

近傍ベクトルを用いた甲骨文字の特徴点と線の検出

岸雅大[†] 石井康史[†] 孟林[†] 山崎勝弘[†]

立命館大学大学院 理工学研究科[†]

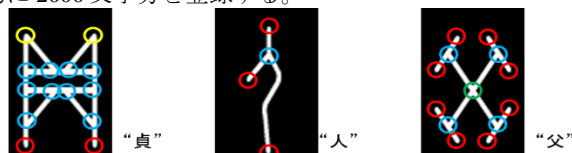
1. はじめに

甲骨文字は中国殷代の象形文字であり、漢字の祖形とも言われている。我々は1つの原画像に対し、その文字に対応するテンプレート画像とのマッチングにより認識を行ってきた[1]。本研究では、甲骨文字に含まれる特徴点を用いて甲骨文字の認識を行う。まず、ノイズ除去を行い、次に、甲骨文字の骨格を抽出する。その後、細線化を行った画像から各画素の近傍のベクトルを算出・合成し、合成後のベクトルの長さや角度を用いて特徴点の検出と分類を行う。さらに、抽出した特徴点と細線化画像を用いて線の本数を検出する。最後に、それらの特徴を用いて、甲骨文字のテンプレート画像と特徴が保存してある甲骨文字データベースから、類似する特徴をもつテンプレート画像の検索を行う。本稿では、近傍ベクトルを用いた特徴点と線の検出、それらを用いた候補テンプレートの検索について述べる。

2. 特徴量を用いた甲骨文字認識

2.1 甲骨文字の特徴量とデータベース

甲骨文字データベースは、甲骨文字の原画像、テンプレート、及び甲骨文字の特徴量から構成される。テンプレートは文字の傾きがなく線分の太さを統一し正規化した甲骨文字画像である。特徴量として文字に含まれる端点、角点、分岐点、交差点それぞれの総数、特徴点の間に存在する線の本数、線の角度を保存する。図1に甲骨文字のテンプレートと特徴量の構成を示す。現在、甲骨文字データベースには1000文字分登録されており、最終的に2000文字分を登録する。



文字	端点	角点	分岐点	交差点	特徴点	線	0~90°	91~180°
真	2	2	10	0	14	18	16	12
人	3	0	1	0	4	3	1	2
父	8	0	4	1	12	12	6	10

図1 甲骨文字テンプレートと特徴量の構成

2.2 甲骨文字認識システム

本研究での甲骨文字認識システムを図2に示す。原画像からガウシアンフィルタ、2値化、ラベリングによって、ノイズを除去し、細線化、ハフ変換を行って骨格を抽出する。次に、細線化画像を用いて甲骨文字の特徴量を抽出し、それらを用いて、甲骨文字データベース内の全テンプレートの特徴量とのマッチングを行い、認識対

象と類似するテンプレートを自動的に抽出する[2]。さらに、原画像と候補となるテンプレートの差異を少なくする正規化の処理を行う。それらと原画像との類似度を計算して、最大値を認識結果とする。

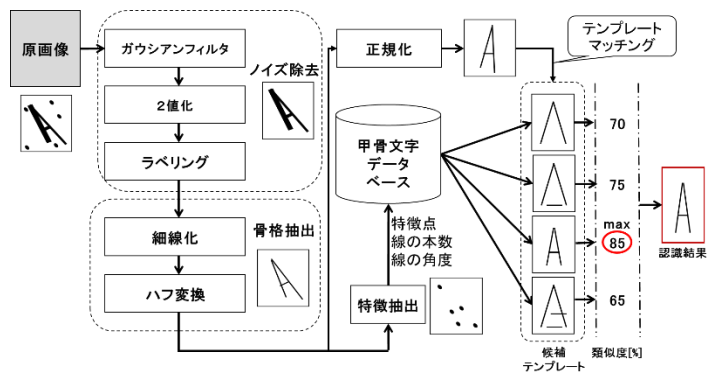


図2 甲骨文字認識システム

3. 甲骨文字の特徴点と線の検出

3.1 近傍ベクトルを用いた特徴点の検出

甲骨文字画像の各画素の近傍ベクトルを算出、合成することで特徴点の抽出を行う。まず、ノイズを除去した、細線化画像に対しラスタスキャンをして、注目画素の7×7ピクセルの範囲に存在する点との間で、向きと大きさの計算を行う(図3(a))。次に、それらのベクトルの和を取って合成し(図3(b))、合成したベクトルの向きと大きさを用いて特徴点か直線の一部かを判断し、直線の一部の場合は画素を除去し(図3(c))、特徴点の場合は残す(図3(d))。さらに、抽出した変化点に7×7ピクセルの範囲でクラスタリングを行い、点の集合を1点に定めることで特徴点の抽出を行う(図4(d))。クラスタリングでは、注目画素の周囲を参照して画素があるかどうかを判断し、ある場合はその点を一つにまとめる。

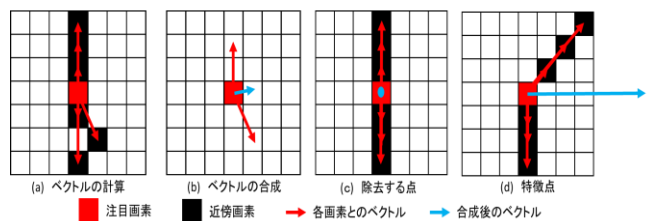


図3 近傍ベクトルの合成と点の判断例

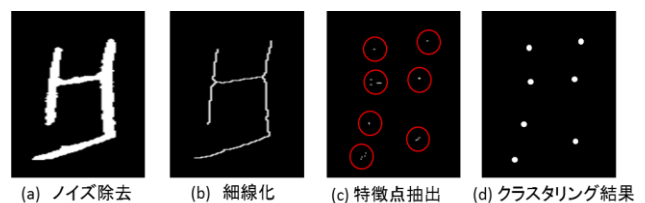


図4 特徴点の抽出例

Detecting Features and Lines of Oracle Bone Inscriptions using Neighboring Vectors

Masahiro Kishi[†], Koji Ishii[†], Lin Meng[†] and Katsuhiko Yamazaki[†]

[†]Graduate School of Sciences and Engineering, Ritsumeikan University

3.2 特徴点の分類

特徴点の分類例を図5に示す。抽出した特徴点と細線化された原画像を用いて、図5(d)の赤枠のように特徴点の周囲7×7ピクセルの範囲を参照し、その特徴点と接している線の本数を調べる。接している本数が1本なら端点、2本なら角点、3本なら交差点、4本以上なら交差点として分類を行う。図5(d)の場合は接している線が3本となるので分岐点である。

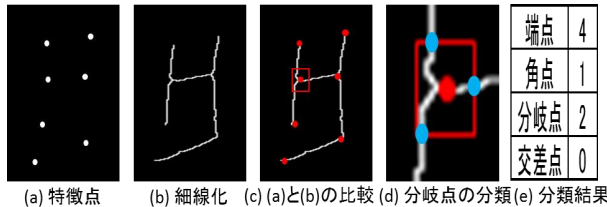


図5 特徴点の分類例

3.3 特徴点を用いた線の検出

特徴点を用いた線の検出方法を図6に示す。特徴点の分類と同様に抽出した特徴点と細線化された原画像を用いて、図6(c)のように、ある特徴点と隣接する特徴点の間を区切りとしてラベリングを行い、最大のラベル番号を線の本数として検出する。

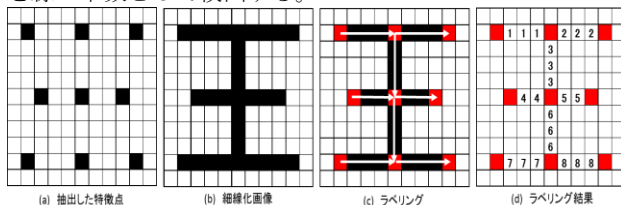


図6 線の検出例

4. 実験と考察

4.1 実験内容

甲骨文字の拓本より抽出し、細線化後の画像20文字計200枚の画像に対して特徴点と線の抽出を行い、それらを用いて類似テンプレートの検索実験を行った。本研究で用いるCPUは、Intel Corei7-3820プロセッサで、動作周波数は3.60GHz、64bitの命令セットである。OSはubuntu14.04、C言語(gcc4.9.1)を用いる。

4.2 実験結果と考察

図7に特徴点と線の本数の検出結果の一部を示す。また、特徴点と線を正確に検出した画像での候補テンプレートの検索結果を図8に示す。

今回の実験では、原画像200枚の画像中78枚に対して、特徴点と線の抽出を行えた。原画像を細線化することにより、文字の歪みやノイズなどを軽減できていると考えられ、特徴点を正確に検出できなかった画像は図9のように、文字のノイズが非常に大きく、ノイズが文字と繋がっているか文字が断裂しているために細線化を行ってもノイズが線として検出され、点の数が増えている。

これらの改善点を以下に示す。

- (1) 近傍ベクトルを算出する範囲と、合成後のベクトルの長さや角度の閾値をより正確に設定する。
- (2) 現在のノイズ除去に、文字と接しているノイズに対しても除去が行える処理を追加する。また、断裂している文字には線の肥大化を行ってから細線化を行う。

線の検出に関しては、特徴点が正確に検出されていれば線の本数も正確に検出できている。

	ノイズ除去画像	細線化画像	特徴点抽出	クラスタリング	検出結果	
人					端点	5
					角点	0
					分岐点	1
					交差点	1
					線の本数	6
					亡	
角点	1					
分岐点	2					
交差点	0					
線の本数	6					
雨						
					角点	2
					分岐点	1
					交差点	0
					線の本数	8
					田	
角点	4					
分岐点	4					
交差点	1					
線の本数	12					

図7 特徴点と線の検出結果

	入力画像	抽出した特徴	候補テンプレート	
己			端点	2
			角点	4
			分岐点	0
			交差点	0
			線の本数	5
			候補テンプレート	
雨			端点	9
			角点	2
			分岐点	1
			交差点	0
			線の本数	8
			候補テンプレート	

図8 候補テンプレートの検索結果



図9 実験対象の比較

候補テンプレートの検索では、抽出した特徴点と線の本数を用いて検索を行った。特徴が正確に抽出できている場合は、候補テンプレートの中に正答テンプレートを検出できている。また、文字によっては、図7(雨)のように正答テンプレートのみが検出されている。しかし、単純な文字に関しては候補テンプレートが複数検出される。今後は、文字の角度を追加しての検索を行うことでより正確に候補テンプレートを絞りこめる。

5. おわりに

本稿では、細線化画像の注目画素近傍の画素からベクトルを算出し、特徴点の抽出を行って、特徴点から線の本数を検出する方法を述べた。また甲骨文字データベースから原画像に類似するテンプレート画像の検索を行った。今後の課題として、合成ベクトルの長さや角度の適切な閾値の設定、甲骨文字原画像からの角度の抽出、及び、甲骨文字データベースの完成などがあげられる。

参考文献

- [1] L.Meng, Y.Fujikawa, A.Ochiai, T.Izumi and K.Yamazaki, "Recognition of Oracular Bone Inscriptions Using Template Matching", International Journal of Computer Theory and Engineering, Vol.8, No.1, pp.53-57, 2016
- [2] 石井, 藤川, 孟, 山崎, "特徴量を用いた甲骨文字の候補テンプレート抽出と認識", 情報処理学会第78回全国大会, 1N-08, 2015.
- [3] 岸, 石井, 孟, 山崎, "甲骨文字原画像からの特徴点抽出と類似テンプレートの検索", FIT2016, H-014, 2016.