

# 空中手書き文字入力における DP マッチングと HMM の比較に関する研究\*

西田 好宏<sup>\*1</sup>, 花房 大佑<sup>\*2</sup>

## Studies on the Comparison between DP Matching and HMM for Aerial Handwritten Hiragana Character Recognition

Yoshihiro NISHIDA <sup>\*1</sup> and Daiyuu HANAFUSA <sup>\*2</sup>

<sup>\*1</sup> Faculty of Engineering, Department of Electrical and Electronic Engineering

This paper presents a method to recognize a character handwritten in the air. We had prototyped an aerial handwritten hiragana character recognition system that detect single character period by hovering in the air using DP (dynamic programming) matching. In this time, we investigate an aerial handwritten character recognition system using Hidden Markov Model (HMM). HMM theory is an extension of the Markov Model process. It has found uses in such areas as speech recognition, target tracking and word recognition. We implemented HMM using Baum–Welch algorithm. The Baum–Welch algorithm uses the well known EM algorithm to find the maximum likelihood estimate of the parameters of a hidden Markov model given a set of observed feature vectors. We compared correct recognition rate of using DP matching and using HMM. We estimated features of both methods.

**Key Words** : Character Recognition, Handwriting in the Air, DP Matching and HMM(Hidden Markov Model)

### 1. 緒 言

クラウド・コンピューティングや SNS の普及により、いつでも、どこでもふと思いついたアイデアなどをメモとしてクラウドに保存したり、出来事を呟きあったりする機会が増えている。そのためにはテキストデータの入力が必要不可欠であるが、パソコンのキーボードによる文字入力には両手での操作が必要になる。また音声認識は、公共施設や騒音が激しい場所では利用が困難な場合がある。そのような背景の中、新たな文字入力方法の選択肢の一つとして、相対的な移動方向情報のみを利用した空中手書き文字入力システムの開発を行っている<sup>(1)</sup>。

空中手書き文字入力システムは、①文字を書く動きを検出するセンサ、②相対的な移動方向情報から文字を認識するエンジン、③認識した文字を変換・出力するアプリケーションで構成される。

今回は②の文字認識エンジンについての検討を行う。先行研究では、文字認識は端点フリー DP マッチングを用いて行われていた。理由としてはアルゴリズムが容易であること、文字を 8 種類のコードのみで表現するため、比較的単純な DP マッチングでも十分な認識率が期待できること、端点フリーの為、文字の書き始めのノイズを軽減できることが挙げられる<sup>(2)(3)</sup>。本研究ではこれとは異なる認識方法として、隠れマルコフモデル (Hidden Markov Model : HMM) を用いたマッチング方法の開発・検討を行う。DP マッチングと隠れマルコフモデルを用いたマッチング方法を比較することで、それぞれの利点を把握して、空中手書き文字入力システムの更なる発展を目標とする。

### 2. 隠れマルコフモデルによるマッチング

隠れマルコフモデル (Hidden Markov Model : HMM) を用いたマッチングは、音声認識において広く利用されている。実際の音声には、発声の時間的な揺らぎと、発声や音質の違いといったスペクトル的な揺らぎが含まれ

\* 原稿受付 2018 年 2 月 28 日

<sup>\*1</sup> 工学部, 電気電子工学科

<sup>\*2</sup> 工学部, 電気電子情報科学学生

E-mail: nishida@fukui-ut.ac.jp

る. 隠れマルコフモデルには, そのような揺らぎを吸収しやすいという利点がある. 手書き文字入力においても, 音声と同じように筆記者や筆記方法による揺らぎが生じる特徴があるので, 隠れマルコフモデルを文字認識に用いる研究が行われている<sup>(4)(5)</sup>.

隠れマルコフモデルは, 確率付き有限状態オートマトンと同じパラメータを持つ. モデルと入力記号のマッチングは, 入力記号を生成する確率によって求められるが, 任意の時刻にどの状態に居るのかを定義できない.

隠れマルコフモデルのパラメータは, ①初期状態確率  $\pi_i$ , ②最終状態, ③状態遷移確率  $a_{i,j}$ , ④シンボル出力確率  $b_i(O)$  から定義されるが, 音声認識や文字認識で使用する Left-to-Right HMM では, 初期状態確率  $\pi_1=1$ , 最終状態は 1 つとなる. Fig. 1 に状態数 2, シンボルの種類が 2 の Left-to-Right HMM を示す.

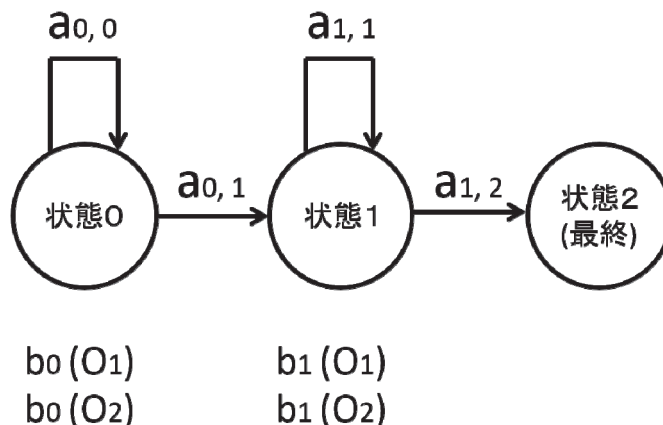


Fig.1 An example of Left-to-Right HMM

状態遷移確率  $a_{i,j}$  は, 状態  $i$  から状態  $j$  に遷移する確率を意味し, ある状態  $i$  から遷移する可能性のある全ての状態  $j$  への, 状態遷移確率の総和は 1 となる. シンボル出力確率  $b_i(O)$  は状態  $i$  におけるシンボル (特徴を現す値) の出力確率を表す. 状態遷移確率と同様に, ある状態  $i$  から遷移するときに出力される可能性のある全てのシンボルの, シンボル出力確率の総和は 1 になる.

隠れマルコフモデルを用いた認識は Fig.2 に示すように行う. まず認識対象の情報を基に作成されたモデルから, 入力されたシンボル系列が出力される確率をそれぞれ求める. 次にそれらと比較して最も高い確率を出力したモデルを, 入力されたシンボル系列と推定する. シンボル系列とモデルが一致する確率を尤度, 尤度を比較して最も尤もらしい情報を推定することを, 最尤推定という.

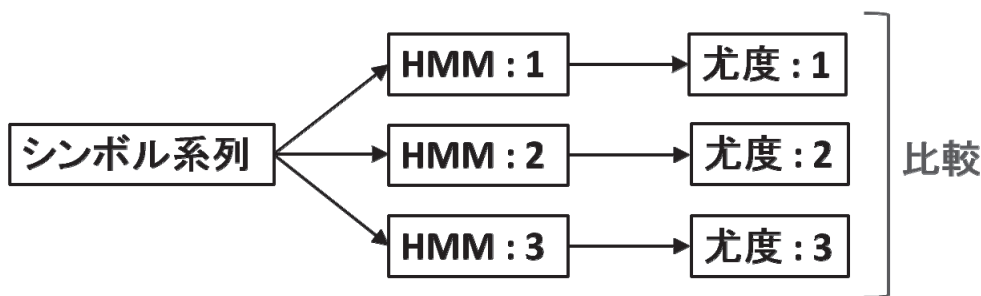


Fig.2 Recognition Algorithm using HMM

### 3. 隠れマルコフモデルの作成

#### 3.1 パラメータの設定

隠れマルコフモデルの作成は、状態数、状態遷移確率、シンボル出力確率等のパラメータを決定していく必要がある。今回は、認識対象の文字をひらがな（46文字）と数字（10文字）とし、DP マッチングで使用していた辞書データを流用して、文字ごとの状態遷移確率及び状態ごとのシンボル出力確率を算出した。なお、これらの算出において、状態数は5と仮定し、Baum-Welch アルゴリズム<sup>6)</sup>の繰り返しの回数は20回とした。

全56文字に対して隠れマルコフモデルを作成した。例として「あ」のパラメータを Fig.3 に示す。

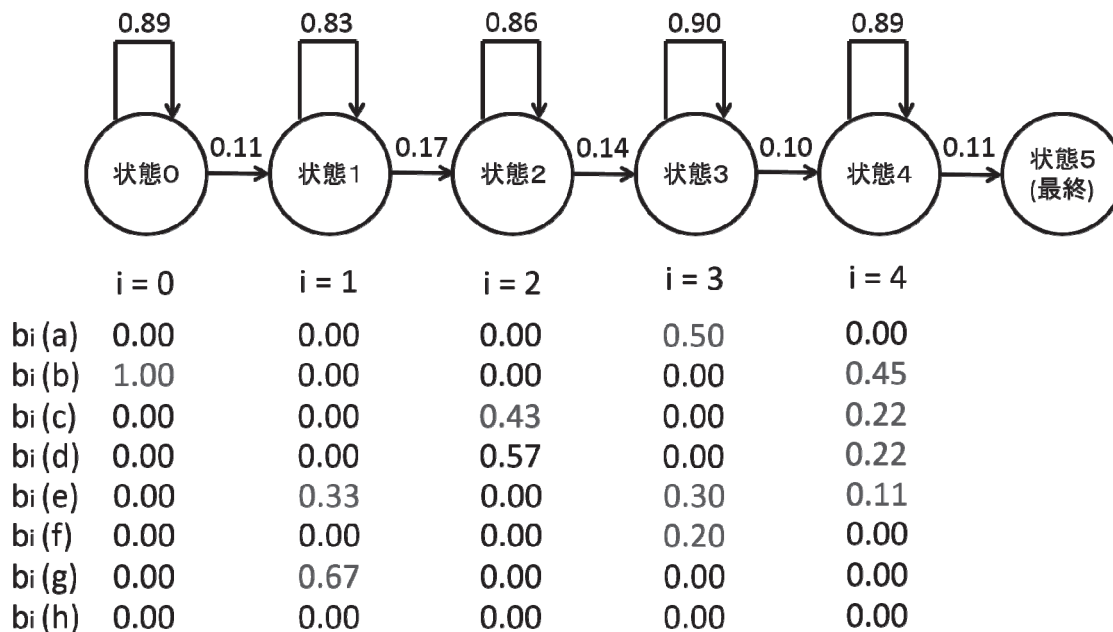


Fig.3 An example of structure and state of character “あ”

このように作成した隠れマルコフモデルを用いて、被験者1名に認識対象の56文字を筆記してもらい認識率を測定した結果を Table 1 に示す。「第一候補」は入力文字と第一候補として出力された文字が一致した割合を、「～第三候補」は第一候補、第二候補、第三候補のうち、どれかひとつが入力文字と一致した割合を示す。

Table 1 Initial correct recognition rate

	第一候補	～第三候補
被験者 A	0.18	0.23

Table 1 に示す通り、文字の認識率は非常に低かった。理由としては、隠れマルコフモデルの尤度推定に使用した forward アルゴリズムは、状態遷移確率とシンボル出力確率の積を合計するため、値に0が含まれる場合は尤度が大幅に低下してしまうからだと考えられる。

認識率を改善させるために、シンボルの出力確率の最低値を0より大きな値に設定することにした。また、ある状態においてシンボル出力確率が「1」の場合、近接する2つのコードの出力確率に特定の数値を加えることで、文字の崩れや筆記者の癖に対応できるのではないかと考えた。前者の加算値を add X、後者の加算値を add Y とし、それぞれの値が add X = (0.001, 0.005, 0.01, 0.05), add Y = (0.0, 0.1, 0.2, 0.3) の場合において、認識率が最も高くなるパラメータを検討した。

この時のパラメータの値の様子を Fig.4 に、パラメータを加算した際の認識率を Table 2 に示す。パラメータに値を加えることにより、認識率が大幅に向上した。表より、add X = 0.005, add Y = 0.1 を採用した。

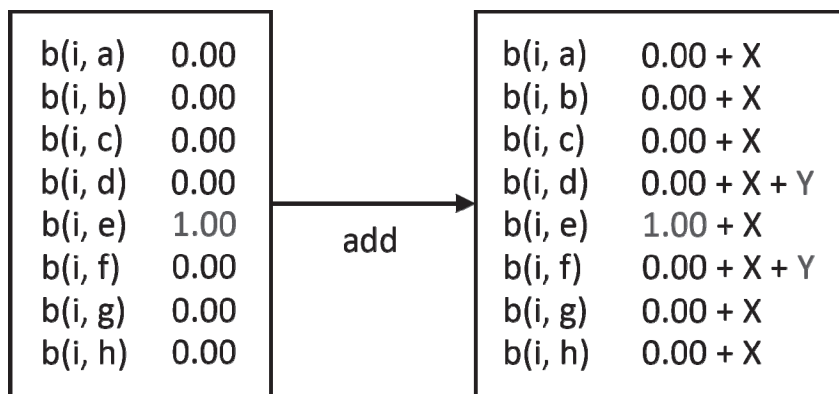


Fig.4 Parameter tuning

Table 2 Correct recognition rate with parameter tuning

被験者 A 状態数 5		add X							
		0.001		0.005		0.01		0.05	
		第一候補	～第三	第一候補	～第三	第一候補	～第三	第一候補	～第三
add Y	0.0	0.68	0.89	0.73	0.89	0.75	0.89	0.75	0.88
	0.1	0.75	<b>0.93</b>	0.79	<b>0.93</b>	<b>0.80</b>	0.89	0.79	0.88
	0.2	0.71	<b>0.93</b>	0.79	0.89	0.78	0.89	0.75	0.86
	0.3	0.68	0.89	0.77	0.88	0.75	0.88	0.71	0.84

### 3.2 状態数の設定

次に状態数が 5, 7, 9, 11, 13, 15 の隠れマルコフモデルを作成し、モデルごとの認識率を比較する。コード化・評価方法、加算パラメータ (add X, add Y) は 3.2 と同様である。

また、モデルの状態数が増えれば認識率は向上するが、認識に必要な計算回数が増えることが考えられる。状態数を抑えながら認識率を上げるために、文字ごとに異なる状態数を持つモデルも作成した。具体的には、Fig.5 に示すように、文字に含まれる同一な筆記方向の纏まりを基に状態数を定めた。例えば「1」は同一な筆記方向の纏まりが 1 つ、「く」は纏まりが 2 つ、「い」は纏まりが 3 つと考えて状態数を設定した。同一な筆記方向の纏まりの数をそのまま状態数とした「flex\_1」と、同一な筆記方向の纏まりの数を 2 倍したものを状態数とした「flex\_2」の 2 種類のモデルを作成した。

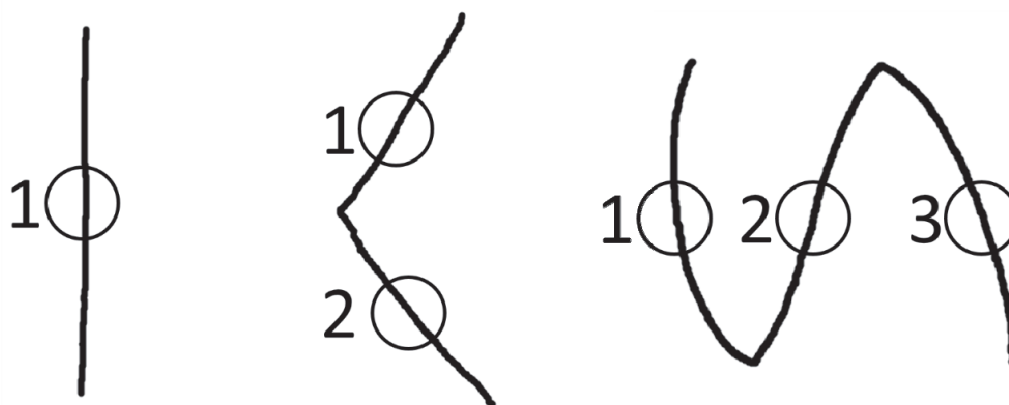


Fig.5 Number of state (direction groups)

状態数ごとの認識率の結果を Table 3 に示す. 状態数に応じて認識率の向上が見られたが, 状態数 13 以降は大きな向上は見られなかった. また, 文字の筆記方向の種類に応じて作成したモデル「flex\_1」「flex\_2」は, 特に目立った成果を得ることが出来なかった. 認識率, 総状態数, モデル化の手間を考慮して, 今回の検討では状態数「11」のモデルを採用した.

Table 3 Correct recognition rate of each state

X : 0.005 Y : 0.1	状態数(総状態数)					
	5(285)		7(399)		9(513)	
	第一候補	～第三候補	第一候補	～第三候補	第一候補	～第三候補
被験者 A	0.79	0.90	0.86	0.91	0.91	0.95
被験者 B	0.73	0.93	0.84	0.93	0.84	0.95
被験者 C	0.68	0.82	0.79	0.93	0.82	0.93
平均値	0.73	0.88	0.83	0.92	0.86	0.94

X : 0.005 Y : 0.1	状態数(総状態数)					
	11(627)		13(741)		15(855)	
	第一候補	～第三候補	第一候補	～第三候補	第一候補	～第三候補
被験者 A	0.93	0.96	0.93	0.96	0.93	0.96
被験者 B	0.88	0.96	0.88	0.96	0.86	0.96
被験者 C	0.82	0.95	0.80	0.96	0.88	0.96
平均値	0.88	<b>0.96</b>	0.87	<b>0.96</b>	<b>0.89</b>	<b>0.96</b>

X : 0.005 Y : 0.1	状態数(総状態数)			
	flex_1(369)		flex_2(738)	
	第一候補	～第三候補	第一候補	～第三候補
被験者 A	0.84	0.96	0.93	0.96
被験者 B	0.78	0.93	0.86	0.98
被験者 C	0.77	0.91	0.84	0.93
平均値	0.80	0.93	0.88	<b>0.96</b>

### 3.3 パラメータの再設定

状態数が 11 と定まったので, 更なる認識率の向上を目指して, 加算パラメータ (add X, add Y) の再設定を行った. 始めに add X = (0.001, 0.005, 0.01), add Y = (0.0, 0.1, 0.2, 0.3, 0.4) を組み合わせた複数の隠れマルコフモデルを作成する. 次に 3.2 節の実験で収集した筆跡データを使用して, モデルごとの認識率を比較した.

今回は, 加算パラメータと認識率の明らかな関係を見出すことが出来なかったが, どの被験者・候補においても比較的認識率が良かった add X = 0.001, add Y = 0.2 を採用することにした.

#### 4. DP マッチングと隠れマルコフモデルの比較

上記の通り作成した隠れマルコフモデルを用いた文字認識と、従来の DP マッチングを用いた文字認識の比較を男女9名（被験者 A~I）に対して行った結果を Table 4 に示す。

Table 4 Correct recognition rate by both methods

	回	DP マッチング		隠れマルコフモデル	
		第一候補	～第三候補	第一候補	～第三候補
A	1	0.95	1.00	0.89	0.95
	2	0.95	1.00	0.89	0.98
	3	0.93	1.00	0.88	0.98
B	1	0.91	0.96	0.86	0.98
	2	0.93	1.00	0.79	1.00
	3	0.91	0.98	0.84	0.95
C	1	0.95	0.96	0.84	0.98
	2	0.86	0.95	0.70	0.91
	3	0.95	0.96	0.84	0.93
D	1	0.95	1.00	0.86	0.98
	2	0.93	0.98	0.80	0.91
	3	0.95	1.00	0.86	0.98
E	1	0.86	0.93	0.82	0.91
	2	0.91	0.98	0.88	0.96
	3	0.89	0.98	0.82	0.98
F	1	0.84	0.96	0.75	0.89
	2	0.75	0.91	0.75	0.89
	3	0.80	0.96	0.73	0.98
G	1	0.89	0.95	0.77	0.95
	2	0.89	0.98	0.84	0.96
	3	0.91	1.00	0.82	0.98
H	1	0.86	0.96	0.82	0.95
	2	0.77	0.91	0.70	0.86
	3	0.84	0.95	0.71	0.88
I	1	0.77	0.95	0.80	0.91
	2	0.86	0.98	0.80	0.96
	3	0.80	0.93	0.79	0.93
	平均	0.88	0.97	0.81	0.95

第一候補として出力された文字の平均認識率は、DP マッチングを用いた場合は 0.88 で、隠れマルコフモデルを用いた場合は 0.81 となった。第一候補から第三候補までの合計の認識率は、DP マッチングを用いた場合は 0.97 で、隠れマルコフモデルを用いた場合は 0.95 となった。どちらの場合においても、DP マッチングを用いた場合の認識率が高く、特に第一候補の差が顕著であった。

## 5. 検討と考察

### 5.1 共通の特徴

DP マッチングを使用した文字認識と、隠れマルコフモデルを使用した文字認識で、共に認識率が高い文字は、「い」「か」「く」「け」「し」「せ」「そ」「つ」「に」「ひ」「へ」「ほ」「ゆ」「ろ」「ん」「2」「8」であった。認識率が高い理由として、画数が少ないためコードが単調になりやすい、直線で構成されるため筆記者の癖が出にくい、書き順を間違えにくいいため誤入力が少ない、他に似た文字が存在しないため誤認識されにくいことが考えられる。

共通して認識率が低い文字は、「う」「ち」「ね」「ら」「0」「3」「7」「9」であった。認識率が低い理由として、他に似た文字が存在するため誤認識されやすい事が考えられ、全ての文字がこれに該当した。具体的に「う」「ち」「ら」「3」は「ろ」に、「ね」は「れ」に、「0」は「6」に、「7」は「り」と「の」に、「9」は「4」に間違えられることが多かった。

### 5.2 異なる特徴

DP マッチングを用いた文字認識では文字は 8 割以上の認識率を得られたが、隠れマルコフモデルを用いた文字認識では 7 割以下の認識率しか得られなかった文字は「あ」「き」「ぬ」「よ」「わ」の 5 文字であった。隠れマルコフモデルで誤出力された文字に注目すると、「あ」は「す」と「め」に、「き」は「ま」に、「ぬ」は「め」に、「よ」は「た」と「せ」に、「わ」は「め」と「れ」に間違えられる場合が多く見られた。これらの共通点として、文字の筆跡に「回転」が含まれることが挙げられる。

認識率が低い理由として、隠れマルコフモデルの場合、同じ状態のシンボル出力確率は、出力するシンボルの順番までは区別できないことが考えられる。つまり、文字が回転する筆跡が一つの状態の中に含まれる場合は、回転の方向や開始場所の特定が困難である。「あ」と「す」、「き」と「ま」の様に片方の文字に含まれる回転の方向や開始場所を変えると、もう片方の文字に似るのはそのためである。

この弱点を改善するために、隠れマルコフモデルで認識率が低かった文字「あ」「き」「ぬ」「よ」「わ」の状態数を増加させた。具体的には、「あ」「き」「ぬ」「よ」「わ」の状態数を 22、他の文字の状態数を 11 としたモデルを作成し、全ての文字の状態数が 11 のモデルと認識率を比較した。筆跡データと認識方法は第 4 章と同一のものを使用した。その結果を Table 5, Table 6 に示す。

Table 5 Improvement of correct recognition rate

	あ	き	ぬ	よ	わ
全ての状態数 11	0.44	0.67	0.33	0.67	0.37
一部の状態数 22	0.93	0.74	0.85	0.85	0.26

Table 6 Keeping of correct recognition rate

	す	せ	た	ま	め	れ
全ての状態数 11	0.89	1.00	0.74	0.81	0.85	0.89
一部の状態数 22	1.00	1.00	0.78	0.81	0.85	0.89

Table.5 に示すように、隠れマルコフモデルにおいて認識率が低い文字の中で、「わ」以外の認識率の向上を確認できた。一方 Table.6 に示すように、認識率の低い文字の代わりに出力されやすい文字「す」「せ」「た」「ま」「め」「れ」の認識率の低下は見られなかった。

これらのことから、隠れマルコフモデルを用いた文字認識の利点として、文字の選ばれやすさを文字ごとに設定できることが挙げられる。

## 6. 結 言

隠れマルコフモデルを作成して、DP マッチングを用いた文字認識との比較を行った。結果は、認識率やアルゴリズムの明快さ、モデル化の手間など多くの面で DP マッチングの方が上だった。理由としては 8 種類のコード情報のみ比較する場合、隠れマルコフモデルの強みを生かせなかったこと、隠れマルコフモデルの辞書データを DP マッチングのために作成されたものを流用したことが考えられる。しかし、隠れマルコフモデルならではの弱点や、双方のモデルに共通する課題が見つかることが出来たので、どちらのモデルを使用する場合においても、更なる認識率の向上が期待できる。今後は辞書データの改善、隠れマルコフモデルに適した辞書データの作成、誤認識の学習などに取り組みたい。

今回の検討では、空中手書き文字入力システムの中の文字の認識方法のみに着目した。そのため実際の入力方法や、文字区切りの方法、使い勝手向上のための特殊キーなどは考慮しなかった。これからは認識方法以外にも目を向け、空中手書き文字入力システムの更なる発展を目指したい。

## 文 献

- (1) 西田好宏, “二次元移動方向に基づく空中手書きひらがな文字認識に関する研究”, 和歌山大学大学院博士後期課程学位論文, 甲第 44 号 (2011).
- (2) 吉田大志, “センサを利用した空中手書き文字入力システム及び文字区切りに関する研究”, 福井工業大学大学院博士前期課程学位論文 (2014).
- (3) 油尾尚輝, “空中手書き入力において認識可能な文字追加および利便性向上の研究”, 福井工業大学卒業論文 (2016)
- (4) 野中利成, 小沢慎治, “Hidden Markov Model を用いたオンライン手書きひらがな認識”, 電子情報通信学会論文誌 Vol. J74-D-II, No.12, pp1810-1813 (1991).
- (5) 中井満, 嵯峨山茂樹, 下平博, “サブストローク HMM を用いたオンライン手書き文字認識”, 電子情報通信学会論文誌 Vol. J88-D-II, No.9, pp1825-1835 (2005).
- (6) 村上仁一, “Baum-Welch アルゴリズムの動作と 応用例” 電子情報通信学会, 基礎・境界ソサイエティ Vol. 4, No. 1, pp.48-56 (2010).

(平成 30 年 3 月 31 日受理)