

## オンライン手書き文字認識の最新動向

Recent Trends in Online Handwritten Character Recognition

朱 碧蘭 中川正樹

### A bstract

筆記過程の情報に基づいて手書き文字パターンを認識するオンライン手書き文字認識は、スマートフォンやタブレット端末の爆発的普及とともに、今後広く実用化されていくことが期待されている。筆記済みの画像情報を用いるオフライン文字認識との共通点や相違点等も含めて、オンライン手書き文字認識技術の最新動向について解説する。  
キーワード：手書き文字認識，オンライン認識，文字列認識，確率モデル

#### 1. はじめに

近年、スマートフォンやタブレット端末の爆発的普及に伴い、筆記過程から手書きを認識するオンライン手書き認識に再び注目が集まっている。実は、オンライン手書き認識のプラットフォームはそれ以前から多様化が進み、ペン PC、ゲーム機、そして、実際の紙にペンで書かれた筆跡を時系列情報として取り込むデバイスも数種類実用化されている。大形の電子ボードも含めると、直接指示・直接操作、そして、書（描）ける情報環境は着実に拡大している。

一方、そのアプリケーションに目を向けると、文字入力、帳票入力、アノテーションやメモ、デザイン、学習・教育、ゲーム、医療（電子カルテ）、電子書籍、コミュニケーション、情報発信等に及ぶ。

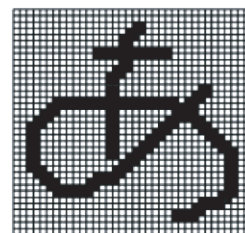
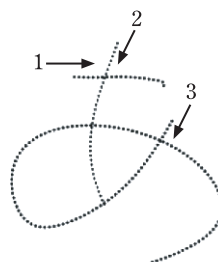
オンライン手書き認識は、いわゆるアナログ行為としての手書きを情報処理できる対象に変換するデジタル化のプロセスである。1964年にRANDタブレットが発明されて以来、英数字の認識から始まり、1970年代、1980年代には日本語の文字認識も多く発表され、1984年には手書きワープロ、1990年代には携帯情報端末、そして、ペン PC と何度かの波を経験してきた。この間

に認識技術は着実に進歩してきている。

本稿では、オンライン手書き認識のうち、特に手書き日本語文字認識に焦点を絞り、2000年頃までの研究紹介は先行文献(1)~(3)に委ね、最近の技術動向や、先行文献で取り上げられなかった実用化にまつわる課題とその克服について、平易な解説を試みる。

#### 2. オンライン認識とオフライン認識

時系列の筆点座標列から手書きを認識する手法をオンライン認識と呼び、画像から認識するオフライン認識と対比される。図1はそれぞれの手法が対象とするパターンを示し、表1は各手法の長所・短所を示している。しかし、オンライン認識では、筆点座標列を画像情報に変換してオフライン認識を適用することも可能であり、両者を融合することもできる。その先駆的な発表が田中らによりなされている<sup>(4)</sup>。また、どちらかの認識方式に、他方で用いる特徴を含める（例えば、オンライン認



(a) オンラインパターン

(b) オフラインパターン

図1 オンライン手書きパターンとオフライン手書きパターン

朱 碧蘭 正員 東京農工大学大学院情報工学科

E-mail zhubilan@cc.tuat.ac.jp

中川正樹 正員：フェロー 東京農工大学大学院情報工学科

E-mail nakagawa@cc.tuat.ac.jp

Bilan ZHU, Member and Masaki NAKAGAWA, Fellow (Department of Computer and Information Science, Tokyo University of Agriculture and Technology, Koganei-shi, 184-0012 Japan).

電子情報通信学会誌 Vol.95 No.4 pp.335-340 2012年4月

©電子情報通信学会 2012

表1 オンラインとオフライン認識手法の長所・短所

手法	長所短所	長所	短所
オンライン認識手法	筆画の続けや崩しに頑健		筆順誤り, 重ね書きに不安定
オフライン認識手法		筆順誤り, 重ね書きに影響されない	続けや崩しに不安定

識に筆点近傍の画像特徴を入れる, または, オフライン認識に運筆の方向情報を入れる) ことも行われている。

オフライン認識手法としては, 欧米では音声認識で発達した隠れマルコフモデル (HMM: Hidden Markov Model) が主流であるが, 漢字文化圏では, 木村らによって提案された修正二次識別関数<sup>(5)</sup>が主流である。

### 3. 文字ごとの認識

オンライン手書き文字認識では一般的に, まず最初に, 入力パターンを標準サイズに線形正規化あるいは非線形正規化<sup>(6)</sup>し, 次に, 筆点列から特徴点を抽出することが多い。それには Ramner の方法<sup>(7)</sup>やその変形が利用される。例を図2(a)に示す。まず, 各ストロークの最初と最後の筆点を特徴点として抽出する。そして, 2隣接特徴点の連結線からそれらの中にある筆点への距離を求め, しきい値を超える最遠の筆点を特徴点として選択する。全ての2隣接特徴点に対して, その連結線からその間の筆点への距離がしきい値を超えなくなるまで特徴点抽出を再帰的に繰り返す。

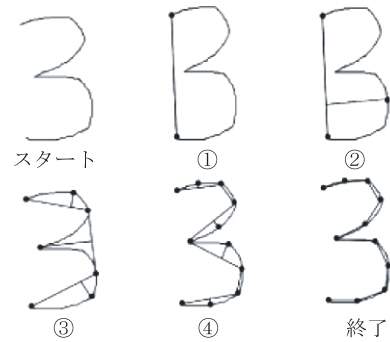
特徴点を抽出した後, 図2(b)に示すように特徴点の対応付けモデル化と認識を行う (図は一番単純化した一面の場合を示す。多画の場合は, それらを連結して, 運筆が表出する確率を考慮した対応付けを行う)。

特徴点の対応付けには, DP マッチング, そして HMM<sup>(8)-(10)</sup>が使われてきた。しかし, 特徴点間の関係を考慮しないため, その効果には限界があった<sup>(11)</sup>。

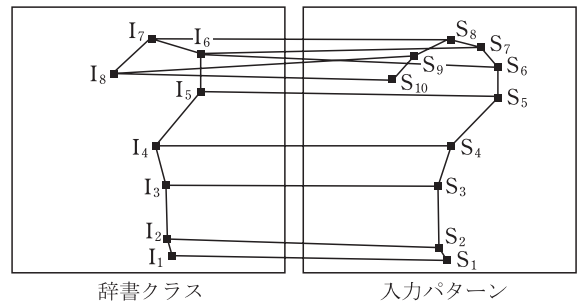
MRF (Markov Random Field) は HMM を包含するモデルであり, 対応付けの評価において, 筆点間の2項間, 3項間等の特徴も確率的に考慮できる<sup>(12)</sup>。MRF は既にオフライン手書き文字認識とオンラインのストローク (筆画) の分類において成功を取っている<sup>(13), (14)</sup>。

Cho らは BN (Bayesian Network) に基づくオンライン手書き文字認識手法とその効果を示した<sup>(15)</sup>。BN は MRF と類似していてグラフモデルの一つであるが, BN は有向グラフ構造により隣接する特徴点間の関係を条件付き確率でモデル化するのに対して, MRF は無向グラフ構造により隣接する特徴点間の関係を事前確率でモデル化する。

生成的手法である MRF モデルは, 重みパラメータを導入し CRF (Conditional Random Field)<sup>(16)</sup> や MCE



(a) 特徴点の抽出



(b) 特徴点の対応付け

図2 特徴点の抽出と対応付け オンライン筆点列から特徴点を抽出する。入力パターンの特徴点と辞書クラスの特徴点を対応付けて評価することで認識する。

(The Minimum Classification Error)<sup>(17)</sup>等の識別的な手法により, それらの重みパラメータを最適化することで更なる認識性能向上を図れる<sup>(18)</sup>。オンライン手書き日本語文字認識においても, CRF により最適化される重みパラメータを導入した MRF モデルは, 伸縮マッチング等の従来手法や HMM より高い認識率を達成できる<sup>(19)</sup>。

### 4. 文字列認識

文字列を認識する場合, 文字ごとの認識だけではなく, 文脈を利用することができる。図3に示すように, 文字ごとの認識により, 各文字パターンに対して, 類似度が高い順に複数の候補文字クラスが得られ, 文字認識候補ラティスとして表現される。文脈処理は, 文字クラスから文字クラスへの接続を評価する。候補内の文字クラスだけでなく, 候補外のもを復活させることもある。文字列認識は, このラティスにおいて, 類似度と文字接続の累積値の小さい経路を探索することになる。

文脈処理の手法には, 文字接続情報を用いる方法と単語接続情報を用いる方法の2種類がある<sup>(20)</sup>。図4に示すように, 文脈処理は文字列  $C$  の確からしさ  $P(C)$  を評価する。ここでは,  $C$  は文字系列  $\{c_1, c_2, \dots\}$  か単語系列  $\{w_1, w_2, \dots\}$  からなる。

文字接続情報を用いる方法は, 候補文字列に対して文

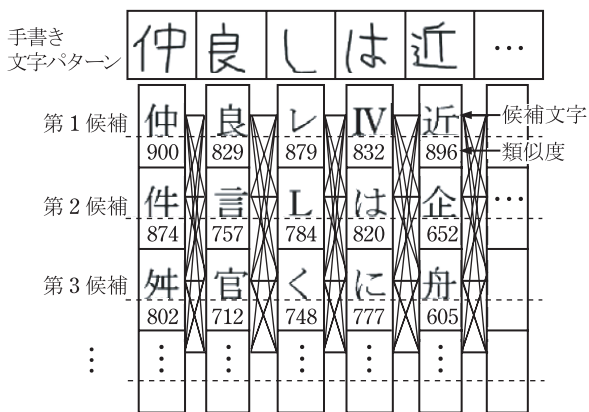
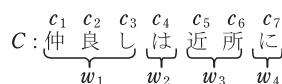


図3 文字認識候補ラティス 各文字パターンに類似度が高い順に最大  $m$  個の候補文字が、それぞれの類似度とともに与えられ、文字認識候補ラティスを構成する。



文字接続情報を用いる方法(バイグラム):  
 $P(C) = P(c_1)P(c_2|c_1)P(c_3|c_2)P(c_4|c_3)P(c_5|c_4)P(c_6|c_5)P(c_7|c_6)$

単語接続情報を用いる方法(バイグラム):  
 $P(C) = P(w_1)P(w_2|w_1)P(w_3|w_2)P(w_4|w_3)$

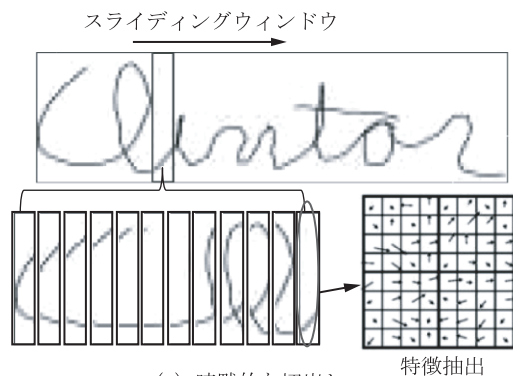
図4 文脈処理の手法 文字接続情報を用いる方法は文字と文字の接続の確からしさを評価する。単語接続情報を用いる方法は単語と単語の接続確からしさを評価する。

字と文字の接続の確からしさを評価する。1文字だけの出現確率(ユニグラム)、2文字のそのバイグラム、3文字のトリグラム、一般に  $n$  文字の  $n$ -グラムが利用できる。実用化においてはユニグラムとバイグラム、トリグラムの重み付きの線形統合により推定誤差とメモリ容量の問題を効率的に解決できている。

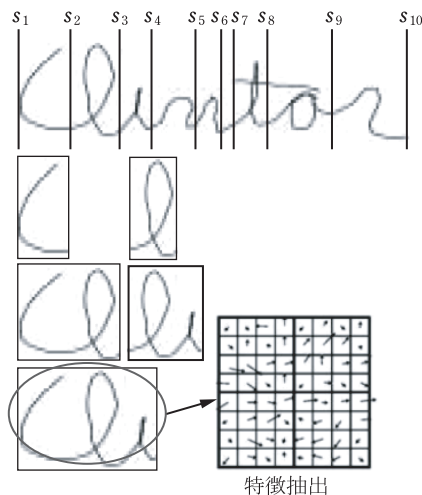
我々の実験では、文字筆記枠がある条件で、元の認識率 92.9% から、バイグラムを利用することで5ポイントの認識率向上、トリグラムを利用することで更に1ポイントの向上が得られている。また、後述する文字枠なしの条件で、元の認識率 81.3% から、バイグラムを利用することで10ポイントの向上、トリグラムを利用することで更に3ポイントの向上が得られている。

単語接続情報を用いる方法は、形態素解析等により候補文字列を単語に分解し、単語と単語の接続確からしさを評価する。ここでも、1単語だけの出現確率(ユニグラム)、2単語のバイグラム、3単語のトリグラム、一般に  $n$  単語の  $n$ -グラムが利用できる。この方法では、未知語への対応や単語辞書の容量等が問題になるが、永田は、未知語にも対応し、手書き OCR のシミュレーションで誤認識の 2/3 以上を救済している<sup>(21)</sup>。

また、電子ドキュメント処理システムにおいては、特定語句の認識が頻出する。住所、人名、日付、部署名等



(a) 暗黙的な切出し



(b) 明示的な切出し

図5 切出し方法 暗黙的な切出し手法は文字列パターンを一定幅のフレームに分割し、分割されたフレームから特徴値を抽出しフレーム列に字種クラスのラベルを付ける。明示的な切出し手法は明示的に文字列を基本切出しパターンに過分割し、そのまま、あるいは結合して生成された候補文字特徴値から文字として認識する。

はその例である。上に述べた一般的な文脈処理を用いるとある程度の認識率向上が見られるが、特定語句の語彙集合に含まれないものへの誤認識がよくあるため、特定語句の情報を利用することで、それらの認識精度を著しく向上させることができる。

特定語句認識において、研究が最も進んでいるのは住所認識である<sup>(22)</sup>。その方式では住所のトライ辞書の中でビームサーチにより検索スペースを展開する文字同期方式が採用されている。一方で、時間同期方式で手書き病名認識を行う研究もある。共に、高い認識率と高速化が報告されている。

## 5. 文字枠なしの認識

タブレット PC、パッド PC、電子白板、そして、Anoto ペンや e-pen 等のペン入力インタフェースの発展に伴い、人々はより大きい記入面に自由に書けるよう





