

中文社会科学引文索引 (CSSCI) 来源集刊

中文学术期刊 (光盘版) 全文收录集刊

万方数据全文收录辑刊

维普网全文收录辑刊

# 中國文字研究

THE STUDY OF CHINESE CHARACTERS

教育部人文社会科学重点研究基地

华东师范大学中国文字研究与应用中心

华东师范大学语言文字工作委员会

主办

Vol.29 第二十九辑

1999  
创刊二十周年  
2019



上海書店出版社  
SHANGHAI BOOKSTORE PUBLISHING HOUSE

# 中国文字研究

---

教育部人文社会科学重点研究基地  
华东师范大学中国文字研究与应用中心 主办  
华东师范大学语言文字工作委员会

第二十九辑

---



上海書店出版社  
SHANGHAI BOOKSTORE PUBLISHING HOUSE

## 图书在版编目(CIP)数据

中国文字研究.第二十九辑/教育部人文社会科学  
重点研究基地华东师范大学中国文字研究与应用中心,华  
东师范大学语言文字工作委员会主办. —上海:上海书  
店出版社,2019.6

ISBN 978-7-5458-1803-1

I. ①中… II. ①教… ②华… III. ①汉字-文字学  
-文集 IV. ①H12-53

中国版本图书馆 CIP 数据核字(2019)第 086081 号

责任编辑 邹 烨

封面设计 郇书径

技术编辑 丁 多

## 中国文字研究(第二十九辑)

教育部人文社会科学重点研究基地  
华东师范大学中国文字研究与应用中心 主办  
华东师范大学语言文字工作委员会

出 版 上海书店出版社  
(200001 上海福建中路 193 号)  
发 行 上海人民出版社发行中心  
印 刷 上海商务联西印刷有限公司  
开 本 889×1194 1/16  
印 张 15  
字 数 400,000  
版 次 2019 年 6 月第 1 版  
印 次 2019 年 6 月第 1 次印刷  
ISBN 978-7-5458-1803-1/H·37  
定 价 65.00 元

# THE STUDY OF CHINESE CHARACTERS

Vol.29

---

**Organized by**

Center for the Study and Application of  
Chinese Characters at East China Normal University  
(Key Research Institute in University Authorized by  
the Ministry of Education of  
People's Republic of China)  
East China Normal University Language Work Committee

---



SHANGHAI BOOKSTORE PUBLISHING HOUSE

## 编委会委员

(按音序排名)

---

### 顾 问

董莲池 李宇明 裘锡圭 王 宁 许嘉璐 曾宪通

### 编 委

阿辻哲次〔日〕 白于蓝 出野文莉〔日〕 丁克顺〔越〕 大形徹〔日〕 海村惟一〔日〕  
何莫邪〔挪〕 河永三〔韩〕 黄德宽 黄天树 雷启立 李圭甲〔韩〕 李国英  
李家浩 刘 钊 刘志基 潘玉坤 朴兴洙〔韩〕 蒲芳沙〔法〕 阮俊强〔越〕  
史克礼〔德〕 苏培成 王贵元 王 辉 王 平 王元鹿 吴振武 夏含夷〔美〕  
徐富昌 徐世猛 殷国光 喻遂生 臧克和 张涌泉 张再兴 赵平安 朱歧祥

---

### 主 编

臧克和

### 执行主编

潘玉坤

## Editorial Board

---

### Advisors

Dong Lianchi Li Yuming Qiu Xigui Wang Ning  
Xu Jialu Zeng Xiantong

### Members of Editorial Board

Tetsuji Atsuji[Japan] Bai Yulan Fumiri Deno[Japan] Dinh Khac Thuan[Vietnam]  
Tohru Ohgata[Japan] Amamura Yuiji[Japan] Christoph Harbsmeier[Norway]  
Ha Young-sam[Korea] Huang Dekuan Huang Tianshu Lei Qili  
Lee Kyoo-kap[Korea] Li Guoying Li Jiahao Liu Zhao Liu Zhiji Pan Yukun  
Park Heung-soo[Korea] Françoise Bottéro[France] Ngugen Tuan Cuong[Vietnam]  
Shi Keli[Germany] Su Peicheng Wang Guiyuan Wang Hui Wang Ping  
Wang Yuanlu Wu Zhenwu Edward L. Shaughnessy[USA] Xu Fuchang Xu Shimeng  
Yin Guoguang Yu Suisheng Zang Kehe Zhang Yongquan Zhang Zaixing  
Zhao Ping'an Zhu Qixiang

---

### Chief Editor

Zang Kehe

### Executive Editor

Pan Yukun

# 目 录

## 古文字研究

- 基于编码的甲骨文识别技术研究 ..... 陈婷珠 吴少腾 吴 江 李 淋 ( 1 )  
侯南卜辞性质新论  
——兼谈何组、黄组间的关系 ..... 韩文博 ( 13 )  
卜辞重文省略现象补正 ..... 展 翔 ( 23 )  
三例屯南甲骨缀合分歧评议 ..... 周妮娜 ( 27 )  
从大保簋“𠄎”字说到“道”“造”二字的起源 ..... 张世超 ( 34 )  
说毛公鼎“翩翩四方大猷不静” ..... 董莲池 ( 43 )  
《积微翁回忆录》所见杨树达金文研究资料 ..... 曾庆炳 ( 46 )  
灵岩寺古文诗刻考释 ..... 李春桃 ( 53 )

## 简牍碑刻文字研究

- 释“造” ..... 姚道林 ( 59 )  
《楚居》“涅”字补考 ..... 赵思木 ( 62 )  
浅谈草书的萌芽形成与结构特点 ..... 朱葆华 ( 66 )  
从字频统计看秦简用字的相关性 ..... 高 魏 张显成 ( 73 )  
秦汉简帛文献中病愈义字词计量考察 ..... 刘艳娟 王斯泓 张再兴 ( 81 )  
连云港海州西汉墓 M1:60 衣物疏词语诂解 ..... 吕志峰 ( 89 )  
《李君碑》《裴君碑》释文补正 ..... 伊 强 ( 95 )  
《睡虎地秦简》介词系统新探 ..... 向贤文 ( 99 )

## 传统语言文字研究

- 《龙龕手鏡》疑难字例释 ..... 梁春胜 ( 111 )  
《五音集韻》重文考辨六则 ..... 邱龙升 ( 116 )  
杜牧《阿房宮賦》异文辨证 ..... 冷卫国 梁秋芬 ( 122 )  
《本草綱目·檀》异文考辨一则 ..... 袁开惠 赵怀舟 ( 127 )  
“床前明月光”新解质疑  
——与周同科先生商榷 ..... 沈 伟 ( 131 )  
论汉字分化中的母字归属机制  
——对王凤阳先生“常用占简”理论的反思 ..... 陈 青 ( 140 )

## 汉字规范与应用研究

- 《汉字类推简化问题研究综论》订补 ..... 苏培成 ( 147 )

|                          |               |
|--------------------------|---------------|
| 蒲松龄《日用俗字》的用字类型和特点 .....  | 刘中富 (152)     |
| 草书楷化字研究 .....            | 吴立业 陈双新 (158) |
| 俗字考释十六则 .....            | 赵静莲 (167)     |
| 明刊本《花关索传》符号轮廓类俗字研究 ..... | 高 岩 (174)     |

**少数民族文字·海外汉字研究**

|                                    |               |
|------------------------------------|---------------|
| 木里县甲波村东巴文墓碑译释及研究 .....             | 杨亦花 (179)     |
| 贵州古彝文的造字方式考探 .....                 | 金 瑞 (188)     |
| 利玛窦《西国记法》对汉字的分析识记 .....            | 曹保平 李 莹 (201) |
| 朝鲜王朝碑石书迹拓印史初探 .....                | 刘元春 游程程 (209) |
| 外国人汉字习得的难度等级<br>——一项基于数据库的研究 ..... | 王 骏 (217)     |

# Contents

|   |  |     |
|---|--|-----|
| Study on Oracle-bone Script Recognition Technology Based on Coding<br>.....   | Chen Tingzhu Wu Shaoteng Wu Jiang Li Lin | 11  |
| The New View on Nature of Houjia Village South Oracle-bone Inscriptions<br>—Also on the Relationship Between He Group and Huang Group .....               | Han Wenbo                                | 22  |
| Correction about the Variant Form of a Chinese Character of the Oracle Inscriptions .....   | Zhan Xiang                               | 26  |
| Three Controversial Examples' Judgment of Xiaotun South Bones' Rejoining .....  | Zhou Nina                                | 33  |
| The Study about Character 𠄎 in Dabao Gui(大保簋) and Researching on the Origin of 道 and 造<br>.....   | Zhang Shichao                            | 42  |
| The Research of “翩翩四方大慨不静” on the Duke Mao Tripod .....   | Dong Lianchi                             | 45  |
| Materials of Yang Shuda's Bronze Inscriptions Study in <i>The Memories of Jiweiweng</i> .....   | Zeng Qingbing                            | 52  |
| Research and Explain the Poetry Stele by Chuanchao Guwen of Lingyan Temple .....  | Li Chuntao                               | 58  |
| Explaining the Meaning of “Cuo(造)” .....  | Yao Daolin                               | 61  |
| Additional Research on the Character “Ying(溟)” in the Bamboo-slip Text <i>Chuju</i> .....   | Zhao Simu                                | 65  |
| About the Formation and Structure Characteristics of Cursive Script .....   | Zhu Baohua                               | 72  |
| Research on the Character Using Relativity of Qin Bamboo Slips Based on Character Frequency<br>.....  | Gao Wei Zhang Xiancheng                  | 80  |
| Quantitative Research on the Words and Characters of Healing in Qin and Han Manuscript Literature<br>.....  | Liu Yanjuan Wang Sihong Zhang Zaixing    | 88  |
| Definitions for the Main Words of M1: 60 Yiwushu Unearthed from the Western Han Dynasty Tomb<br>in Haizhou Distinct, Lianyungang City .....               | Lv Zhifeng                               | 94  |
| Revisions to <i>Lijunbei</i> (《李君碑》) and <i>Peijunbei</i> (《裴君碑》) .....   | Yi Qiang                                 | 98  |
| A New Study of the Adposition System in <i>Bamboo Slips of the Qin Dynasty Unearthed at Shuihudi</i><br>.....   | Xiang Xianwen                            | 110 |
| Investigation on the Knotty Characters in <i>Longkan Shoujing</i> .....   | Liang Chunsheng                          | 115 |
| Six Research Cases on Chongwen(重文) in <i>Wuyin Jiyun</i> .....  | Qiu Longsheng                            | 121 |
| The Dialectic Differentiation of Du Mu's <i>E-Panggong Fu</i> .....   | Leng Weiguo Liang Qiufen                 | 126 |
| Textual Research and Analysis of Variant Text of <i>Sandalwood</i> from <i>Compendium of Materia Medica</i><br>.....                                      | Yuan Kaihui Zhao Huaizhou                | 130 |
| A Criticism of the New Interpretation about “So Bright a Gleam on the Foot of My Bed” .....   | Shen Wei                                 | 139 |
| On the Graphic Attribution Mechanism in the Differentiation of Chinese Characters: A Reflection<br>on the Frequency Determinism of Mr.Wang Fengyang ..... | Chen Qing                                | 146 |
| The Revision of <i>Summarize of Study on Chinese Character Simplify by Analogy</i> .....  | Su Peicheng                              | 151 |
| Types and Features of Characters in <i>Generally Used Chinese Characters</i> by Pu Songling .....   | Liu Zhongfu                              | 157 |
| The Study of Cursive Script Turn into Regular Script .....  | Wu Liye Chen Shuangxin                   | 166 |
| Research on Sixteen Folk Characters .....   | Zhao Jinglian                            | 173 |
| Study on the Symbolized Characters in <i>The Story of Hua Guansuo</i> Printed in the Ming Dynasty .....   | Gao Yan                                  | 178 |
| Interpretation and Research of Dongba Script Tombstone in Jiabo Village, Muli County .....  | Yang Yihua                               | 187 |
| Discussion of the Creation Methods of Ancient Yi Characters of Guizhou .....  | Jin Rui                                  | 200 |

*Western Mnemonics* of Matteo Ricci's Methods to Analyse and Memorize Chinese Characters  
..... Cao Baoping Li Ying 207

Study on the History of the Rubbings of Stone Inscriptions in Joseon Dynasty  
..... Liu Yuanchun You Chengcheng 216

The Difficulty Level of Chinese Characters in Acquisition by Foreigners; A Study Based on Data Analysis  
..... Wang Jun 227

## 基于编码的甲骨文识别技术研究\*

陈婷珠 吴少腾 吴江 李淋

**【摘要】**随着 AI 领域的崛起,如何利用机器视觉高效、准确地识别图像目标,近年来已成为国内外学者研究的重点。目前,国内外学者对甲骨文的图像识别进行了大量研究,但在识别精度、识别时间和抗干扰能力等方面仍有发展空间,尤其是现有甲骨文识别技术对专家的依赖性高而甲骨文信息共享普及率则相对较低。针对上述问题,本文以殷墟小屯村中村南甲骨为实验基础,提出基于编码的甲骨文识别技术,即在图像预处理计算过程中,把甲骨文图像转换为编码,并在此基础上提出甲骨文识别系统的设计与实现方法。

**【关键词】**编码;甲骨文;图像识别技术

**【作者简介】**陈婷珠,女,上海交通大学人文学院副研究员,研究方向为甲骨文字学。吴少腾、吴江和李淋为上海机械与动力工程学院本科生和研究生,研究方向为视觉与图像识别。(上海 200030)

### 一 问题的提出

甲骨文作为一种古老文字,对于它的识别处理与研究常常遇到种种困难和挑战。目前,国内外学者对甲骨文识别技术的研究主要有以下几类。(1)有的学者从字形出发,利用笔画方向和长度等字形特征,采用无向或有向笔段和笔元技术以描述甲骨文字形,主要代表有李锋等《甲骨文自动识别的图论方法》(1996)<sup>①</sup>、《甲骨文计算机识别方法研究》(1996)<sup>②</sup>,王晓丽《高精度曲线轮廓甲骨文字形生成系统的研究与实现》(2010)<sup>③</sup>,栗青生等《基于有向笔段甲骨文输入方法的设计与实现》(2012)<sup>④</sup>、《甲骨文字形动态描述库及其字形生成技术研究》(2013)<sup>⑤</sup>、《甲骨文识别的图同构方法》(2011)<sup>⑥</sup>。这种方法将甲骨文字形进行矢量描述,建立了甲骨文字形描述库,为古文字的数字化编辑提供了一个新思路。不过,由于甲骨文字形线条极其复杂,不存在现代汉字一般意义上的笔画概念,人类在识别甲骨文上往往是整体输入方式,而非一个个笔画。因此,采用这种方法的分类率并不高,尚无法真正意义上为甲骨文进行机器识别分类。(2)有的学者利用拓扑结构或图形分类进行分析,如顾绍通《基于拓扑配准的甲骨文字形识别方法》(2016)<sup>⑦</sup>、吕肖庆等《一种基于图形识别的甲骨文分类方法》(2010)<sup>⑧</sup>。(3)有的学者基于语义进行分类,如高峰等《基于语义构件的甲骨文模糊字形的识别方法》(2014)<sup>⑨</sup>。(4)有的学者利用甲骨文字结构等特征,采用语料库和支持向量机(Support Vector

\* 基金项目:该项目获得教育部重点研究基地重大项目古文字专题数据库系列(课题编号:18JJD740004)、2018年度孔子学院建设课题(课题编号:CIKT016)支持。

① 李锋、周新伦:《甲骨文自动识别的图论方法》,《电子科学学刊》1996年,第41—47页。

② 周新伦、李锋、华星城等:《甲骨文计算机识别方法研究》,《复旦学报(自然科学版)》1996年第5期,第481—486页。

③ 王晓丽:《高精度曲线轮廓甲骨文字形生成系统的研究与实现》,硕士学位论文,苏州大学,2010年。

④ 吴琴霞、栗青生:《基于有向笔段甲骨文输入方法的设计与实现》,《计算机应用》2012年,第2374—2377页。

⑤ 栗青生、吴琴霞、杨玉星:《甲骨文字形动态描述库及其字形生成技术研究》,《北京大学学报(自然科学版)》2013年,第61—67页。

⑥ 栗青生、杨玉星、王爱民:《甲骨文识别的图同构方法》,《计算机工程与应用》2011年,第112—114页。

⑦ 顾绍通:《基于拓扑配准的甲骨文字形识别方法》,《计算机与数字工程》2016年,第2001—2006页。

⑧ 吕肖庆、李沫楠、蔡凯伟等:《一种基于图形识别的甲骨文分类方法》,《北京信息科技大学学报(自然科学版)》2010年,第92—96页。

⑨ 高峰、吴琴霞、刘永革等:《基于语义构件的甲骨文模糊字形的识别方法》,《科学技术与工程》2014年,第67—70、86页。

Machine, SVM<sup>①</sup>)的理论建立知识库,实现简单的甲骨文字形相似性分析分类和识别,主要代表有:史小松《基于支持向量机的甲骨文字结构分析研究》(2010)<sup>②</sup>、孙莹莹《基于混合核 LS-SVM 的古汉字图像识别》(2015)<sup>③</sup>、刘永革等《基于 SVM 的甲骨文字识别》(2017)<sup>④</sup>。不过,甲骨文虽然是成熟的文字系统,但毕竟处于汉字早期阶段,除了异体字众多,低频字高度集中外,还有很多甲骨文字属于未考释字。该方法主要适用于模式识别领域,在解决图画特征明显,结构不清晰的甲骨文字上还有困难。另外,该方法对甲骨文字的旋转、缩放、平移、遮挡等识别率不高,算法处理尚有空间,同时在存储和计算时耗费大量的机器内存和运算时间,运行效率较低。(5)还有的学者尝试深度学习(核心算法是卷积神经网络)技术,辅以草图识别技术,建立甲骨文文字数据库。主要代表有:高峰等《Recognition of Fuzzy Characters on Oracle-bone Inscriptions》(2015)<sup>⑤</sup>、王长虎《Building Hierarchical Representations for Oracle Character and Sketch Recognition》(2016)<sup>⑥</sup>、微软亚洲研究院(2017)<sup>⑦</sup>采用 Mind Finder(草图识别技术)系统,建立起一个拥有包括 260 类、2 万甲骨文文字的数据库。不过,该方法适合大数据,简单应用神经网络无法在甲骨文识别上取得较好的效果,模型的过拟合严重,无法泛化。数据集中分类数目过多,且在分类的类目下面仍然存在大量的异体字,能够为网络提供特征提取的样本过少。因此,目前甲骨文识别研究尚无法完全提取甲骨文字的主要特征,尚未得出完全符合甲骨文字实际情况的算法,且使用范围受到一定的限制。高速、精确及立体化、智能化应该是未来甲骨文识别技术发展的重要方向。

## 二 甲骨文识别技术

甲骨文识别技术可以定义为一种视觉识别技术,它是基于甲骨文的特征信息进行搜索定位的一种视觉识别技术。用户使用手机或相机等带有摄像头的工具采集含有甲骨文字的图像,随后系统自动在该图像上进行检测,进而对检测到的甲骨文进行识别,最后提供该甲骨文的原始拓片或近似拓片信息以及相关研究成果。这一系列相关技术,包括程序和具体甲骨材料,有机地构成了甲骨文识别技术。该技术的智能性体现在基于图像搜索而非基于文本搜索,因此不需过多依赖用户的甲骨文已有知识。同时,在识别时不受用户拍摄时可能产生的个体行为或客观外在因素的影响,如甲骨文拓片的旋转、缩放、平移(RST: rotation, scaling and translation),或用户视点即图像仿射/投影变换(view-point),或光照影响(illumination),或杂物场景(clutter)与噪声,甚至于甲骨文字本身被部分遮挡(occlusion)等等。

本文认为,如何将甲骨文图像高速、有效地转化为计算机编码是目前甲骨文识别技术的关键性问题。该问题的合理解决将有助于甲骨文识别技术得到重要发展。其中,如何进行有效的图像预处理,

① SVM 技术是在统计学习理论的基础上发展而来的一种新的模式识别方法,在解决有限样本、非线性及高维模式识别问题中表现出许多特有的优势。

② 史小松:《基于支持向量机的甲骨文字结构分析研究》,硕士学位论文,华东师范大学,2010年。

③ 孙莹莹:《基于混合核 LS-SVM 的古汉字图像识别》,硕士学位论文,安徽大学,2015年。

④ 刘永革、刘国英:《基于 SVM 的甲骨文字识别》,《安阳师范学院学报》2017年第2期,第54—56页。

⑤ Feng G, Jing X, Yong-ge L. Recognition of Fuzzy Characters on Oracle-bone Inscriptions[C]//Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing(CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on. IEEE, 2015:698—702.

⑥ Rui Y, Romanrangel E, Chao H, et al. Building Hierarchical Representations for Oracle Character and Sketch Recognition [J]. IEEE Trans Image Process, 2016, 25(1):104—118.

⑦ 王长虎:《Mind Finder 基于轮廓的图像搜索》,2010年8月10日,微软研究院官网(<https://www.msra.cn/zh-cn/news/features/29d51ea4-f4bd-44a3-973d-86f007c2d589>)。

特别是有效的图像分割是研究的重中之重。图像分割是图像识别的前提和基础。在甲骨文图像处理过程中,图像分割的质量直接影响着最终的甲骨文图像识别效果。

随着 AI 技术的崛起,基于机器视觉的图像识别技术再一次成为研究热点,前人关于甲骨文识别研究在识别精度、识别时间和抗干扰能力等方面仍有一定的发展空间,尤其是目前现有甲骨文识别技术对专家的依赖性高而甲骨文信息共享普及率则相对较低。针对上述问题,本文在殷墟小屯村中村南甲骨实验基础上,提出基于编码的甲骨文图像识别方法,并在此基础上,提出甲骨文识别系统的设计与实现方法。

### 三 实验过程

图像识别技术是对研究对象进行图像获取,然后根据研究对象自身特征信息进行分类或识别。本文以《殷墟小屯村中村南甲骨》<sup>①</sup>(简称《村中南》)为研究对象,该书著录中国社会科学院考古研究所安阳工作队历年来在小屯村中、村南发掘发现的甲骨文 498 片,编为 514 号(16 片分正反);收入小屯村北 12 片,花园庄东地 3 片,苗圃北地 1 片,大司空村 1 片。全书共计甲骨 515 片。选择这本书为研究对象,主要有以下三点考虑:1.资料整饬,考古学属性完整,地层情况清晰。2.拓片总量适中,对数据采样和实验结果具有可推广性。3.该书编著体例科学完备,采取拓本、摹本、照片、释文四位一体的方式,科学丰富且真实准确。

基于编码的甲骨文识别实验过程可分解为如下三步:

首先,有关《村中南》的单字图像的获取过程。

这一步是对研究对象《村中南》甲骨进行最初的图像采集工作,这样做的目的是将最原始的对象转换为图像信息,并将最初的数据材料带入图像信息中。具体做法如下:将《村中南》的原始拓片采用扫描方式,全部录入电脑后,对其进行单字切分,得到一个单字图片库,每个单字,不计重复,都有唯一的一张图片。最后得到与甲骨文单字一一对应的 6230 张甲骨文单字图像,本文把图片统一放在 C:\5X\5 目录下。

其次,对获取图像进行甲骨文图像识别的预处理过程。

这一步是如何将甲骨文图像高速、有效地转化为计算机编码的关键性所在。本文采用图像分割法对甲骨文识别图像进行预处理,并对预处理后的图像进行相应的编码。具体是指对原始图像中的噪声以及非相关特征信息进行过滤,减少所获取原始图像的信息量,减少后续算法的复杂度并提高识别效率,从而为后期的最终特征信息的分类减少干扰。

本文先采用了 6 张甲骨文单字图像作为初步实验对象,以找出最佳的图像分割比例,然后再进行大规模实验。本文以  $32 * 32$ 、 $16 * 16$ 、 $8 * 8$ 、 $4 * 4$  的模板进行了图像处理,并转成灰度值的结果如图 1—4 所示,即将图像分为多个区域,并且每个区域只包含一定数量的特征信息。上一行为甲骨文单字图像,下一行为按照不同模板图像分割后转成灰度值的计算结果。



图 1  $32 * 32$  的结果

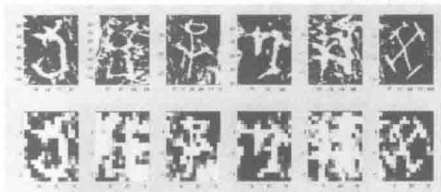


图 2  $16 * 16$  的结果

<sup>①</sup> 中国社会科学院考古研究所编:《殷墟小屯村中村南甲骨》,云南人民出版社,2012 年。

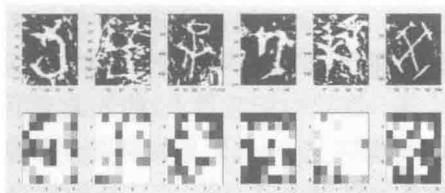


图3 8\*8的结果

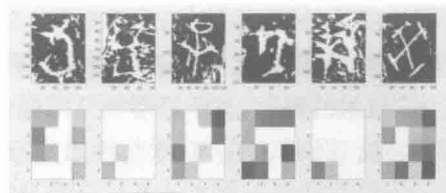


图4 4\*4的结果

虽然肉眼分辨不出来,但实际上看似一片白的地方,数据也是不同的。按照4\*4的模板,一共是将每个单字分成16个格子,在每个处理结果中对应16个数字,每一个横排与纵竖相交叉的数据就是对应每个格子的不同的灰度值。对6幅图计算的结果显示,数据具有明显的唯一性。所显示的数据是0—255范围内的数值,代表灰度值。上述6张甲骨文单字对应的处理结果,如图5—10所示:

| result{1, 1} |         |          |          |         |
|--------------|---------|----------|----------|---------|
|              | 1       | 2        | 3        | 4       |
| 1            | 44.6426 | 43.8084  | 90.8040  | 53.5198 |
| 2            | 28.5781 | 130.5910 | 131.3824 | 41.2201 |
| 3            | 38.8243 | 46.8459  | 78.3546  | 60.2152 |
| 4            | 87.7879 | 87.0606  | 148.5807 | 47.8299 |

图5 第一个字的结果

| result{1, 2} |          |          |          |          |
|--------------|----------|----------|----------|----------|
|              | 1        | 2        | 3        | 4        |
| 1            | 148.8006 | 101.2877 | 87.0221  | 70.2321  |
| 2            | 206.1173 | 81.7581  | 127.6663 | 69.2341  |
| 3            | 139.4271 | 54.4793  | 144.3389 | 76.1223  |
| 4            | 44.0296  | 63.9114  | 152.1468 | 161.9312 |

图6 第二个字的结果

| result{1, 3} |         |          |         |          |
|--------------|---------|----------|---------|----------|
|              | 1       | 2        | 3       | 4        |
| 1            | 45.0422 | 82.5215  | 48.1230 | 6.3122   |
| 2            | 46.8143 | 78.1087  | 83.4480 | 22.4458  |
| 3            | 23.0481 | 129.6485 | 55.6166 | 87.7102  |
| 4            | 38.6374 | 51.6324  | 99.6280 | 169.1429 |

图7 第三个字的结果

| result{1, 4} |         |          |          |          |
|--------------|---------|----------|----------|----------|
|              | 1       | 2        | 3        | 4        |
| 1            | 44.1298 | 27.2744  | 23.6473  | 21.8338  |
| 2            | 43.0275 | 175.2036 | 125.3131 | 130.7182 |
| 3            | 56.3269 | 38.3336  | 104.8661 | 5.7251   |
| 4            | 5.5829  | 18.4200  | 84.5614  | 43.8098  |

图8 第四个字的结果

| result{1, 5} |          |          |          |          |
|--------------|----------|----------|----------|----------|
|              | 1        | 2        | 3        | 4        |
| 1            | 70.9291  | 96.6916  | 156.8689 | 128.5122 |
| 2            | 101.5351 | 149.1911 | 181.4016 | 105.1491 |
| 3            | 77.7971  | 168.4305 | 178.8074 | 121.8692 |
| 4            | 23.7980  | 47.6710  | 94.0823  | 85.3099  |

图9 第五个字的结果

| result{1, 6} |         |         |         |         |
|--------------|---------|---------|---------|---------|
|              | 1       | 2       | 3       | 4       |
| 1            | 51.3238 | 41.3299 | 39.5048 | 10.4208 |
| 2            | 37.8563 | 77.1255 | 49.7784 | 43.8909 |
| 3            | 78.2736 | 79.4511 | 31.7922 | 17.3680 |
| 4            | 42.7429 | 16.5437 | 54.9004 | 0.7654  |

图10 第六个字的结果

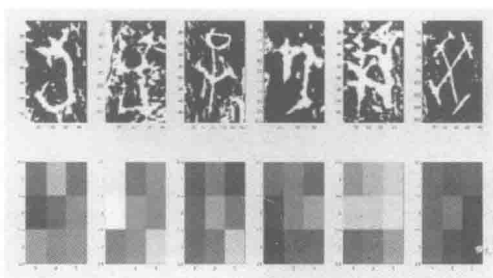


图11 3\*3的结果

本文又以3\*3的模板做了图像分割和灰度值计算。实验证明,在3\*3的模板下,上述6张图像依然能唯一表达,如图11所示。因为数字太大之后,都显示为白色,肉眼识别不明显。于是,我们把数据按比例限定在0—99之间,然后按灰度显示,这样看起来明显点。

最后,根据《村中南》甲骨对图像中的对应特征信息进行查找识别过程。

这一步是对前期获取的图像信息预处理结束后,根据相关的先验知识,利用计算机技术快速实时对图像中的特征信息进行提取分类。以3\*3作为图像

分割比例,扩大样本量,对所有《村中南》甲骨进行计算。仅仅经过 2 分钟,6230 张甲骨文单字图像全部计算结束,效率很高。通过程序运行,本文得到如下结果,如图 12 所示:

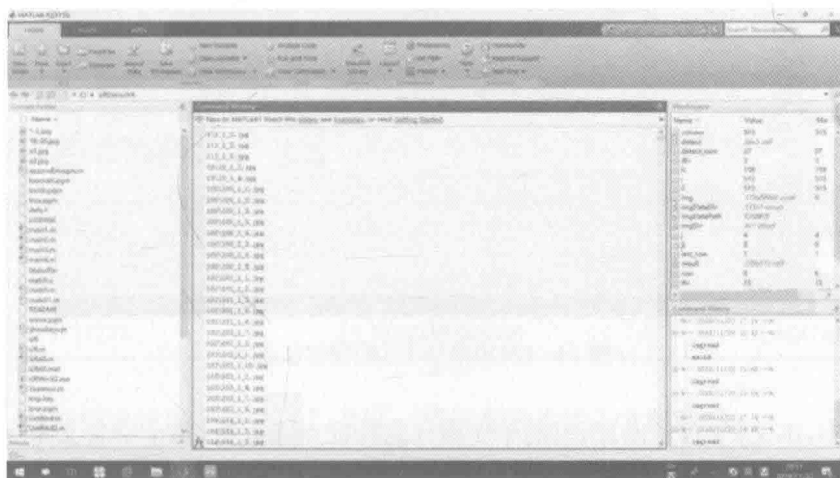


图 12 《村中南》所有切图

在图 12 中,左边一栏为命令文件夹,中间为命令栏,右边一栏上面是工作平台,下面是历史记录。经过编写与运行程序,本文得到的运算结果如图 13 所示:

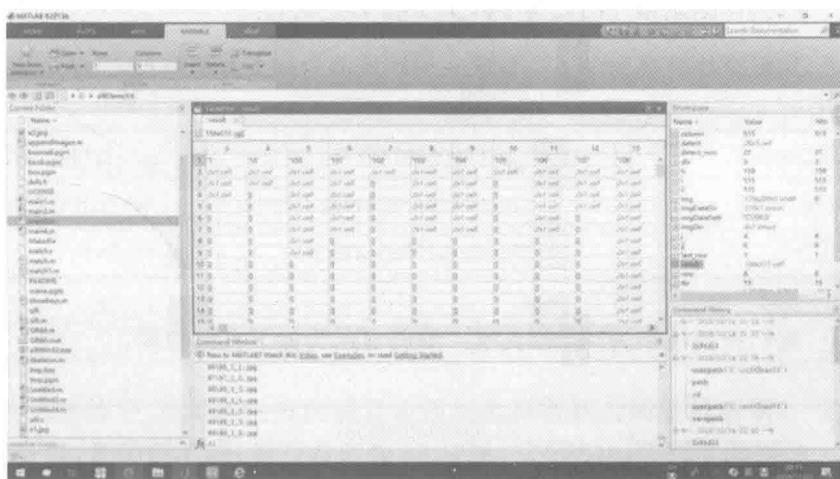


图 13 《村中南》所有切图的相关信息点

在图 13 中,中间一栏 result 就是《村中南》所有切图的相关信息点,横向为文件夹,纵向为具体图像,点击每张图像,可以得到具体信息点。如,点击文件夹 1 下,纵向为 2 的图像,即横 3 纵 2 的  $2 \times 1$  cell,可以得到图 14 所示:

在图 14 中,中间一栏 result{2, 3}就是我们在第一步中截图并保存的文件夹《村中南》1-1-2,它的相关关键信息点显示为:[20.7969348659004; 21.2137931034483; 28.3371647509579; 34.4613026819923; 45.0942528735632; 34.4674329501916; 49.8482758620690; 8.16551724137931; 27.3318007662835],即  $3 \times 3$  共九组数据。<sup>①</sup>

① 本文用  $3 \times 3$  的方式提取了《殷墟小屯村中村南甲骨》各个拓片的信息后发现每张图片都得到了较大差别的九组数据。

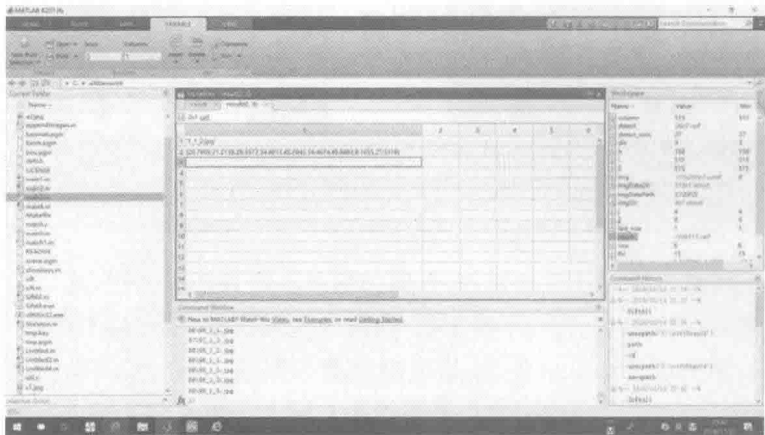


图 14 《村中南》1-1-2 图像信息点

在上述实验中,本文把《村中南》所有的单字图片的特征点提取后,每个甲骨文单字图像都拥有了一个具有唯一且高度显著相关的特征点集合。在此基础上,用户随机抽取《村中南》任意一张甲骨文单字的原始拓片,则包含该单字信息,也就是包含单字的特征点信息,电脑都可以通过这些特征点信息进行在线实时识别。

在本次实验中,按照 3 \* 3 的模板计算可以得到每个字的唯一编码,本文选取了以下几组数据展开具体对比。

第一组是“乙”字,具体原始拓片出处为 243\_1\_1 和 452\_6\_1。

从 243\_1\_1 上,我们查找该字为“乙”,其唯一编码显示为如下九组数据:

|                   |                  |                  |
|-------------------|------------------|------------------|
| 0.283524099548462 | 18.2820539745878 | 2.39945395358606 |
| 0.273023206972593 | 35.3145017326473 | 5.01417620497742 |
| 0                 | 22.5034127900872 | 0                |

从 452\_6\_1 上,我们查找该字为“乙”,其唯一编码显示为如下九组数据:

|                   |                  |                  |
|-------------------|------------------|------------------|
| 0                 | 23.6996336996337 | 0                |
| 0.329670329670330 | 33.2967032967033 | 4.28571428571429 |
| 0                 | 23.7728937728938 | 0                |

第二组是“于”字,具体原始拓片出处为 14\_2\_1 和 21\_2\_12。

从 14\_2\_1 上,我们查找该字为“于”,其唯一编码显示为如下九组数据:

|                  |                  |                  |
|------------------|------------------|------------------|
| 3.20268756998880 | 26.4874953340799 | 11.0787607316163 |
| 10.5188503172826 | 36.9764837625980 | 13.2960059723778 |
| 7.45054124673386 | 16.7674505412467 | 4.10600970511385 |

从 21\_2\_12 上,我们查找该字为“于”,其唯一编码显示为如下九组数据:

|                  |                  |                  |
|------------------|------------------|------------------|
| 6.30160799652325 | 27.8835289004781 | 11.4776184267710 |
| 5.77574967405476 | 37.3359408952629 | 15.2890047805302 |
| 6.87527162103433 | 16.2624945675793 | 5.41069100391134 |

第三组是“卜”字,具体原始拓片出处为 409\_1\_3 和 51\_3\_3。

从 409\_1\_3 上,我们查找该字为“”,其唯一编码显示为如下九组数据:







|                   |                  |                   |
|-------------------|------------------|-------------------|
| 0.327272727272727 | 7.60727272727273 | 7.34545454545455  |
| 1.30909090909091  | 47.3963636363636 | 3.62909090909091  |
| 0.552727272727273 | 11.1563636363636 | 0.458181818181818 |

从 51\_3\_3 上,我们查找该字为“”,其唯一编码显示为如下九组数据:

|                   |                  |                  |
|-------------------|------------------|------------------|
| 0.518433179723502 | 7.69969278033794 | 9.17818740399386 |
| 0                 | 45.6029185867896 | 5.47235023041475 |
| 0                 | 8.21812596006144 | 1.20967741935484 |

在程序中,用图像分割方法计算后,两组数据的绝对差值显示如表 1 所示。







表 1

|    | 出处 1  | 出处 2  | 绝对差值             |
|----|---|---|------------------|
| 1. | 243_1_1.jpg  | 452_6_1.jpg  | 12.1729462388918 |
| 2. | 14_2_1.jpg   | 21_2_12.jpg  | 14.3742751702970 |
| 3. | 409_1_3.jpg  | 51_3_3.jpg   | 11.3045691942466 |

这几组字均为异体字,在笔画的方向或长短粗细上有着较为明显的区别。“乙”字的两个字形呈镜像式,经过计算,其编码不同,绝对差值为 12.1729462388918,数值差异明显。对比“于”的两个字形的数据,我们可以看到虽然两字字形比较接近,但是经过计算,其编码不同,绝对差值为 14.3742751702970,特别是在左边的三个数据相差比较大,这也如实反应了两个字形的情况。

为了比较近似度更高的字形,我们再看“卜”和“一”两字在表 2 的三组数据所示。

表 2

|    | 出处 1   | 出处 2   | 绝对差值             |
|----|--|--|------------------|
| 1. | 39_1_3.jpg    | 39_1_35.jpg   | 11.6272736856228 |
| 2. | 340_1_30.jpg  | 341_2_14.jpg  | 6.70049007487431 |
| 3. | 341_11_7.jpg  | 341_2_14.jpg  | 13.6537791710206 |