

文字の形状モデルを利用したオフライン手書き文字認識

永崎 健 中川 正樹

東京農工大学 情報コミュニケーション工学科

本稿は、文字パタンの形状モデルを用いた、オフライン手書き文字の識別手法について述べる。形状モデルに基づく文字識別はパターン変形への対応能力に優れるとして、これまで幾つかの手法が提案されてきた。本手法は、オフライン手書き文字パタンの生成確率を、筆順構造を持つ骨格文字パターンと骨格周辺に確率的に生成されるドットによってモデル化し、未知パタンの生成確率を最尤法により推定することで識別を行う。形状モデルによる認識は、これまで単独文字の識別にのみ適用されてきた。今回我々は、動的計画法を利用した初期位置推定法と段階整合を提案し、これを文字列中の文字識別に適用することを検討した。

An Off-line Handwritten Character Recognition Method using a Character Shape Model

Takeshi NAGASAKI and Masaki NAKAGAWA

Tokyo University of Agriculture and Technology
Department of Computer Science

This paper describes an off-line handwritten character recognition method using a relaxation process based on a character shape model. The probability of unknown character pattern is calculated from a skeleton pattern model which consists of a writing sequence of feature points, and the recognition process is formalized as a likelihood estimation problem of the probability. We propose rough pattern matching method based on dynamic programming and examined a relaxation process which applied to a character recognition in a string.

1. はじめに

伝票に書かれた氏名等の情報を自動認識し、処理の効率化を計ることへの要求は大きい。そのためには手書きパタンの変形にロバーストで、文字列の中からターゲットとする文字だけ識別するようなアルゴリズムが必要となる。文字の形状モデルに基づく整合法は、この要求に答え得ると我々は考える。

文字の形状モデルを用いた認識手法の研究には、若原[1]、Revow[3]、加藤[4]、内田[5]らの研究が挙げられる。若原は、入力された手書きパターンを局所アフィン変換で変形して整合する手法を提案した。Revow、加藤の伸縮変形モデルは文字パタンの生成・変形をスプライン曲線の確率モデルで表現したものである。内田はパターン照合法として2次元パターン同士のDPマッチングを提案した。またWebster[2]らは、力学的モデルに基づく認識手法の提案している。

本稿では、文字パタンの形状モデルを用い

た、オフライン手書き文字の照合及び識別手法を提案する。本手法はWebsterの主張する力学的モデルと、Revowの主張する確率モデルとの融合を試みたものである。更に、これまで単独文字の識別にのみ適用された形状モデルを、本稿で提案する段階整合と、動的計画法を利用した初期位置推定法により、文字列中の文字照合と識別を試みた。

2. 手書き文字の形状モデルと認識

2.1 手書き文字の確率モデル

文字パターンは、2次元空間上の特徴点列で構成された骨格パターンと、その周辺に確率的に打たれた黒点の集合としてモデル化される。このアプローチはRevow、加藤らの論文で提案されたものと同様である。但し、これらの論文では、手書き文字の変形をスプライン曲線で記述する。一方、本稿は手書き文字を特徴点間を結ぶ短い線分の集合として記述する。これには2つの理由がある、1つは文字パタンの解像度が低い場合は直線近似

で十分であるという理由, もう1つは1次元特徴点列という構造を利用した初期位置推定を適用するためである.

カテゴリ ω_j から手書き文字パターン H が生成される確率は, カテゴリ ω_j を代表するパターン S_j の関数として定義される.

$$P(H | \omega_j) = P(S_j | \omega_j) \prod_{k=1}^{n_h} P(h_k | S_j) \quad (1)$$

ここで, $P(S_j | \omega_j)$ はカテゴリ ω_j から骨格文字 S_j の変形が生じる確率, $P(h_k, \lambda_k | S_j)$ は骨格文字 S_j から手書き文字の黒点 h_k が生じる確率である. パターン認識の問題は, ある入力画像 H に対して式(1)で定義される確率を最大にするようなカテゴリ ω_j を見つけることと定式化される.

2.2 文字の骨格パターン

カテゴリ ω_j を代表するパターンは, 筆順に沿って一列に特徴点が並んだ文字として定義される. これを骨格パターン S_j と言う.

骨格パタンの生成確率は, 特徴点の数 n_r , k 番目の特徴点の初期座標 a_k , 同特徴点の移動後の座標 r_k に対して, 次式で定義される.

$$P(S_j | \omega_j) = \prod_{i=2}^m P(r_{k-1}, r_k | a_{k-1}, a_k) \quad (2)$$

2.3 弾性ストロークモデル

骨格文字の生成確率 $P(S_j | \omega_j)$ を負の対数で変換したものを, 尤度あるいはエネルギーと呼ぶ. ここで骨格パタンの確率をエネルギーに置き換えたものを弾性ストロークモデルと称することにする. 弾性ストロークモデルでは, 骨格パタンの変形度をバネのエネルギーで測る. 変形エネルギーの取り得る値は0以上で, エネルギー値が低いほど変形の度合いは小さい. 変形エネルギーは次のように定義される.

$$E(S_j | \omega_j) = -\log P(S_j | \omega_j) = \sum_{k=2}^m E(r_{k-1}, r_k) \quad (3)$$

ここで $E(r_{k-1}, r_k)$ を, 特徴点間にバネが張られたことにより生じたエネルギーであると解釈して, 2種類のバネ - 特徴点間バネと相対初期位置バネを考える. 弾性ストロークモデルとバネの概念図を図1に示す.

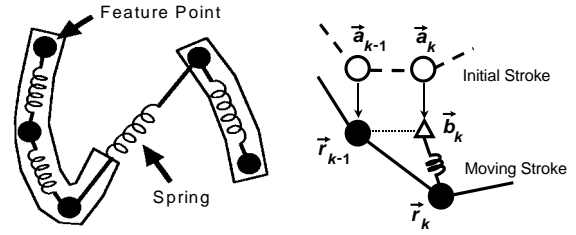


図1 弾性ストロークモデル

2.4 文字画像の生成確率

カテゴリ ω_j から, 入力画像中のある座標 h_k に黒点が生じる確率を次式で定義する.

$$P(h_k | \omega_j) = \frac{\pi_{noise}}{S} + \frac{\pi_{dot}}{B} P(h_k | S_j) \quad (4)$$

ここで π_{noise} , π_{dot} はノイズと画素発生器の分岐確率とされるもので $\pi_{noise} + \pi_{dot} = 1$ を満たす. S は文字画像の面積, B は黒画素発生器の個数である. 更に, 骨格パターン上の連続する2特徴点 (r_{i-1}, r_i) を考え, 2点間を楕円の長軸とするような正規分布から黒画素が生じるとする. 従って, 骨格パターンから黒画素 h_k が生じる確率 $P(h_k | S_j)$ は:

$$P(h_k | S_j) = \sum_{i=2}^{n_r} \{c(r_i) P(h_k | r_{i-1}, r_i)\} \quad (5)$$

$$P(h_k | r_{i-1}, r_i) = -C_1 \exp^{-d(h_k, r_{i-1}, r_i)^2} \quad (6)$$

となる. ここで $c(r_i)$ は特徴点の属性に対する関数で, 特徴点がストローク始点の場合は0, それ以外は1になるとする. また C_1 は確率を1にするための正規化係数. σ は分布の広がりを表す係数で0から $|r_i - r_{i-1}|/2$ までの値を取る. $d(h_k, r_{i-1}, r_i)$ は2特徴点間を楕円の長軸とする距離関数である.

3. 照合アルゴリズム

3.1 目的関数

パターン認識の問題は, ある入力画像 H に対して定義された確率を最大にするようなカテゴリ ω_j を見つけることと定義される. 但し, 確率のままでは計算上扱い難いので, 確率を負対数で変換した関数にして解く. この関数は特徴点集合 $R = \{r_k | k = 1, \dots, n_r\}$ と, 黒画素生成の分布の広がりを示す係数 σ に関するエネルギー関数と考えることができる. そこで問題を, パラメータ群 $\{R, \sigma\}$ に対して次式のエネルギー関数を最小化することと定式化する.

$$E(H|\omega_j) = E(S_j|\omega_j) + \sum_{k=1}^{n_h} \{-\log P(h_k|S_j)\} \quad (7)$$

この問題は確率論における混合分布のパラメータ推定問題として解釈できる。混合分布のパラメータ推定アルゴリズムとしては EM アルゴリズム, 最急降下法を用いた最尤推定などがよく使われる。本稿では, エネルギー関数の最小化問題を自動微分による最急降下法で解く。

3.2 段階整合

一般に, 単純な最急降下法(ここでは単純整合と呼ぶ)では多数のパラメータが同時に更新されるため, その解が極小解に陥りやすい。本稿が対象とする問題は, 文字パタンの照合という独特の構造を持っている。そこで最初に文字パターン全体の大きさを, 次に偏や旁, ストロークを合わせるといった, 段階的な整合プロセスを考える。本稿では次の 4 段階からなる段階整合 (CFR: coarse to fine relaxation) を行う。
 Mode1: 文字パタンの大きさ・傾きを合わせる
 Mode2: 各ストロークの大きさ・傾きを合わせる
 Mode3: 各ストローク内の折れ曲がり合わせる
 Mode4: 全ての特徴点を合わせる

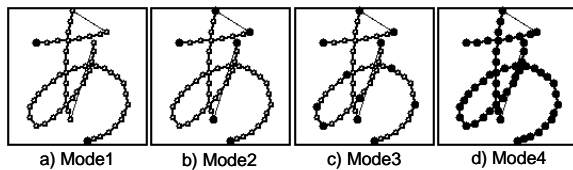


図 2 各段階における制御点

3.3 実装の方針

最急降下法では目標関数の偏微分係数が必要になる。これまでは関数の定義式から, 手計算により偏微分方程式を導いて, これを実装していた。しかし, 研究段階ではエネルギーの定義式をしばしば変更する。また段階整合では変微分対象の変数が動的に変更される。そこで, 実装を容易にするために, ボトムアップ型自動微分を用いた照合エンジンを実現した。

3.4 動的計画法による初期位置の推定

一般に最急降下法のような弛緩的解法は初期値に対する依存性が強く, 初期値によりその解が大きく変わることが頻繁に起きる。段階整合は初期値への依存性を減らすための 1 つの手段である。しかし, 文字列中の 1 文字を認識するようなケースを考えた場合, 段階整合だけで解決を図るのは難しい。そこで文字列中の文字を認識するようなケースでは, 骨格パタンの 1 次元性を利用して, 動的

計画法により初期値を推定する。

初期値推定の考え方は次のようになる。パターン照合の問題は, 次式をパラメータ群 $\{R, \sigma\}$ に関して最小化することであった。

$$E(H|\omega_j) = E(S_j|\omega_j) + \sum_{k=1}^{n_h} \{-\log P(h_k|S_j)\} \quad (8)$$

このときノイズの生成分岐確率 π_{noise} を 0 とすると, 上式右辺は次のように展開できる。

$$E(S_j|\omega_j) + \sum_{k=1}^{n_h} -\log \sum_{i=2}^{n_r} \{P(h_k|r_{i-1}, r_i)\} \quad (9)$$

更に分布 $P(h_k, \lambda_k|r_{i-1}, r_i)$ の拡がり σ が極めて狭いと仮定する。そうした場合, 画素の得る確率値は, 画素 h_k に近い特徴点ペア (r_{i-1}, r_i) にのみ支配されることになる。そこで特徴点ペア (r_{i-1}, r_i) に近い画素の集合を H_i と置いて, 上式を

$$E(S_j|\omega_j) + \sum_{h \in H_i} -\log P(h, \lambda|r_{i-1}, r_i) \quad (10)$$

と展開する。一見して, この式は特徴点列 $R = \{r_k | k = 1, \dots, n_r\}$ に関して動的計画法で解けることが分かる。

動的計画法を適用できるということは, 上記問題の最適解が求まるということである。しかし, 実際には, 特徴点が入力画像上の全ての座標に配置できるとした場合, その計算量は全画素数の 2 乗のオーダーになる。そこで特徴点 r の配置できる位置を黒画素近傍に絞り, さらに入力画像の解像度を落として計算することにする。これにより初期位置の推定を行う。但し, 精度が落ちることが予想されるため, 初期位置の推定結果は複数を使うこととする。

4. 認識実験

4.1 実験環境

本稿では実験対象字種を平仮名 46 字種に限定する。データベースとして, 通産省電子技術総合研究所により収集された ETL8B を使用した。骨格パターンは, ETL8B の見本セットを元に, 各字種につき 1 パターンを作成した。但し, 「き, さ, そ, ふ, ゆ, り」の 6 文字については, ストロークを続けて書かれるものや書かれないものがあるため 2 パターンを作成する。骨格パタンの初期形状は, 学習用 80 セットによって学習される。学習の際の照合では, 始めに外接矩形による正規化を行

った後に、初期値による誤照合の影響を避けるため上下左右に一定量移動したパターンも用意して、これら全てに対して照合を行う。

事前に与えられる係数としてはノイズとドットの分岐確率 π_{noise} , π_{dot} がある。これについては加藤論文と同様に、それぞれ 0.1, 0.9 と設定した。

4.2 単独文字に対する識別性能

認識性能の実験では、弾性ストロークモデルにおけるバネの設定が、認識率にどのように関係するかを調べた。特徴点間バネに加えて、相対初期位置バネを併用した場合の認識実験の結果を表 1 示す。これらは段階整合を Mode3 まで行っている。最高認識率は相対初期位置バネのみ使用したときの 98.5% であった。この認識率はパターン形状に基づく他の整合手法とほぼ同程度の精度である。

同じ条件下で、段階整合を Mode2 まで行った場合の認識率は 96.9%、Mode4 まで行った場合は 98.1% であった。従って、文字の整合はストロークの屈曲点を合わせる所まで行うのが一番良いということが分かる。

表 1 相対初期位置バネによる認識率 [%]

k1 \ k2	0	0.1	1	10	100
0	96.9	98.5	96.9	95.9	94.4
0.1	96.9	98.1	96.8	96.0	94.3
1	97.0	97.0	96.7	95.6	94.6
10	95.6	95.7	95.6	95.5	93.6
100	94.5	94.4	94.5	94.2	93.2

k1:特徴点間バネ k2:相対初期位置バネ

4.3 文字列中文字に対する識別性能

文字単体に対する識別精度は 98.5% と十分な精度が出た。しかし、本稿で提案した手法が文字列中の文字に対して正しく機能するかという問題が残る。そこで ETL8B の文字パターンから、部分文字列を切り出したような擬似的なパターンを合成し、これに対する認識実験を行った。擬似部分文字列の合成では縦横解像度を 64×128 とし、中央に識別対象とする文字パターンを配置し、その左右にランダムに選んだ文字の右半分・左半部分を配置する。このような合成パターンに対して、中央に置かれた文字が正しく認識できれば、ロバストな手法と言える。

実験は ETL8B 平仮名 46 字種 20 セットの各文字に対して、各パターンに 1 つずつの擬似文字列を作成して照合・認識を行った。照合手法は、単純整合、段階整合、動的計画法による初期位置推定の 3 つを比較した。単純整合

と段階整合に関しては、その初期形状に対して擬似文字列パターンとの外接矩形による正規化を行っている。また、バネとして係数 0.1 の相対初期位置バネを用いた。認識実験の結果を表 2 に掲げる。動的計画法による初期位置推定により候補数 1 の場合で 2 ポイント、候補数 5 の場合で最高 4.5 ポイントの向上が見られることが分かる。このことから、動的計画法による初期位置の推定が有効に働いていることが分かる。

表 2 照合手法による認識精度

単純整合	段階整合	動的計画法による初期位置推定				
		1	2	3	4	5
79.6%	88.7%	90.9%	92.2%	92.6%	93.0%	93.2%

5. おわりに

本稿では、文字パターの形状モデルに基づくオフライン手書き文字認識手法について述べた。提案した手法について計算機実験を行った結果、単体文字識別に対して段階整合を行うことで、ETL8B の平仮名 46 字種に対し認識率 98.5% を達成した。また、擬似的な部分文字列パターンに対する実験を行い、動的計画法による初期位置推定が効果的であることを示した。

しかし、残された課題も多い。文字列中の部分パターンに対する照合能力という点に関して言えば、動的計画法による初期位置推定を用いても認識率が 5 ポイント低下している。今後は、確率モデルの精緻化と、動的計画法による最適解探索及び弛緩的解法の効果的な組み合わせにより、よりロバストなアルゴリズムの実現を目指したい。

参考文献

- [1] T.Wakahara, K.Odaka : " Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation, " Proc.ICDAR-97, vol.1, pp28-33, Aug. 1997.
- [2] R.G.Webster,M.Nakagawa:"The Feasibility of Parallel Processing Oriented Character Recognition Method Based on a Dynamic Model" Proc.ICDAR-93, pp.714-717, Oct. 1993.
- [3] M.Revow,C.Williams,G.Hinton:"Using generative models for handwritten digit recognition," IEEE Trans. PAMI, vol.18, no.6, pp.592-606, July 1996.
- [4] 加藤毅,大町真一郎,阿曾弘具:"伸縮変形モデルを用いた手書き文字認識", 信学論, vol.J83-D-, no.12, pp.2578-2586, Dec.2000
- [5] 内田誠一, 迫江博昭:"動的計画法に基づく単調連続 2 次元ワーブ法の検討", 信学論, vol.J81-D-II, no.6, pp.1251-1258, 2000.