

ストローク HMM によるオンライン手書き文字認識の性能評価

中井 満[†] 嵯峨山茂樹[†] 秋良 直人[†] 小場 久雄^{††} 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
^{††} 金沢学院大学 美術文化学部
〒 920-1392 石川県金沢市末町 10

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

あらまし 文字の構成要素の HMM を用いたオンライン手書き文字認識方式の評価実験について報告する。まず、辞書のサイズを最小にするストローク(方向線分)をモデルの単位とした認識実験を行い、本手法のベースラインとなる認識性能を示す。評価用資料には新たに学生を筆記者として2種類のデータベース(α セット, β セット)を収集した。 α セットは筆順違いによる性能劣化要因を省くために正しい筆順で新旧教育漢字 1016 字種を 108 人に筆記してもらった。また β セットは筆順の自由な JIS 第一・第二水準漢字の 6353 字種のデータセットで、1 人あたり 1200 字程度の分担で 80 人分収集した。これらを用いた認識実験により、 α セットを用いた 1016 字種カテゴリの実験で 94% 以上、 β セットの未学習字種を含む 6353 字種カテゴリの場合で約 80% の 1 位認識率が得られた。また、少量の字種サンプルによる筆記者適応によって認識率の向上も確認した。

キーワード オンライン手書き文字認識, HMM, 筆記者適応, 未学習文字, データベース

Evaluation of HMM Based On-line Handwritten Character Recognition with Subcharacter Model Units

Mitsuru NAKAI[†], Shigeki SAGAYAMA[†], Naoto AKIRA[†], Hisao KOBAYASHI^{††}, and Hiroshi SHIMODAIRA[†]

[†] School of Information Science, Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Tatsunokuchi, Ishikawa, 923-1292 JAPAN
^{††} Faculty of Fine Arts and Informatics, Kanazawa Gakuin University,
10 Sue, Kanazawa, Ishikawa, 920-1392 JAPAN

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

Abstract This paper describes a new on-line handwritten character recognition scheme based on subcharacter HMM units and we report base accuracies of proposed method by using simple stroke units that minimize a memory requirement for models and dictionaries. For evaluation, we collected two kinds of handwritten character databases (α set and β set) from students. The α set is a fixed stroke order Kanji database with 1016 character categories of an old and new educational Kanji from 108 people. The other β set consists of about 1200 characters per person by free stroke order from 80 people and includes whole 6353 Kanji of JIS 1st level and 2nd level sets. In experiments, we achieved correct recognition rate of above 94% by using the α set database and achieved about 80% correct rate in 6353 categories recognition by using the β set database that includes 5337 untrained characters. We also show an improvement of recognition accuracy by using writer adaptation technique.

Key words on-line handwritten character recognition, HMM, writer adaptation, untrained character, database

1. はじめに

オンライン手書き文字認識は文字入力デバイスより得られる筆跡特徴量の時系列パターンの認識と捉える事ができる．同じく時系列パターンの認識である音声認識では HMM (Hidden Markov Model) による認識が現在の主流となっており，文字認識の分野でも字形の変形などの揺らぎに有効な手法として着目されている．とりわけ英文字の場合，アルファベットを認識カテゴリとすると，その種類はせいぜい数十程度であり，1モデルあたりの学習サンプルを大量に要する HMM とも相性が良い．そのため，欧米のオンライン文字認識では比較的早くから HMM を利用している [1]~[4]．一方，漢字仮名文字についても，46 字種の平仮名認識 [5], [6]，881 字種の教育漢字認識 [7], [8] に HMM が導入されてきた．これらの手法では基本的に 1 字種に対して 1 つの HMM を用いている．筆順違いも含めるとモデル数は更に増え，例えば高橋ら [7] のモデルでは辞書サイズが約 2 MB に達している．

これに対し，我々は漢字の構成要素をストローク (方向線分) \in 複合ストローク (画) \in 偏傍冠脚 \in 字種のような階層を持った辞書を作成し，なるべく低い階層を用いる事により，高々数十種類程度のモデルであらゆる字種の認識が可能となるオンライン手書き文字認識方式を提案した [9]．また，従来手法の多くは入力特徴量を方向チェーンコードで量子化した離散分布出力型の HMM を用いているが，我々の手法では，筆跡の速度ベクトルに対する揺らぎが正規分布の形状になるという仮定から連続混合分布出力型の HMM を用い，量子化による情報量の損失を避けている．このようなモデルを音声認識と対応させるならば「ストローク」は「音素」であり「文字」は「単語」であり「文字列」は「文音声」に相当する．したがって，文字認識は字種を定義するストロークのネットワーク内を探索する連続ストローク認識となる．この手法の利点としては

- 辞書サイズが小型である
- 効率の良いネットワーク探索により認識が高速
- 未学習字種の認識が可能
- 少量の学習サンプルでの筆記者適応が容易
- 辞書登録で筆順違いが解消できる

などが挙げられる．

本稿では提案する手法のベースラインとして，最もモデルサイズが小さくなる条件となるストローク (方向線分) を単位とした単一正規分布出力型の HMM を用いた場合の認識性能について評価する．また，新旧教育漢字 1016 字種で学習したモデルを用いて学習に用いられなかった JIS 第一・第二水準のその他の字種を含む 6353 字種の認識実験，および不特定

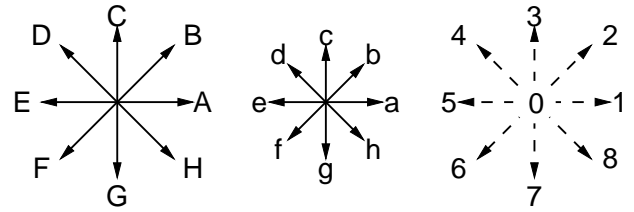


図 1 ストロークの種類と対応するモデルのラベル：A ~ H (a ~ h) は pen down の長い (短い) 線分のストローク，0 ~ 8 は pen up の方向

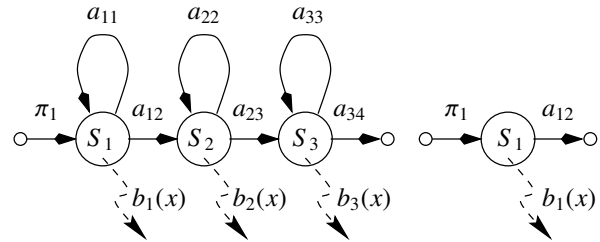


図 2 ストローク HMM : (左) pen down モデル (右) pen up モデル

筆記者 HMM を特定筆記者に適応する筆記者適応実験について，その結果を報告する．

2. ストローク HMM を用いた手書き文字認識方式

2.1 文字特徴量

入力デバイス (タブレット) から得られる筆跡情報は 2 次元座標，筆圧，ペンの傾きなどがあるが，本報告ではこれらのうちの最も基本的な特徴量である 2 次元座標情報 (x, y) のみを用いる．ただし，ストロークを単位としてモデル化する場合，字種によってそのストロークの出現する座標が異なり，文字内でも同一のストロークが複数個所に出現しうるので，座標値は用いずに連続する 2 点間の差分 (dx, dy) を特徴量とする．ペンがタブレット上に接している時 (pen down) の筆跡は時間サンプリングされているので，これは速度ベクトルとなる．また，ペンが離れている時 (pen up) の軌跡はサンプリングされないので，pen up は画間の移動量を表すベクトルとなる．なお，本稿の認識実験では pen down / up の状態にあるという情報は明示的には用いないが，モデルの学習では文字資料をストロークにセグメンテーションするための情報として使用する．

2.2 ストローク HMM

ストロークの単位として図 1 のような 8 方位の方向線分を定義する．更に細分して，pen down の移動距離が長い方向線分を A ~ H，短い方向線分を a ~ h，pen up の移動方向を 1 ~ 8，および移動しないモデル 0 を合わせた合計 25 種類のモデルとする．これ

らのモデルを図 2 に示されるように left-to-right 方向のみに遷移する HMM に対応させる．すなわち，pen up の場合，連続するフレーム（時間サンプリング単位）数は常に 1 フレームであるので自己遷移確率のない 1 状態で表わし，それ以外の pen down のモデルを自己遷移確率のある 3 状態で表わす事にする．ここで，ストローク k の HMM を以下のパラメータを用いて $\lambda^{(k)} = (A^{(k)}, B^{(k)}, \pi^{(k)})$ と表す．

$A^{(k)} = \{a_{ij}^{(k)}\}$: 状態 S_i から S_j に遷移する確率の集合，
 $B^{(k)} = \{b_j^{(k)}(x)\}$: 状態 S_j が x を出力する確率の集合，
 $\pi^{(k)} = \{\pi_i^{(k)}\}$: 初期状態確率の集合

また，出力確率は連続分布型とし，

$$b_j(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp\left(-\frac{1}{2}(x - \mu_j)' \Sigma_j^{-1} (x - \mu_j)\right)$$

で表される平均ベクトル μ ，共分散行列 Σ の n 次元正規分布で与える．

2.3 認識

観測される筆跡の時系列特徴量 $O = O_1 O_2 \cdots O_T$ ，($O_t = (dx_t, dy_t)$) がストローク列 $W^* = W_1^* W_2^* \cdots W_{N^*}^*$ で表される文字であると認識する原理について述べる．

まず，ストローク列 $W = W_1 W_2 \cdots W_N$ から時系列データ O が観測される確率は各時刻に対応する状態の系列を $q = q_1 q_2 \cdots q_T$ として

$$P(O|W) = \sum_{\text{all } q} P(O, q|W)$$

$$P(O, q|W) = \prod_{n=1}^N P(O_{T_{n-1}+1} \cdots O_{T_n}, q_{T_{n-1}+1} \cdots q_{T_n} | W_n)$$

$$= \prod_{n=1}^N \left(\pi_{q_{T_{n-1}+1}}^{(W_n)} \prod_{t=T_{n-1}+1}^{T_n} a_{q_{t-1}, q_t}^{(W_n)} b_{q_t}^{(W_n)}(O_t) \right)$$

である．ここで， $T_0 = 0$ ， $T_N = T$ である．一方，観測ベクトル O が得られた時に，それがストローク列 W である確率は

$$P(W|O) = \frac{P(O|W)P(W)}{P(O)}$$

であり，確率最大を認識基準とした場合，大小比較に影響を及ぼさない $P(O)$ を無視した

$$W^* = \arg \max_W P(O|W)P(W)$$

が認識ストローク列となる．ここで $P(W)$ はストローク列の出現確率であり，辞書登録された字種定義上，存在し得ないストローク列の場合は $P(W) = 0$ とし，それ以外の場合は等確率とする．ただし，実際には $P(O|W)$ を計算しないで， $P(O, q|W)$ が最大となる状態系列 q を探索する Viterbi アルゴリズムを用いる．

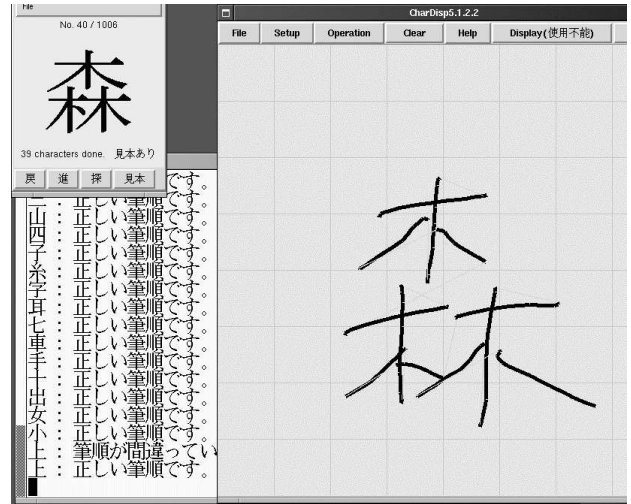


図 3 文字データ収集画面：(左)入力文字の提示部 (右)筆跡表示部

2.4 学習

モデル λ に対して学習資料セット O の出力確率 $P(O|\lambda)$ を最大にするようにモデルのパラメータである正規分布の平均 μ ，共分散 Σ ，および遷移確率 a_{ij} を学習する．本報告では最も基本的な学習法である Viterbi 学習を用い，状態毎のデータの区分化とパラメータの更新を繰り返す事により決定する．例えば，“口”という文字はストローク列 “g 3 a g 5 a” と定義されており，一画目 (g)，三画目 (a) は pen down / up 情報のみで区分できるが，二画目 (a g) はストローク境界が不明なので，モデル a とモデル g を時間方向に接続した連結 Viterbi 学習によって，ストローク境界および状態境界を最適化することができる．

3. JAIST-IPL (北陸先端科学技術大学院大学・知能情報処理学研究室) オンライン手書き文字データベース

筆順を指定したデータベース (α セット) と筆順を指定しないデータベース (β セット) の 2 種類のデータベースを新たに作成した． α セットを利用する事により，学習時には方向線分ラベルと文字データの対応画を求める作業を省く事ができ，認識時には性能評価の上で筆順違いの要因を考慮しなくても済むようになる．また β セットは JIS 第一・第二水準漢字の全てを網羅し，異筆順文字や未学習字種などの認識実験に利用できる．

収集環境は図 3 のような linux の X Window System 上に Tcl/Tk で構築された入力キャンバスを表示し，ペンタブレット (Wacom intuos i-400) を用いて入力した．また，条件として

- 入力字種の順は筆記者毎に異なるようにランダムに提示する
- 文字の大きさ，書く領域は特に指定しないが，タ

プレットの回転は認めない
ようにした。

3.1 α セット：筆順の正しい教育漢字

データの収集は北陸先端大の在学学生 108 人を対象に 1999 年 11 月～12 月の期間に行った。対象字種は教育漢字であり、平成 4 年改定前の 996 字と改定後の 1006 字を包含する 1016 字とした。各筆記者にはこれらの全ての文字を一通りずつ筆記してもらった。ただし、筆順が不確かであると判断した字種については続けて再入力を要求した。

この筆順チェックは「筆順指導の手びき」に準拠した漢字辞典 [10] [11] を参考にして視察で確認しながら予め見本となるデータを筆記しておき、収集時には収集文字と見本文字との対応画を求める筆順探索を行う事で自動化した。筆順チェックアルゴリズムは JKanji (漢字練習サーバ) [12] を参考とし、改良を加えて画数の多い漢字でも瞬時に判定できるようになっている。休憩を入れながら筆記してもらったが、連続して入力した場合、書き始めから筆順判定まで含めても、およそ 10 文字/分の速さで入力できる位で、筆順チェックにはほとんど時間を要さない。事前に正しい筆順は教えず、2～3 度書き直しても正しい筆順でなかった場合に筆順見本を表示して覚えてもらい、再度、筆記してもらうようにした。また、見本と比べて字形の変動が著しい場合は筆順チェックが正しく動作しない事もあるので、筆記者に任せて字種毎に筆順チェック機能を解除する事を許した。それらの文字については収集後に視察で確認し、最終的に全ての字種の筆順が正しいデータセットは 96 人分となった。

[筆順チェックアルゴリズム]

1. 画の標本化
参照漢字(筆順の正しいデータ)と入力漢字(字種は既知)、それぞれの各画を等距離間隔で N 等分する。
2. 座標の正規化
サンプル点の (x, y) の最小値を 0, 最大値を 100 になるように拡大/縮小する。文字高と文字幅の比率は無視した。
3. 筆順の探索
参照漢字の画の全ての順列の中から入力漢字の画系列に最も歪みの小さい候補を探索する。全探索は効率が悪いので (m 画の文字の場合, $m!$ の組み合わせ), 画を展開する度に歪みの小さい n 個の候補を残すビーム探索を行う。また、画の方向が正しくないもの(例えば“北”の四画目は、正しくは右から左に筆記するが、左から右に筆記する人もいる)についても同時に判定を行い、

方向の違うものも筆順違いとして処理する。

3.2 β セット：筆順フリーの JIS 第一・第二水準漢字

データの収集は北陸先端大の在学学生 80 人を対象に 2000 年 2 月に行った。このうち α セットと重複する筆記者は 38 人である。対象字種は JIS 第一・第二水準の漢字 6353 字 (JIS X0208-1983) とした。各筆記者はこのうちの 1200～1300 字ずつ分担し、一通りずつ筆記してもらった。収集文字の配分には以下の点に注意した。

- 教育漢字, その他の JIS 第一水準漢字, JIS 第二水準漢字の比率が筆記者間でほぼ等しい。
- 80 人分の合計で 1 字種あたりの文字数は 16 字。
- 同一の字種集合を書く筆記者はいない (任意の 2 筆記者を選んだ場合, 少なくとも 14 個, 字種が異なる。)
- 同一の筆記者集合を持つ字種は無い (任意の 2 字種を選んだ場合, 少なくとも 2 人, 筆記者が異なる。)

3.3 辞書

漢字の辞書は

語	=	言 2 吾
言	=	、 6 一 6 一短 6 一短 6 口
吾	=	五 6 口
五	=	一 5 ノ 3 短 5 一
口	=	短 3 短 5 一短
短	=	a g
一	=	A
一短	=	a
ノ	=	F
短	=	g
、	=	h

のように階層的に定義されており、各字種の構成要素について共通する要素の定義は字種毎に異ならない。また、本稿の実験では最も下の階層である方向線分 (A～G, a～g, 0～8) をモデルの単位としたが、今後、適当な階層 (画, 偏旁冠脚) をモデルの単位として扱う事もできる。

現在もまだ辞書の整備中であり、本報告の評価に使用した辞書^{注1)}では少なくとも新旧の教育漢字 1016 字種に関して、筆順と画数がほぼ正しく登録されている事を特定筆記者を用いた認識実験と視察を繰り返して確認した。また、“田”と“由”のように筆順では区別がつかない字種については

(注1): 便宜上の研究室内部バージョンは Ver.26 (2000 年 3 月 2 日版)

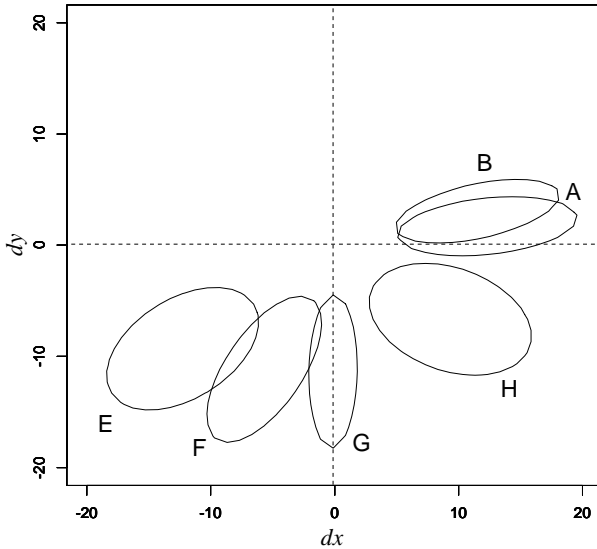


図4 各ストローク HMM の第 2 状態の分布 (確率密度関数の最大値 $\times 0.5$ の等高線)

田 = G 3 A G 4 G 4 A 6 A

由 = g 3 A g 4 G 4 A 6 A

のように字種定義が異なるように長短のモデルを使い分けた。教育漢字以外の字種については誤定義が多々ある事は承知しており、それらについては次の版で修正する。いずれは辞書を人手によらず自動作成する予定である。

4. オンライン手書き文字認識実験

4.1 実験 1: ストローク HMM vs 字種 HMM

階層化辞書の最小の構成要素であるストロークと最大の単位となる字種について、それぞれを HMM の単位とした場合の認識性能の比較を行う。

4.1.1 実験条件

認識カテゴリ	: 新旧教育漢字 1016 字種
学習資料	: α セットの 奇数 ID, 47 筆記者
評価資料	: α セットの 偶数 ID, 49 筆記者
特徴量	: 速度ベクトル (dx, dy)
HMM 共通部	: 全共分散行列単一正規分布型
ストローク HMM	: 3 状態自己遷移あり / pen down 1 状態自己遷移なし / pen up 23 モデル (C, D を除く)
字種 HMM	: 画数 N に対して $4N - 1$ 状態 自己遷移あり, 1016 モデル
学習法	: Viterbi 学習
認識法	: Viterbi 探索 (全探索)

使用データベースは α セットであり, 108 人の筆記者を収集順に a0001 ~ a0108 の ID で管理してい

表 1 ストローク HMM と字種 HMM の 1 位認識率, および 10 位累積認識率の比較 (open / closed は筆記者について)

	1-best open / closed	10-best open / closed
字種 HMM	95.92 / 97.73	99.84 / 99.95
ストローク HMM	94.37 / 94.87	97.38 / 97.62

る。この実験では前述した筆順の正しい 96 筆記者について, 奇数 ID 筆記者を学習資料に, 偶数 ID 筆記者を評価資料にした。奇数, 偶数で筆記者数が異なるのは ID の欠番があるためである。また, 現在の登録字種の定義にはストローク C および D は使用していないので, 実質的なストロークモデルの数は 23 個である。字種 HMM については, 一画あたり 3 状態と画間の移動に 1 状態を割り当て, 画数 N に対して $4N - 1$ 状態の HMM とし, 全ての状態の自己遷移を可とした。

HMM の学習結果について

図 4 は実際に学習した HMM セットのうち, pen down のストロークの長い A ~ H の第 2 状態についてその形状をプロットしたものである。図 1 の定義ではストローク E は右から左への方向線分に対してラベル付けられていたが, 漢字でそのようなストロークが使われる事は無いので, “毛” の一画目などのように比較的水平に近い斜め左下方向の線分などで用いた。それに対し, ストローク F は “人” の一画目のような傾斜の急な線分に割り当てた。また, ストローク A は “一” などの構成要素であるが, 文字を右上がりぎみに筆記する人が多いので, 統計的に右上方向の分布となっている。一方, ストローク B は “虫” の五画目など, やや右上がりのものに用いたため, A と B はかなり類似した分布となっている。

4.1.2 実験結果

ストローク HMM と字種 HMM の認識性能を表 1 に比較する。表中 1-best は 1 位認識率 (正解認識文字数 / 認識文字数) を, 10-best は 10 位までの累積認識率を示す。また, 学習に使用した筆記者の平均認識率を closed の欄に, 評価用の筆記者の平均認識率を open の欄に表記した。全字種を学習に使用しており, 各筆記者は字種あたり 1 文字のみのデータセットであるので, closed の場合, 学習資料と評価資料は同一である。当然ながら closed 認識率はいずれの場合も open 認識率よりも高くなるが, その差は小さいので, 十分な量の資料で学習できていると考えられる。

認識率について

字種 HMM の場合, HMM の各状態が特定の文字

表2 字種 HMM 使用時の認識率ワースト 5 字種 (10-best, open 基準)

字種 HMM						ストローク HMM	
字種	認識候補例 (括弧内は 49 筆記者の 10-best の内数)					10-best	累積認識率
貝 (44)	具 (49)	見 (47)	県 (44)	典 (34)	映 (29)	90%	94%
紀 (45)	絶 (49)	緑 (49)	続 (46)	統 (46)	綿 (41)	92%	92%
早 (45)	車 (47)	軍 (37)	専 (36)	律 (35)	草 (34)	92%	100%
七 (46)	毛 (47)	士 (32)	犬 (29)	亡 (27)	元 (25)	94%	69%
入 (47)	人 (47)	八 (47)	文 (39)	久 (37)	大 (37)	96%	18%

表3 ストローク HMM 使用時の認識率ワースト 5 字種 (10-best, open 基準)

ストローク HMM						字種 HMM	
字種	認識候補例 (括弧内は 49 筆記者の 10-best の内数)					10-best	累積認識率
補 (0)	複 (48)	練 (48)	福 (42)	簡 (38)	観 (36)	0%	100%
評 (1)	誌 (49)	詩 (43)	語 (41)	聞 (33)	試 (32)	2%	100%
垂 (3)	乗 (49)	善 (46)	鉱 (39)	産 (37)	笛 (24)	6%	98%
入 (9)	人 (46)	本 (41)	区 (40)	少 (39)	八 (38)	18%	96%
万 (11)	功 (49)	形 (46)	示 (39)	方 (36)	参 (26)	22%	100%

の特定のストロークの特徴を良く捉えているので、認識率が非常に高い。字種毎の認識率もほとんどの場合が 10 位候補までに正解文字が含まれ、全字種平均の累積認識率が 99.8 % 以上であり、最も認識率の低い字種“貝”でも 90 % であった。一方、ストローク HMM の場合の全字種平均では累積認識率 97 % 以上となり、字種 HMM に比べて 2 ~ 3 % 低い程度であるが、認識率の高い字種と低い字種の差が顕著である。表 2 および表 3 はストローク HMM および字種 HMM を用いた場合の 10 位累積認識率の低い 5 字種について、その認識候補の例をまとめたものである。それぞれ 49 筆記者の 10 位候補までに現れた字種であり、括弧内の数値は、特に認識順位を考えずに矢印()の左のものが正解認識されうる筆記者数、あるいは右の各候補に誤認識されうる筆記者数を表す。字種 HMM を用いた場合には字形が良く似ている字種が候補として現れており、HMM の各状態が各字種のストロークに特化したモデルになっている事が分かる。例えば“紀”の字種 HMM の前半の状態は「糸偏」を良く表していると推測できる。これに対し、ストローク HMM の方は一見すると似ても似つかない候補が出現し得る。これは辞書登録の際のストロークの分類が主観による手作業である事、辞書の階層化によってストロークの長短が実際の長さとは必ずしも一致しなくなっている事、また、連続するストローク間の影響を考えた場合には 25 種類のストロークモデルでは少なすぎる事などが理由として考えられる。

辞書・モデルサイズについて

比較実験は認識カテゴリが 1016 字種であったが、これを 6353 字種に拡張すると、字種を定義する辞書およびモデルを蓄えるのに必要な容量が字種 HMM

では約 7 MB となる事に対し、ストローク HMM ではわずか 115kB 程度に小型化できる。参考までに富士通研究所の提案するオンライン文字認識辞書生成法 [13] では JIS 第一水準および第二水準の一部で 166kB である。字種数とテンプレート数が異なるので直接比較する事はできないが、筆順違いを考慮していないストローク HMM の場合で同程度に小型である事が分かる。

認識速度について

ストローク HMM の場合、予め全ての字種をストロークのネットワークとして展開する事により、字種の共通構成要素の計算が一回で済む。例えば、字種 HMM の認識候補の例の“紀”、“絶”、“緑”、“続”、“統”、“綿”の文字は偏の“糸”までは字種定義が共通であるので傍に関らずスコアは同じであり、一度の計算で良い。現在の 6353 字種のストローク列を木状に展開した場合では理論上約 2 倍の高速化、今回使用したネットワークの場合では約 6 倍以上の高速化が達せられると見込んでいるが、実装に至らなかったため速度比較は行っていない。更には登録されていないストローク列を許容した、効率の良いネットワークが構築できると考えている。

以上のように、字種 HMM とストローク HMM では長短があり、今後はその中間の階層に位置する認識単位の最適化、および前後のストロークの影響を考慮した環境依存型のモデル作成が必要である。

4.2 実験 2: 未学習字種認識

ストローク HMM の利点として、どのような文字でも (例え図形であろうとも) 辞書の追加登録のみで対応できる点がある。ここでは 1016 字種から学習した HMM を用いて、学習に使用されなかった字

表 4 学習字種および未学習字種の認識率の比較 (open / closed は筆記者について)

	1-best open / closed	10-best open / closed
認識カテゴリ 1016 字種, 筆順固定		
α セット学習字種	94.10 / 95.40	97.08 / 97.74
認識カテゴリ 6353 字種, 筆順自由		
β セット学習字種	82.31 / 81.92	86.90 / 86.14
β セット未学習字種	79.75 / 78.46	86.05 / 85.00

種, すなわち未学習字種を認識する実験を行う。

4.2.1 実験条件

認識カテゴリ : 6353 字種 (JIS X0208-1983)

評価資料 : β セット中, α セットでの
収集経験のある 38 筆記者
奇数 ID, 17 筆記者
偶数 ID, 21 筆記者

実験には筆順フリーで JIS 第一・第二水準の全字種を網羅した β セットを使用する。学習資料は実験 1 と同様に α セットの奇数 ID の 47 筆記者を用いた。したがって教育漢字を除く 5337 字種が未学習字種である。

4.2.2 実験結果

表 4 に実験 2 の結果を実験 1 の結果と併わせて示す。そもそも、認識カテゴリ数が異なるので、直接比較する事はできないが、認識率の差のその他要因として筆順違いの影響と収集環境の違いがある。筆順については、収集時期に 2 ヶ月の間があるため、ほとんどの人は本来自分が書いていた筆順で書いたと思われる。また、収集環境については、評価用筆記者の全員が入力経験がある事と、データ収集ツールから筆順チェック機能を除いた事により、筆記速度が上がった文字もあれば、見慣れない JIS 第二水準漢字では筆記速度が落ちるなど、特徴量の変動が大きくなった事が考えられる。これらの要因が関わらない β セット内で学習字種と未学習字種の認識率を比較すると、若干、未学習字種の方が認識率が低い程度である。現在の辞書では教育漢字以外については整備が不十分であるので、この差もおそらく辞書の修正次第で解消すると思われる。

4.3 実験 3: 筆記者適応実験

大勢の資料で学習した不特定筆記者モデルは平均して良好な認識率となるが、筆記者が限定される場合は、その特定筆記者に合わせたモデルを用いた方が一般に認識精度が良い。しかしながら、特定筆記者について大量の学習資料を用意するのは手間を要するので、できるだけ少量のサンプルから特定筆記者

表 5 筆記者あたりの適応字種に含まれるストロークの出現頻度

方向 (down)	A	B	E	F	G	H			
出現頻度	16	0	0	4	14	1			
方向 (down)	a	b	c	d	e	f	g	h	
出現頻度	7	0	1	2	0	3	6	8	
方向 (up)	1	2	3	4	5	6	7	8	0
出現頻度	0	8	9	7	8	10	0	0	0

の筆跡の特徴を学習し、不特定筆記者のモデルを変形させる筆記者適応技術が必要となる。ここでは最大事後確率 (MAP) 推定法を用いた筆記者適応実験を行う。

筆記者適応原理

MAP 適応法では HMM の状態 S_j の平均値 μ_j を適応しようとする筆記者で観測される平均値 $\bar{\mu}_j$ を用いて、重みを付けた修正を行う。すなわち、適応後の平均 $\hat{\mu}_j$ は

$$\hat{\mu}_j = \frac{N_j}{N_j + \tau} \bar{\mu}_j + \frac{\tau}{N_j + \tau} \mu_j$$

で定義され、ここで N_j は状態 S_j が占める適応データのフレーム数であり、 $\bar{\mu}_j$ はそれらのフレームの特徴量の平均値である。 τ は事前知識の確からしさに関する係数である。

4.3.1 実験条件

適応法	: MAP ($\tau = 15.0$)
初期モデル	: α セットの奇数 ID, 47 筆記者
評価用筆記者	: α セットの偶数 ID, 49 筆記者
適応字種	: “一右雨円王音下火花貝”
評価字種	: 適応字種を除く 1006 字種

予備実験により事前知識に関する係数を $\tau = 15.0$ に固定した。今回使用した適応字種はいずれも小学一年で学ぶ漢字であり、JIS コード順に先頭の 10 字種を選んだ。この 10 字種に含まれるストロークの内訳は表 5 の通りであり、本適応法では個数が 0 のストローク (B, E, b, e, 1, 7, 8, 0) については適応前後でモデルのパラメータは変化しない。また、比較実験として個々の筆記者のデータのみから学習した特定筆記者 HMM による認識実験も行った。

4.3.2 実験結果

評価用筆記者のうち適応前の認識率が低い 5 人についての筆記者適応実験の結果を表 6 に示す。ちなみに適応前の実験条件は実験 1 と同じであるが、表 1 の数値と若干異なるのは母数が適応データを除く 1006 字となるためである。

表 6 適応前の認識率ワースト 5 筆記者 (open) に対する適応後の認識率

筆記者	筆記者適応前		筆記者適応後		特定筆記者 HMM	
	best-1 /	~10	best-1 /	~10	best-1 /	~10
a0052	88.97 /	95.33	91.35 /	96.82	95.83 /	98.11
a0072	90.16 /	95.53	92.45 /	96.12	94.33 /	97.51
a0012	91.25 /	95.92	93.44 /	96.92	95.63 /	97.71
a0056	91.35 /	96.22	92.74 /	96.62	95.03 /	97.91
a0062	91.35 /	96.02	93.24 /	96.82	94.83 /	97.81
49 人平均	94.37 /	97.37	94.94 /	97.67	95.32 /	97.74

平均認識率では 3 種類のモデル (筆記者適応前, 適応後, 特定筆記者の HMM) による認識率の差が僅か 1 % 未満である。元々筆記者適応を行わなくとも高い認識率を挙げていた筆記者も含まれるため、このような筆記者については適応後の認識率が逆に下がる事や、特定筆記者モデルを用いても認識率がほとんど上がらない事もあり得る。筆記者間のストロークの変動よりも筆記者内の字種間のストロークの変動の方が大きいと考えられ、多くの筆記者の場合には、より多くの資料で学習された不特定筆記者 HMM と自分の資料のみで学習した特定筆記者 HMM では大差が無い傾向にある。ただし、認識率の低い筆記者については適応後の認識結果は適応前と特定筆記者 HMM のほぼ中間の値をとり、適応効果が現れている事が分かる。

5. ま と め

我々の提案する文字の構成要素を単位としたオンライン手書き文字認識方式のベースラインとなる認識性能について報告した。また、字種モデルに比べて、辞書の小型化、認識速度の高速化、未学習字種への対応、筆記者適応などの面で有利である事を示した。今後はストロークと字種の間接的な認識単位の検討、前後のストロークを考慮した環境依存型のモデル、データ収集環境によらない特徴量の正規化、字種の多重登録による筆順違いへの対応、客観的な辞書の自動作成など、様々な研究課題が挙げられ、更なる高精度化を目指す。

謝辞 文字データ収集を実施して頂いた北陸先端大修士生の川口弘光氏、ならびに文字データ収集に協力して頂いた学生諸氏に感謝する。なお、本研究は文部省科学研究費補助金 (基盤研究 B) 課題番号 11480074 の助成を受けて行われた。

文 献

- [1] R. Nag, K. H. Wong and F. Fallside, "Script Recognition Using Hidden Markov Models," *Proc. ICASSP '86*, vol.3, pp.2071-2074 (1986-04).
- [2] Amlan Kundu and Paramvir Bahl, "Recognition of Handwritten Script: A Hidden Markov Model Based Approach," *Proc. ICASSP '88*, vol.2, pp.928-931 (1988-04).
- [3] Thad Starner, John Makhoul, Richard Schwartz and George Chou, "On-Line Cursive Handwriting

- Recognition Using Speech Recognition Methods," *Proc. ICASSP '94*, vol.5, pp.125-128 (1994-04).
- [4] Jianying Hu, Michael K. Brown and William Turin, "HMM Based On-Line Handwriting Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.18, no.10, pp.1039-1045 (1996-10).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] ケーレ・ローデリック, 松本隆, "HMM オンライン手書き文字認識における Pruning アルゴリズムについて," 信学技報 PRMU97-5, pp.33-39 (1997-05).
- [9] 嵯峨山茂樹, 中井満, 下平博, "ストローク HMM に基づくオンライン手書き文字認識方式," 信学技報 PRMU2000 (2000-06).
- [10] 江守賢治編, "楷行草 筆順便覧," 三省堂 (1999).
- [11] "小学漢字 1006 字の正しい書き方," 旺文社.
- [12] <http://fl176.hyper.chubu.ac.jp/wwkanji/JKkanji/>
- [13] 秋山勝彦, 石垣一司, "学習標本による高品質なオンライン文字認識辞書生成の手法," 信学技報 PRMU99-235, pp.31-36 (2000-02).
- [14] C. H. Lee, C. H. Lin and B. H. Juang, "A Study on Speaker Adaption of the Parameters of Continuous Density Hidden Markov Models," *IEEE Trans. Signal Processing*, vol.39, no.4, (1991-04)