

ストローク HMM に基づく オンライン手書き文字認識方式

嵯峨山茂樹^{†,††} 中井 満[†] 下平 博[†]

[†] 北陸先端科学技術大学院大学 情報科学研究科
〒 923-1292 石川県能美郡辰口町旭台 1-1 / 0761-51-1225

^{††} 東京大学大学院 工学系研究科
〒 113-8656 東京都文京区本郷 7-3-1 / 03-5841-6900

E-mail: †{sagayama,mit,sim}@jaist.ac.jp

あらまし オンライン手書き文字(特に漢字)認識の手法として,サブワード単位の連続音声認識の手法を最大限活用するアプローチを提案する.ストローク(字画)(あるいはさらに小さいサブストローク)など,文字より小さい単位を HMM でモデル化する.連続音声認識における文音声が音素モデルを単位とした語彙と文法で定義されるのと同様に,漢字もストロークを単位とする語彙的・文法的な組み合わせであると見なすことにより,様々な特徴のある文字認識手法が考えられる.基本構成は,速度ベクトル列を文字の特徴量とし,ペン先の動きを 25 種類の方向線分でモデル化するサブストローク HMM と,サブストロークを最小単位とする文字構造の階層的な定義,および尤度(事後確率)最大経路の探索からなる.利点としては,辞書やモデルの記憶容量が小さい,少数の文字の学習で辞書登録されている全字種の認識が可能,少量の文字で筆者適応が可能な事が挙げられる.また,非目視文字の認識が可能となり,新しいヒューマンインタフェースとしてのモダリティとして活用できる.最後に,さらに発展させるための今後の課題を詳細に論じる.

キーワード オンライン手書き文字認識,連続音声認識,HMM,ストロークモデル,階層化漢字辞書

Online Hand-Written Character Recognition Based on Stroke HMMs

Shigeki SAGAYAMA^{†,††}, Mitsuru NAKAI[†], and Hiroshi SHIMODAIRA[†]

[†] Graduate School of Information Science, Japan Advanced Institute of Science and Technology
1-1, Asahi-dai, Tatsu-no-kuchi, Nomi-gun, Ishikawa 923-1292 Japan

^{††} Graduate School of Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656 Japan

E-mail: †{sagayama,mit,sim}@jaist.ac.jp

Abstract We propose a new approach of on-line handwritten character recognition that fully utilize continuous speech recognition technology. Basic idea of this approach is that we treat a Kanji character as a sequence of strokes, and this is similar to a relation of sentence speech and phonemes in continuous speech recognition. Therefore, recognition system basically consists of input features, HMMs, dictionaries and decoder. In our method, velocity vectors of pen movement are input features, the pen movements in the 8 directions are modeled by 25 kinds of substroke HMM, Kanji dictionary is hierarchically defined by substroke unit, and recognition results are obtained as maximum likelihood path searched by Viterbi decoder. The advantages of this approach are the following: The memory requirement for the dictionary and models is very small. All characters defined in the dictionary can be recognized by stroke HMMs trained by small amount of data. The writer adaptation of HMM is easily performed by writing a few sample characters. Furthermore, it can recognize characters written with no sight and therefore it will be possible to use it as a new modality of human-interface.

Key words on-line handwritten character recognition, continuous speech recognition, HMM, stroke model, hierarchical Kanji dictionary

1. はじめに

オンライン手書き文字認識の研究の歴史は長く、すでに携帯端末を中心にペン入力的手段として実用的に広く用いられている。また今後もモバイル機器や、さらにはウェアラブル・コンピュータなどの要素技術として重要性を大きく増す可能性がある。そこでは、多様な形態で使用できること、多様な文字のクラスを認識できること、そしてその認識性能の向上が要求されるであろう。そのため、従来の手法とは異なる新しい手法も模索して行く必要がある。

同じく時系列パターンの認識である音声認識の分野に目を向けると、近年、その進展は著しい[1]～[5]。特に、隠れマルコフモデル(Hidden Markov Model: HMM)のような確率統計モデルを中心とした認識手法が高度に洗練されてきた。この音声認識の基本手法を文字認識に応用した研究として、アルファベット文字認識[6]～[9]を中心に成果が上がっており、HMMを用いたオンライン文字認識技術が急速に向上する兆しがあるものの、連続音声認識で培われた技術を十分に活用するには至っていないようである。また、国内でも仮名漢字文字の認識に HMM を利用した研究[10]～[12]が行われているが、欧米に比べるとまだまだ HMM は主流ではない。なぜならば、従来手法では字種を単位としてモデル化する、あるいはテンプレートを用意する事が多く、1モデルあたりの学習に大量のデータを要する HMM はカテゴリ数の多い手書き漢字認識には不利であると考えられていたからである。だが、音声認識を省みると HMM は少数個の音素モデルで大語彙単語や文音声の認識が可能なモデルであり、文字認識との同型性を考えるとこのモデルは文字の構成要素に相当している事に気付くであろう。文字認識において漢字の部首に着目した研究例[13]はあるが、オンライン手書き文字認識において、しかも HMM と組み合わせた研究例はほとんど行われていない。

そこで我々はサブワード単位の連続音声認識の手法を活用したオンライン手書き文字認識方式を提案し、その基本方式と利点を述べると共に、本方式における研究課題と解決のための方策について論じる。

2. 新しい利用形態の文字認識

2.1 非目視手書き文字認識

従来の文字認識入力では、筆記する領域や筆跡を目で見ながら丁寧に文字を書くとその認識される、という形態が取られていた。例えば複数の枠の中にペンで一文字づつ順次記入して行く、などの形態が取られることが多かった。しかし、このような方法では利用場面が限られる。

今後の可能性の一つとしては、見ないで書く文字(非目視手書き文字)入力が、携帯機器などへの入力、

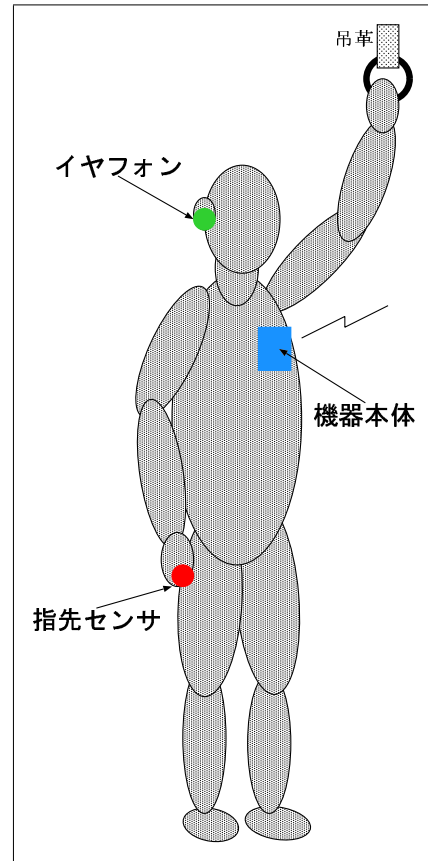


図1 満員電車の中でも使える文字認識技術

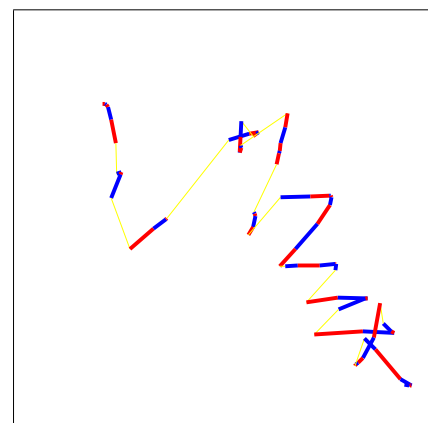


図2 非目視手書き漢字の例(“漢”)

あるいは視覚障害者用途などに有望であろう。このような指先の文字入力には広い用途が見込める。例えば満員電車内に立ったまま、あるいは歩きながら、ポケットに手を入れて指先でコンピュータを操作したり、周囲に気付かれないようにデータを入力したりするなど、将来のウェアラブル・コンピュータへの利用が可能である。図1は満員電車内での非目視手書き文字入力のイメージである。吊革につかまりながら、自分の膝の上などで文字を書いて携帯機器を操作することにより、電子メールやインターネットにアクセスし、音声合成によって入力の確認や端末の情報をモニターしているところである。

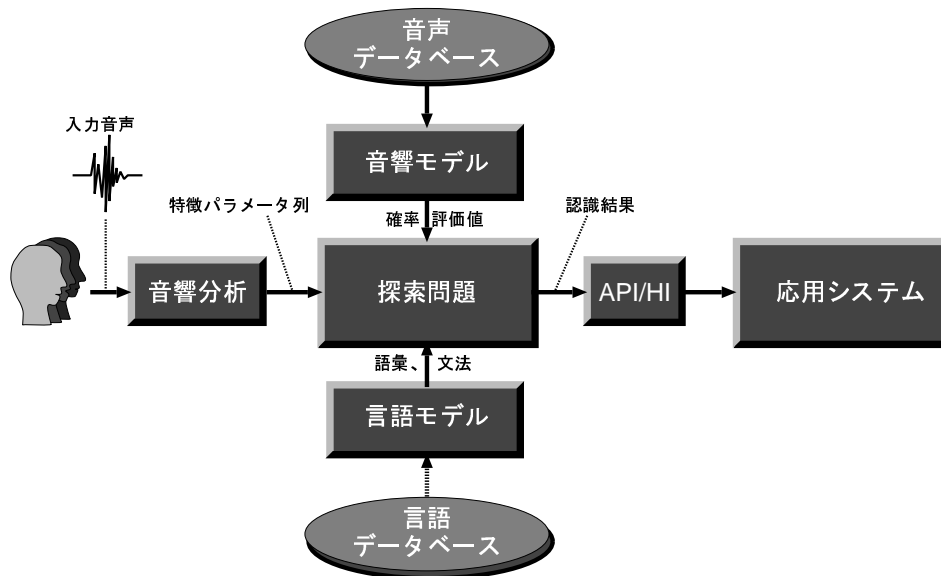


図3 連続音声認識の構成要素

非目視手書き文字では視覚フィードバックがないので、書かれた文字自体は読めないくらいに崩れた字となることが予想される。図2は“漢”という漢字を非目視で書いた時の筆跡例である。視覚的には、人間でさえ読み難く、いわゆる位相幾何学的特徴（交差や閉曲線など）は必ずしも保存されないが、指先の動き（特に速度ベクトル列）に着目すれば、ある文字を書こうとする意志を持った運動と考えることができる。すなわち、書かれた文字を認識するより、書こうとしている意志を認識する、と考えることにより、文字認識の方法論が広がる可能性がある。

3. 手書き文字認識の新アプローチ

3.1 連続音声認識の手法

音声認識の性能はHMMおよび統計的言語モデルの研究の進展により、近年著しく進歩した。朗読調の音声ならば数十万種類の単語を含む連続音声認識を家庭用のパソコンでもほぼ実時間で行うことができ、現在、多数のソフトウェア製品に音声認識が組み込まれたり、電話などのサービスに広く用いられるようになった。

連続音声認識の構成を図3に示す。音声波形から認識に有効な特徴量（メルケプストラムが一般的）を抽出する音響分析部、言語音声構成要素（通常は環境依存音素）として尤度を計算できる音響モデル（通常はHMM）部、発話内容を語彙と文法（通常は単語 bigram, trigram 確率）で表現し探索空間を規定する言語モデル部、およびそれらを統合して効率的に探索（ビーム探索、A*探索など）を行う探索問題部からなる。原理を簡単に言えば、規定された解探索空間の中で、音響モデルと言語モデルが与える尤度の総積（あるいは対数尤度の総和）が最大になるような経路を探索するのである。

表1 連続音声認識とオンライン手書き文字認識の対応

連続音声認識	手書き文字認識
音素	画（ストローク）
調音結合モデル	連続画モデル
単語音声	部首
文音声	文字
話者適応	筆者適応

このような手法を採ることにより、連続音声認識では、サブワードモデル（通常は音素モデル）を組み合わせて単語を表現し、単語列の組合せである文音声を認識する。以前は、単語テンプレートとして登録された単語のみを対象に、DPマッチングに基づいて単語音声認識が行われていた（この手法は今も使われる。）それに比べ、不特定話者連続音声認識では、入力文章音声のパターンを、与えられた語彙と文法に従って並べることのできる全ての可能な音素モデル列の中から、尤度最大となる経路を探索する問題を解くことにより認識することができる。

3.2 連続音声認識アルゴリズムに基づく手法

オンライン手書き文字（走り書きの文字、続け字、崩れた字、非目視手書き文字などを含む）の認識の問題は、連続音声認識の問題と同型であると言ってよいほど、問題の性質と構造が似通っている。音声認識は音声を短時間ごとに分析して得られるスペクトル特徴ベクトル（多次元）の運動軌跡を認識するものである。文字認識は、特徴ベクトルが2次元の場合に相当する。

特に漢字の場合、音素を画（ストローク）に、単語を部首に、文音声を漢字に対応させることにより、探索手法や文法記述まで含めて完成度の高い音声認識技術を適用する事ができる。ペンの位置とペンの

表2 漢字を表現する基本構成要素の比較

単位	構成要素	利点	必要状態数	連続音声認識対応
一字種	切 守	「揺らぎ」に強い	6,353 字種 × 平均 15 サブ ストローク (pen down) × 3 状態 = 285,885 状態	単語音声 shimekiri (切) geNshu (守)
扁旁冠脚	匕 刀 冫 寸		約 280 種 × 平均 5 状態 = 1,400 状態	音節 i, e, me, shi, N, ..
画	一 丨 冫 一 一 一			音素 i, e, sh, m, N, ..
方向線分	✳ ✳	・モデル数少 ・学習必要量小	25 方向 × 1~3 状態 = 57 状態	音響イベント, 定常母音, 破裂, 閉塞, 摩擦, ..

速度は、音声認識の特徴量として広く用いられているケプストラムおよびその時間微分成分であるデルタケプストラムに対応付けられる。画(ストローク)を書くペンの動きのモデルは、音素モデルに対応し、確率統計モデルである HMM が適している。画の組み合わせにより部首が、それらの組み合わせにより字が成り立つが、これらには音声における単語辞書(音素連鎖)と、単語の組み合わせとしての文の文法に対応する。また、画はつながったり(続け字)、前後の画の影響で変形するが、これは調音結合に起因する音素の環境依存変形のモデル化がよく対応し、これらに対処するモデル化と環境依存パーザの手法など、音声認識における技術が活用できる。さらに、話者適応アルゴリズムなどは筆者適応に応用できる。表1に概念の対応を示す。

このように、サブワードベースの連続音声認識の手法を用いて、サブキャラクター HMM を用いたオンライン手書き漢字認識を行えば、次のような利点が得られる。

- ・モデル数、記憶容量が削減できる
- ・少量のデータでモデルの学習が効率良く行なえる
- ・筆順違いには辞書登録で対応できる
- ・未学習の字種でも辞書登録により認識できる
- ・サブキャラクターを筆者適応することにより全字種の認識性能の向上がはかれる
- ・音声認識の研究・経験が活用できる

4. サブストローク HMM に基づく手書き漢字認識の基本構成

4.1 構成要素に基づく漢字の表現

典型的な漢字は、扁旁冠脚の連鎖で構成され、扁旁冠脚は画の連鎖、画は方向線分の連鎖で、階層的に構成される。複雑な文字は階層が増す。これは連続音声認識の音素、音節、単語音声の関係に似ている。漢字の構成要素は表2に示すような階層的な構造になり、どの階層をモデル化の単位とするかにより、様々な特性の様々な方式が考えられる。以下では、連続音声認識手法の特徴が端的に現れる例として、8 方向線分を基本要素とする手書き文字認識の

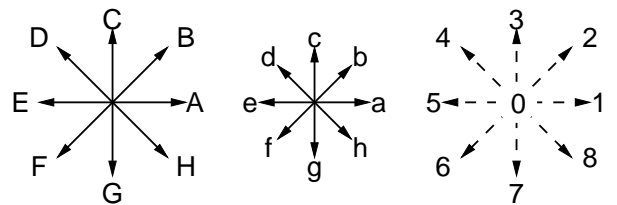


図4 25 方向の方向線分 (Sub Stroke)

方式の基本構成について述べる。

4.2 階層的漢字辞書の構成

基本構成では、図4に示すように、長く書かれる線分 8 種 (A-H) , 短く書かれる線分 8 種 (a-h) , 空中のペン移動 8 種 (1-8) と、ほぼ同じ位置でのペンの上下 (0) によりすべての漢字を表現する。これら 25 種類の最小要素 (サブストローク) により、以下に一部を示すような書き換え規則ですべての漢字を表現する。

```

J      G f
\      H
又     J 3 \
矢頭   J 2 一
夕     夕頭 5 又
至     一 5 厶 5 土
致     至 3 夕
糸     幺 4 小
緻     糸 3 致
    
```

このように、必ずしも従来の漢字部首や扁旁冠脚に捕らわれないで、時系列としての漢字の構成を表現することができ、JIS 第二水準漢字である“緻”も容易に表現できる。例えば我々は、JIS X0208-1983 の第一・第二水準全漢字 6353 字種を表現するのに、367 種の間要素(上の例では「矢頭」など)を加えて、合計 6720 の書き換え規則を用いている。

4.3 サブストローク HMM

漢字の最小構成要素 (サブストローク) は、書くたびに、また筆者により、確率的な変動を伴って観測されると考え、これを HMM を用いた確率モデル

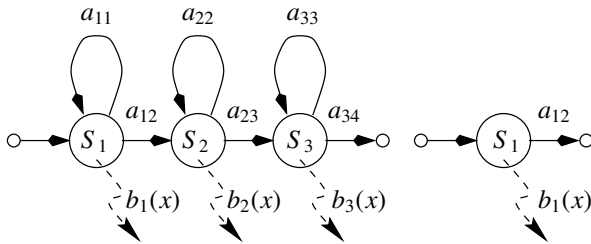


図5 サブストローク HMM

により表現する。

基本構成では、簡単のためペン先の位置座標 (u, v) を等時間間隔でサンプリングし、その変化分（ここでは速度ベクトルと呼ぶ）を観測特徴量 $\vec{x}_t = (\Delta u, \Delta v)$ として用いる。図5に示すように、これを3状態の left-to-right 型の HMM によりモデル化する。HMM のモデルパラメータとしては、出力確率の集合 $B = \{b_j(\vec{x}_t)\}$ はある状態 S_j から入力速度ベクトル \vec{x}_t が生成される確率の集合、遷移確率行列 $A = \{a_{ij}\}$ はサブストローク内の状態 S_i から状態 S_j へ遷移する確率を扱う。ペンアップの状態は速度ベクトルが観測されず、次の着地点までの移動ベクトルのみが得られるので、1状態の自己ループなしの HMM を用いる。

サブストロークモデルの学習には、25種類のペン移動のデータより、各々のモデルについて Viterbi 学習あるいは Baum-Welch 学習アルゴリズムを用いる事によりモデルのパラメータが学習できる。さらに、これを初期モデルとして、手書き漢字データと対応する漢字構造を展開したサブストローク列を用いて連結学習する事により、より良いモデルパラメータが推定できる。

4.4 解探索

サブストロークモデル $\lambda = (A, B)$ から入力速度ベクトル時系列 $X = \vec{x}_1 \cdots \vec{x}_T$ が生成される確率は、対応する状態系列を $q = q_1 \cdots q_T$ として

$$P(X|q, \lambda) = \prod_t a_{q_{t-1}, q_t} b_{q_t}(\vec{x}_t)$$

であり、最大確率を与える状態系列は Viterbi 探索により求める事ができる。すなわち、漢字構造を展開した全てのサブストローク列の中から最大確率を与える候補を求める事ができ、これが認識文字候補となる。

4.5 ストローク HMM 方式の利点

HMM によりモデル化したストロークあるいはサブストロークを基本単位とするオンライン手書き漢字認識はさまざまな利点を持つが、ここでは辞書サイズを見積もる。

JIS 第一・第二水準漢字 6353 字種に対して、表2

の最大構成要素である一字種を単位とした場合、漢字構造辞書およびモデルを蓄えるのに必要な容量は約 7 MB にも達し、そのほとんどはモデルの蓄えに消費される。一方、最小構成要素のサブストロークモデルでは辞書サイズの方が大きくなるが、モデルと合わせてもわずか 115 kB 程度である。

5. ストローク HMM 方式の研究課題

本方式は以上に述べた基本方式に対し、性能向上のための様々な研究課題がある。以下では研究課題とその解決の方策を述べる。

5.1 基本要素の実現

基本方式の構成要素としては、以下のような要素が必要である。

- モデル
当初は 8 方向線分であるサブストロークを単位とした HMM のモデル化を行う。また、HMM のパラメータ学習のため、少なくとも Viterbi 学習、さらには連結学習の実装が必要である。
- 階層化辞書・ネットワーク化辞書
— 木 林 麻 摩のような階層化辞書が必要であり、当面は人手によって辞書を作成する。また、ストローク HMM の利点である探索効率の向上のためにはさらに階層化辞書からネットワークを構築する必要がある。
- 解探索
辞書が規定するストロークのネットワークを探索する認識エンジンとして、木探索、ビーム探索、ネットワーク HMM Viterbi 探索などの実装を行う。
- データ収集
学習および評価用のデータ収集を行う。研究目的に沿い、筆順指定の漢字、筆順フリーの漢字、走り書き文字、非目視手書き文字などのデータベースを順次構築する。また、効率良くデータ収集を行うための収集ツールも作成する。

5.2 辞書構成法の検討

- 非字を含めた効率的なネットワークの生成
漢字を定義する文法において、例えば、

```

文字  偏旁
扁  人扁|木扁|さんずい
旁  主|反|木|古|可|目

```

とすると「住、柱、注、仮、板、派、休、林、沐、枯、沽、估、何、河、柯、相、泪」の 17 字が表現でき、ネットワーク探索の効率を高くすることができる。また、この例だと JIS 漢字に存在しない「人扁に目」のような字（非字）も含まれてしまうが、複数候補を探索して、非字を除去する後処理を行えば良い。

- 漢字サンプルからの辞書の自動生成
基本方式では人手で辞書を作成したが、辞書の作成は一意的ではないので文法表現の最適性や誤りの残留の問題がある。そこで、手書き文字データサンプルから自動的に辞書を生成することが考えられる。これは音声認識で行われている baseform の生成問題であり、基本的には実際のデータを Viterbi アルゴリズムなどによって基本単位の列として decode し、それをそのまま辞書とする方法であるが、これでは同一の文字が何通りもの表現を持ち、同一部首が共有できないなどの問題を生じ、表現効率が低くなる恐れがある。漢字の階層的な構造を生かしながら、辞書表現を最適化する研究が必要である。

5.3 特徴量の検討

オンライン手書き文字の入力情報はペン先の速度などの豊富な情報を含み、原理的にはオフライン文字認識で扱える特徴量よりも有利で、その有効的な活用も重要な課題である。基本方式では、入力時系列の中で、同一の画を同一の時系列モデルで扱うように速度ベクトルを基本とした 2 次元の特徴量のみを用いているが、今後、以下の検討が必要である。

- 特徴量の高次元化
加速度ベクトル、ペン上下、筆圧などの情報を特徴量に加えた高次元のパターン空間内の点の運動軌跡を HMM でモデル化する。
- 位置ベクトルの利用
漢字中の基本単位が現れる位置に依らないモデル化のためには、速度や圧力など位置依存性のない特徴量を用いるのが基本であるが、位置情報にも利用価値の高い情報が含まれている。それらを画の重心位置などの形で取り込み、漢字固有の辞書に追加して、尤度計算に付加することが出来る。
- 特徴量の新たな表現
従来より多く用いられているペンの運動方向を、例えば、角度 θ と速度 v により表現する。これは x, y 方向ベクトルと等価な情報であるが、確率モデルで表現した場合の性質については不明なので、両者の比較が必要である。
- 特徴量時系列の正規化
ペン位置のサンプリングに関して、等時間間隔サンプリング、等距離間隔サンプリング、一定数サンプル点でのストロークサンプリングなどの性能比較を行う。

5.4 認識単位の検討

どの単位を HMM によりモデル化するかという選択は、学習量と頑健性の trade-off、必要記憶容量、計算量などに関わる問題である。すでに表 2 に示した

ように、漢字の構造の持つ自然な階層に基づいて様々な階層を考えることも可能である。しかし、階層が一様であることが必ずしも有利ではなく、部首の使用頻度と崩れやすさも考慮した、データに基づく客観的な単位選定手法が望ましい。以下にその手法の候補を挙げる。

- 環境依存モデル
手書き漢字中の各画は、その前後関係に依存して変形する。例えば、文字“三”と“未”の第 2 画の最後の部分は、第 3 画の影響を受けて前者では下方向へ、後者では上方向へ曲げられることが多い。このような現象は、音声認識では「調音結合」に当たり、その結果生ずる様々な「異音 (allophone)」に対応すると考えられる。これを、文字構造辞書と手書き文字データを用いて効率よくクラスタリングし、ネットワーク状の HMM である隠れマルコフ網 (HMnet) として前後画依存モデルを表現する。生成方法として、逐次状態分割 (SSS) 法 [14] が利用できる。
- 可変長モデル
全てを基本的な 25 種類の方向線分モデルのみで表現するのではなく、これらがある程度まとまった扁旁冠脚などの単位や、さらには常識的な画や部首に捕われない漢字の時系列パターン的一部分なども認識単位として組み合わせた可変長モデルの利用が考えられる。モデル集合を構築する際に、それらのモデルの表現しにくさに相当する量 (尤度の低さ) や使用頻度などを基準にする事により、可変長のモデルが自動獲得できる。

5.5 ストローク HMM の検討

基本方式としては、全共分散単一正規分布を出力確率とする単純な HMM を用いているが、音声認識の分野で様々な開発されている HMM を利用することが考えられる。例えば、

- 連続混合分布 (Continuous Mixture) HMM
予備実験の段階ではあるが、オンライン手書き文字認識においても性能向上の効果を確認している。
- 非同期遷移型 HMM [15]
各特徴量の変化が同期しない場合のモデル化に有効である。文字認識の場合は x, y の各方向の速度の変化が必ずしも同期しないので、親和性が高いと予想される。などの手法が考えられる。

5.6 解探索法の検討

連続音声認識では解探索法の研究が大きく進歩している。文字認識に関連すると考えられる事項としては、

- 漢字辞書からのネットワークの自動生成
- 非字を含めた効率的なネットワーク設計
- A* 探索, ビーム探索, stack decoder
- N-best 探索
- 事後処理としてのフィルタ

などが挙げられる。LR パーザと組み合わせて, LR テーブルから文法展開を行いながら経路をビーム探索する手法 [16] などにも利用できる。また, ビーム探索に前向き A* 探索アルゴリズムを組み合わせる方法も有望である。

5.7 尤度計算高速化の検討

本方式では基本単位が少なければ, 出力尤度計算は共通化され, 計算効率が高い。しかし, 環境依存サブストロークモデルを多用する場合は, 計算効率の向上を図る必要があり, その一例としては特徴量のスカラ量子化とテーブル参照による計算の効率化 [17] がある。

5.8 筆者適応アルゴリズムの検討

画に含まれる筆者の書き癖(横棒が右上がりなど)を学習することにより, 少数のサンプルを用いて全ての漢字に適用できる筆者適応を行なう。このために, ベクトル場平滑化法(VFS) [18], MAP(Maximum A Priori) 推定法 [19], MAP/VFS 法アルゴリズム [20], MLLR(Maximum Likelihood Linear Regression) [21] などの話者適応で開発された手法を用いることができる。文字の場合, VFS 法は, 筆者の癖は平均からのズレのベクトル場で表現され, それは滑らかな場であると仮定することにより少数のサンプルからでも精度良く求められる, という意味を持つ。

5.9 画(ストローク)の連結への対応

連続音声認識では音素が連続して境界が曖昧になることは前提であり, 走り書きで生じる画の連結も扱える。画と画の間のペン移動が, pen up されない場合も確率的分岐により扱える。画の連結が著しくなると, もはや前後画に影響された画のパターンとして扱うより, 別のマクロパターンへの分岐と捉える方がよい。このための辞書項目を拡大し, モデル化の精度を向上させる。

5.10 文字列(単語・文)認識の研究

音声認識では, 言語モデルの利用は不可欠であるが, 文字認識でもその種の研究は古くから行われている。もし書かれた文字あるいは文字列が文あるいは単語を構成しているならば, 文字認識結果の複数候補の中から正解を選んだり, 再順位付けをすることが可能である。このためには, その文が特定の文法あるいは統計文法に合致することを仮定でき, その文法から候補文字列を生成する言語確率を計算することが可能になっていることが必要である。この

ような文法としては, 同様に音声認識で用いられている

- 単語語彙定義
- 文字 n -gram 確率
- 単語 n -gram 確率
- 文脈自由文法などの文法

などが考えられる。

5.11 走り書き・非目視手書き文字データの収集

この種のデータは殆んど存在しないので, 新たに収集し, データベースを構築する。例えば, タブレットを肩から脇腹に吊し, 立った姿勢で文字を書くなど, データ収集方法を検討する。

5.12 認識性能評価, ヒューマンインタフェース評価, システム評価

- 認識結果信頼性自己判断
正しく(自信を持って)認識できているかどうかを, システム自身が評価する方法を, 音声認識の発話検証(utterance verification)手法を用いて組み込む。
- 認識結果確認修正方式, ジェスチャ認識
認識した結果は, ユーザへ音声合成あるいはヘッドマウントディスプレイでフィードバックして結果を確認し, 誤っている場合には修正する。これらをすべてペン(指先)入力で統一的に行うヒューマンインタフェースを検討し, 実際にシステムを実現することも大きな課題である。
- 走り書き・非目視手書き文字認識の HI 評価
認識性能, ヒューマンインタフェース, アプリケーションと連動したシステムのそれぞれのレベルで性能評価を行なう必要がある。特に, 走り書き・非目視手書き文字認識という新しい機能が, ユーザビリティの向上としてどのように評価されるか, 効率向上(文字入力速度向上), 主観評価(快適さ), システム適合性(既存アプリケーションに, 新しい使用方法が組み合わせられるか)などの観点から評価することが必要である。

6. 結 び

オンライン手書き文字認識の手法として, サブキャラクターを単位とした HMM に基づく手法を提案した。基本方式として, ストロークあるいはサブストロークを単位としたモデル化を行い, さらに発展させるための今後の研究方針について述べた。今後, これらの掲げられた課題について研究を進め, その成果については随時報告していく予定である。

謝辞 オンライン手書き漢字認識の状況についてご教示下さった東北大 大町真一郎助教授, オンライン手書き文字データを提供して下さい下さった東京農工大

中川正樹 教授，東京高専 市村洋 教授・鈴木雅人 助教授に感謝します。また，当大学にて，漢字構造辞書作成に協力下さった富永絵美，小場久雄，筆者適応実験を行って下さった川口弘光，100名を超える大量のデータ収集を担当された秋良直人，井波暢人の諸氏に感謝する。なお，この研究は文部省科学研究費補助金(基盤研究B)課題番号11480074の助成を受けた。

文 献

- [1] D. Childers, R. V. Cox, R. DeMori, S. Furui, B.-H. Juang, J. J. Mariani, P. Price, S. Sagayama, M. M. Sondhi, and R. Weischedel: "The Past, Present, and Future of Speech Processing," *IEEE Signal Processing Magazine*, vol.15, no.3, pp.24-48 (1998).
- [2] 嵯峨山茂樹: "数理統計モデルを用いた音声認識の現状と将来," 日本音響学会誌, vol.48, no.1, pp.26-32 (1992-01).
- [3] Chin-Hui Lee, Frank K. Soong, Kuldip K. Paliwal, editors: "Automatic Speech and Speaker Recognition — Advanced Topics —," Kluwer Academic Publishers (1996).
- [4] Laurence Rabiner, Biing-Hwang Juang: "音声認識の基礎(上,下)," (古井貞熙 監訳), NTT アドバンステクノロジー (1995).
- [5] 国際電気通信基礎技術研究所編: "翻訳電話," ATR 先端技術シリーズ, オーム社 (1994).
- [6] W. G. Aref, P. Vallabhaneni and D. Barbara: "On Training Hidden Markov Models for Recognizing Handwritten Text," *Proc. of the 4th IWFHR*, pp.275-284 (1994).
- [7] K. S. Nathan, H. S. M. Beigi, J. Subrahmonia, G. J. Clary, and H. Maruyama: "Real-time On-line Unconstrained Handwriting Recognition Using Statistical Methods," *Proc. ICASSP95*, vol.4, pp.2619-2623 (1995).
- [8] H. Bunke, M. Roth, E. G. Schukat-Talamazzini: "Off-line Cursive Handwriting Recognition Using Hidden Markov Models," *Pattern Recognition*, vol.28, no.9, pp.1399-1413 (1995).
- [9] G. A. Saon, A. Belaid: "Recognition of Unconstrained Handwritten Words Using Markov Random Fields and HMMs," *Progress in Handwriting Recognition*, World Scientific, pp.75-80 (1996).
- [10] 野中利成, 小沢慎治: "Hidden Markov Models を用いたオンライン手書き平仮名認識," 信学論(D-II) vol.J74-D-II, no.12, pp.1810-1813 (1991-12).
- [11] 伊藤等, 中川正樹: "Hidden Markov Model に基づくオンライン手書き文字認識," 信学技報 PRMU97-85, pp.95-100 (1997-07).
- [12] 高橋賢一朗, 安田英史, 松本隆: "Hidden Markov Model を用いたオンライン手書き文字認識," 信学技報 PRMU96-211, pp.143-150 (1997-03).
- [13] A. B. Wang, K. C. Fan: "A Radical-Based OCR System for Handwritten Chinese Character Recognition," *Progress in Handwriting Recognition*, World Scientific, pp.127-132 (1996).
- [14] 鷹見淳一, 嵯峨山茂樹: "逐次状態分割法による隠れマルコフ網の自動生成," 信学論(D-II), vol.J76-D-II, no.10, pp.2155-2164 (1993-10).
- [15] Shigeki Matsuda, Mitsuru Nakai, Hiroshi Shimodaira and Shigeki Sagayama: "Asynchronous-Transition HMM," *Proc. ICASSP2000*, vol.2, pp.1005-1008 (2000-06).
- [16] 永井明人, 鷹見淳一, 嵯峨山茂樹, ハラルドシンガー: "隠れマルコフ網と一般化 LR 構文解析を統合した連続音声認識," 信学論(D-II), vol.J77-D-II, no.1, pp.9-19 (1994-01).
- [17] Shigeki Sagayama and Satoshi Takahashi: "On the Use of Scalar Quantization for Fast HMM Computation," *Proc. ICASSP95*, pp.213-216 (1995-05).
- [18] 大倉計美, 杉山雅英, 嵯峨山茂樹: "混合連続分布 HMM 移動ベクトル場平滑化話者適応方式," 信学論(D-II), vol.J76-D-II, no.12, pp.2469-2476 (1993-12).
- [19] R. M. Stern and M. J. Lasry: "Dynamic Speaker Adaptation for Feature-based Isolated Word Recognition," *IEEE Trans. Acoust., Speech & Signal Process.*, ASSP-35, 6, pp.751-763 (1987).
- [20] Jun-ichi Takahashi and Shigeki Sagayama: "Vector-Field-Smoothed Bayesian Learning for Fast and Incremental Speaker/Telephone-Channel Adaptation," *Computer Speech and Language*, vol.11, no.2, pp.127-146 (1997-04).
- [21] C. J. Leggetter, P. C. Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," *Computer Speech and Language*, vol.9, no.2, pp.171-185 (1995).