

石造遺物画像の文字解析のためのノイズ除去手法の開発

上相英之^{†1} 上相真之^{†2} 多仁照廣^{†3}

概要: 斜光撮影による影のみを抽出する光拓本技術により作成された画像は、石碑に刻みこまれた文字の抽出が可能である。一方で石造遺物表面には、文字以外の凹凸が存在することが多く、その形状・大きさも様々である。これらの文字以外の陰影と文字の陰影との分離は、拓本画像の判読性向上の為に必要な処理であり、且つ状況に応じた処理方法が求められる。本発表では文字の陰影サイズより小さな凹凸を、画像処理技術の一つであるクラスターラベリングを用いて除去した結果を報告する。

Development of noise removal method for character analysis of stone monument images

HIDEYUKI UESUGI^{†1} MASAYUKI UESUGI^{†2}
TERUHIRO TANI^{†3}

1. はじめに

津波デジタルライブラリ[1]をはじめ、津波碑の収集・アーカイブ化が進むなど、石造遺物はその価値が見直されつつある。その一方で、従来の石造遺物研究においても課題とされてきた、刻まれた文字「金石文」の撮影画像の可読性が低いことに起因する、第三者の検証が困難、若しくは不可となる問題が顕在化してきている。

これまで、金石文を可視化する為には、多くの場合拓本が選択されてきた。拓本それ自体は優れた手法であるが、データ取得の時間、熟達までの期間、試料への接触が不可欠、破損・汚損の可能性などの課題を抱えている。一方で、IT技術の進歩により三次元化の手法も多く誕生し、石造遺物への適用事例も見られる。これらの手法は3次元形状を保存するという点に於いて、石造遺物のデジタル化に最適の方法とも言える。一方で比較的高度な技術や機材やソフトを必要とし、データ取得・解析にも時間がかかる。

石造遺物は日本中に無数に存在する。防災・減災に活かす為には、速やかなデジタル化と平易に利用できるコンテンツが必要であるが、上記の手法では時間やコストがかかり過ぎ、風化作用を受け続け、常時情報が減衰し続けている石造遺物の迅速なアーカイブには不向きである。

筆者らはこれまで、石造遺物表面の金石文を、撮影画像上で可視化する手法を研究してきた。結果、フィールドワークでの作業を前提とした「高効率」「軽量」「平易」というコンセプトの下、斜光撮影画像から影を抽出し複数の影を合成する光拓本の技術を完成させ、風化などで通常撮影では読むことが困難な金石文画像の可読性を向上させるこ

とに成功した[2]。

本報告では、光拓本技術における基本的な画像処理ステップによって抽出された金石文画像の可読性や文字の自動認識への障害となっている「文字の凹凸以外の影」の除去について報告する。

2. 共通画像処理後の課題

光拓本では全ての試料に最初に適用される共通画像処理と、共通画像処理後の画像状況に応じて適用される個別画像処理の二つの画像処理ステップがある。石造遺物表面が比較的平坦な場合、共通画像処理で十分な可読性が確保できる(図1)。



図1 共通画像処理後例
左画像：処理前 右画像：処理後

^{†1} 国文学研究資料館
National Institute of Japanese Literature

^{†2} 高輝度・光科学研究センター
Japan Synchrotron Radiation Research Institute

^{†3} 若狭路文化研究センター
Wakasa-ji Culture Research Institute

一方で表面に文字以外の凹凸が多数存在する場合、これに起因する影が可読性に大きく影響を与える。これまで、これらのノイズの内、ピクセル数の少ない(小さな)ノイズに対しては、各ピクセル近傍のピクセルの最小輝度値で置き換える(erosion)で対応してきたが、これらの方法では金石文の線幅が十分な幅の場合は問題ないが、線幅が狭い場合その文字の線まで消してしまうことがあった(図2)。

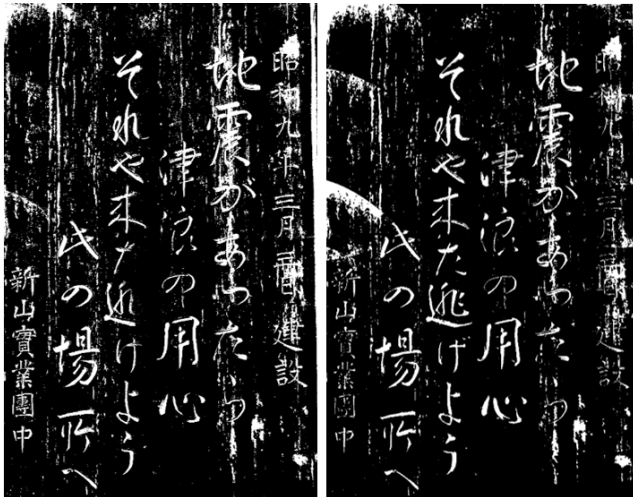


図2 左図ノイズ未除去 右図 erosion 後結果崩れた字形

3. クラスタラベリングによるノイズ除去

本報告では、字形を維持したままノイズを除去する為、画像処理の手法であるクラスタラベリング法を採用する。

クラスタラベリングは、指定された範囲の画素値を有する隣接したピクセルを1つの塊、すなわちクラスタと見なす方法である。画像上の全てのクラスタを認識し、クラスタ内の画素値をクラスタサイズに従って変化させる。すべてのクラスタの画素値はクラスタのサイズを示すので、画像を閾値処理して抽出するだけで小さいクラスタ(ノイズ)を除去でき、大きなクラスタ(文字)が表示される(図3)。

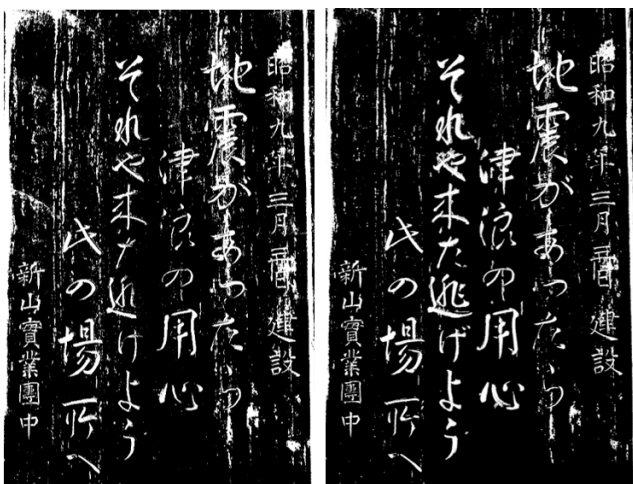


図3 左図ノイズ未除去 右図 クラスタラベリング後

図2の右図では、ある程度小さいノイズの除去は行われている一方で、「昭和九年三月三日建設」「新山実業團中」字形が崩れている。しかし、図3右図のクラスタラベリング後では、同じ様に小さなノイズが除去され、さらに字形を保つことにも成功している。

4. おわりに

クラスタラベリングによるノイズ除去は一定の効果が確認できた。これらは、将来の金石文の自動認識を可能にするための処理である。しかし、図4-5に見られるように風化など表面の凹凸の状況によって、その効果は大きく変動する。

今後は異なるノイズ除去の手法も組み込んでいきたい。



図4 表面にほぼ凹凸が無い場合の字形

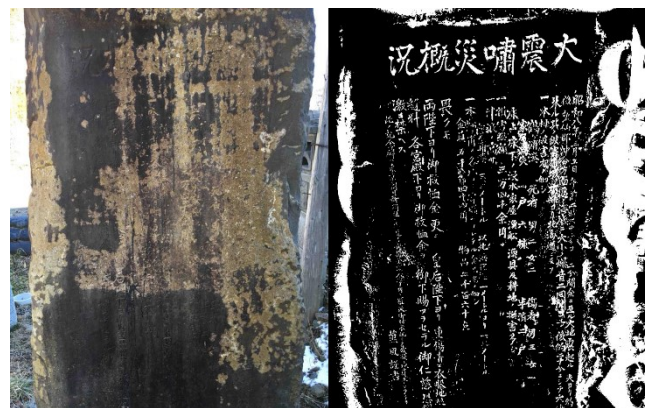


図5 風化による凹凸の影響を受けた字形

参考文献

- [1] “津波デジタルライブラリ”. <http://tsunami-dl.jp>, (参照 2019-01-20).
- [2] H. Uesugi, and M. Uesugi, “Stone Monument Text Image Database,” Theory and Practice in Modern Computing, 2015, pp. 203-207.