

非決定性形状分析モデルによる飾り文字認識

女川 俊一 大町 真一郎 阿曾 弘具

東北大学大学院工学研究科

E-mail: mega@aso.ecei.tohoku.ac.jp

Decorated Character Recognition Using Non-Deterministic Shape Analysis Model

Shunichi MEGAWA, Shin'ichiro OMACHI, and Hiroto ASO

Graduate School of Engineering, Tohoku University

あらまし 我々が普段目にする文書体は一様なものではなく、特にカタログ等の企業名・商品名に用いられる飾り文字は、文字本来の目的である“情報伝達”に加えデザイン性を高め、消費者の注意を引くといった特別な目的を持つため、既成のフォントにとらわれない様々な形状を形作っている。このような飾り文字に対して、複数のパラメータでぼかしを与えた画像を用い、濃淡画像から求まる幾何的な特徴を利用して大局構造を表す画像を抽出する方法が提案されている。本論文では、抽出された大局構造を用いて飾り文字を認識するためのモデルを提案する。さらに、様々な飾り文字に対して認識実験を行い、提案モデルの有効性を確認する

キーワード 文字認識, OCR, 飾り文字, 構造解析

1 まえがき

文字認識システムにおいて、印刷物のデータベース化や、マンマシンインターフェースとしての有効性を考慮すると、多様な文書に対応できることが重要である。我々が普段目にする字体は一様なものではなく、多数のフォントが存在する。特に、カタログ等の企業名・商品名に用いられる飾り文字は文字本来の目的である“情報伝達”に加え、デザイン性を高め、消費者の注意を引くといった特別な目的を持つため、既成のフォントにとらわれない多様な形状を有する。

文字認識手法は、大きくパターン整合法と構造解析法の2種類に分類できる。

パターン整合法は、パターン同士の重なり具合で評価し認識を行なう手法であり、入力された未知の文字パターンとあらかじめ作成しておいた標準パターン（以後、辞書と呼ぶ）との重ね合わせを行ない、最も近いものを認識結果とする。このため、文字の多少の変形やノイズには強く、計算機上で容易に高速に実現できるが、類似文字の識別は難しいという問題がある。パターンの特徴量を用いた製合法として、文字画像から方向線分を抽出して辞書との近さをもとに認識を行う手法 [1][2] は日本語手書き文字認識において高い認識率を得ている。

構造解析法は、イメージデータから文字の線分などの構造情報を抽出し線分の位置関係や接続関係などの

文字構造に着目して識別を行う手法である [3][4]。この手法は類似文字の識別、トポロジが異なる文字の識別には有効であるが、特徴量の定義や抽出が難しく、また計算機上での処理に時間がかかる等の特徴がある。構造解析法として、山本による文字パターンの輪郭線を折線近似し文字自体を多角形近似することで線分を抽出し、線分間の対応付けを行う手法 [3] や、Rocha らによる文字画像からストロークを抽出しグラフ表現に変換して形状を解析する手法 [4] 等が提案されている。前者は漢字のようなほぼ直線からなる文字に適した手法であり、後者は曲線のストロークも柔軟に扱うことのできるモデルになっている。

これらの手法ではいずれも文字画像の黒画素の連結性に注目することにより特徴抽出を行っている。しかし、飾り文字の場合、黒画素の連結性に注目しただけではその文字の形状を表すとは限らない。このような飾り文字に対し、複数のパラメータ（スケール）でぼかしを与えた画像を用い、濃淡画像から求まる尾根や谷等の幾何的な特徴 [5] を利用し大局構造を表す画像を抽出する手法 [6] が提案された。ここで得られる画像は同一字種ではほぼ同じトポロジとなる。このことは、構造解析法による飾り文字認識の可能性を示唆している。

本論文では、飾り文字認識の手法として、ストロークのトポロジに着目し、Rocha らの形状分析モデル [4] に基づく新しい手法を提案する。飾り文字から抽出さ

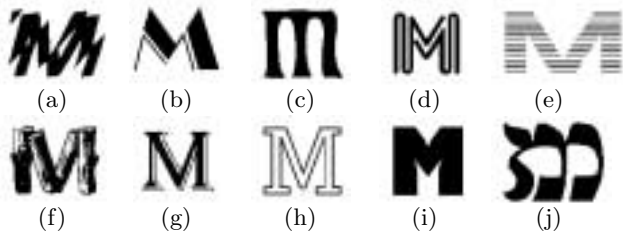


図 1: 飾り文字の例



図 2: 構造抽出結果

れた大局構造は非常に変形が大きいので、まず、より大きな変形に対応できるような非決定性形状分析モデルを提案し、そのモデルを組み込んだ認識手法を提案する。提案手法により飾り文字の認識実験を行い、提案手法の有効性を示す。

2 文字認識手法

提案する飾り文字認識手法の概要を述べる。

認識対象となる飾り文字の例を図 1 に示す。これらの飾り文字に対して、文献 [6] で提案された手法で文字構造を抽出する。図 1(a) の画像から構造を抽出した結果を図 2 に示す

ついで、パターン整合法で大分類を行う。このため、 64×64 に正規化し、次元圧縮と文字画像の変形の吸収のために、隣接する 4 画素を 1 つの小領域とし、 32×32 個の小領域を設け、各小領域内の黒画素数をその小領域の値とする 1024 次元の特徴ベクトルを用いる識別関数としてユークリッド距離を使用し、上位 K 個の候補を選出する。

最後に、提案する形状分析モデルにより細分類を行う。

3 非決定性形状分析モデル

本章では、Rocha らの提案した形状分析モデル [4] を基に、飾り文字認識に対応できる方法を提案する。

形状分析モデルは未知入力形状を表すマッチンググラフと呼ばれるモデルを用いている。そのマッチングにおいて、

- ストロークの間隙を考慮するためにブリッジを

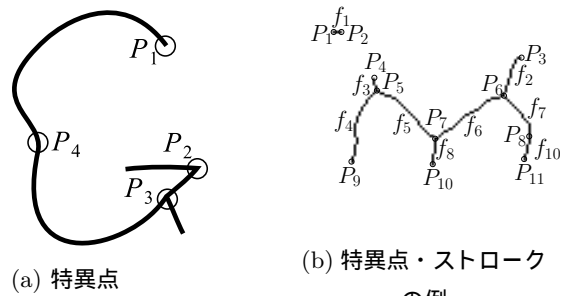


図 3: 特異点・ストロークの抽出

導入する

- 未知入力の複数のストロークでパスを構成し、パスと辞書のストロークとを整合する (多対一対応)

- パスを辞書のストロークに対応して変形させる

いう方法を用いている。識別の基準として、未知入力のパスと辞書のストロークとの間の変形コスト・整合コストを用いる。

Rocha らのモデルではマッチンググラフを作成していく際にその時点でのコスト最小のもののみを考慮する。この手法では、比較的小さい文字には対応できるが、飾り文字のような変形が大きい文字に対しては局所解に陥り易い。そこで、提案する方法では、作成可能なマッチンググラフを全て作成し、その中でコスト最小のものを選ぶという方法を採用した。

3.1 特徴抽出

最初に入力画像から構造抽出した画像を細線化し、ストローク、特異点を抽出する。特異点としては、

- ストロークの端点 (図 3(a) 点 P_1)
- 鋭角の頂点 (図 3(a) 点 P_2)
- ストロークの交点 (図 3(a) 点 P_3)
- 凹形の頂点 (図 3(a) 点 P_4)

を抽出する。以下では、2 つの特異点を結ぶ線分や曲線を改めてストロークと呼ぶ。具体例を図 3(b) に示す。図 3(b) は図 2 の画像を細線化したものである。図 3(b) における f_1, f_2, \dots, f_{10} は抽出されたストロークを、 P_1, P_2, \dots, P_{11} は特異点を示す。

3.2 ブリッジ

図 4(a) の実線は文字 “A” より、(b) の実線は文字 “H” より抽出された大局構造であるが、一部のストローク

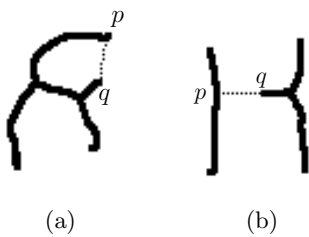


図 4: ブリッジ

が欠けてしまい、ストロークの対応が取れず誤認識の原因となる。これに対処するため、ストロークの欠如にも対応できるように、ブリッジと呼ぶ仮結合を導入する(図4, 点線)。ブリッジは次の条件を満たす2点 p, q が存在する場合に p, q 間に導入する。

- 点 p はストロークの端点であり, 点 q は黒画素上の点である
- $d(p, z) < d(p, q)$ かつ $d(q, z) < d(p, q)$ を満たす黒画素上の点 z が存在しない

但し $d(p, q)$ は2点 p, q 間のユークリッド距離である。

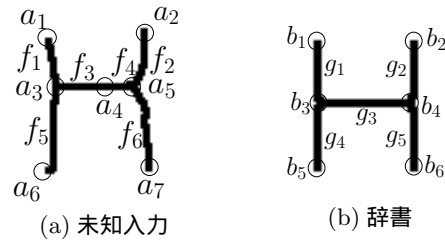
3.3 マッチンググラフの作成

標準文字パターンに対して, その構造を表すグラフを辞書として用意しておく。未知入力の複数の連続するストロークの組をパスと呼ぶ。マッチンググラフとは, 未知入力のパスの集合を辞書の1つの文字の枝集合 E に, パスの端点の集合を節点集合 V に対応させたグラフである。マッチンググラフを作成することで, 未知入力と辞書との, パス-ストローク間の対応関係, パス-ストローク間のコスト, 特異点の対応関係などを表すことができる。

マッチンググラフの生成は, Match_Join, Match_extension, Match_paths という3つの手続きでなされる。

3.3.1 Match_Join

Match_Join はある特異点の周りのストロークの対応関係を表す部分マッチンググラフを作成するための手続きであり, 与えられた未知入力のストローク f , 辞書のストローク g , 未知入力の特異点 a , 辞書の特異点 b に対して, 複数の部分マッチンググラフ N_1, N_2, \dots, N_N を求める。すなわち, 図5(a)の未知入力と(b)の辞書の組において, ストローク f_1, g_1 , 特異点 a_3, b_3 に着目した場合, そのマッチンググラフとして図5(c)を得



(c) 部分グラフ

図 5: 部分グラフ

る。ストローク $f_1 - g_1, f_3 - g_3, f_5 - g_4$ がそれぞれ対応している。

図6に, 対応するストロークの組が複数ある場合の例を示す。提案手法では, 特異点の周りに現れるストロークの順番は保存される(すなわち, 図6の例ではストローク f_1 と g_2 が対応している場合に f' は g_3 としか対応しない)という仮定を満たす全てのグラフを作成する。図6の例では仮定を満たすグラフとして N_1, N_2, N_3 が作成される。

また, 図6における未知入力のストローク f' は最適な部分グラフを作成する妨げとなることが予想できる。従って, あるストロークが次の条件を満たす場合, そのストロークを削除した部分グラフも作成する。

- 両端が共に3本以上のストロークの交点
- 接しているストロークの中で最短

図6の例では f' は f, f_1, f_2, f_3 より短く, かつ, その両端の点 a, a' がそれぞれ3本のストロークの交点となっているので, f' を削除し, 点 a, a' を同一の点とみなした部分グラフ N_4 も作成する。

3.3.2 Match_extension

前節で説明した手続き Match_Join によって得られる部分グラフでは, 未知入力と辞書のストロークが一对一に対応している。しかし, 本来一本のストロークとして抽出されるべきものが, 複数本のストロークとして抽出される場合がある。従って, 未知入力の複数のストロークからなるパスと辞書のストロークとを対応(多対一対応)させる必要がある。

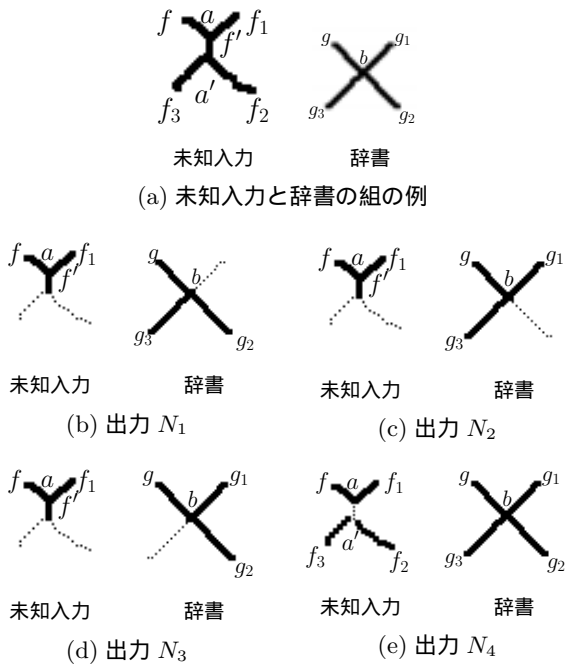


図 6: 手続き Match_Join

Match_extension は部分グラフのパスを延長し，多対一対応させ，パスの終点の周りの対応関係も表す部分グラフに拡張する手続きであり，与えられた部分マッチンググラフ M に対し， M を拡張した複数の部分マッチンググラフ M'_1, M'_2, \dots, M'_N を求める．

図 7 に具体例を示し説明する．図 7(a) の部分マッチンググラフ M が与えられた場合，最初に，辞書のストローク g_3 (終点は b_4) に対応する未知入力のパス $\alpha_1 = \{f_3\}$ (終点は a_4) を構成する．次に，パスの終点とその周りに着目し，Match_Join(f_3, g_3, a_4, b_4) を実行し， $N_{11} = (\{a_4 - b_4\}, \{f_3 - g_3, f_4 - g_2\})$ ， $N_{12} = (\{a_4 - b_4\}, \{f_3 - g_3, f_4 - g_5\})$ を得る．そして， $M'_1 = M \cup (\{\}, \{\alpha_1 - g_3\}) \cup N_{11}$ (図 7(b))， $M'_2 = M \cup (\{\}, \{\alpha_1 - g_3\}) \cup N_{12}$ (図 7(c)) を出力する．図 7(b)(c) の丸数字は Match_Join(f_3, g_3, a_4, b_4) で新しく対応付けされたストロークの組である．次にパス α_1 を延長し $\alpha_2 = \{f_3, f_4\}$ (終点は a_5) とする．このパス α_2 に対し，Match_Join(f_4, g_3, a_5, b_4) により部分グラフ $N_2 = (\{a_5 - b_4\}, \{f_4 - g_3, f_2 - g_2, f_6 - g_5\})$ を得て， $M'_3 = M \cup (\{\}, \{\alpha_2 - g_3\}) \cup N_2$ (図 7(d)) を出力する．この際，辞書のストローク g_3 に対応する未知入力のパスは $\{f_3\}, \alpha_2, \{f_4\}$ があるが，延長したパス α_2 を g_3 と対応するパスとする．同様にパスを延長するのだが，このとき，可能なパスを全て作成するように延長する．すなわち， α_2 を延長する際には，2 本のパス $\alpha_3 = \{f_3, f_4, f_6\}$ (図 7(e)) と $\alpha_4 = \{f_3, f_4, f_2\}$ (図 7(f)) を作成し，この 2 本のパスに対して，同様の手続きを繰り返

し， $N_3 = (\{a_7 - b_4\}, \{f_6 - g_3\})$ ， $N_4 = (\{a_2 - b_4\}, \{f_2 - g_3\})$ を得て， $M'_4 = M \cup (\{\}, \{\alpha_3 - g_3\}) \cup N_3$ (図 7(e))， $M'_5 = M \cup (\{\}, \{\alpha_4 - g_3\}) \cup N_4$ (図 7(f)) を出力とする．ここで，パス α_4, α_5 が延長不可であるので繰り返しを終了する．

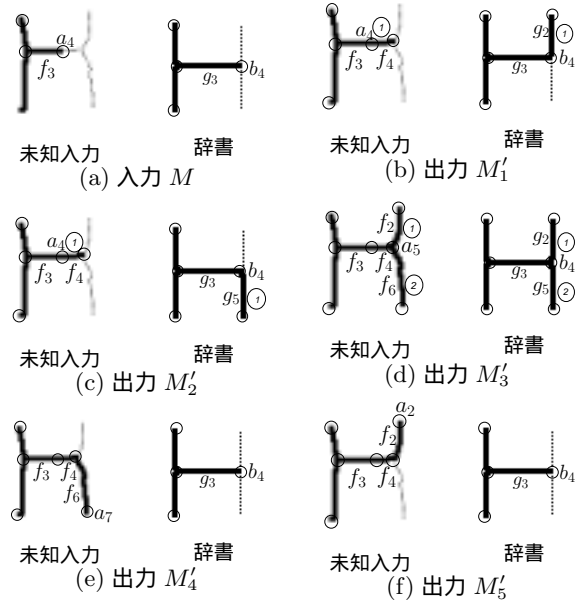


図 7: マッチンググラフの拡張

3.3.3 Match_paths

Match_paths は手続き Match_Join, Match_extension を組合せて未知入力の画像と辞書の画像に対しコスト最小のマッチンググラフを求める．

最初に，辞書のあるストローク g (終点 b) と未知入力の全てのストローク f_i (終点 a_i) に対して，手続き Match_Join(f_i, g, a_i, b) を実行し，初期部分マッチンググラフ M_1, M_2, \dots, M_{N_1} を作成する．具体例を図 8 に示す．図 8(a) は未知入力，(b) は辞書であり，ストローク g_1 (終点 b_3) に対して実行した結果である．(c) ~ (h) は初期部分マッチンググラフ M_1, M_2, \dots, M_6 である．図 8(c) ~ (h) 各図の丸数字は (b) の辞書と対応するストロークを示している．次に， M_1, M_2, \dots, M_{N_1} に対し Match_extension(M_i) を実行し，グラフ $M_{N_1+1}, M_{N_1+2}, \dots, M_{N_2}$ を得る．更に $M_{N_1+1}, M_{N_1+2}, \dots, M_{N_2}$ に対し Match_extension(M_i) を実行し，新しいグラフが作成されなくなるまで繰り返す．

最後に全てのマッチンググラフ M_1, M_2, \dots, M_N のコストを調べコスト最小のグラフを解とする．コストの計算方法については次節で述べる．図 8(a) が入力の場合，手続き Match_paths は図 7(d) を出力する．

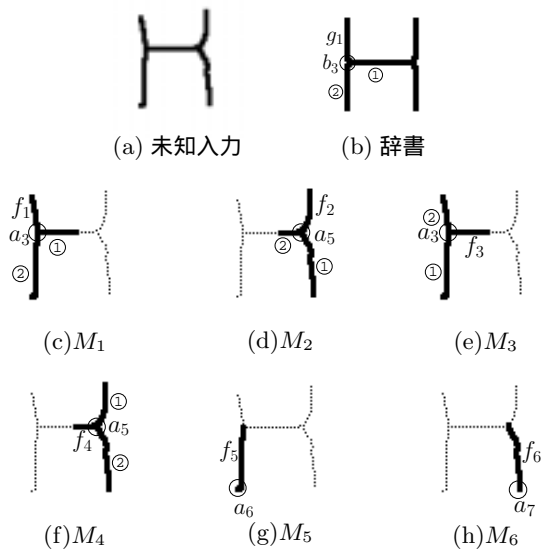


図 8: 初期部分グラフ

3.4 コスト計算

マッチンググラフのコストは認識の基準になるように、次のように定義する。

$$\begin{aligned}
 cost = & w_1 \sum_i trans(\alpha_i, g_i) + w_2 \sum_i match(\alpha_i, g_i) \\
 & + w_3 \sum_i length(b_i) + w_4 \sum_i length(f'_i) \\
 & + w_5 \sum_i length(g'_i)
 \end{aligned}$$

上式の第 1 項の $trans(\alpha_i, g_i)$ は未知入力のパス α_i を辞書のストローク g_i に合わせて変形する際の変形コスト、第 2 項の $match(\alpha_i, g_i)$ は未知入力のパス α_i と辞書のストローク g_i との整合コストを与える。第 3 項は使用したブリッジ b_i の長さ、第 4 項は使用されなかった未知入力のストローク f'_i の長さ、第 5 項は使用されなかった辞書のストローク g'_i の長さである。また、 $w_1 \sim w_5$ は重み定数で実験により定める。

上式の第 1 項、第 2 項の変形コストと整合コストについて、それぞれ 3.4.1、3.4.2 において詳しく説明する。

3.4.1 変形コスト

未知入力のパスの対応の良さを知るために対応する辞書のストロークに合うようにパスを変形し、変形の度合を示すコストを定める。

図 9(a) は対応する辞書のストロークが線分である未知入力の曲線を持つパス、折線パスにおける変形を示す。その変形コストは図の d に比例した値とする。すなわち、パスの始点と終点を結んだ線分とパスの各点との距離の最大値 d を $trans(\alpha_i, g_i)$ の値とする。図 9(b)

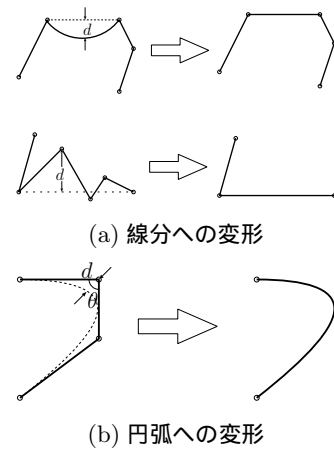


図 9: 変形コスト

は対応する辞書のストロークが円弧の場合のパスの変形を示す。その変形コストはパス上の特異点を囲むストロークが成す角を θ とおいて、 $d = \cos \frac{\theta}{2}$ と定める。特異点が複数存在する場合はそのような d の総和を変形コストとする。

3.4.2 整合コスト

整合コストは、辞書のストロークと未知入力のパスとの一対一対応に対して整合の度合を定めるものである。

未知入力のパスの長さを l_α 、辞書のストロークの長さを l_g 、未知入力と辞書の始点の水平方向に対する角度をそれぞれ θ_α, θ_g 、未知入力と辞書の終点の水平方向に対する角度をそれぞれ ϕ_α, ϕ_g として、以下のように整合コストを定める。

- 未知入力のパスと辞書のストロークが共に線分の場合

$$(l_\alpha^2 + l_g^2 - 2l_\alpha l_g \cos(\theta_\alpha - \theta_g))^{\frac{1}{2}}$$

- 未知入力のパスと辞書のストロークが共に円弧の場合

$$\frac{1}{2} \varepsilon_{arc}(|\theta_\alpha - \theta_g|) + \frac{1}{2} \varepsilon_{arc}(|\phi_\alpha - \phi_g|) + \varepsilon_{st}(|l_\alpha - l_g|)$$

ただし、

$$\begin{aligned}
 \varepsilon_{arc}(\theta) &= \begin{cases} 0 & \text{if } \theta < 1 \\ \theta - 1 & \text{if } \theta \geq 1 \end{cases} \\
 \varepsilon_{st}(l) &= \begin{cases} 0 & \text{if } l < l_g/4 \\ l - l_g/4 & \text{if } l \geq l_g/4 \end{cases}
 \end{aligned}$$

- パスが円弧、辞書が線分の場合

円弧を弦に変形し、その弦と辞書の線分とのコストとする。すなわち、整合コストは 3.4.1 で説

明した変形コストと線分どうしの整合コストの和となる。

- パスが線分，辞書が円弧の場合

曲線に歪みが生じて線分になることは無いという仮定をし，この場合はコストを ∞ とする。

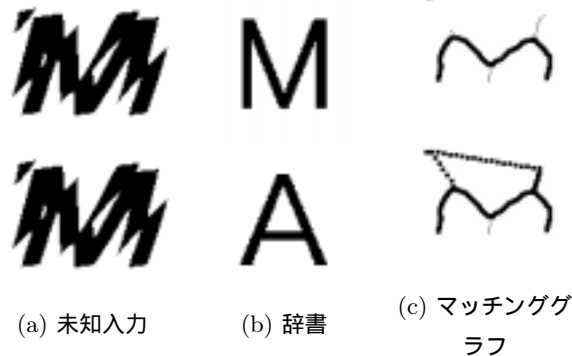


図 10: マッチンググラフの例

4 実験

本手法の有効性を検証する為に，飾り文字を認識対象とする性能評価実験を行う。テストサンプルには，アルファベット大文字 26 文字・各 10 種の飾り文字を用いた (図 1 に一部を示した)。辞書作成用にはアルファベット大文字 26 文字・各 10 種のゴシック体を用いた。

本認識手法において，大分類による絞り込み数を $K = 4$ とした場合，認識率 65.4% を得ることができた。パターン整合法のみで飾り文字を認識した場合の認識率は 57.7% であり，第 4 位累積認識率は 85.4% であった。これは本認識手法が飾り文字に対して有効であることを示している。

また，大分類を行わず，最初から形状分析モデルのみを用いて認識した場合の認識率は 56.5% であった。モデルが柔軟に様々な文字画像と整合されるため，概形の異なる文字とも対応してしまい，認識率が低下したものと考えられる。これは，パターン整合法と構造解析法との融合が有効であることを示す証拠でもある。

図 10(c) に提案モデルで作成したマッチンググラフを示す。図の細破線は使用しなかったストローク，太破線はブリッジを表す。図 10 の上の図は飾り文字 “M” に対して辞書 “M” を対応させた例を示す。グラフの形状は “M” に近い形状であり，辞書のストロークは全て使用されている。未知入力側に使用されていないストロークはあるもののその長さは短くマッチングコストは小さくなっている。一方，図 10 の下の図は飾り文字 “M” に対して辞書 “A” を対応させた例を示す。マッチンググラフは文字 “A” と同じトポロジを持つものの，形状は大きく変形し，使用されるブリッジの長さも長くマッチングコストが大きくなっている。“M” 以外の文字との対応も “A” と同様であり，図 10(a) は “M” と認識された。

5 むすび

本論文では，飾り文字認識のための形状分析モデルを提案した。飾り文字を用いた認識実験を行い，本手法の

有効性を確認した。その結果提案手法により 65.4% の認識率が得られた。比較実験により，パターン整合法単独，形状分析モデル単独の認識実験より高精度であることを示した。

しかし，変形の極端に大きい文字に関してはまだ改良の余地があり，今後の課題である。また，今回はパターン整合法で絞り込む際の候補数を実験的に求めた。何らかの基準を設けることも今後の課題である。

参考文献

- [1] 鈴木雅人，大町真一郎，加藤寧，阿曾弘具，根元義章，“混合マハラノビス関数による高精度な類似文字識別手法”，信学論 (D-II)，vol.J80-D-II，no.10，1997。
- [2] 若林哲史，鄧 巖，鶴岡信治，木村文隆，三宅康二，“非線形正規化と特徴量の圧縮による手書き漢字認識の高精度化”，信学論 (D-II)，vol.J79-D-II，no.5，pp.765-774，1996。
- [3] 山本和彦，“弛緩整合法による手書き教育漢字認識”，信学論 (D)，vol.J65-D，no.9，pp.1167-1174，1982。
- [4] J. Rocha and T. Pavlidis，“A shape analysis model with application to a character recognition system,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.4, pp.393-404, 1994.
- [5] R. M. Haralick, L. T. Watson, and T. J. Laffey，“The topographic primal sketch,” International Journal of Robotics Research, vol.2, no.1, pp.50-72, 1983.
- [6] 大町真一郎，井上政樹，阿曾弘具，“マルチスケール画像を用いた飾り文字の構造抽出”，信学技報，PRMU98-30，1998。