

木簡画像の特徴の抽出

2C-2

*園田 浩一 **上島 紳一 **大庭 脩

*姫路獨協大学一般教育部

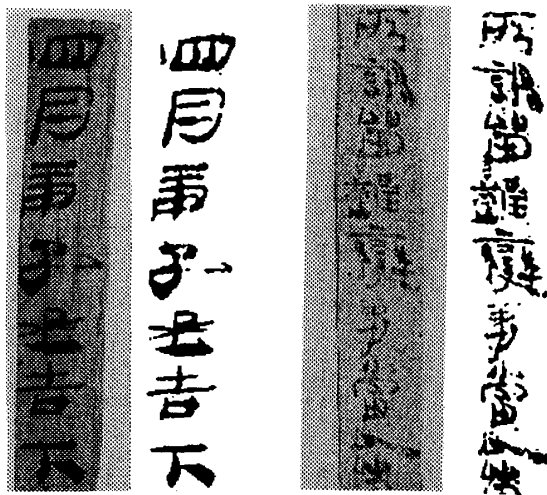
**関西大学文学部

1. はじめに

中国古代の木簡研究は中国、台湾及び英国で発表される写真集や著録に依存しているが、写真集には頁単位の木簡が内容に無関係に多数本ずつ並べられている。つまり、その木簡の配列が冊書としての簡牘の前後関係に一致しているわけではない。この問題の解決として、まず、写真集の各ページの木簡一葉毎を切り出し、画像処理手法を用いて簡質を比較する。更に書かれている文字を一字ずつ切り出して文字の形態を比較することで、冊書を構成していた簡の共通性の探索を試みる。

2. 前処理

木簡の写真集から一頁単位にスキャナを用いてコンピュータにデータを取り込み、木簡と背景のコントラストを利用して木簡一葉毎に切り出す。次に、書かれている文字一字ずつを切り出すために画像の二値化を行なう。コントラストの強い画像については単純二値化が可能であり、しきい値の決定は判別分析法⁽¹⁾による。また、判読しにくい画像についてはエッジ強調や濃度補正等の画像処理を施した後、同分析法を用いてしきい値を決定する。続いて二値化した画像の雑音除去、不用な画素の除去、文字内の欠落した画素の補填などを行なって漢簡文字の比較を容易にする(図1)。



(図1) 入力画像と二値化画像

3. 特徴パラメータの検討

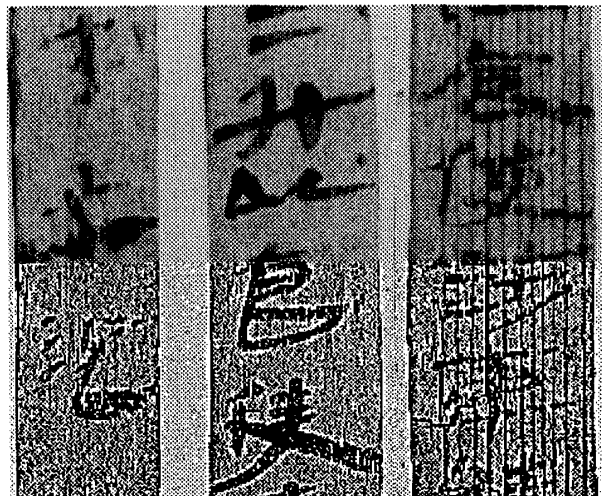
冊書を作成した人は多くが当時の官吏であり、一つの冊書は同一人が、同種類の木に作成する場合が多く、文字の形態や内容による書き方や用いている木の質は似ているものと考えられるので、この仮定を基本に置き、木目の比較をした上、一文字ずつを切り出している。更に次の(1)-(5)について調べた。

(1) 長さ、幅

複数の木簡からなる冊書について各々の長さを見てみるとほぼ同じ長さのものが多い。そこで二値化画像に対して木簡の長さ幅を計測し、特徴パラメータファイルに記憶する。木簡は必ずしも長方形の整形された木が使われているとは限らないし、発掘されるまでに土砂や水などの外部圧力を受けて折れたり、捻れたりしているものも多い為に、幅は木簡の各部分の幅を計測して最大値を求めている。

(2) 木目

木簡は木目に対して平行に切り、整形したものが多く、鮮明な写真画像が得られているものについては木目が縦に流れている。複数の木簡で一つの冊書となっているものについては同じ種類の木を用いていた可能性が高いものと考えられる。そこで、画像処理によって鮮鋭化を図る(図2)。



(図2) 鮮鋭化による木目の強調

Classification of wooden strips

*Koichi SONODA, **Shinichi UESHIMA, **Osamu OBA
*HIMEJI DOKKYO University, **KANSAI University

(3) 文字間隔

二値化した木簡画像に対して、書かれている一文字毎に外接四角形を求めて文字を切り出す。この時、簡によっては文字の多様性に加えて、文字どうしが重なり合ったり、一葉に二行ある簡（両行）があり、切り出しが難しい文字があるので研究者との対話的な作業により実行する。次に各文字の外接四角形の重心を求めて文字の位置とし、文字間隔を計測する。外接四角形に重心を書き込んだものが図3である。同時に次の文字の広がりパラメータ σ も求めている⁽³⁾。

$$\sigma^2 = \frac{\sum_i \sum_j [(i-i_0)^2 + (j-j_0)^2] f(i, j)}{\sum_i \sum_j f(i, j)}$$

但し $f(i, j) = 1$ (文字), $= 0$ (背景)
 (i_0, j_0) は文字の重心

また、文字の太さや縦横の長さの特徴を捉える為に、文字の面積、外接四角形に対する文字部分の占有度を調べている。更に文字の円形度や周囲長、伸長度、2次モーメントを求め、同じ部首を持つものやハネに対して主軸方向を考える。

(4) データベースとの連結

写真集は紙面の大きさの関係があって木簡が出土順に並んでいるとは限らないが、冊書が同一の出土地から発見される可能性が非常に高い。従って多数の木簡の分類の試みの第一歩は木簡画像を発掘当時の出土順に配列を戻し、改めて配列し直すことが肝心である。また、同一冊書に同じ文字が現れていれば当然似ているはずである。そこで釈読されている範囲において同じ文字について検索し、画像を比較検討することが考えられる。そこで敦煌漢簡データベース(5)を用いて一字検索を行なっている。

(5) ちぎれた木簡について

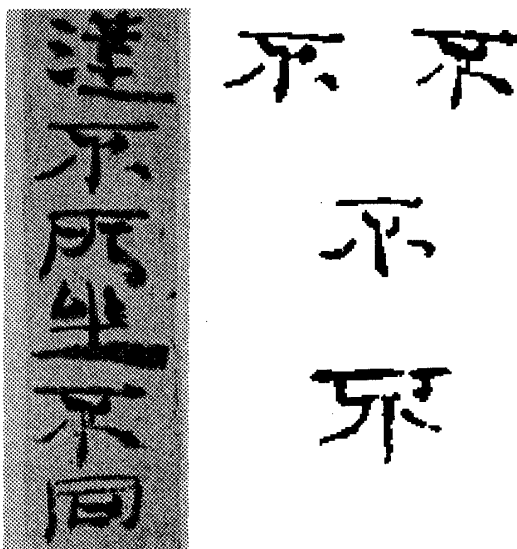
木簡の相方さがしの場合、漢字が一字でも抜け落ちると意味が全く代わる為に、釈読についてよほどの確信がある場合を除いて、切れ口が完全に合致する場合のみを考える。

4. 適用例

敦煌漢簡⁽⁴⁾の中の一例として「不」について、パラ

メータの検討を行なった。サンプルとして、確実に作者が同じである同一の簡に存在する2つの「不」と、その簡と冊書を構成していると考えられている簡に存在する「不」、そして、明かに作者が違うと分かる「不」の4つを選んだ。図4の上段の2文字は同一簡の文字で、中段が上段の簡と冊書構成していると考えられているもの、下段が明かに違う作者のものであると分かる文字である。

この4つの文字から、外接四角形の縦横比、文字部分の面積、広がりパラメータを計測した。その結果、外接四角形の縦横比はすべて同程度の値を示した。しかし、文字部分の面積、文字の広がりパラメータが明かに違うと思われる文字と同一作者の文字と考えられている文字とに差異が認められた。



(図4)「不」による作者の判定例

5. おわりに

冊書復原の為に木簡の並べ換えには内容の把握と構成の論理が必要であり、画像的に類似した木簡が得られたとしてもその中の配列の決定はやはり研究者に判断してもらわねばならない。例に見られる簡は漢代の法令の執行命令と考えられるが、構成は上奏部と命令内容に分かれることが知られている。

本研究は文部省科学研究費(課題番号63450050)の補助を受けて行った。

参考文献

- (1) 田村:FAのための画像処理技術(工学研究社)(1988)
- (2) A. Rosenfeld, A. C. Kak: Digital picture processing, academic press (1982)
- (3) 樋口他:文字の間隔的大きさの定量化, 電子情報通信学会総合全国大会p6-171(1987)
- (4) 大庭:敦煌漢簡(同朋社)(1989)
- (5) 上島他:敦煌漢簡画像データベース、情報処理学会第40回全国大会予稿集(1989)