるが、古い技術であり、分離した文字であっても上下にオーバーラップしていた場合などで失敗するという問題点を有している。

手法 [B] は、行パターン全体から検出した情報を利用し、行パターン中の文字の大きさを特定し、推定された文字の大きさを利用して、切断箇所を探索する手法である。ここでは、行パターン内の各文字の大きさはほぼ均一であるという仮定に基づいている。これは、日本語印刷文字のように文字のピッチが一定の場合の切断箇所を特定する場合や、等幅で書かれたパターンには有効であるが、図 2の例のように縦書き漢字等では文字の大きさにばらつきが大きいため良好な結果は得られない。

パターンから検出した輪郭線情報を利用すれば、各部のストロークの幅を検出することができる。手法 [C] は、文字のストローク幅の変化から接触パターン中の接触箇所を探索する手法である。これは、英数字のように、一つの文字パターンが数ストロークであるもの同士の接触箇所探索には有効であるが、漢字パターンのように一文字内のストローク数が多く複雑なものでは良好な結果が得られない場合が多いと予想される。

上記の理由から、[A] は問題外として除くと、従来手法を用いても、[B] [C] の二手法では比較的単純な構造である英数字パターンでは良好な成果が



得られるが、本研究で対象とする日本語文字、特に漢字のように複雑なパターンに対しては良好な成果が得られないと予想される。

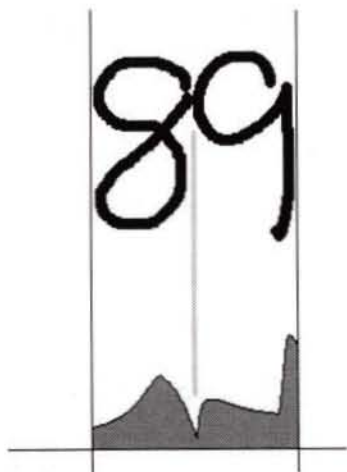


図 1：周辺分布からの切断箇所推定



図 2：文字の大きさを用いた切断箇所推定（失敗例）

## 2. 3. 1 Oversegmentation

提案する文字切り出し手法は、Oversegmentation を利用した認識系での利用を仮定している。ここで、Oversegmentation を利用した認識について説明する。

はじめに、接触しているパターンに対して接触箇所と推定された複数の箇所でもパターンを切断し、パターンを複数の連結成分の集合として扱う。次に、複数の連結成分をいくつか組み合わせて統合したパターンを多重に作成する。最後に、多重に作成された統合パターン全てに対して文字認識を行い、その文字認識結果が最適となる組み合わせで、各組み合わせに位置的矛盾がないものを接触パターンの認識結果とする。

ここでいう位置的矛盾を図 3 で説明する。図 3 上のように接触パターンが 6 つの連結成分に分割されたとき、A のように左から連続に  $a_n$  個（図 3 では  $a_1 = 3, a_2 = 3$ ）の連結成分の組み合わせを  $b$  個（図 3 では  $b = 2$ ）過不足なく作る場合や、B のように組み合わせに使われない連結成分が僅かに残っても雑音のように小さな成分である場合などは位置的矛盾がないとする。しかし、B で使われない連結成分が大きい場合や、C のように連結成分 3 と 4 で入れ替えがある組み合わせは、位置的矛盾があるとするとする。

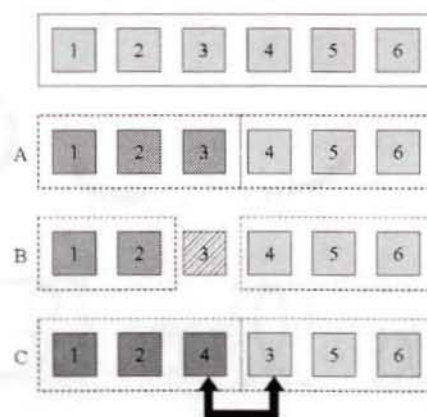


図 3：組み合わせの例



具体的には、図 4上段の接触パターンで、矢印1, 2, 3の箇所での切断の可能性が検出されたとする。そのときに、パターンを図 4中段のように、切断の可能性のある全ての箇所での切断を行い、複数の連結成分に切り出す。次に、連結成分を統合した図 4下段の3組6個のパターンが認識部に送られ、それぞれの認識結果が得られる。このようにして全ての組み合わせで文字認識を行い、リジェクトされたものは、適当な組み合わせでないとして単語認識の候補からはずす。最後に、単語全体として過不足なく切断された部分パターンの中で、十分信頼できる認識結果を与えたものを採用すれば、文字の切断と認識が同時に完成したことになる。

この手法は数字に対しては十分効果があるが[2]、連続筆記体英単語認識では、認識部の性能があまり高くないので、認識部のみでは確定できないという問題がある。そこで、単語照合も協調させ、認識結果の中で単語として成立する組み合わせを採用する手法が提案され、効果を上げている[3]。

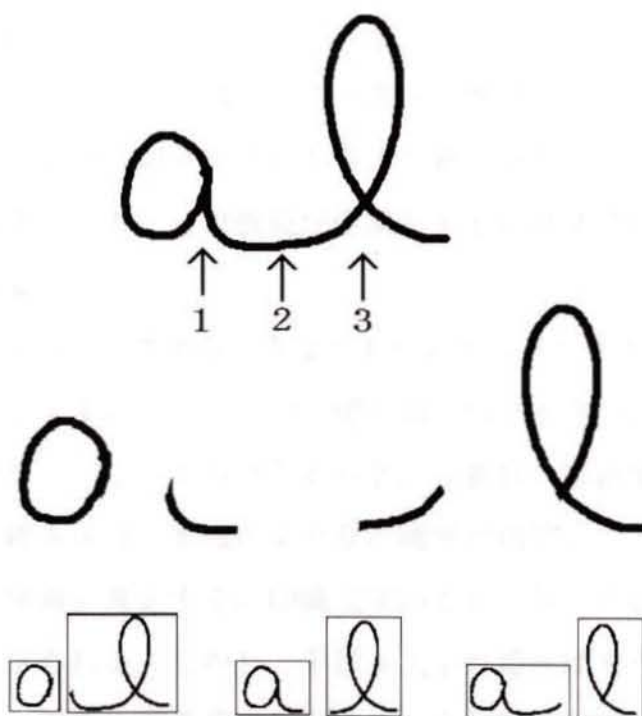


図 4 : Oversegmentation の例

## 2. 3. 2 接触がないときの日本語住所の文字切り出し

提案する漢字パターンに対する文字切り出し手法は、接触がない日本語住所認識システムに前処理として組み込むことが最終目標である。ここで、接触がない日本語住所認識システムについて説明し、接触文字が認識系に与える不具合を示す。

前節で述べた接触パターン（続け書きパターン）に対する文字切り出し手法は、接触のない日本語住所の切り出し・認識と原理的には同一である。

日本語の郵便書状には、行書体あるいは草書体の住所が希に出現する。しかし、大半は楷書体の漢字で書かれている。漢字は複数の連結成分（黒画素が連なってできる、他とは隔離された領域）からなるので、文字切り出しの問題は、どの連結成分の組み合わせを一つの文字と解釈するかに帰着する。

すなわち、漢字住所の個別文字への分割の問題は、続け書き英単語があらかじめ連結成分に分割されていたとき、どの連結成分を組み合わせで一つの文字にするかという問題と同じ問題である。

言い換えれば、この二つの問題は（続け書き英単語の場合、適切な箇所では分離しなくてはならないが）、切り出された要素を組み合わせで一つの文字とする仮説が多重に存在するうち、どの仮説が認識系あるいは単語照合から見て妥当かということである。

具体的例で説明すれば、図 5 のようなパターンが与えられたとき、(人間の目では一つと思われる文字に対し) 二つの切り出し仮説が立つ。それぞれに対する文字認識結果が妥当でないもの（左から 2, 3 番目の連結領域）については棄却される。この結果図 5 の太線のような認識率が閾値以上の経路を作成し、「長野市」という候補が確定する。印刷文字のとき、第二の仮説は中央の二つの文字がリジェクトされる。しかし、手書き文字の場合は大きさの変動を許容しなければならないので、認識系の第二の切り出し仮説はリジェクトできないであろう。しかし、「長・野・市」「長・里・予・市」という二つの仮説は、認識系ではいずれとも決定できないが、地名の知識を用いて「長野市」に確定で



きる。

しかし、図 6のように、[野市] の間に接触がある場合、[野] [市] の文字で正しい認識が行われず、住所読み取りが失敗する。このような問題を解決するため、接触文字を接触箇所でも分離し、連結領域を振り分けることで文字認識を可能にすることが本研究の目標である。すなわち、図 6の例で、もし適切な箇所でも文字を分離できれば、図 5と同様な処理をすることで、認識可能になるのである。

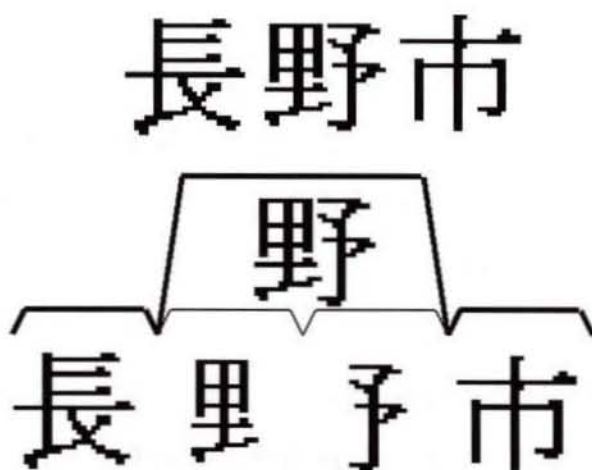


図 5：接触文字を含まない住所読み取り流れ図

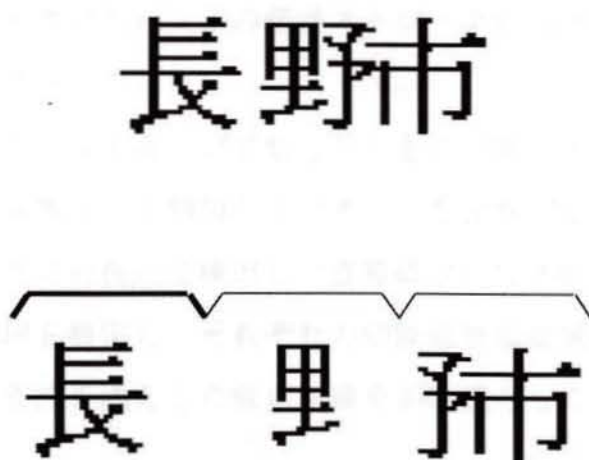


図 6：接触文字を含む住所読み取り流れ図

## 2. 4 本研究の方針

本研究では日本語宛名書きデータを対象としている。はじめに、一般日本語文字データの特徴を挙げる。

- [A] 行について、縦書き・横書き（左から右）が存在する
- [B] 比較的複雑な構造で直線的な字体である文字（漢字）と、構造が比較的単純ではあるが曲線を含むことが多い文字（平仮名、アラビア数字、片仮名）が混在する
- [C] 数字と漢字などで大きさに隔たりが大きい、平仮名に比べ漢字が大きい、漢字のなかでも大きさに隔たりが大きいものがある（縦書きの漢数字「一」など小さく書かれる場合がある漢字がある）など、文字の大きさが均一ではない
- [D] 漢字は、英数字と比較して一文字が複雑で文字を構成するストローク数が多い

本研究では日本語文字一般ではなく、郵便宛名書きデータを対象としているため、上記の特徴 [B]、「構造が比較的複雑な文字と単純な文字の混在」は少なく、ほとんどの文字が比較的複雑な構造の字体である。すなわち、筆記されている文字は直線的で構造が比較的複雑な漢字が占める割合が高く接触文字も漢字同士であることが多い。実際に郵便宛名書きデータの接触部を解析したところ、接触文字の大半は先行文字の最終ストロークが延びすぎて後続文字に接触する場合が大半であった。

そこで、本研究では漢字同士が接触したときの切断を対象とし、パターン中の主要成分である直線成分を利用してパターンを分割する手法を提案する。

具体的には、マクロ的視点で検出した直線成分の交差箇所パターンを分割する切断線候補領域を推定し、それぞれの切断線候補領域内で仮切断線を推定する。ある一定の条件を満たした仮切断線を切断線として接触パターンの切り出しを実現した。



## 2. 5 統合的手法による手書き文字切り出し法

2.3 節の方針に従い、図 7 の流れに示した文字切り出しシステムを構築した。前提として、ここで扱うパターンは郵便宛名書きパターンから接触文字を含む行パターンとして切り出したものとした。すなわち、実際のパターン中に接触箇所が必ず存在すると仮定している。各処理について次節以降で説明する。

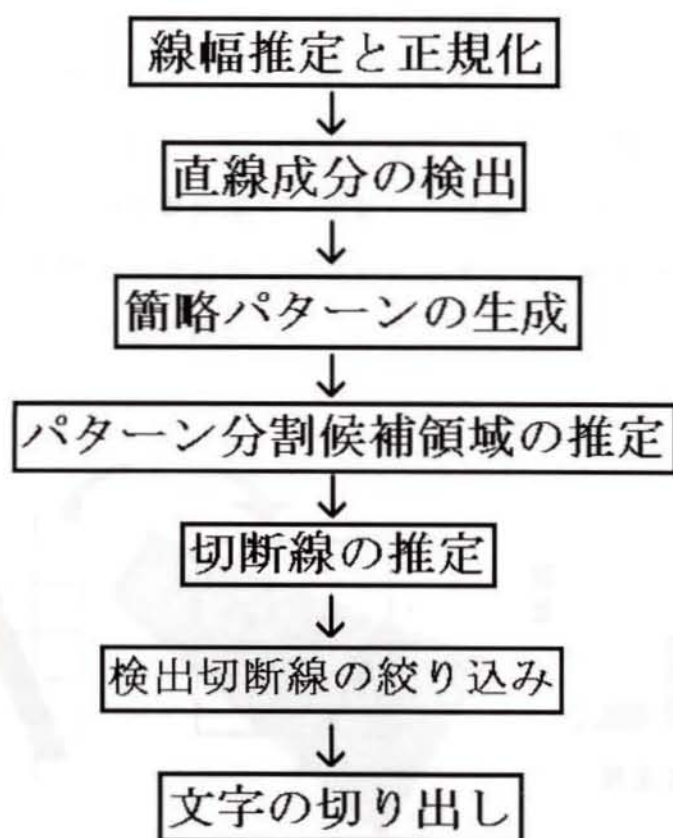


図 7：手書き文字切り出しアルゴリズムの流れ

## 2. 5. 1 線幅の推定と正規化

郵便書状の宛名は、文字の大きさ、筆記具が多様であるため、書状ごとの文字パターン中の線幅は多様である。同一書状であっても種々の要因で、線幅は変化する。線幅が異なると、同一の処理を行っても結果に大きな差が現れることがある。そのため線幅を推定し、パターンの正規化を行う。

線幅を推定するため、図 8に示す手順で処理を行う。線幅推定は、文字パターン内の線幅の変化に対応するため、パターン全体を幾つかの領域に分割し、それぞれの領域で独立に行うものとした。具体的には図 8左に示すようにパターンを格子状の領域に分割し、図 8中央に示すように各領域内にあるパターンの輪郭線に垂直な方向の黒画素連長（連続した黒画素の個数）を調べる。各領域それぞれで調べた黒画素連長から図 8右に示すように頻度分布を求め、最も頻度の高いものをその領域の線幅として推定する。

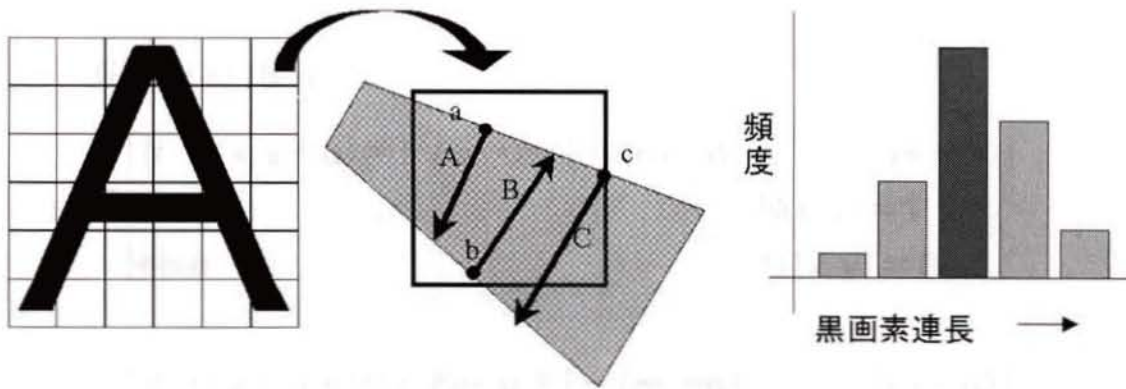


図 8：筆記線幅の推定

線幅の太いパターンでは、次節以降で提案する手法では過剰検出等の誤りを生じやすい。それを回避するための前処理として、推定線幅がある閾値以上であれば、パターンを細めて基準線幅に正規化する。この正規化処理は、線幅を一画素にするいわゆる細線化ではなく、太すぎるパターンの線幅を基準値に近

づけるための処理である。具体的には、縦、横の両方向で独立に原パターンの各黒画素に対して以下の判定を行い線幅の正規化を行う。

### 正規化アルゴリズム

実測筆記線幅の画素数：	m
基準筆記線幅（既定値）：	n
画像横方向：	X 軸方向
画像縦方向：	Y 軸方向
入力画像：	O
正規化画像：	N
入力画像上の画素：	$O(x, y)$
正規化画像上の画素：	$N(x, y)$
画像上の黒画素：	1
画像上の白画素：	0

If (m > n) then

縦  $\left\{ \begin{array}{l} \text{If } O(x, y+u)=1 \text{ For } u=\{-(m-n), \dots, (m-n)\} \\ \hspace{10em} N(x, y)=1 \\ \text{else} \hspace{10em} N(x, y)=0 \end{array} \right.$

横  $\left\{ \begin{array}{l} \text{If } O(x+u, y)=1 \text{ For } u=\{-(m-n), \dots, (m-n)\} \\ \hspace{10em} N(x, y)=1 \\ \text{else} \hspace{10em} N(x, y)=0 \end{array} \right.$

このアルゴリズムは斜めのストロークに対しては必ずしも正確ではないが、漢字ストロークの大半が縦または横の直線であること、厳密に全てのストロークを同一線幅にする必要はないことなどから、このように正規化した。



## 2. 5. 2 直線成分の検出

提案する文字切り出し手法では漢字パターン中の直線的ストロークの情報を利用している。そのため、本研究では文字パターンの輪郭をフリーマンコード列で表現し、フリーマンコード列の連続性を利用して直線成分の検出を行った。

フリーマンコードとは、パターンの連続する輪郭線情報を図 9左のようにコード化して表現するもので、図 9右のパターンで黒点から始まる輪郭をフリーマンコード列で表現すると (6 7 0 0 7 0 1 2 2 2 3 4 4 4 5 5) となる。但し、輪郭は時計回りに追跡する。

同じフリーマンコードが続いている部分は直線と考えられ、これを用いて直線部分を検出できる。しかし、フリーマンコードの方向分解能は8方向であり、これ以上詳細な角度の直線検出には不十分である。また、輪郭に僅かなノイズがあり、異なるフリーマンコードが混入すると直線を見落とす恐れがある。

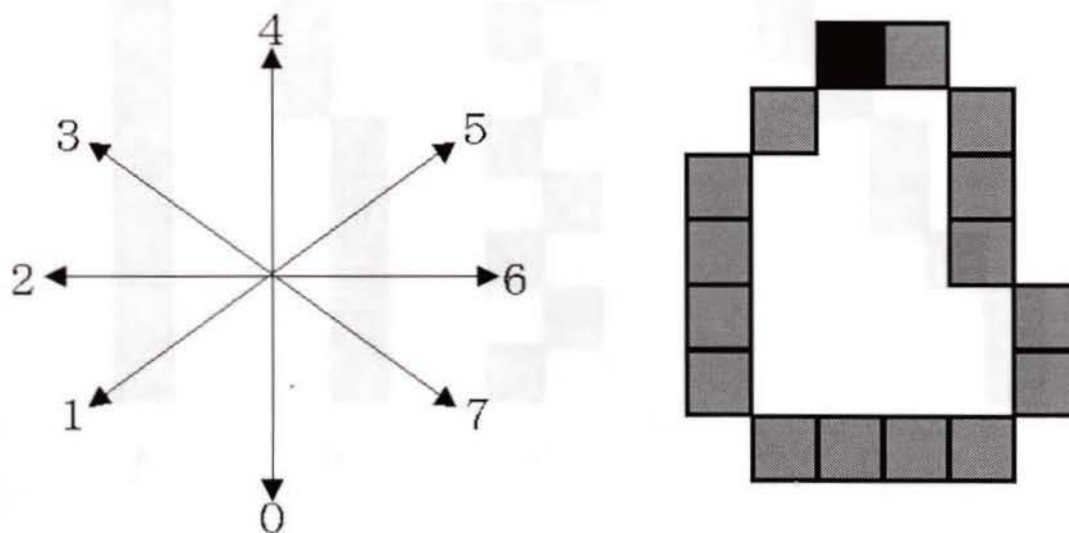


図 9：フリーマンコードの例

そのため、フリーマンコード列における多少の変動を許容し、8方向以上の

分解能を有する直線成分の検出法を開発した。具体的には図 10に示すように、以下のようなフリーマンコード列を検出し直線成分としている。

- [A] 同一コードの並びがある  
例：フリーマンコード列 (0 0 0 0 0 0 0 0...)
- [B] 同一コードの並びが閾値以下の数の異なるコードを挟んでつながっている  
例：フリーマンコード列 (0 0 0 7 0 0 0 0...)
- [C] 閾値以下の個数のコードが反復的につながっている  
例：フリーマンコード列 (7 1 7 1 7 1 7 1 ...)
- [D] 同一コードの並びが周期的につながっている  
例：フリーマンコード列 (7 0 7 7 0 7 7 0 7...)

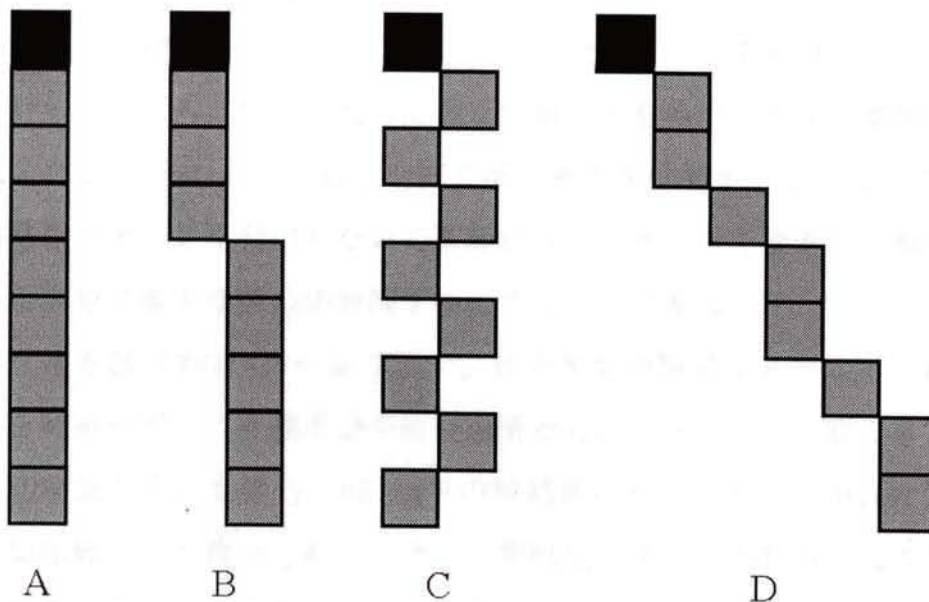


図 10：直線成分検出の例

## 2. 5. 3 簡略化パターンの生成

簡略化パターンとは、2.5.2 節で検出した直線成分を描画して太めた後、直線方向に引き延ばしを行って生成したパターンである。漢字などの複雑な文字パターンに対しては、マクロ的な視点で文字の切り出しを行うことが有効であると考え簡略化パターンを生成した。

このとき直線成分を、8 方向を超える分解能で描画したのではパターンが複雑になってしまうため、検出した直線成分の方向を 8 方向に量子化して処理している。この結果、簡略化パターンは推定線幅（線幅の正規化を行ったパターンでは、正規化後の基準線幅）を有する 8 方向の直線成分だけによって描画されたパターンとなる。図 11 に実画像から検出された簡略化パターンの例を示す。図 11 で左は原パターン、中央は直線成分検出で検出された直線成分のみを描画したパターン、右は簡略化パターンである。

研究の初期に、複雑なパターンに対して簡略化パターンを利用せずに切断箇所推定を行ったところ、膨大な数の余剰な切断箇所を検出したが、簡略化パターンの導入により、かなりの余剰な切断箇所の検出を抑制することに成功した。これは簡略化パターンを利用しないで直線成分の交差箇所を探索した場合、漢字の複雑な部分で膨大な数の切断線を検出するためである。

具体的な例を図 12 のパターンで示す。研究初期の開発システムで、簡略化パターンを利用せずに、直線成分の直交箇所から、パターンを切断できる適当な方向に切断線を引いた場合、図 12 下の輪郭線パターン内に引かれた短い直線を全て切断線として検出した。しかし、簡略化パターンの利用や 2.5.5 節に示す切断線条件を与えることによって、矢印で示す 2 箇所のみで切断線を検出するようになった。



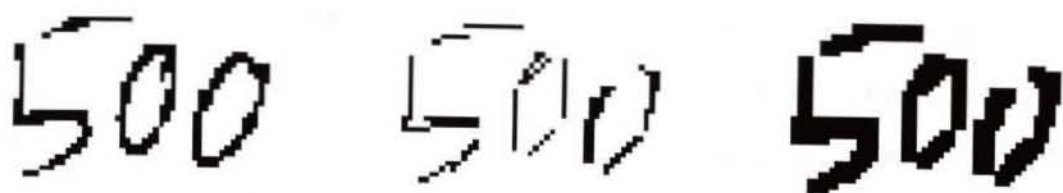


図 11：簡略化パターンの生成例



図 12：簡略化パターンの効果

## 2. 5. 4 パターン分割候補領域の推定

パターンの分割は、簡略化パターンの中で連続領域（黒画素が連結している領域）ごとに実行し、実パターンを分割する。ある切断線で連続領域を切断する場合、その連続領域が2個以上の場所で他の連続領域に接触していた場合などでは、一つの切断線ではその連続領域を分割できない。

そのため、切断線推定処理ではパターンが分割されるかどうかテストをする必要があるが、パターン全体から一度に計算を行った際には膨大な計算量となり、実用化に耐えないと予想された。そこで、切断線を検出する候補領域を推定し、切断線の推定はこの領域内で行うものとした。

そこで、2.5.3節の方法で作成した簡略化パターン中から直線同士の交差箇所を求め、その部分の面積がある閾値より大きければ、交差箇所を中心としたある一定の大きさの領域をパターン分割候補領域として検出する方法を取った。

このとき、近接する閾値以内の距離の領域で複数のパターン分割候補領域が検出されたときには、パターンを過剰に切り出す可能性があるため、パターン分割候補領域内の直線交差成分を多く含む領域のみを採用する。

次のようなアルゴリズムで交差箇所を求める。

- ・ 簡略化パターン内で、ある閾値以上の長さの直線を検出する
- ・ 検出した直線上の各画素から、3画素×3画素の領域で、直交する直線成分を探索し、直交する直線を検出する
- ・ 検出された直交する直線が、閾値以上の長さを持っているならば、その画素を交差箇所とする

具体的には、図13でAは現在交差箇所を検査している直線とする。直線Aの各画素を中心に、3×3画素の領域で他の直線の画素がないか、パターン上方から検査していくとする。はじめに、直線Bとの接触が検出されるが、直線Bは直線Aと直交ではないので、交差箇所としない。次に直線Cとの接触が検出されるが、直線Cは閾値以上の長さを持たないため、交差箇所としない。最後に直線Dとの接触が検出され、直線Dは閾値以上の長さを持ち、直線Aと

直交しているので交差箇所として検出する。

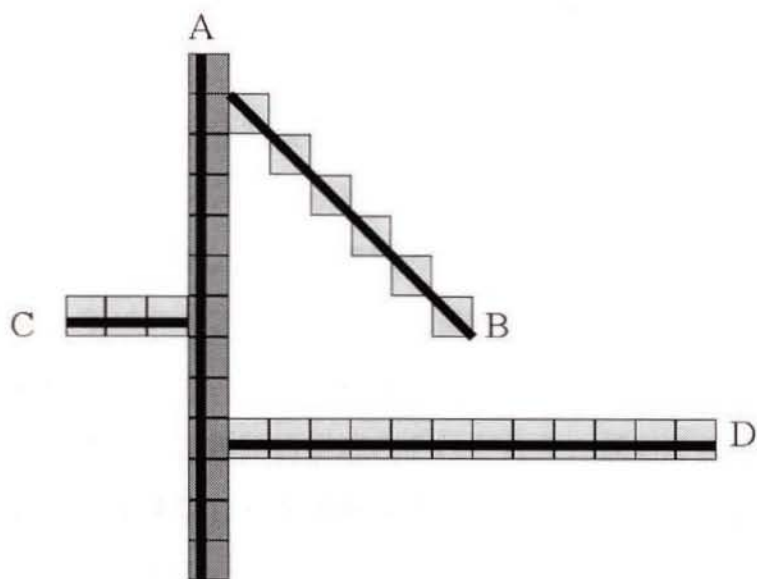


図 13：交差箇所の判定

図 14に実際の画像に対する例を示す。図 14で、左は原パターン、中央は2.4.3節で求めた簡略化パターン、右は簡略化パターンと検出されたパターン分割候補領域を重ねて表示したものである。

説明した処理により、妥当な分割候補領域が求まっているのが解る。



図 14：パターン分割候補領域の推定



## 2. 5. 5 切断線の確定

2.5.4節の手法により求めたパターン分割候補領域内に存在する、原パターンの輪郭線上の点を、端点と呼ぶ。行方向が横（縦）であるとき、この端点は2つに分類される。パターン分割候補領域の上（左）から、はじめの輪郭線上の点を上（左）端点、パターン分割候補領域の下（右）から、はじめの輪郭線上の点を下（右）端点と呼ぶ。この2種類の端点を結ぶ直線を仮切断線と呼ぶ。この仮切断線によりパターンを切断すれば接触文字の分離ができるが、一般に仮切断線は複数個あるので、一つに確定する必要がある。

したがって、パターン分割候補領域のそれぞれで、仮切断線の探索を行い、後述する切断線条件を満足した仮切断線の中で、切断線の直線距離が最小のものを切断線として確定する。

仮切断線の推定のため、住所の筆記方向に直交する走査線で、パターン分割候補領域中のパターンの端点を検出する。次に、検出した上（左）端点と下（右）端点を結び、仮切断線として推定する。すなわち、この仮切断線は検出した上下（左右）の端点の組み合わせ数だけ推定される。

次に、各仮切断線で次の切断線条件を調べる。

### [切断線条件]

[A] 対象連続領域を分割すること

[B] 分割により生成される連続領域各々の大きさがある閾値より大きいこと

この条件を満足した仮切断線のうち、仮切断線の端点間の直線距離が最小のものを、そのパターン分割候補領域中の切断線として推定する。同様の処理を、すべてのパターン分割候補領域で行い、仮切断線を検出する。

具体的な処理を、図 15の例で説明する。図 15左で、横書きの住所であったとすれば走査線は縦方向となるので、左から右に走査線を動かしてパターン内

の端点（画素点）を図中の黒丸点のように検出する。次に黒丸点の上下について、全ての組み合わせを結ぶ仮切断線を推定する。次に仮切断線全てについて切断線条件を満足するか調べ、満足するもののみ残す。最後に、図 15の例のように切断線条件を満たした複数の仮切断線のうち、長さが最小のもの  $a$  をそのパターン分割候補領域中の切断線として断定する。

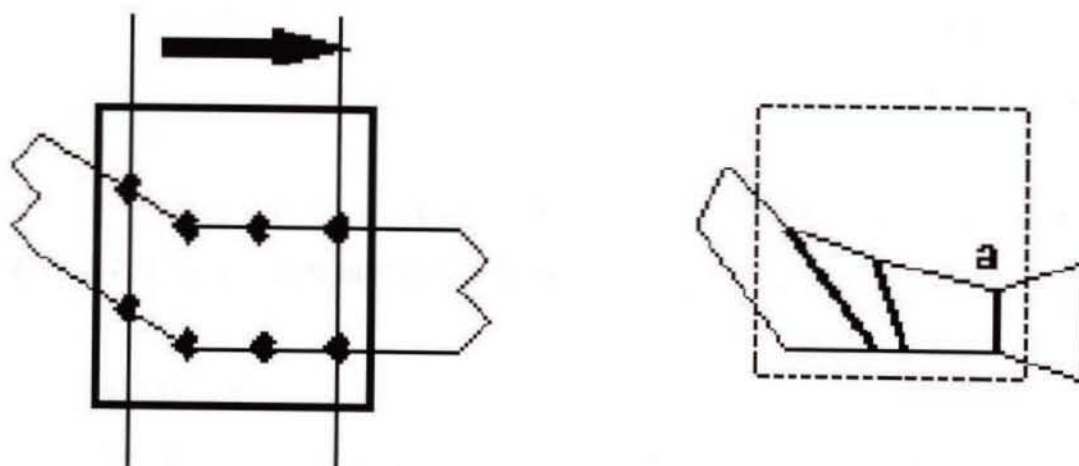


図 15：切断線推定

## 2. 5. 6 検出切断線の絞り込み

前節までの処理により、パターンを切り出すための切断線が推定できるが、このままでは余剰切断線が多く、余剰にパターンを切り出してしまうことがある。Oversegmentation を利用する認識手法[2][3]では、認識部に送られるパターンは連結成分の組み合わせであるが、その組み合わせの個数は切断線が多いと指数関数的に増大するため、処理時間の長大化等の悪影響を及ぼす恐れがある。したがって、認識部の速度の観点からは、余剰と思われる切断線を削減することが望ましい。

そこで、切断線それぞれに得点付けを行い、その得点を利用して検出切断線の絞り込みを行う。切断線に付与する得点Pは次式によって決定した。

得点付与アルゴリズム

切り出しによって分割された連続領域数	:	R
切り出しによって分割された第 i 領域の画素数	:	$L_i$
切り出しを行う前の連続領域の画素数	:	M

$$P = \left( \sum_{i=1}^R L_i \right) \times M$$

この得点方式を採用した理由は次の二点である。

- [1] 切り出し前の連続領域が大きいほど、その領域の分割が必要な場合が多い
- [2] 連続領域を中央付近で分割したほうが良好な結果が得られる場合が多い

このようにして各切断線に得点を付与し、検出する切断線を絞り込んで出力することを可能にした。今後、認識部が十分に高速になった場合には、絞り込みを行わず、検出した切断線をすべて出力することも考えられる。



## 2. 5. 7 文字の切り出しの評価法

前節までの方法により、パターンの切り出しを可能にする切断線が検出できる。切り出し処理の最終部分では、実際に切断線を用いて文字を切り出す。

文字を切り出す場合、切り出した文字をどのような仕様の認識系で利用するかによって適当な切り出し方法は異なる。例えば、**oversegmentation** を利用した認識系[2][3]では連続領域に分割することで認識可能になる。本研究では、この **oversegmentation** を利用した認識部で利用できるような文字の切り出しを想定した。

しかし、切り出し結果を評価するのに実際の認識系を利用したのでは、認識系の性能に依存して大きく評価は異なる。2.1 節で述べたように最終的には認識系と協調することが目標であるが、切り出し単体で評価できることが望ましい。特に、本研究は企業との共同研究により行ったため、認識系と協調した評価を行ったのでは、日々進歩・変化している認識系の開発を阻害する恐れがあった。

そこで本研究では、与えられたパターンから切断箇所の画素を消去することで、切り出しパターンを作成することとし、切り出しパターンを視察することでその切り出し処理を評価することを可能とした。

## 第3章 Hidden Markov Model を利用した文字認識

### 3.1 従来の研究

HMM を利用した認識手法は、音声認識において有効な手法として確立されている。[9][10]。また、音声と同じく1次元の性質を有するオンライン文字認識においても有効性が示されている[11]。

音声認識、オンライン文字認識で有効性が確認され、高い表現能力が実証された HMM を利用した認識系を、オフライン文字認識にも適用しようという研究が近年進められてきている[11][12][13][14][15][16][17]。

HMM を利用したオフライン文字認識では、音声認識、オンライン文字認識と同様なモデルである「1次元 HMM を利用した文字認識手法」がまず確立された[14][15][16][17]。しかし、方向性パターンマッチングやニューラルネットワークを利用した文字認識系と比較して、十分な性能を発揮するまでに至っていない。

2次元のパターンを1次元の系列に変換して利用する1次元 HMM を利用したのでは、2次元パターンであるオフライン文字の情報を十分にモデルに反映できないとする考えから、HMM 自体を拡張し2次元シンボルの入力を可能にして認識系に利用しようという研究が現在盛んに行われている[12][23][24]。しかし、これら2次元 HMM を利用した認識系は、モデルの未確立やモデルの複雑さ等の理由から、高い認識性能の可能性はあるものの、実用に向けた課題が多く残されている。

## 3. 2 1次元 HMM の解説

1次元 HMM は、音声認識、オンライン文字認識で確立した手法である。本節では、1次元 HMM の原理について、[9][10]に基づいて解説する。

### 3. 2. 1 Hidden Markov Model の解説

Hidden Markov Model はマルコフ過程 (Markov Process) の拡張である。マルコフ過程では、一つの状態遷移に対して一つの出力シンボルが対応しているため、ある出力シンボル系列が与えられたとき状態遷移は一意に決定される。

これを拡張し、Hidden Markov Model では、一つの状態遷移に対して  $M$  個のシンボルのうち一つが確率的に出力されるものとする。そのため、ある出力シンボル系列が与えられても、これを出力するような状態遷移は一意に決定されない。言い換えれば、同一出力シンボル系列を出力するモデルパラメータは多数存在する。それらのパラメータ群の中で、ある学習サンプルの集合が与えられたとき、これを出力するようなモデルの最適パラメータがあると考えられる。

このようなモデルにおいて、あるシンボル系列出力の集合を仮定し、そのシンボル系列集合を出力する出力確率が最大となるようにモデルのパラメータ推定を行うのが Hidden Markov Model の学習である。

また、学習によって決定したモデルを利用して、クラス未知のシンボル系列に対する状態遷移確率を求め、出力確率を比較して最も確からしいクラスを定めることが Hidden Markov Model を利用した認識である



### 3. 2. 2 1次元HMMの定義

音声認識やオンライン文字認識で利用されている HMM と同様のモデルを本論文では、2次元 HMM と区別するため 1次元 HMM と記述する。

1次元 HMM は図 16 のように、いくつかの状態と状態遷移によって表現され、各状態遷移は状態遷移確率と状態遷移に対するシンボルの出力確率を持っている。状態遷移と、それに伴い出力されるシンボルは、モデルごとに定めたパラメータに依存する。

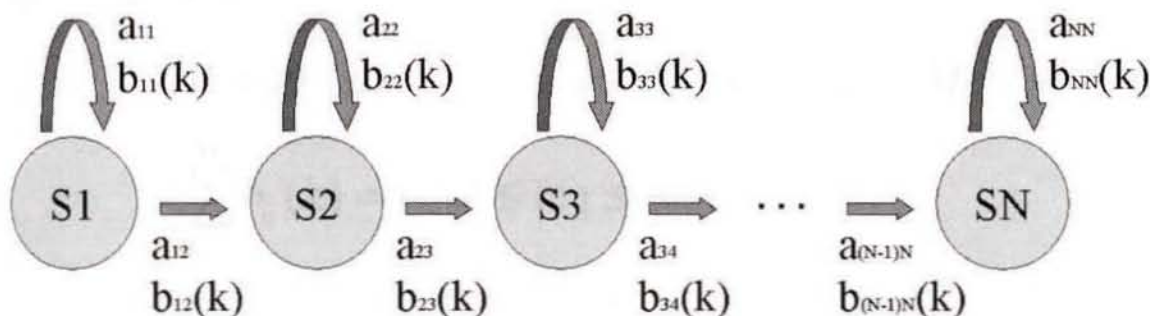


図 16：1次元 HMM 状態図の例

#### left-to-right モデル

HMM において、一般的には全ての状態間の遷移は可能であるが、図 16 では状態  $i$  から状態  $i + 1$  へしか遷移しない。このような HMM を left-to-right モデルと呼ぶ。

音声のような時間関数の認識では left-to-right モデルが広く使われている。本研究でもこの left-to-right モデルを利用した。

1次元 HMM を決定するパラメータについて説明する [9][10]。

**状態遷移確率の集合** 1次元 HMM は  $S = \{S_1, S_2, \dots, S_N\}$  の  $N$  個の状態を持ち、各状態は状態遷移によって結ばれている。 $a_{ij}$  は状態  $S_i$  から  $S_j$  へ遷移する確率である。集合  $\{a_{ij}\}$  を  $A$  と表す。

$$A = \{a_{ij}\} \quad 1 \leq i, j \leq N$$

$$\sum_{j=1}^N a_{ij} = 1 \quad i = 1, 2, \dots, N$$

シンボル出力確率の集合      1次元 HMM は  $V = \{v_1, v_2, \dots, v_M\}$  という  $M$  個のシンボルがあり、状態  $S_i$  から  $S_j$  へ遷移のときにシンボル  $v_k$  を出力するとする。 $b_{ij}(k)$  は状態  $S_i$  から  $S_j$  への遷移のときにシンボル  $v_k = k$  を出力する確率である。集合  $\{b_{ij}(k)\}$  を  $B$  と書く。

$$B = \{b_{ij}(k)\} \quad 1 \leq i, j \leq N \quad k = 1, 2, \dots, M$$

$$\sum_{k=1}^M b_{ij}(k) = 1 \quad 1 \leq i, j \leq N$$

初期状態確率の集合       $\pi_i$  は初期状態が  $S_i$  である確率である。本研究では left-to-right モデルを採用したので初期状態を  $S_1$  のみとしたが一般には任意の状態  $S_i$  ( $i=1, \dots, N$ ) であってもよい。集合  $\{\pi_i\}$  を  $\pi$  と書く。

$$\pi = \{\pi_i\} \quad i = 1, 2, \dots, N$$

$$\pi_i = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases}$$

最終状態の集合      最終状態は、一般には任意の複数の状態であってもよいが、本研究では  $S_N$  のみとした。

以上により、本研究で使用する 1次元 HMM は上記のパラメータを用いて次式で表せる。

$$\lambda = (A, B, \pi)$$

### 3. 2. 3 1次元 HMM の確率計算と最適パスの算出

1次元 HMM でモデルパラメータ  $\lambda$  が与えられたとき、シンボル系列  $O=O_1O_2\dots O_T$  の出力確率（尤度） $P(O|\lambda)$  を計算する。これは、モデルパラメータ  $\lambda$  から  $O$  が発生する確率であり、シンボル  $O$  を観測したとき、これがモデルパラメータ  $\lambda$  から出力された尤度に相当する。この出力確率は、シンボル  $O$  を出力することが可能な状態遷移系列  $Q=q_0q_1\dots q_T$  と  $Q$  の下で  $O$  が発生する事象の同時確率の和である。

$$P(O|\lambda) = \sum_{all\ Q} P(O, Q|\lambda) = \sum_{all\ Q} P(O|Q, \lambda) P(Q, \lambda)$$

状態遷移はマルコフ過程なので、

$$\begin{aligned} P(Q|\lambda) &= P(q_0q_1\dots q_T) \\ &= \prod_{t=1}^T P(q_t | q_{t-1}, \lambda) \\ &= \pi_{q_0} a_{q_0q_1} a_{q_1q_2} \dots a_{q_{T-1}q_T} \end{aligned}$$

また、非決定性（Hidden）の過程から、

$$P(O|\lambda) = \sum_{q_0q_1\dots q_T} \pi_{q_0} a_{q_0q_1} b_{q_0q_1}(O_1) a_{q_1q_2} b_{q_1q_2}(O_2) \dots a_{q_{T-1}q_T} b_{q_{T-1}q_T}(O_T)$$

以上より

$$\begin{aligned} P(O|Q, \lambda) &= P(O_1O_2\dots O_T | q_0q_1\dots q_T, \lambda) \\ &= \prod_{t=1}^T P(O_t | q_t, q_{t-1}, \lambda) \\ &= b_{q_0q_1}(O_1) b_{q_1q_2}(O_2) \dots b_{q_{T-1}q_T}(O_T) \end{aligned}$$



この計算を効率よく行うための手法である、Forward-Backward アルゴリズムを説明する。

### Forward-Backward アルゴリズム

前向き確率 時間  $t$  における状態  $S_i$  の前向き確率 (Forward Variables)

$\alpha(t, i)$  を定義する

$$\alpha(t, i) \triangleq P(O_1 O_2 \dots O_t, q_t = S_i | \lambda) \quad \begin{matrix} t = 0, 1, \dots, T \\ i = 1, 2, \dots, N \end{matrix}$$

初期化 ( $t = 0$ )

$$\alpha(0, i) = \pi_i \quad i = 1, 2, \dots, N$$

反復

$$\alpha(t+1, j) = \sum_{i=1}^N \alpha(t, i) a_{ij} b_{ij}(O_{t+1}) \quad \begin{matrix} t = 0, 1, \dots, T-1 \\ j = 1, 2, \dots, N \end{matrix}$$

終止

$$\sum_{S_i \in F} \alpha(T, i) = P(O_1 O_2 \dots O_T | \lambda) = P(O | \lambda)$$

後向き確率 前向き確率と双対な後向き確率 (Backward Variables)

$\beta(t, i)$  を定義する

$$\beta(t, i) \triangleq P(O_t O_{t+1} \dots O_T, q_t = S_i | \lambda) \quad \begin{matrix} t = 0, 1, \dots, T \\ i = 1, 2, \dots, N \end{matrix}$$

初期化 ( $t = T$ )

$$\beta(T, i) = \begin{cases} 1 & \text{if } S_i \in F \\ 0 & \text{otherwise} \end{cases}$$

反復

$$\beta(t, j) = \sum_{j=1}^N a_{ij} b_{ij}(O_{t+1}) \beta(t+1, j) \quad \begin{array}{l} t = T-1, \dots, 1, 0 \\ j = 1, 2, \dots, N \end{array}$$

定義から明らかなように、 $\alpha$  と  $\beta$  は以下の関係で結ばれる。

$$P(O|\lambda) = \sum_{S_i=F} \alpha(T, i) = \sum_{i=1}^N \beta(0, i) \pi_i$$

前向き確率または後向き確率によって、与えられたシンボルに対する、1次元 HMM の最適な出力確率である受理確率は計算される。これらは、1次元 HMM のパラメータ推定で利用する。

### Viterbi のアルゴリズム

先に挙げた Forward-Backward アルゴリズムにより、1次元 HMM の受理確率は計算される。しかし、この手続きの計算時間は大きい。

ところが、Viterbi のアルゴリズム[9][10]が知られており、計算時間を短縮することができる。この手法は、動的計画法を用いて最も可能性の高い状態遷移系列（最適パス）を探索し、そのパス上での生起確率を1次元 HMM の受理確率とするものである。各状態の時刻  $t$  に対する尤度は、次式で与えられる。

$$\delta(t, i) \triangleq \max_{q_0 q_1 \dots q_{t-1}} P(q_1 q_2 \dots q_t = S_i, O_1 O_2 \dots O_t | \lambda)$$
$$\begin{array}{l} t = 0, 1, \dots, T \\ i = 1, 2, \dots, N \end{array}$$

初期化 ( $t = 0$ )

$$\begin{cases} \delta(0, i) = \pi_i \\ \varphi(0, i) = 0 \end{cases} \quad i = 1, 2, \dots, N$$

反復

$$\begin{cases} \delta(t, j) = \max_{1 \leq i \leq N} [\delta(t-1, i) a_{ij} b_{ij}(O_t)] & t = 1, 2, \dots, T \\ \varphi(t, j) = \arg \max_{1 \leq i \leq N} [\delta(t-1, i) a_{ij} b_{ij}(O_t)] & j = 1, 2, \dots, N \end{cases}$$

終止

$$\begin{cases} P^* = \max_{S_i \in F} \delta(T, i) \\ q_T^* = \arg \max_{S_i \in F} \delta(T, i) \end{cases}$$

パス（状態遷移系列）のバックトラッキング

$$q_t^* = \varphi(t+1, q_{t+1}^*) \quad t = T-1, T-2, \dots, 0$$

$F = \{S_N\}$ なので、時刻  $t = T$  における最終状態  $S_N$  での尤度  $\delta(T, N)$  が 1 次元 HMM の受理確率  $P^*$  となる。状態遷移系列は、各  $\delta(t, j)$  に対して最大値を与える状態  $S_i$  を  $\varphi(t, j)$  に保存し、時刻  $t$  に対して逆トレースすることにより得られる。すなわち、 $q_0^* q_1^* \dots q_T^*$  が最適状態遷移系列となる。

### 3. 2. 4 1次元 HMM のパラメータ推定

パラメータ推定は、パターン認識の用語でいえば学習に相当する。これは、学習サンプルセットが与えられたとき、それが確率モデルから出力される確率（尤度）を最大にする確率モデルのパラメータを推定することである。

1次元 HMM  $\lambda = (A, B, \pi)$  が与えられたとき、シンボル系列  $O = O_1 O_2 \dots O_T$  に対して、

$$P(O | \bar{\lambda}) \geq P(O | \lambda)$$



となるようなパラメータ  $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$  を有する 1 次元 HMM を推定することができる[9][10]。したがって、1 次元 HMM の定義を満たす初期パラメータを与え、上式によるパラメータ推定を繰り返せば、(局所) 最適値に収束することが保証される。

### パラメータ推定アルゴリズム

1 次元 HMM  $\lambda = (A, B, \pi)$  のパラメータ推定アルゴリズムを示す。そのうち、 $\pi$  については初期状態のままとし、パラメータ推定はしない。パラメータ推定方法は、最尤推定法を基本とする EM (Expectation-Maximization) アルゴリズムを適用する。EM アルゴリズムは、観測可能な変数のみでなく、観測できない変数をも含む場合の確率モデルのパラメータ推定に用いられるアルゴリズムである。

本方法は、Forward-Backward アルゴリズムと呼ばれる。それは前節で紹介した前向き確率と後向き確率を利用することに由来する。次式で示される再推定式を用いて状態遷移確率  $a_{ij}$  とシンボル出力確率  $b_{ij}$  の推定値  $\bar{a}_{ij}, \bar{b}_{ij}(k)$  をそれぞれ計算する。

$$\bar{a}_{ij} = \frac{\sum_{t=0}^{T-1} \alpha(t, i) a_{ij} b_{ij}(O_{t+1}) \beta(t+1, j)}{\sum_{t=0}^{T-1} \alpha(t, i) \beta(t, i)}$$

$$\bar{b}_{ij}(k) = \frac{\sum_{t=0}^{T-1} \alpha(t, i) a_{ij} b_{ij}(O_{t+1}) \beta(t+1, j)}{\sum_{t=0}^{T-1} \alpha(t, i) a_{ij} b_{ij}(O_{t+1}) \beta(t+1, j)} \quad s.t. O_{t+1} = v_k$$

複数シンボル系列の集合に適用する場合

複数のシンボル系列の集合  $\mathbf{O} = \{O^{(1)}, O^{(2)}, \dots, O^{(R)}\}$  に対してパラメータ推定を行う[9]。そのときの前向き確率、後向き確率をそれぞれ  $\alpha^{(r)}(t, i)$ ,  $\beta^{(r)}(t, i)$  とすれば、再推定式は次式となる

$$\bar{a}_{ij} = \frac{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=0}^{T_r-1} \alpha^{(r)}(t, i) a_{ij} b_{ij}(O_{t+1}^{(r)}) \beta^{(r)}(t+1, j)}{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=0}^{T_r-1} \alpha^{(r)}(t, i) \beta^{(r)}(t, i)}$$

$$\bar{b}_{ij}(k) = \frac{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=0}^{T_r-1} \alpha^{(r)}(t, i) a_{ij} b_{ij}(O_{t+1}^{(r)}) \beta^{(r)}(t+1, j)}{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=0}^{T_r-1} \alpha^{(r)}(t, i) a_{ij} b_{ij}(O_{t+1}^{(r)}) \beta^{(r)}(t+1, j)} \quad \text{s.t. } O_{t+1}^{(r)} = v_k$$

ここで、 $P_r = P(O^{(r)} | \lambda)$  である。

### 3. 3 1次元 HMM を利用したオフライン文字認識

#### 3. 3. 1 1次元 HMM のオフライン文字認識への適用

3.2 節でも述べたように、1次元 HMM は音声認識やオンライン文字認識のための有効な手法として確立された。その成功の主な原因は、音声やオンライン文字の情報は、時系列に対する状態変化として捕えることで1次元のシンボル系列に変換することが容易なためであった。パターンを1次元のシンボル系列に変換できれば1次元 HMM を使った認識システムで扱うことが容易である。

しかし、オフライン文字の情報は2次元のパターンとしてしか与えられないため、2次元情報を1次元のシンボル系列に変換する必要がある。図 17に1次元 HMM を利用したオフライン文字認識の流れを示す。

本節では、本研究で文字パターンに対して適用した、1次元 HMM 認識系の各処理について説明する。

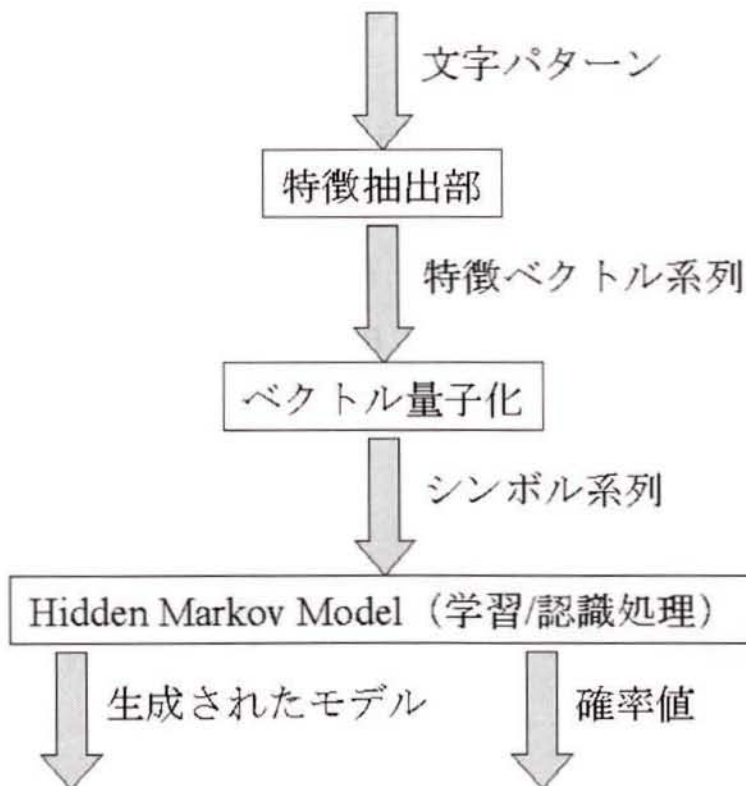


図 17：1次元 HMM を利用したオフライン文字認識系の流れ



**特徴抽出** 一般的に、1次元 HMM のオフライン文字認識では、文字パターンを短冊状に分割しそれぞれの短冊で独立に特徴ベクトルを求める。具体的には、図 18左のパターンが与えられたとき、それを左から右に走査し、パターンを図 18右のように短冊化する。

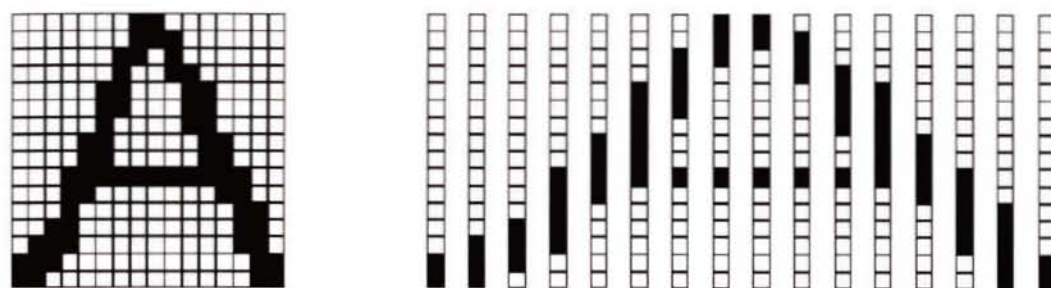


図 18：パターンの短冊化

短冊化されたパターンそれぞれに対して、図 19に示すように、以下のような特徴を求め、特徴ベクトルを求める。

特徴の例

[特徴A (B)] 短冊内の上から一番目の連続黒画素の始点 (終点)

[特徴C (D)] 短冊内の上から二番目の連続黒画素の始点 (終点)

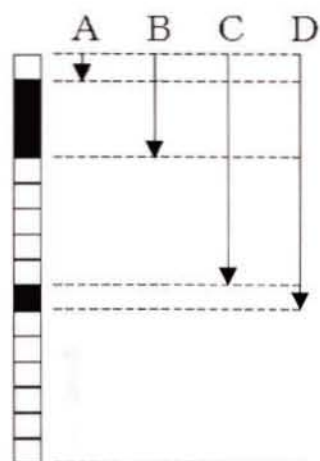


図 19：特徴抽出例

同様に、特徴の値を各短冊で求めることで特徴ベクトル系列を求める。  
 文献[13]で用いられた特徴ベクトルに含まれる特徴を列挙する。各特徴の単位は、上端を0とした画素の座標、および画素数である。

- [特徴 1] 短冊内の上から一番目の黒画素連の始点
- [特徴 2] 短冊内の上から一番目の黒画素連長
- [特徴 3] 短冊内の上から二番目の黒画素連の始点
- [特徴 4] 短冊内の上から二番目の黒画素連長
- [特徴 5] 短冊内の下から一番目の黒画素連の始点
- [特徴 6] 短冊内の下から一番目の黒画素連長
- [特徴 7] 短冊内の下から二番目の黒画素連の始点
- [特徴 8] 短冊内の下から二番目の黒画素連長

ベクトル量子化      パターンから特徴抽出部によって生成された特徴ベクトル系列  $\{f_1, f_2, \dots, f_N\}$  は図 20のように並んでいる。

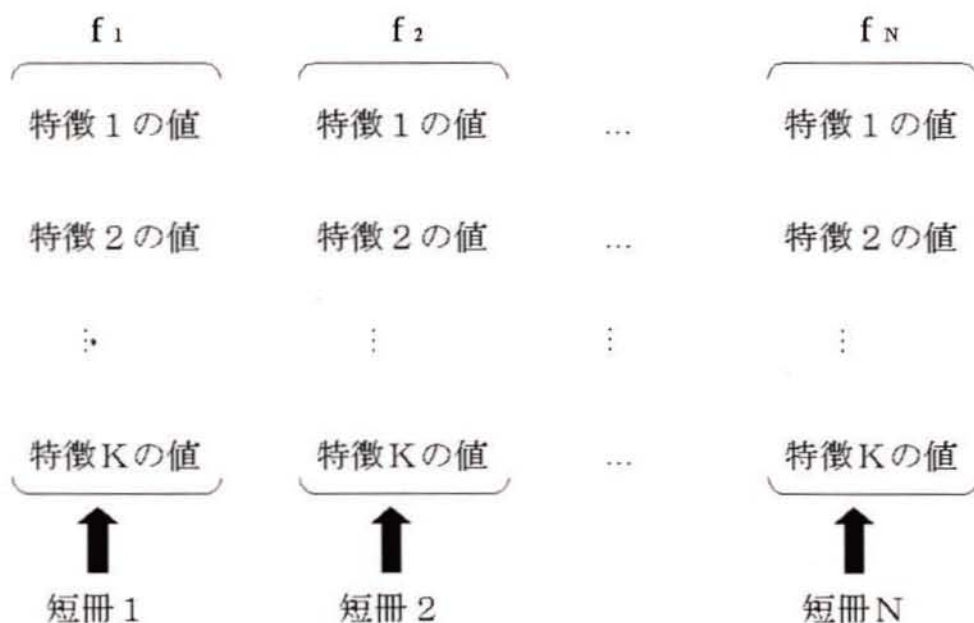


図 20：特徴ベクトル系列

この特徴ベクトルを1次元 HMM で利用するためには、ベクトル系列をシンボル系列に変換する必要がある。そのために1次元 HMM では、一般的に、ベクトル量子化と呼ばれる手法で、特徴ベクトルを一つのシンボルに変換することでシンボル系列を生成する。

ベクトル量子化では、あらかじめ作成したコードブックと呼ばれるベクトル集合と各特徴ベクトルを比較し、 $L_2$  距離が最小となるベクトルのインデックスによって特徴ベクトルを置き換える。

本研究でも上記の手法により2次元文字パターンを1次元シンボル系列に変換した。

**学習** 1次元 HMM の認識系では、ベクトル量子化によって作成したシンボル系列の出力確率が最大となるようにパラメータ推定を行う。通常は、各文字クラスごとにシンボル系列をまとめ、各クラスに対して一つの HMM を作成する。

**認識** 学習により構築した各文字クラスの HMM で、未知の文字パターンより生成したシンボル系列の出力確率を求める。ここで求められた HMM 出力確率（クラスの個数だけある）を比較し、最も高い確率を出力した HMM に対応する文字クラスを認識結果とする。

### 3. 3. 2 1次元 HMM を利用した文字認識の問題点

従来の1次元 HMM を利用した文字認識系の問題点は、方向性パターンマッチングやニューラルネットワークを利用した文字認識系と比較して、認識性能が劣ることである。

パターンを短冊化して独立に処理するため、2次元的な特徴がモデルに反映されにくいことが原因していると考えられる。



### 3. 4 2次元 HMM を利用した文字認識

文字パターンのように、2次元のパターンを HMM 認識系で扱う場合、3.2.2 節のような 1次元 HMM を利用したのでは、パターンの 2次元の特徴をモデルに反映することが難しい。そのため、HMM を拡張し、2次元の出力シンボル系列を扱えるようにしたものを 2次元 HMM と呼ぶ。

2次元 HMM を利用したオフライン文字認識の研究は現在盛んに進められているが、2次元 HMM にはいろいろなモデルがあり、認識系の構築手法は、まだ完全には確立していない。

本論文では、既に報告されている 2次元 HMM の一つである MRF-HMM を利用した文字認識手法[12]を、2次元 HMM を利用した文字認識系の代表として追試し、1次元 HMM と同一データで性能を比較した。MRF-HMM 文字認識系は、1次元 HMM 文字認識系とは異なり、各画素の値をシンボルとして入力し、文字認識系を構築する。本節では MRF-HMM について解説し、2次元 HMM 文字認識系の問題点を挙げる。

#### 3. 4. 1 MRF-HMM の解説

追試に利用した MRF-HMM についてその定義を説明する。

2次元パターンである文字パターンは、すべての要素が勝手に動けるのではなく、文字として成立するための周囲の画素に対する制約があると考え、その制約を確率モデルで表現しようとするのが MRF(Mesh Random Fields)である。このような状態表現を HMM に採用し、拡張したものが MRF-HMM である。

次節ではこの MRF-HMM について、2次元 MRF パターンを表現するのに適したモデルである MRF を HMM への入力として扱えるように HMM を拡張する方法について解説する。

### 3. 4. 2 MRF の解説

MRF-HMM は Belaid[12]により文字パターンを表現する確率モデルとして提案されたものである。本研究では提案した認識手法と比較するため、追試を行った。

$X = \{X_{ij} \mid (i, j) \in L\}$  を  $H \times W$  の格子  $L$  上で定義されたランダム場とする。ここで、 $\Omega_{ij} = \{(k, l) \in L \mid 1 \leq k < i \text{ または } 1 \leq l < j\}$  とし、 $\Theta_{ij} \subset \Omega_{ij}$  とする (図 21)。 $\Theta_{ij}$  を台 (Support) という。

$X$  は以下の式を満たすとき MRF という。

$$P(X_{ij} \mid X_{\Omega_{ij}}) = P(X_{ij} \mid X_{\Theta_{ij}}) \quad \forall (i, j) \in L$$

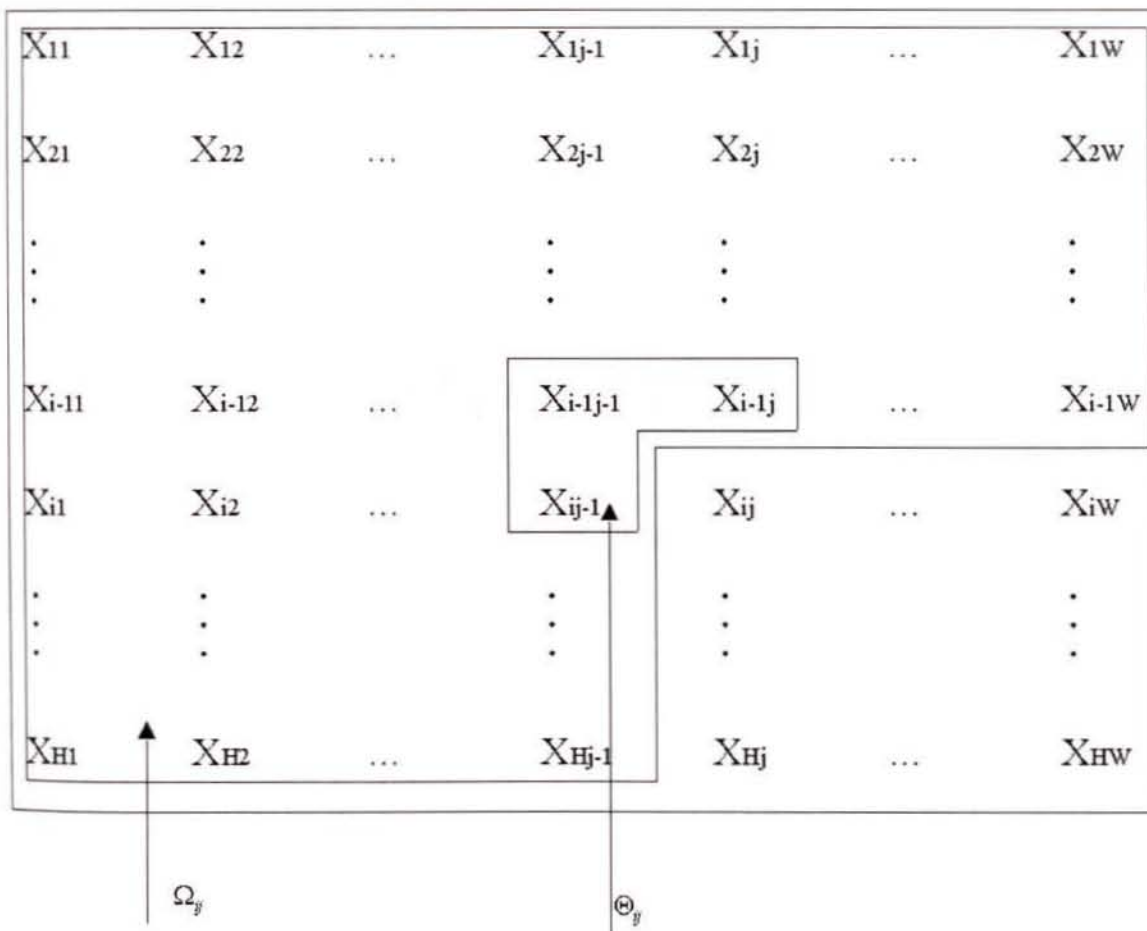


図 21 : 2次元座標と MRF-HMM

ここで、 $P(X_{ij} | X_{\Omega_{ij}})$  は部分集合  $A \subset X$  に対して  $P(X_{ij} | X_{kl}, (k, l) \in A)$  の意味である。MRF  $X$  の結合確率  $P(X)$  は以下の式で計算される。

$$P(X) = \prod_{j=1}^W \prod_{i=1}^H P(X_{ij} | X_{\Omega_{ij}}) = \prod_{j=1}^W \prod_{i=1}^H P(X_{ij} | X_{\Theta_{ij}})$$

MRF  $X = X_1 X_2 \dots X_W$  を、各列を一まとまりとした系列とみなす。状態遷移は1次のマルコフ連鎖であると仮定し、 $Q = q_1 q_2 \dots q_W$  を  $X$  の系列に対する状態遷移系列とする。モデル  $\mu$  に対する MRF  $X$  の尤度は

$$\begin{aligned} P(X | \mu) &= \sum_Q P(X, Q | \mu) \\ &= \sum_Q P(X | Q, \mu) P(Q, \mu) \end{aligned}$$

マルコフ連鎖の仮定より

$$P(Q | \mu) = \prod_{j=1}^W P(q_j | q_{j-1})$$

1次元 HMM と同様な非決定性を仮定して

$$\begin{aligned} P(X | Q, \mu) &= \prod_{j=1}^W P(X_j | X_{j-1}, \dots, X_1, q_j, \mu) \\ &= \prod_{j=1}^W P(X_j | X_{j-1}, q_j, \mu) \end{aligned}$$



MRF の行に対する独立性を仮定して

$$\begin{aligned} P(X_j | X_{j-1}, q_j, \mu) &= \prod_{i=1}^H P(X_{ij} | X_{\Omega_{ij}}, q_i, \mu) \\ &= \prod_{i=1}^H P(X_{ij} | X_{\Theta_{ij}}, q_i, \mu) \end{aligned}$$

以上から

$$P(X | \mu) = \sum_Q \prod_{j=1}^W P(q_j | q_{j-1}, \mu) \prod_{i=1}^H P(X_{ij} | X_{\Theta_{ij}}, q_j, \mu)$$

### 3. 4. 3 MRF-HMM の定義

図 22 に追試に利用した MRF-HMM を示す。MRF-HMM を決定するパラメータについて説明する[12]。

状態遷移確率の集合 1次元 HMM と同様で、 $N$  個の状態  $S = \{S_1, S_2, \dots, S_N\}$  を持ち、 $a_{ij}$  は状態  $S_i$  から  $S_j$  へ遷移する確率である。

$$\begin{aligned} A &= \{a_{ij} \mid 1 \leq i, j \leq N\} \\ \sum_{j=1}^N a_{ij} &= 1 \quad i = 1, 2, \dots, N \end{aligned}$$

シンボル出力確率の集合 シンボルの集合は  $V = \{0, 1\}$  の 2 値とする。近傍集合を以下のように定義する。

$$\begin{aligned} \Theta &= \{\Theta_{ij} \mid (i, j) \in L\} \\ \Theta_{ij} &= \{(i - i_k, j - j_k) \mid 1 \leq k \leq P, (i_k > 0 \text{ または } j_k > 0)\} \cap L \end{aligned}$$

ここで、 $P$  は近傍の次数である。シンボル出力確率は、第  $i$  行において、状態  $S_k$  で、近傍が  $\mathbf{x}$  であるとき、シンボル  $x$  が出力される条件付き確率である。

$$B = \{b_{ik}(x, \mathbf{x})\} \quad 1 \leq i \leq H,$$

$$b_{ik}(x, \mathbf{x}) = P(X_{ij} = x | X_{\Theta_{ij}} = \mathbf{x}, q_j = S_k) \quad 1 \leq k \leq N,$$

$$\sum_{x \in V} b_{ik}(x, \mathbf{x}) = 1 \quad \begin{array}{l} x \in V, \\ \mathbf{x} \in V^P \end{array}$$

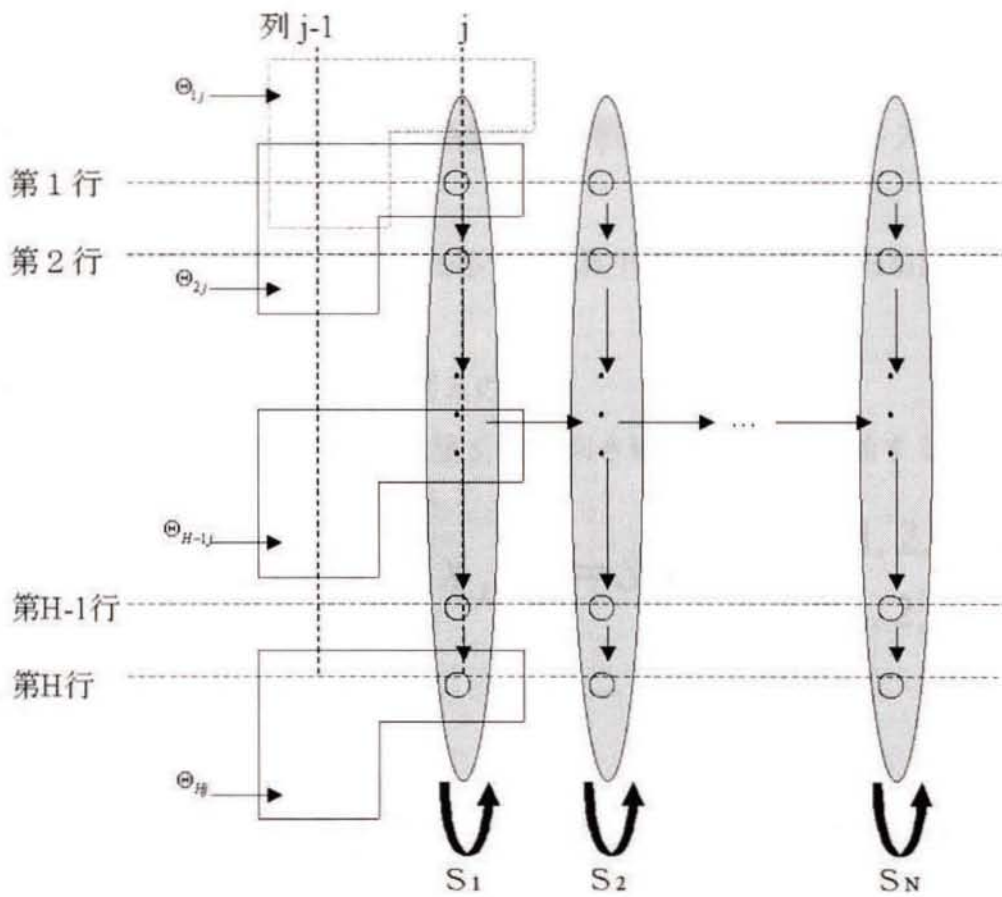


図 22 : MRF-HMM

初期状態確率の集合 1次元 HMM と同様で、 $\pi_i$  は初期状態が  $S_i$  である確率である。ここでも初期状態は  $S_1$  のみとする。

$$\pi = \{\pi_i\} \quad i = 1, 2, \dots, N$$

$$\pi_i = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases}$$

最終状態の集合 1次元 HMM と同様に、最終状態は  $S_N$  のみとする。

以上から、MRF-HMM のパラメータを  $\mu = (A, B, \pi)$  で表す。

### 3. 4. 4 MRF-HMM の確率計算と最適パスの算出

3.2.3 節で示した確率計算と最適パスの算出方法を、MRF-HMM にあわせて拡張する。

**Forward-Backward アルゴリズムの拡張**

前向き確率 列  $j$  における状態  $S_i$  の前向き確率  $\alpha(j, i)$  を定義する。

$$\alpha(j, i) \triangleq P(X_1 X_2 \dots X_j, q_j = S_i | \mu) \quad \begin{array}{l} j = 1, 2, \dots, W \\ i = 1, 2, \dots, N \end{array}$$

初期化 ( $j = 1$ )

$$\alpha(1, i) = \pi_i \prod_{k=1}^H b_{ki}(X_{k1}, X_{\Theta_{k1}}) \quad i = 1, 2, \dots, N$$

反復

$$\alpha(j, i) = \left[ \sum_{l=1}^N \alpha(j-1, l) a_{li} \right] \prod_{k=1}^H b_{kj}(X_{kj}, X_{\Theta_{kj}}) \quad \begin{array}{l} j = 2, 3, \dots, W \\ i = 1, 2, \dots, N \end{array}$$



終止

$$\sum_{S_i \in F} \alpha(W, i) = P(X_1 X_2 \dots X_W | \mu) = P(X | \mu)$$

後向き確率 前向き確率と双対な後向き確率  $\beta(j, i)$  を定義する。

$$\beta(j, i) \stackrel{\Delta}{=} P(X_j X_{j+1} \dots X_W, q_j = S_i | \mu) \quad \begin{array}{l} j = 1, 2, \dots, W \\ i = 1, 2, \dots, N \end{array}$$

初期化 ( $j = W$ )

$$\beta(W, i) = \begin{cases} 1 & \text{if } S_i \in F \\ 0 & \text{otherwise} \end{cases}$$

反復

$$\beta(j, i) = \sum_{l=1}^N a_{li} \beta(j+1, l) \prod_{k=1}^H b_{kl}(X_{kj+1}, X_{\Theta_{kj+1}}) \quad \begin{array}{l} j = W-1, \dots, 2, 1 \\ i = 1, 2, \dots, N \end{array}$$

$\alpha$  と  $\beta$  は以下の関係で結ばれる。

$$\beta(X | \mu) = \sum_{S_i \in F} \alpha(W, i) = \sum_{i=1}^N \beta(1, i) \pi_i$$

Viterbi のアルゴリズムの拡張

$$\delta(j, i) \stackrel{\Delta}{=} \max_{q_1 q_2 \dots q_{j-1}} P(q_1 q_2 \dots q_j = S_i, X_1 X_2 \dots X_j | \mu) \quad \begin{array}{l} j = 1, 2, \dots, W \\ i = 1, 2, \dots, N \end{array}$$

初期化 ( $j = 1$ )

$$\begin{cases} \delta(1, i) = \pi_i \prod_{k=1}^H b_{ki}(X_{k1}, X_{\Theta_{k1}}) & i = 1, 2, \dots, N \\ \varphi(1, i) = 0 \end{cases}$$

反復

$$\begin{cases} \delta(j, i) = \max_{1 \leq l \leq N} [\delta(j-1, l)] a_{li} \prod_{k=1}^H b_{ki}(X_{kj}, X_{\Theta_{kj}}) & j = 2, 3, \dots, W \\ \varphi(j, i) = \arg \max_{1 \leq l \leq N} [\delta(j-1, l)] a_{li} \prod_{k=1}^H b_{ki}(X_{kj}, X_{\Theta_{kj}}) & i = 1, 2, \dots, N \end{cases}$$

終止

$$\begin{cases} p^* = \max_{S_i \in F} \delta(W, i) \\ q_W^* = \arg \max_{S_i \in F} \delta(W, i) \end{cases}$$

パス (状態遷移系列) のバックトラッキング

$$q_j^* = \varphi(j+1, q_{j+1}^*) \quad j = W-1, W-2, \dots, 2, 1$$

$F = \{S_N\}$  なので、列  $j = W$  における最終状態  $S_N$  での尤度  $\delta(W, N)$  が MRF-HMM の受理確率  $P^*$  となる。状態遷移系列は、各  $\delta(W, N)$  に対して与える状態  $S_i$  を  $\varphi(j, i)$  に保存し、列  $j$  に対して逆トレースすることにより得られる。すなわち、 $q_1^* q_2^* \dots q_W^*$  が最適状態遷移系列となる

### 3. 4. 5 MRF-HMM のパラメータ推定

MRF-HMM  $\mu=(A, B, \pi)$  が与えられたとき、MRF  $X$  に対して

$$P(X | \bar{\mu}) \geq P(X | \mu)$$

となるようなパラメータ  $\bar{\mu}=(\bar{A}, \bar{B}, \bar{\pi})$  を有する MRF-HMM を推定することができる。

MRF の集合  $\mathbf{X}=\{X^{(1)}, X^{(2)}, \dots, X^{(R)}\}$  に対しては、

$$\prod_{r=1}^R P(X^{(r)} | \bar{\mu}) \geq \prod_{r=1}^R P(X^{(r)} | \mu)$$

となるようなパラメータ  $\bar{\mu}=(\bar{A}, \bar{B}, \bar{\pi})$  を有する MRF-HMM を推定することができる。

パラメータ推定アルゴリズム

MRF-HMM のパラメータ  $\mu=(A, B, \pi)$  に対して、 $A$  と  $\pi$  についてはパラメータ推定を行わない (パラメータの更新をしない)。出力確率  $b_{il}(x, \mathbf{x})$  の推定方法は、次式で示される再推定式を用いて出力確率の推定値  $\bar{b}_{il}(\bar{x}, \bar{\mathbf{x}})$  を計算する [12]。

$$\bar{b}_{il}(x, \mathbf{x}) = \begin{cases} \frac{\sum_{r=1}^R \frac{1}{P_r} \sum_{\substack{j=1 \text{ s.t.} \\ X_{ij}^{(r)}=x, X_{\theta_{ij}}^{(r)}=\mathbf{x}}}^{W_r} \alpha^{(r)}(j, l) \beta^{(r)}(j, l)}{\sum_{r=1}^R \frac{1}{P_r} \sum_{\substack{j=1 \text{ s.t.} \\ X_{\theta_{ij}}^{(r)}=\mathbf{x}}}^{W_r} \alpha^{(r)}(j, l) \beta^{(r)}(j, l)} & \begin{array}{l} i=1, 2, \dots, H, \\ l=1, 2, \dots, N, \\ x \in V \\ \mathbf{x} \in V^P \end{array} \\ b_{il}(x, \mathbf{x}) \text{ if } \sum_{r=1}^R \frac{1}{P_r} \sum_{\substack{j=1 \text{ s.t.} \\ X_{\theta_{ij}}^{(r)}=\mathbf{x}}}^{W_r} \alpha^{(r)}(j, l) \beta^{(r)}(j, l) = 0 & \end{cases}$$

ここで、 $P_r = P(X^{(r)} | \mu)$  である



### 3. 4. 6 2次元 HMM を利用した文字認識の問題点

2次元 HMM を利用した文字認識手法では、既に報告されている手法全てに共通しているのが、2次元のシンボルを扱えるようにHMMを拡張して利用しているということである。このため、拡張されたモデルでは HMM のパラメータ数が膨大化し、学習に必要となるサンプルの数の膨大、学習時間の長大化が問題となる。

しかし、オフライン文字認識では、大量のサンプルを収集することは困難であるため、学習に必要となるサンプル数の膨大化は大きな問題である。すなわち、2次元 HMM を利用したオフライン文字認識手法は、高い認識性能の可能性はあるものの、実装において効果的であるとはいえない。

## 3. 5 本研究の方針

3.3.2 節で述べたように1次元 HMM を利用したオフライン文字認識は認識性能の問題から、3.4.6 節で述べたように2次元 HMM は必要とする学習サンプルの量の問題から、双方とも実用的な手法として確立していない。

しかし、音声認識やオンライン文字認識で1次元 HMM の表現能力が高いことは実証されている。そこで本研究では、1次元 HMM を利用した認識系の改良、新しい認識系構築方法の提案、Bagging アルゴリズムとその改良アルゴリズムを HMM に適用し、それらの効果を検証した。

3.6 節では、1次元 HMM の特徴を改良し、より高い認識性能を実現する方法を提案する。

3.7 節では、パターンを複数の方向から短冊化、特徴抽出を行い、独立に1次元 HMM の学習をした各クラス複数の HMM を組み合わせて一つのクラスのモデルを構築する手法を提案する。

3.8 節では、ニューラルネットワークや決定木を利用した認識系など、認識系が不安定な場合に効果が確認されている Bagging アルゴリズム[18][19][20]を HMM に適用し効果を検証するとともに、HMM を利用した文字認識系で利用するためにさらに効果的な導入方法を提案する。

### 3. 6 1次元 HMM の特徴の改良

3.3.1 節で述べたように、一般的な1次元 HMM はパターンを短冊化し、短冊それぞれで独立に特徴ベクトルを求め、ベクトル量子化によりシンボル系列を生成して処理を行う。しかし、各短冊で独立に特徴ベクトルを求めたのでは、2次元的パターンである文字パターンで、その特徴をシンボルに反映することは難しい。

本研究では、2次元の特徴をシンボルに反映させるため、複数の短冊間にまたがる特徴を、従来の各短冊で独立に求めた特徴とあわせて利用する方法を提案する。具体的に採用した特徴の値は以下のようなものを採用した。

- [A] パターン全体の黒画素に対する、短冊内黒画素の割合
- [B] 短冊内の各画素について右隣短冊と値が異なるものの数
- [C] 短冊内上からはじめの黒画素で、右方向の黒画素連続長

実例で示すと、図 23左のパターンが与えられたとき、図 23右のように短冊化する。矢印で示した短冊について上記 [A,B,C] は以下のように求められる。

- [A]  $[\text{短冊内黒画素数 (2)}] \div [\text{パターン全体黒画素総数 (8)}]$
- [B] 右隣短冊との画素の関係は上から、[白・白]、[白・黒]、[黒・黒]、[黒・黒] となっており、右隣画素と値が異なる数は1
- [C] 短冊内上からはじめの黒画素は、上から3番目の画素。そこから右方向の黒画素連続長は3

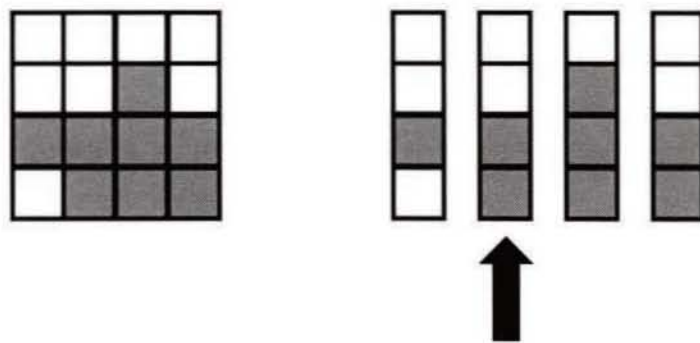


図 23：特徴の改良

このようにして複数短冊間にまたがる特徴を従来までの1次元HMM文字認識系の特徴ベクトルに追加した。ただし、1次元HMMと同一条件で比較するため、採用する特徴の数は同数とした。具体的な特徴を次に列挙する。各短冊で独立に求めた特徴は特徴[1-4]、複数短冊にまたがる特徴は特徴[5-8]であり、各値は上端を0とした画素の座標、および画素数である。

- [特徴1] 短冊内の上から一番目の黒画素連の始点
- [特徴2] 短冊内の上から二番目の黒画素連の始点
- [特徴3] 短冊内黒画素の上端-黒画素の下端
- [特徴4] 短冊内の黒画素連の個数
- [特徴5] [特徴1]の座標における横方向の黒画素数
- [特徴6] [特徴2]の座標における横方向の黒画素数
- [特徴7] 短冊内黒画素数÷パターン全体黒画素数×100
- [特徴8] 短冊内の各画素で右隣短冊と値が異なる数

この改良された特徴ベクトルを用いた1次元HMMを用いて、手書き英文字データを対象とする予備実験を行ったところ、5~10%程度の認識率向上がみられた。

文献で報告されている1次元HMM文字認識系それぞれでは、採用している特徴およびデータが異なるため、その全てと比較してはいないが、提案した特徴を採用することで、より高い性能が実現できたと考える。

以後の実験では、提案した特徴を採用した認識系を、1次元HMM文字認識系として利用する。



### 3. 7 多方向特徴抽出 HMM

3.6 節で示した 2 次元の特徴抽出を行うことで、従来の 1 次元 HMM を利用した認識系と比較し、認識率の向上が得られたが、まだ十分な認識性能とはいえない。

1 次元 HMM 文字認識系での認識結果を解析したところ、誤認識が多いのは、手書きブロック体大文字英字のデータを利用した場合、[C,O] 間、[B,P] 間などで、それらのパターン間に共通した特徴は、横方向の軸に黒画素を投影した場合の差が小さいという点であった。これは、2 次元の特徴を利用してはいるものの、まだ十分に 2 次元的な特徴をモデルに反映させることができていないことを示している。一方、横方向軸（短冊分割方向）に黒画素投影した際の差が大きなパターンに対して、誤認識をしている場合はほとんどなかった。

そこで本研究では、これら誤認識を低下させるため、複数の方向から短冊化、特徴抽出、ベクトル量子化を行ってモデルを構築し、複数のモデルを統合して一つの認識系を生成する手法を提案する。すなわち、2 次元のパターンを複数の方向から走査すれば、1 次元 HMM で誤認識がほとんどなかった状態、すなわち、短冊分割方向に黒画素投影した差が大きい状態で生成したモデルを利用することができ、認識結果を向上できるのではないかと考えた。以下、本論文ではこの手法を多方向特徴抽出 HMM (MD-HMM) と呼ぶ。

文字パターンを 1 次元化して認識する手法として古くから周辺分布法が知られているが、文献[25]では多方向の周辺分布を用いることで文字認識率を高め得ることを報告しており、提案手法はその近代化ともいえる。

MD-HMM について具体的に示す。原理的には、図 24 に示すように、角度  $\theta$  の方向から短冊化特徴抽出を行うが、これはパターンを図 25 のように回転させ、角度  $\theta$  で特徴抽出することと等価である。

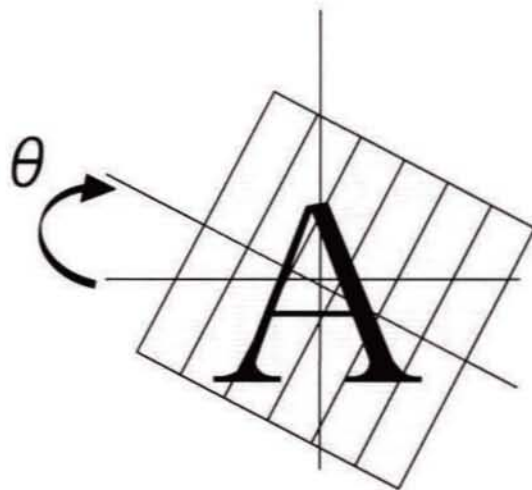


図 24：特徴抽出方向の回転

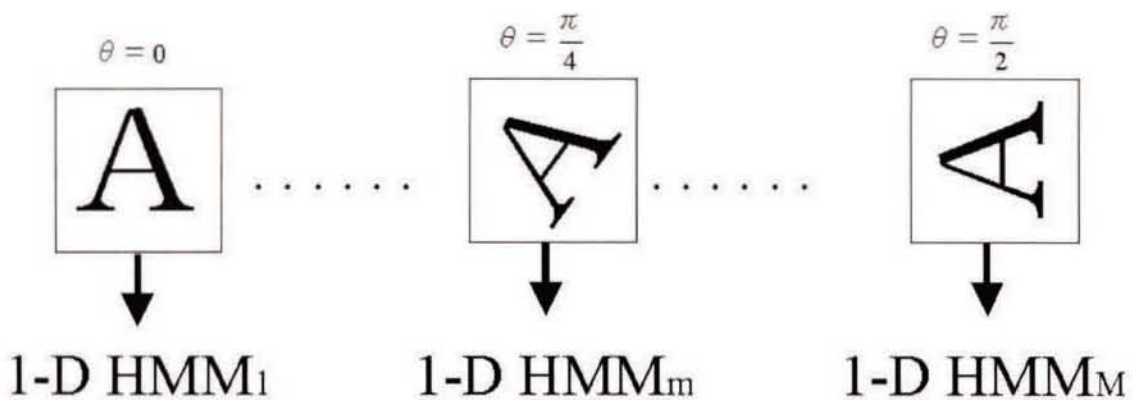


図 25：MD-HMM の構成

次に図 25で示すように、それぞれの回転パターンに対して独立に 1次元 HMM を生成する。ここで、特徴抽出方向数がMであるとすると、それぞれの 1次元 HMM 出力確率は次式で表される。

$$P_m = \max \left( \prod_{k=1}^K P(O^{(k)} | \lambda) \right)$$

これらの1次元 HMM 出力確率を定数パラメータ  $\alpha$  を用いて、線形荷重和を求めたものを MD-HMM 出力確率と定義する。(数学的には、荷重和は確率でないが、便宜上確率と呼ぶ。)

$$0 \leq \alpha_m \leq 1 \quad \sum_{m=1}^M \alpha_m = 1$$

$$P = \sum_{m=1}^M (\max(\prod_{k=1}^K P_m(O^{(k)} | \lambda)) \times \alpha_m)$$

## 3. 8 Bagging 法を利用した Voting 法の適用

本節では、ニューラルネットワークや決定木において有効性が確認されている Bagging 法[18][19][20]を、HMM を利用した認識系に適用できるのではないかとこの観点から考察する。さらに、HMM 文字認識系で利用する場合、認識系の特徴から、Voting 法を改良して利用することも可能であるため、改良 Voting 法の効果についても検討した。

3.8.1 節では、Bagging 法の定義を紹介する。

3.8.2 節では、HMM 認識系で Bagging 法をどのように適用するかを示す。

3.8.3 節では、HMM 文字認識系の特徴を生かすために改良した Voting 法を提案する。

### 3. 8. 1 Bagging 法の解説

Bagging 法とは次のような手法である。

ある識別モデルにおいて、その学習データの集合から重複を許容しランダムにデータを再サンプリングして作成したN個の学習セットを利用し、N個のモデルを作成する。

認識は、そのN個のモデルの結果を多数決したもの（最多得票）をその識別モデルの識別結果とする。

具体的には、Bagging 法を使わない認識モデルの構築方法では図 26左に示すように、学習データから一つの認識モデルを生成するのに対し、Bagging 法では学習データを再サンプリングによりN個生成し、N個の認識モデル構築を行う。

認識では、ここで作成したN個の認識モデルそれぞれで認識結果を求めるが、その結果はN個のモデルの投票に利用され、最多得票が認識結果となる。すな



わち、Bagging 法では、認識モデルは図 26に示す部分となる。

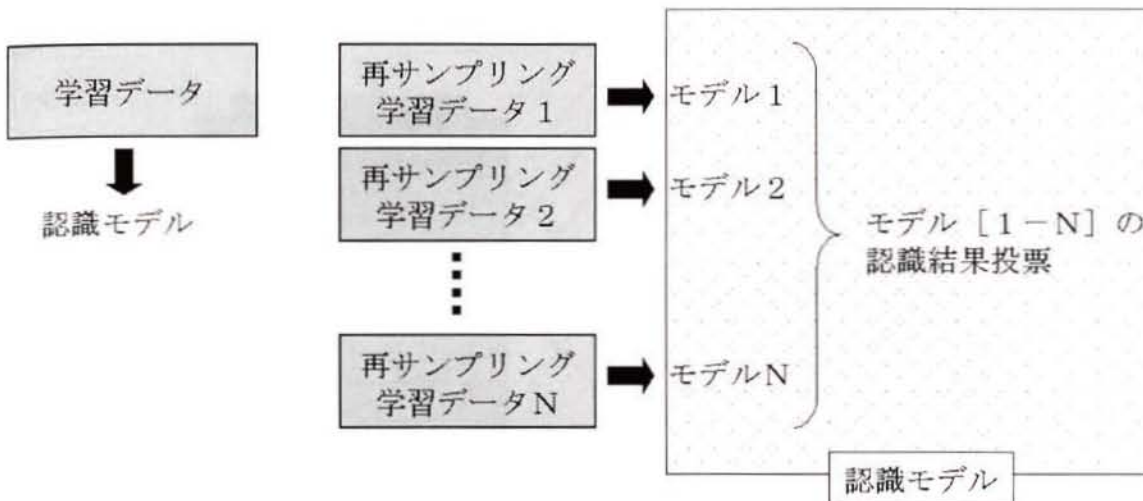


図 26 : Bagging 原理図

### 3. 8. 2 Bagging 法の HMM への適用

HMM の認識系には次のような特徴がある。

- ・一つのクラスに対して一つのモデルをもつ
- ・認識時に各モデルはモデルの適合確率を返す

すなわち、ニューラルネットワークや決定木の認識モデルは、学習サンプル集合に対して一つのモデルを構築するのが一般的であるが、HMM では学習サンプル集合内の一つのクラスに属するサンプル集合に対して一つのモデルを構築し、各モデルがモデル適合確率を返す。具体的には、英字 26 クラスの HMM の認識系では、26 のモデルを構築し、それぞれのモデルで返された適合確率を比較することで認識結果を得るのである。

このような認識系の構成をするため、通常は図 26 のように再サンプリングするが、HMM で Bagging 法を適用する際には、図 27 のように、各クラスごとに再サンプリングを行いモデルを構築する。

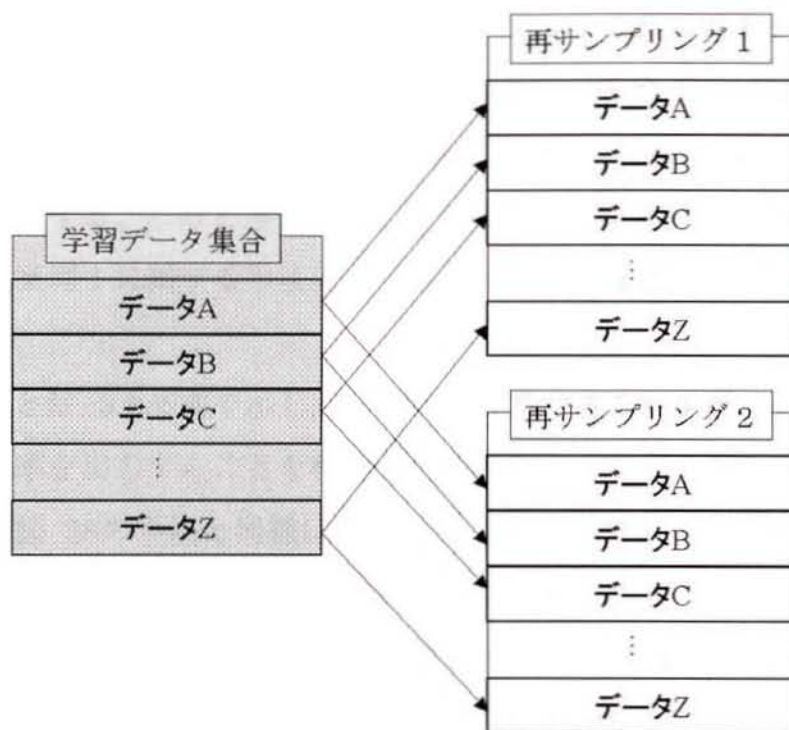


図 27：HMM における Bagging 法の適用

### 3. 8. 3 HMM 認識系の特性にあわせた Voting 法の拡張

文字認識系でよく用いられる評価は、第K番目までの候補の中に正解が含まれる確率、すなわちK位累積正解率である。K位累積正解率で認識系を評価するのは、文字認識を実際に利用する場合に、文脈を用いた知識処理によりK位の候補内から適当なものを選択するという手法が広く使われているためである。しかし、3.9.2 節で示した厳密な Bagging 法は、1位正解率を高める手法である。そこで、累積正解率を向上させるために次のような拡張原理を提案する。拡張原理は2種類あり、それぞれ拡張原理1、拡張原理2と呼ぶ。

拡張原理1とは次のようなものである。

図 28で、学習サンプルセット[A-Z]をランダム再サンプリングし、N倍の学習サンプルセット[A0-Z0][A1-Z0]...[AN-ZN]を生成し、それぞれに学習を行い

[A0-Z0][A1-Z0]...[AN-ZN]モデルを生成する。次に、これらのモデルそれぞれを図 28のように組み合わせた認識系をN個構築し、それぞれにK個の候補を出力する。最後に、各認識系からのK個の候補を順位に基づき投票することで、最終結果を出力する。このとき、 $K=1$ ならば、それは厳密な Bagging 法を意味する。拡張原理 1 に属する手法として具体的には、後述する手法 1、手法 2 を検証した。

拡張原理 2 とは、3.8.2 節で示した HMM 認識系の各モデルが、そのモデルに対する適合確率を出力することを利用した、次のようなものである。

はじめに、図 28の手法と同様にして、ランダム再サンプリングを利用してモデルを N 倍生成する。次に、図 29のように、各クラス N 個のモデルの出力確率をある関数  $f$  で統合し、各クラス一つの出力が得られるモデルを構築する。このように作成したモデルの出力を比較し認識結果を得る。

拡張原理 2 に属する手法として具体的には、次の手法 3 を検証した。

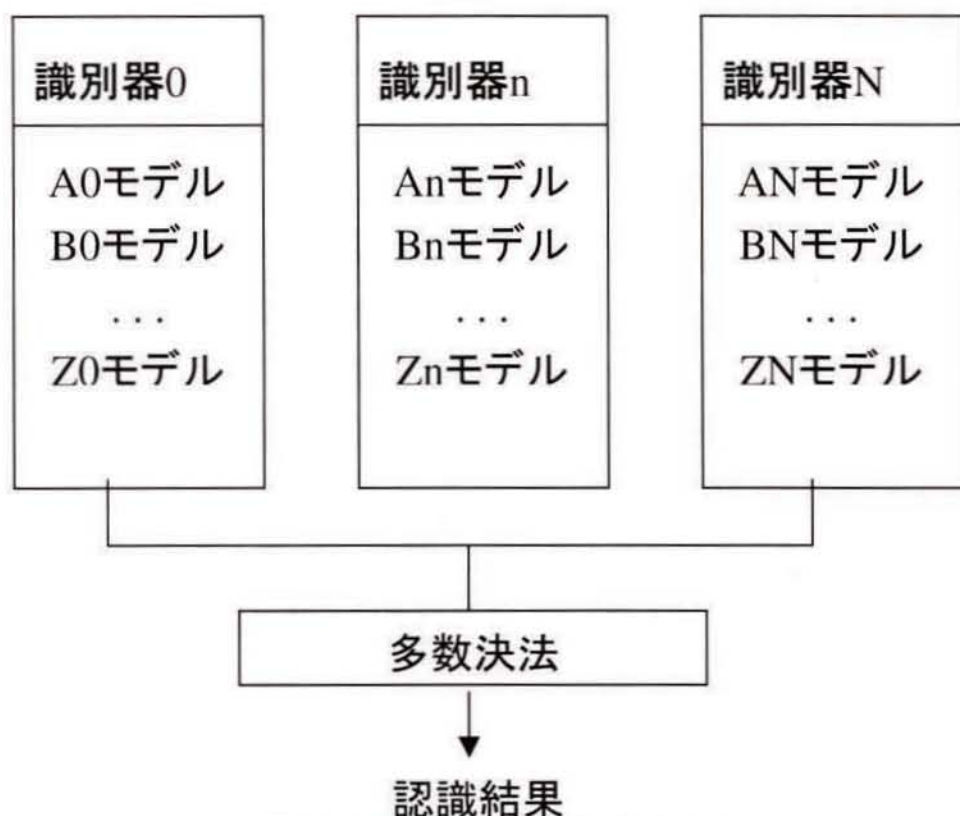


図 28 : Bagging 法の拡張原理 1

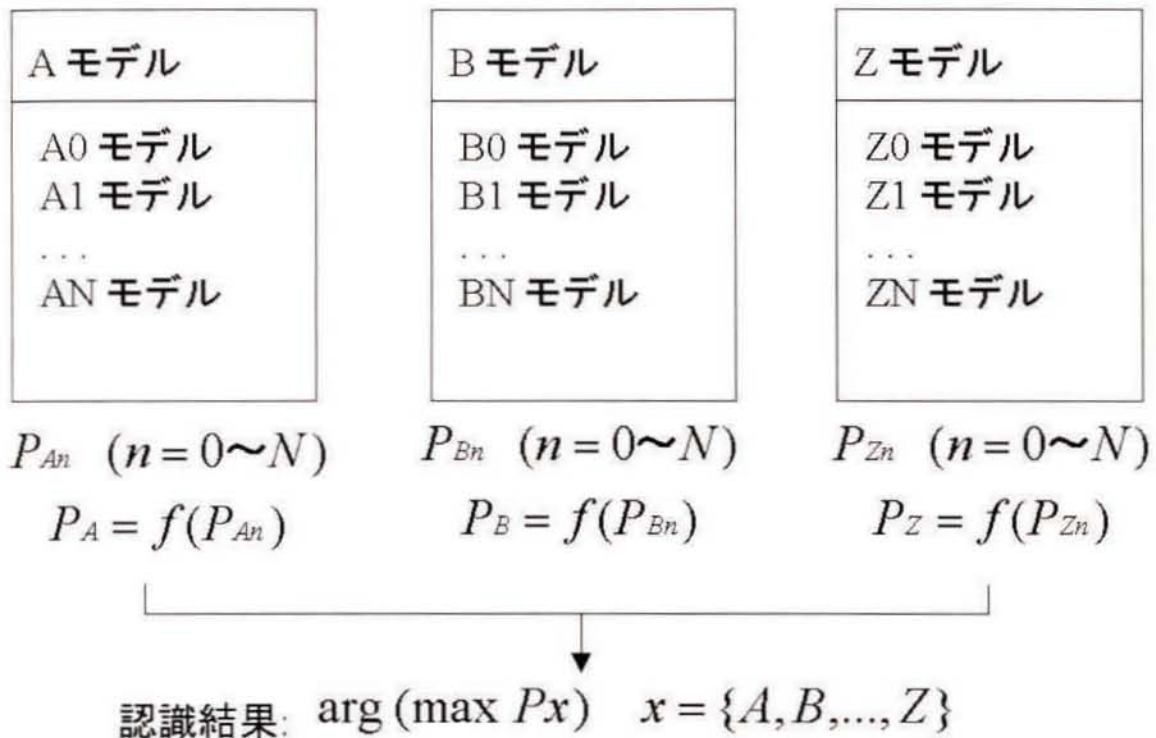


図 29 : Bagging 法の拡張原理 2

#### 実際に検証した手法

[手法 1] 文字クラスごとに複数のモデルを一つずつ採用した識別木を複数作成し、1位の候補の投票数を比較し認識結果とする。厳密な Bagging 法。

[手法 2] 文字クラスごとに複数のモデルを一つずつ採用した識別木を複数作成し、1位の候補を 2 票、2位の候補を 1 票として投票し、投票数を比較し認識結果とする。

[手法 3] 文字クラスごとに、複数のモデルの出力確率上位 3 つを平均し、各クラスでその平均値を比較し認識結果とする。



## 第4章 評価実験

第2章、第3章で提案した各提案手法について行った実験結果および評価を示す。

4.1 節では、提案した文字の切り出し手法についての実験結果、評価方法、評価を示す。

4.2 節では、HMM を利用した文字認識系について、より効果的な実装を目標とした提案手法それぞれについて、従来法と比較した実験結果、およびそれら手法の評価を示す。

## 4. 1 文字切り出し実験

第2章で提案した手法について実験、評価を行った。その実験方法、評価方法、実験結果、評価について本節で示す。

### 4. 1. 1 文字切り出し実験対象

日本語郵便書状宛名書き文字パターンのデータベースは郵政研究所で作成されたものが配布されているが、接触文字がどの程度含まれているか明らかではなく、本研究で使用するには適当でないと考えた。

そこで下記のデータベースを独自に構築し実験に利用した。

#### データベース

実験用宛名書きデータは、10人の筆記者に4回ずつ、ボールペン・サインペンでそれぞれ100住所（都道府県から番地まで）を縦書き、横書きに筆記させたもの8,000画像（ $10 \times 4 \times 2 \times 100$ ）である。特に文字を接触させて書くように指示はせず、文字記入枠を3cm×15cmと小さくすることにより自然に接触が生じるようにした。

画像は200dpiの分解能で2値画像として走査した。採取した画像から接触文字パターンを一個所以上含む領域2487画像を抽出し実験用データベースとした。すべての画像は、日本語の知識のある人間であれば正しく切り出し可能と思われる。一つの文字の大きさは、200dpiで平均60×60画素である。

## 4. 1. 2 文字切り出し結果の評価システム

人間の目視による結果の評価を 4.1.3 節で示すが、定量的な評価方法の一つとして、下記に示す切り出し箇所の正解データベースを利用して、処理によって求めた切り出し箇所の自動評価を行った。

正解データの一つの領域と切断線の一部が一致すれば正解、そうでなければ不正解として返すものとした。具体的には、図 30で、a, b, c の 3 本の切断線が求められたとき、適切と判断されるのは b, c である。

このようにして作成した評価システムによって、検出した切断線の最適性は検査できないが、妥当性は検出でできると考えられる。

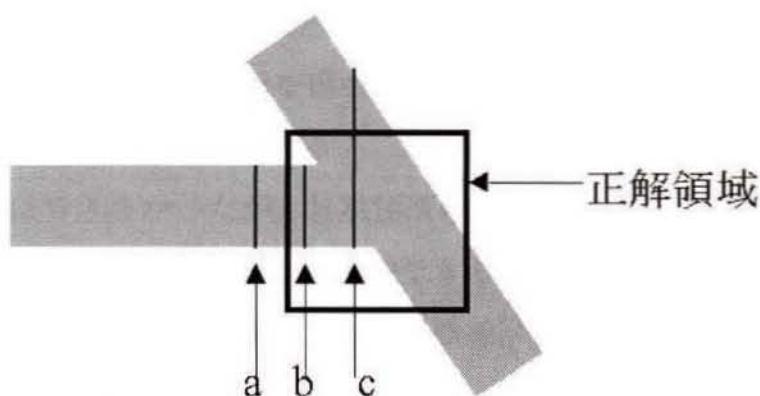


図 30：評価システムの正解・不正解

### 正解データベース

従来の研究では、ある文字切り出しアルゴリズムを提案し、少数のサンプルについて処理結果を紹介するものや、認識部と組み合わせて、切断による認識性能の向上を紹介するものがほとんどであったが、それでは文字切り出しアルゴリズムの妥当性の評価として適当でないと考え、複数の人間の目による正解データの付与を行い、それらを重ねあわせて正解データベースを作成して評価システムで利用した。

### 4. 1. 3 文字切り出し実験結果

表1に、検出切断線の絞り込みの影響とこの評価システムによる結果を示す。表1から解るように、検出切断線の上限を与えないときの評価システムでの正解率は64.3%の性能を示している。したがって、この評価が正しければ、提案した切り出し手法はかなり性能が高いと思われる。

しかし、評価システムが正しく働いている保証はない。そこで、評価システムの妥当性を示すため、表2に被験者による切り出し画像の評価を行い、結果を比較した。

表2は、切断線絞り込み処理は導入していない切り出し結果の画像を、被験者を使い次の判断基準で評価させた結果である。

- [A] 接触パターンを適切な箇所で切断している
- [B] 接触パターンを、適切な箇所ではないが、各文字が正しく認識できるような箇所で切断している
- [C] 切断はできなかったが、接触箇所が隣接文字の小さな点などであるため、原パターンのままでも認識できる
- [D] 不必要な箇所での切断、または切断線が検出できなかった

この被験者による評価基準で、[A, B, C]は切り出した結果を認識部に渡した場合に、良好な結果が得られると考えられるため、提案手法による切り出しが妥当であったと判断し、[D]は失敗と判断した。

切断線本数の上限がないとき、評価システムでの正解率は64.3%で、これは被験者による評価の[A]に極めて近く、多少の出入はあるものの、両者はほぼ対応していると考えられる。これより、評価システムによる評価は妥当であると考えられる。

上記は、提案手法による全ての切断線候補を出力した結果であるが、余剰切断線を多数含むため、認識に要する時間が非常に長くなる。この問題を解決するため、2.5.6節で述べた絞り込みの結果も表1に示す。

この結果、切断線の本数に上限を設けることにより、大きく絞り込めること



が解る。しかし、余剰切断線の絞り込みは、正解率低減のコストを伴う。表1から切断線本数の上限が9のとき、検出切断線の61.8% (100-38.2)の削減を正解率低下12.8%の犠牲の下で実現していることがわかる。したがって、効果的に余剰切断線の絞り込みに成功したと考えられる。

表2から解るように、被験者による判断では、ばらつきはあるものの妥当な切断線 [A, B, C] の割合は約80%であった。

表1：検出切断線の絞り込みと評価システムの評価

パターン内切断線 本数の上限	切断線の絞り込み率	評価システムでの 正解率
3	16.3%	25.1%
5	25.8%	40.9%
7	33.1%	48.0%
9	38.2%	51.9%
∞	100.0%	64.3%

表2：被験者による処理結果の評価

	被験者1	被験者2	被験者3	被験者4
A	66.4%	63.0%	60.3%	53.0%
B	13.3% (79.7%)	13.2% (76.2%)	19.8% (80.1%)	20.6% (73.6%)
C	4.1% (83.8%)	7.5% (83.6%)	4.0% (84.1%)	5.1% (78.7%)
D	16.2%	16.4%	15.9%	21.3%

カッコ内は正解とする範囲をB, Cまで拡大した場合の累積率

#### 4. 1. 4 文字切り出し手法の評価

2.1 節でも述べたように本研究は（株）日立製作所中央研究所との共同研究で行われたものである。提案した切り出し手法の一部は同所に移管され、多少の改良を経て、同所で開発中の郵便宛名読取区分機に組み込まれ、実用化されている。

企業秘密の関係で、具体的な数値は公表されていないが、担当者の直話によれば、実際の局で実験を行ったところ、処理速度を低下させることなく書状単位での通過率（書状がリジェクトされなかった率、ごく少数の誤認識を含む）の向上が見られたそうである。

なお、本研究で述べた内容は、同所に移管した手法をさらに改良し、独自のデータベースを採取して実験したものである。

以上から、提案した文字切り出し手法により多くの手書き日本語接触パターンを、余剰切り出しを抑制しながら、認識可能なパターンに切り出すことに成功したと考える。

## 4. 2 HMM を利用した文字認識の実験

第2章で提案した手法について実験、評価を行った。その実験方法、実験結果、評価方法、評価について本節で示す。

### 4. 2. 1 各手法の学習効率の比較

本研究では、HMM 文字認識系の認識率の向上とともに、実装に向けた効率的な学習方法の実現も目標としている。そこで本節では、提案した各手法で学習をする際のパラメータ数と学習効率について説明する。

1次元 HMM のシンボル数 :  $S$

サンプル画像の画像サイズ :  $W \times H$

とすると、従来手法は

1次元 HMM のパラメータ数 :  $O(S)$

2次元 HMM のパラメータ数 :  $O(W \times H)$

で与えられるのに対して、提案手法は特徴抽出方向数が  $M$  のとき

MD-HMM のパラメータ数 :  $O(M \times S)$

で表せる。

ここで、通常  $M \ll W, H$   $S \leq W \times H$  であるので、提案手法である MD-HMM のパラメータ数は2次元 HMM のパラメータ数よりはるかに小さい。

またパラメータ推定の効率を考えると、2次元 HMM はそのパラメータを一度に決定する学習を必要とするのに対し、特徴抽出方向数が  $M$  の MD-HMM では、1次元 HMM のパラメータ数である  $O(S)$  のパラメータ推定を  $M$  回行う。さらに、 $N$  個の再サンプリング学習セットを利用する Bagging 法を利用した HMM では、1次元 HMM のパラメータ数である  $O(S)$  のパラメータ推定を  $N$  回行う。

これより、本論文で提案した2つの手法、MD-HMM と Bagging 法の応用、

は1次元 HMM のパラメータ推定と同程度の計算をM回、N回行うことで実現でき、それぞれのパラメータ推定は独立に実行可能であるため、効率的な学習が実現できる手法であるといえる。

## 4. 2. 2 文字認識実験対象

本研究で提案した HMM を利用した文字認識系を構築し、評価実験を行った。評価実験は、ETL6 [21]の常用手書き英文字データベース、英文字 [A-Z] を利用した。

常用手書き文字とは、OCR にとって認識しやすく、通常の間人が日常書いているものに近い字形を推奨字形として提示し、筆記者がその字形を見て記入したものである。ETL6 は、そのような文字サンプルを多数集めたもので、文字認識の標準的なデータベースとして世界に広く使われている。

前処理として、2 値化、ノイズ除去、大きさの正規化を行い、その段階でノイズが多いものは目視により実験データから除外した。

各クラスのデータ数は約 1,300 で、そのうち各 600 を学習用サンプルセット、残りを検査用サンプルセットとして利用した。検査用サンプルセットの総数は 18,636 である。

また、Bagging 法やその応用手法では、学習セットからのランダム再サンプリングにより各クラス 600 の学習サンプルセットを 20 セット用意した。組み合わせは、与えられた学習セットを含む 21 セットで行った。



## 4. 2. 3 1次元 HMM の特徴改良の認識実験結果

3.6 節で提案した 1 次元 HMM の特徴の一部を 2 次元の特徴で置き換えたときの認識結果を表 3 に示す。

表 3 で、1 次元の特徴抽出 HMM[13]は、3.3 節で示したように、各短冊独立で特徴ベクトルを求めたもの、2 次元の特徴抽出 HMM は 3.4 節で示したように複数短冊にまたがる特徴の値を追加したものである。同一条件での比較とするため、双方の特徴数が等しくなるようにした。すなわち、追加した 2 次元の特徴量の個数だけ 1 次元特徴量の個数を減少させている。用いた特徴は 3.3 節と 3.6 節を参照されたい。

表 3：1 次元 HMM の特徴改良による認識結果

認識手法	第N位 (累積) 正解率		
	1	2	3
1 次元の特徴抽出 HMM	86.2%	94.3%	96.2%
2 次元の特徴抽出 HMM	94.3%	96.2%	96.6%

表 3 から解るように、2 次元の特徴の採用により、約 8 % の認識率向上が得られ、提案した特徴が効果的であったことが分かる。

#### 4. 2. 4 多方向特徴抽出 HMM の認識実験結果

3.7 節で提案した多方向特徴抽出 HMM の認識結果を表 4 に示す。表 4 中 1 次元 HMM は、3.6 節の 2 次元特徴の値を採用した HMM の結果で、2 次元 HMM は 3.4 節で示した MRF-HMM[12]の追試結果である。MRF-HMM では学習サンプルの数が不足していたため、独自に採取したデータも加えて学習を行っている。それぞれの手法で統一した評価を行うため、それぞれ学習データに対して 1 位正解率 99%以上となるように学習を行っている。

表 4 : MD-HMM の認識結果

認識手法	第N位 (累積) 正解率		
	1	2	3
1次元HMM	94.3%	96.2%	96.6%
2次元HMM	96.5%	98.6%	99.1%
MD-HMM	95.0%	98.3%	99.4%

表 4 から解るように、提案する MD-HMM の認識性能は、1 次元 HMM より高い。さらに、1 位正解率では 2 次元 HMM に劣るが、3 位累積正解率では MD-HMM のほうが 2 次元 HMM よりも高い認識性能を示した。

## 4. 2. 5 Bagging 法と応用した Voting 法の適用実験結果

3.8 節で示した Bagging 法、および HMM での利用に拡張した Voting 法を、1次元 HMM に適用した結果を表5に、MD-HMM に適用した結果を表6にまとめる。

表5, 6中の手法 [1, 2, 3] は3.8.3節で示したように、次の3つを表す。  
[手法1] 文字クラスごとに複数のモデルを一つずつ採用した識別木を複数作成し、1位の候補の投票数を比較し認識結果とする。厳密な Bagging 法。

[手法2] 文字クラスごとに複数のモデルを一つずつ採用した識別木を複数作成し、1位の候補を2票、2位の候補を1票として投票し、投票数を比較し認識結果とする。

[手法3] 文字クラスごとに、複数のモデルの出力確率上位3つを平均し、各クラスでその平均値を比較し認識結果とする。

表5：1次元 HMM への Bagging 法に基づく Voting 法の適用

認識手法	第N位 (累積) 正解率		
	1	2	3
1次元HMM	94.3%	96.2%	96.6%
手法1	94.9%	96.6%	96.9%
手法2	94.6%	96.8%	97.4%
手法3	95.0%	96.9%	97.4%

表6：MD-HMM への Bagging 法に基づく Voting 法の適用

認識手法	第N位（累積）正解率		
	1	2	3
MD-HMM	95.0%	98.3%	99.4%
手法1	96.2%	98.1%	98.5%
手法2	95.3%	98.3%	99.3%
手法3	96.2%	98.8%	99.5%

表5, 6から、手法1の Bagging 法は HMM を利用した文字認識系でも良好な働きをすることを示した。

さらに、本論文で HMM 文字認識の適用にあわせて応用した手法2, 3のほうが厳密な Bagging 法よりも高い性能を発揮することを示した。

#### 4. 2. 6 提案認識手法の評価

4.2.5 節で述べたように、MD-HMM と Bagging 法を組み合わせることにより96%の1位認識率が得られた。やや古い数値であるが、文献[26]によれば常用手書き英字の認識率は97%とされているので、ほぼ実用レベルに近い値が得られているといえる。さらに、実際の商用機では、強制判定ではなく、ある手法では類似していて区別できないパターン（例えばHとMなど）の場合、一旦近接リジェクトとし、二つの候補の差を拡大して判定する処理（二段判定：候補文字のペアごとに別々に用意する）を採用しているのが普通である[27]。

表6の結果では、提案手法での第2位累積正解率が約99%に達している。このことから、1位で誤認識になっているパターンのお大半では、正解クラスの出力確率は、1位候補の出力確率にごく近いものと予想される。したがって、



仮に 1 位候補と 2 位候補の出力確率が近いときに、近接リジェクトとして二段判定で確定するようにすれば、その多くは正解にできると考えられ、実用機に近い性能が発揮できる可能性がある。

## 第5章 結論

本研究ではオフライン文字認識の実用に向けた課題を解決する手法として、接触文字パターンの切り出し手法と Hidden Markov Model を利用した文字認識系の認識性能の向上と効果的な学習を実現する手法について提案し、検証を行った。

文字の切り出し法としては、直線成分を検出し、マクロ的視点で作成した簡略化パターンを利用することで文字の接触箇所を検出して切断する手法を提案した。接触文字を含む日本語住所の 2, 4 7 8 画像を用いた実験では、約 80% の接触パターンに対して認識可能なパターンへの切り出しに成功した。さらに、認識部の速度が十分でない場合に対処するため、検出した切断線に得点付けして絞り込む手法を提案し、正解率低減 12.4% のコストの下で、余剰検出切断線を 61.8% 削減することに成功した。これにより、本研究で提案した手法を導入することで、従来まで認識が困難であった接触文字を含む日本語住所の認識率を大幅に向上できるものと考えられる。

新しい文字認識系の提案として HMM を検討した。

認識率が不十分であることが問題である 1次元 HMM において、2次元的な特徴を採用することで、認識率が向上することを実験により示した。

また、認識率・学習効率の向上を実現するため、MD-HMM を利用した文字認識系の構築方法を提案した。MD-HMM は、2次元 HMM とほぼ同等の認識性能を、少ないパラメータを利用した効率的な学習で実現できることを理論的および実験的に示した。

また、ニューラルネットワークや決定木で有効性が確認されている Bagging 法を HMM 文字認識系に適用し、認識率向上に効果があることを実験的に検証した。さらに、Bagging 法を HMM 文字認識系に対して、より効果的に導入する拡張原理を提案し、その認識率向上効果を実証した。

具体的には、1次元 HMM に 2次元の特徴を採用することで、1位認識率で

8.1% (94.3 - 86.2) の向上を実証した。また、MD-HMM の利用により、認識性能をさらに向上できることを実証した。さらに、Bagging 法の HMM (MD-HMM) 文字認識系に適用し、拡張した Voting 法で 1 位正解率を 94.3% から 95.0% (95.0% から 96.2%) と向上することに成功した。

以上から、本研究で提案した二つの手法の導入により、オフライン文字認識が現在よりはるかに広い分野で実用化されるものと考ええる。

## 謝辞

本論文をまとめるまでに、多くの方々の御指導、御協力を戴いた。

信州大学教授中野康明博士には多大なる御指導、御鞭撻を戴いた。ここに謹んで謝意を表す。本研究テーマを成功裡に終結し、学位論文としてまとめることができたのは氏の適切な御指導に負うところが大きい。

信州大学助教授丸山稔博士、助手宮尾秀俊博士、(株)日立製作所中央研究所 主管研究員藤澤浩道博士、主任研究員酒匂裕博士、池田尚司博士の諸氏には本研究を遂行するにあたり有益な御指導、御助言を戴いた。深く感謝する。

本論文取りまとめにあたりご指導戴いた、信州大学教授師玉康成博士、同岡本正行博士にもここに深く感謝する。

また、データの採取、評価実験にご協力戴いた信州大学中野・丸山研究室諸氏にもここに深く感謝する。



## 参考文献

- [1] S. Rice, J. Kanai and T. Nartker, "A report on the accuracy of OCR devices," ISRI Technical Report, pp. 1-6, Information Science Research Institute, University of Nevada, Las Vegas, 1992
- [2] H. Fujisawa, Y. Nakano and K. Kurino, "Segmentation methods for character recognition: for segmentation to document structure analysis," Proc. of IEEE, Vol. 80 [7], pp. 1079-1092, July, 1992
- [3] H. Yamada, and Y. Nakano, "Cursive handwritten word recognition using multiple segmentation determined by contour analysis," IEICE Trans. on INF&SYSTEM, Vol. E79-D [5], May, 1996
- [4] A. Filatov, V. Nikitin, A. Volgunin and P. Zelinsky, "The AddressScript™ recognition system for handwritten envelopes," Proceedings of Third IAPR Workshop on Document Analysis Systems, pp. 222-236, November, 1998
- [5] S. Srihari, "High-Performance reading machines," Proc. of IEEE, Vol.80 [7], pp. 1120-1132, 1992
- [6] 大井 仁, "郵便物あて名自動読取区分機の文字識別と知識処理," 1990年信学会秋季全大, D-354, 1990
- [7] H. Ino, K. Saruta, N. Kato and Y. Nemoto, "Handwritten address segmentation algorithm based on stroke information," ISSN, Vol. 38 [2], Feb, 1997
- [8] 中島 正臣, 米倉 雄司, "平滑化周辺分布と判別分析を用いた手書き文字切り出し方式," 電子情報通信学会論文誌, Vol. J78-D II [7], pp. 1039-1046, Jul, 1997
- [9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, Vol. 77 [2], pp. 257-

286,1989

- [10] 中川 聖一, “確率モデルによる音声認識,” 電子情報通信学会, 1998
- [11] J. J. Lee, J. H. Kim and M. Nakajima, “A hierarchical HMM network-based approach for on-line recognition of multiple cursive handwritings,” *IEICE Trans. INF&SYSTEM*, Vol. E81-D [8], pp. 881-888, 1998
- [12] G. Saon and A. Belaid, “Off-line handwritten word recognition using a mixed HMM-MRF approach,” *Proc. ICDAR'97*, Vol. 1, pp. 118-122, 1997
- [13] 小林 誠, “HMM を用いた手書き文字認識の研究,” 信州大学大学院工学系研究科情報工学専攻修士論文, 1999
- [14] C. Bose, and S. S. Kuo, “Connected and degraded text recognition using hidden Markov model”, *Pattern Recognition* 27 [10], 1345-286, 1989
- [15] A. Kundu and P. Bahl, “Recognition of handwritten script: a hidden Markov model based approach,” *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, New York City, USA, pp. 928-931, April, 1988
- [16] M. Y. Chen, A. Kundu and J. Zhou, “Off-line handwritten word recognition using a hidden Markov model type stochastic network,” *IEEE Trans. Pattern Anal. Mach. Intell* 15 [5], pp. 481-496, 1994
- [17] O. E. Agazzi and S. S. Kuo, “Hidden Markov model based optical character recognition in the presence of deterministic transformations,” *Pattern Recognition* 26 [12], pp. 1813-1826, 1993
- [18] E. Bauer and R. Kohavi, “An empirical comparison of voting classification algorithms: Bagging, Boosting and Variants,” *Kluwer Academic Publishers Boston “Machine Learning,”* 26 [1], pp. 1-38, 1998

- [19] L. Breiman, "Bagging predictors," *Machine Learning*, 26 [2], pp. 123-140, 1998
- [20] 上田 修功, 中野 良平, "アンサンブル学習における汎化誤差解析," *電子情報通信学会論文誌*, Vol. J80-D II [9], pp. 2512-2521, 1996
- [21] 斎藤 泰一, 山田 博三, 森 俊二, "手書き文字データベースの解析(Ⅲ)," *電総研彙報*, 42 [5], pp. 385-434, 1978
- [22] H. Yamagata, H. Nishida, T. Suzuki, M. Tachikawa, Y. Nakajima and G. Sato, "A handwritten character recognition system by efficient combination of multiple classifiers," *IEICE Trans. INF&SYSTEM*, Vol. E79-D [5], May. 1996
- [23] H. S. Park and S. W. Lee, "Off-line recognition of large-set handwritten character recognition," *Pattern Recognition*, 31 [12], pp. 1849-1864, 1996
- [24] S. S. Kuo and O. E. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models," *IEEE Trans. Pattern Anal. Mach. Intell.* 16 [8], pp. 842-848, 1994
- [25] 中野 康明, 中田 和男, 中島 晃, "周辺分布とそのスペクトルによる漢字認識の改良," *電子通信学会論文誌 D*, Vol. 57-D [1], pp. 15-22, 1974
- [26] 坂井 邦夫, "文字・文書の認識と理解," *電子通信学会誌*, Vol. 71, pp. 1182-1191, 1988
- [27] 中田 和男 編著, "パターン認識とその応用," コロナ社, 1978



## 研究業績

### 著書

Hiromitsu Nishimura, Hisashi Ikeda and Yasuaki Nakano, "A Segmentation Method for Touching Handwritten Japanese Characters," appeared in "Document Analysis: Theory and Practice" ed. Seong-Whan Lee and Yasuaki Nakano, pp. 130-139, Springer, Berlin, Germany, 1999.

### 学術論文誌

"多方向特徴抽出 HMM と Bagging による多数決を利用した文字認識," 西村 広光, 小林 誠, 丸山 稔, 中野 康明, 電子情報通信学会論文誌, Vol. J82 D-II [9], pp. 1429-1434, 1999

### 国際会議 (複数査読付き)

Hiromitsu Nishimura, Hisashi Ikeda and Yasuaki Nakano, "A Segmentation Method for Touching Handwritten Japanese Characters," Proceedings of Third International Association for Pattern Recognition Workshop on Document Analysis Systems, pp. 252-261, Nagano, 1998

Hiromitsu Nishimura, Makoto Kobayashi, Minoru Maruyama and Yasuaki Nakano, "Off-line Character Recognition Using HMM by Multiple Directional Feature Extraction and Voting with Bagging Algorithm," Proceedings of Fifth International Conference on Document Analysis and Recognition, pp. 49-52, Bangalore, India, 1999



Hisashi Ikeda, Yukio Ogawa, Masashi Koga, Hiromitsu Nishimura, Hiroshi Sako and Hiromichi Fujisawa, "A Recognition Method for Touching Japanese Handwritten Characters," Proceedings of Fifth International Conference on Document Analysis and Recognition, pp. 641-644, Bangalore, India, 1999