

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2023)08-2330-13

论文引用格式: Zheng Y T, Li X L, Yin Z X, Gao G and Weng Y. 2023. Multi-feature fusion based automatic reconstruction in related to Chinese ancient manuscript fragments of Dunhuang. Journal of Image and Graphics, 28(08):2330-2342(郑玉彤, 李雪龙, 殷梓轩, 高歌, 翁彧. 2023. 多特征融合的敦煌古籍残片自动缀合. 中国图象图形学报, 28(08):2330-2342)[DOI:10.11834/jig.220896]

多特征融合的敦煌古籍残片自动缀合

郑玉彤, 李雪龙, 殷梓轩, 高歌, 翁彧*

中央民族大学民族语言智能分析与安全治理教育部重点实验室, 北京 100081

摘要: 目的 敦煌遗书作为敦煌学研究的根基, 是华夏多元民族弥足珍贵的文化遗产。现存的敦煌遗书大多为残片残卷, 给整理和研究带来了极大的困难。而人工缀残可谓至难, 费时费力, 对研究者的要求极高。随着计算机技术和计算机图形学的发展, 残片拼接技术也开始进入数字化时代。为此, 本文提出基于分层模型的数字图像缀合方法。**方法** 构建了一个古籍残片数据集。在流程设计上借鉴专家缀合的实践经验, 融入专家知识, 对碎片数字图像进行预处理。在碴口特征匹配的基础上, 融合多种缀合线索, 建立了包含物理层、结构层和语义层3层特征的分层模型, 从低层次到高层次对匹配结果进行评估打分, 完成两阶段的全自动缀合。**结果** 为了验证提出方法的有效性, 在由31张可拼接碎片(11组)和225张孤片组成的256张碎片数据集上进行实验。结果表明, 本文方法能够完成其中8组碎片的完整缀合, 2组不完整缀合, 并找出218张孤片。通过计算, 完整匹配准确率为95.76%, 不完整匹配准确率为95.70%, 缀合准确率都达到了95%。与现有类似任务的3种方法相比, 准确率均有明显提升。**结论** 本文提出的分层模型融合了多方面特征, 能有效完成古籍残片缀合任务, 提升研究人员的缀残效率。

关键词: 古籍残片; 敦煌遗书; 自动缀合; 碴口特征; 分层模型

Multi-feature fusion based automatic reconstruction in related to Chinese ancient manuscript fragments of Dunhuang

Zheng Yutong, Li Xuelong, Yin Zixuan, Gao Ge, Weng Yu*

Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE,
Minzu University of China, Beijing 100081, China

Abstract: Objective The Dunhuang manuscripts are evident for cultural heritage researches of China. Most of preserved manuscripts are restricted of its age-derived fragments and remnants and challenged for their collation and contexts. However, artificial reconstruction is time consuming and difficult to be developed. The emerging computer graphics-derived computer-aided virtual recovery technology has been facilitating in the context of high speed, easy to use and accuracy. **Method** We develop a model-hierarchical digital image reconstruction method. First, a dataset of ancient Dunhuang manuscript fragments is constructed. Second, expertise-relevant digital images of the fragments are pre-processed to assist in the rationalization of fragment features and establishment of a plane for the reconstructing process. Moreover, a three layers model is composed of physical, structural and semantic features via fusing multiple collocation cues. For the physical layer, grey-scale feature similarity measures are based on Jaccard correlation coefficients. For the structural layer, geomet-

收稿日期: 2022-09-16; 修回日期: 2022-11-22; 预印本日期: 2022-11-29

* 通信作者: 翁彧 wengyu@muc.edu.cn

基金项目: 国家自然科学基金重点项目(62236011); 国家社科基金重大项目(20&ZD279, 14ZDB105)

Supported by: National Natural Science Foundation of China(62236011); National Social Science Fund of China(20&ZD279, 14ZDB105)

ric contour matching is based on Freeman coding. For the semantic layer, character column spacing consistency features are based on grey-scale fluctuations. The whole reconstruction process is combined with two matching aspects of local and global contexts. The key to the local matching is to determine whether the two pieces match or not, while the vector similarity calculations are performed on the feature descriptors. The local matching results are evaluated and scored by reasonable thresholds between low and the high level. To realize the whole automation process, global matching strategy is implemented in terms of the Hannotta model, and the two aspect of fully automated reconstruction is performed. **Result** To verify the effectiveness of the proposed method, experiments are carried out on a 256-fragments dataset, which consists of 31 splinterable fragments (which can be reconstructed in 11 groups) and 225 orphaned fragments. The results analysis illustrates that 8 groups of fragments are fully matched, 2 groups are partially matched, and 218 orphaned fragments are identified as well. The accuracy of completed matching is 95.76% while incomplete matching is 95.70%. Both of their accuracies can be optimized and reached to 95%. To be more specific, each of partial accuracy are reached to 20.62%, 63.44% and 23.43%, and the improvement in complete accuracy of each are 39.85%, 68.09% and 23.33%. **Conclusion** The layered model combined with high-speed computing performance of the computer can incorporate multiple features and complete the reconstruction of ancient manuscript fragments effectively. The potential virtual reconstruction is beneficial for secondary damage to the fragile fragments, as well as some irreversible operations. Furthermore, the reconstructed results can provide an important basis for subsequent physical splicing, which can greatly enhance the efficiency of the artificial reconstruction.

Key words: ancient manuscript fragments; Dunhuang manuscripts; automatic reconstruction; curve feature; hierarchical model

0 引言

敦煌遗书于1900年发现于甘肃莫高窟藏经洞,现散藏于世界各地,总计近70 000号,写卷时代涵盖东晋至北宋,跨越600余年。由于经久历年,敦煌遗书绝大多数是残缺不全的,残卷比例达90%以上(张涌泉等,2021)。例如,国图敦煌特藏中,首全尾全的遗书共1 337号,仅占全部遗书17 319号的7.72%(方广钊,2014)。

敦煌遗书的残片缀合工作高度依赖专家,研究者根据残卷或残片的内容、边缘等特征,针对特定残片进行了手工缀合(陈丽萍,2014;景盛轩和陈琳,2017;张小艳,2015;张涌泉和朱若溪,2015)。张涌泉和罗慕君(2016)根据敦煌佛经残卷特点,将影响缀合工作的关键因素概括为内容相连、碴口相合等12项,为计算机辅助拼接中的特征选择提供了参考。

残片缀合本质是二维碎片拼接。众多学者对碎片拼接问题的研究成果,主要在特征利用和匹配策略上。在特征利用上,一部分基于碎片的几何特征(房然然,2015;刘金根等,2002;陶波等,1997;涂春萍等,2011;王磊等,1996;钟家强和王润生,2001;

周丰和黄晓鸣,2007;周鹏等,2002;朱延娟等,2007;Lin等,2015),通过获取碎片边缘信息,得到尖点、尖角、面积、角点特征和角序列等几何特征,再利用几何特征搜索匹配相邻碎片并进行拼接。古籍碎片大都有损坏,并且时有边界不够明晰的情况,所以拼接时如果只利用碎片的边界特征,效果并不理想。另一部分则基于碎片的内容特征,罗智中(2012)提出了基于碎片文字行特征或表格特征的碎片半自动拼接算法,但其基于边界交点距离进行拼接,不适合解决复杂碎片的拼接。刘秋菊等人(2016)提出将边界文字连续点的数目作为文字特征,然后使用连续八连通对边界文字灰度特征进行提取,构造模型进行拼接。刘宪国等人(2016)提出一种对于文档碎纸片基于灰度值矩阵的拼接方法,依据文本行特征进行动态聚类和类间排序,从而完成拼接还原。这些基于碎片文字本身特征的算法,大多适用于残片表面保存完好的情况。

相较于局部特征利用,关于全局匹配策略的研究较少。美国布朗大学的Kong和Kimia(2001)以及McBride和Kimia(2003)提出了一种基于弹性匹配的碎片自动拼接算法,通过多尺度策略来寻找最佳匹配,先进行粗匹配,再进行细匹配。由于经历了多个阶段动态规划,导致过程复杂,因此算法时间复杂度

较高。da Gama Leitão 和 Stolfi(2002)通过碎片轮廓上采样点的曲率表示碎片,也对多尺度方法的匹配问题进行了研究。Zhang 和 Li(2014)提出了一个基于图形的优化框架,利用特征点进行匹配,通过局部匹配平移旋转关系进行全局匹配。

在拼接任务中,还有学者对半自动拼接进行了研究。de Smet 等人(2003)使用链码算法跟踪每张纸的轮廓进行匹配,同时采用了一种人机交互的半自动化拼接方法。陆羽翊(2017)讨论了碎纸片复原问题的一系列算法,提出极小人工干预,并作为向一般问题推广的一次有益尝试。罗智中(2012)也对半自动拼接算法进行了研究,在算法运行过程中,首先由人工选择某一碎片,然后由机器计算此碎片与未拼接碎片的连续点,并根据连续点的个数对未拼接碎片排序,由人工判断是否拼接完成。

在其他研究领域,Tsamoura 和 Pitas(2010)以油画为研究对象,在内容特征的基础上,提出了一种基于颜色匹配的碎片拼接改进算法,通过动态规划为每个保留的图像片段识别匹配的轮廓段。Papaodysseus 等人(2002)针对壁画碎片的修复问题,从任意初始完整平面对象的片段轮廓形状中提取最大可能的信息,以指出可能的片段匹配。针对破碎刚体复原问题,赵夫群等人(2017)改进迭代最近点(iterative closest point, ICP)算法,提出一种先粗配再细配的刚体碎块匹配方法。针对计算机辅助文物虚拟复原,袁洁等人(2018)依据 Morse-Smale 复形理论,提出了一种基于断裂面拓扑特征的破碎文物自动拼接算法。

从之前的研究来看,国内外学者针对二维碎片拼接这一课题已经进行了多方面研究,也都取得了一定成果,并应用到部分领域的碎片辅助拼接中;而针对敦煌遗书残卷自动缀合的应用研究还比较匮乏。

综上,目前古籍残片自动缀合需要解决的主要问题包括:1)针对敦煌古籍残片缀合的专项研究中,图像来源仅为敦煌研究院一家收藏机构,且由实验团队自行拍摄,图像的格式统一、规整。但现实情况中,古籍残片来源广泛,受限于世界各地不同收藏点的能力和技术制约,残片图像存在分辨率不同、拍摄角度不同、缩放比例不同和清晰度不同等问题,不能直接用于自动缀合。2)古籍残片图像质量差,残片边缘残缺,数目不等甚至缺失。采用单一特征进行

匹配时,部分碎片可能会出现特征无法计算、特征缺失的情况,导致匹配精度不够高等问题,使用多种特征进行匹配可以弥补这些缺陷。3)在自动完成缀合算法方面,缺少有效的全局匹配算法。4)通过计算机进行敦煌遗书缀残测试评价困难。因为残片缀合专业性很强,缺少相关实验数据集。

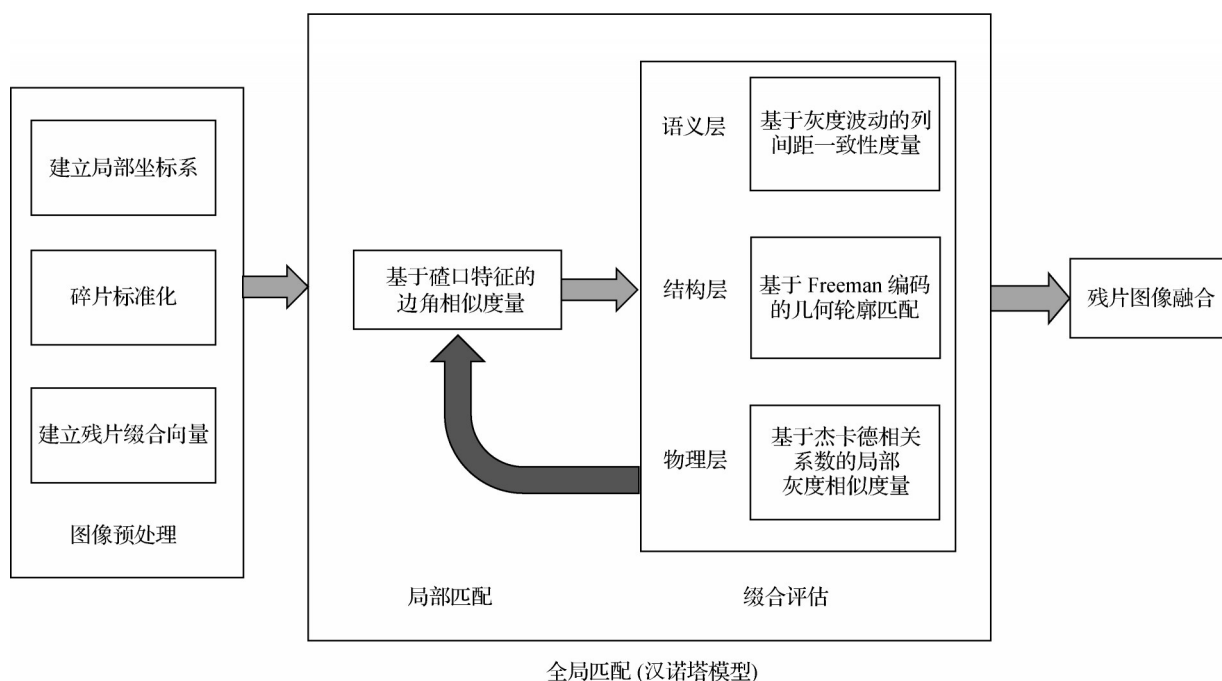
基于此,针对敦煌遗书残片年代久远、数量繁浩等现状,本文提出了一种面向古籍残片的全自动缀合方法,并以敦煌遗书残片为研究实例,通过对古籍碎片进行数字化和预处理,提取碎片的多方面特征,利用特征之间的相似度寻找潜在的可匹配碎片,最后完成拼接,其中显著特征的提取和碎片之间的匹配是该技术研究的重点内容。

本文的主要贡献如下:1)首次提出了基于多层特征融合评估的两阶段自动缀合模型。即基于碴口特征的局部匹配阶段以及基于物理层、结构层和语义层多层特征融合的缀合评估阶段。2)实现了由计算机辅助的古籍残片自动缀合算法,对比已知的专家缀合研究成果,实验取得了令人满意的效果,大幅缩小了古籍文献研究者的搜索、匹配和推测空间。3)构建了敦煌遗书残片缀合数据库,包含 256 幅图像,其中有 11 组(31 张)可缀合碎片及 225 张孤片。该数据库为研究敦煌遗书残片自动缀合提供了研究基础,也为构建更大规模古籍残片缀合数据库提供了参考。4)本文方法具备广泛适用性,并可为古籍残片自动缀合的改进提供有效借鉴。

1 碎片缀合算法

本文算法主要包括残片图像预处理、基于碴口特征的局部匹配及基于多层特征融合的缀合评估等部分,最后根据匹配结果进行残片缀合。总体结构如图 1 所示。

预处理阶段主要包括残片标准化、建立缀合向量等,目的是辅助完成残片特征的提取和建立缀合平面。基于碴口特征的局部匹配包括碴口特征提取、特征匹配和评价,用于完成一次粗匹配。分层模型融合了物理层、结构层和语义层信息,物理层主要利用残片的材质、年代在图像上表现出的颜色特征;结构层主要利用残片边缘的几何特征;语义层则挖掘残片文本列间距一致性等特征,通过多层特征匹配对碴口匹配结果做进一步确认和评估。最后在缀



全局匹配(汉诺塔模型)

图1 总体结构

Fig. 1 General structure

合平面确定匹配残片的位置和方向,完成残片缀合。全局匹配策略采用汉诺塔模型。

1.1 残片预处理

保存在世界各地收藏机构的残片,其规格不同,如图2所示,不同尺寸、分辨率的残片图像直接缀合将影响匹配效果。

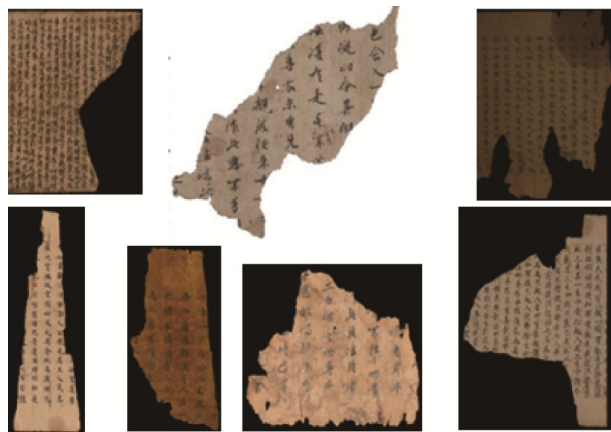


图2 敦煌遗书残片

Fig. 2 Dunhuang manuscript fragments

针对第*i*幅残片图像,预处理步骤如下:

1)建立残片*i*的局部坐标系。通过去噪、形态学处理和边缘提取,获得残片*i*的单像素宽闭合轮廓曲线,对该曲线进行多边形拟合,得到多边形 W_i 及其

边缘点集合,具体为

$$V_i = \{v_1(x_1, y_1), v_2(x_2, y_2), \dots, v_n(x_n, y_n)\}$$

计算 W_i 的质心,并作为残片*i*局部坐标系的原点,基于霍夫变换识别文字走向,将其作为正*Y*轴方向(上下),顺时针垂直*Y*方向的为正*X*轴,建立局部坐标系,目的是方便后续获取残片的相对位置,以及统一自动化摆正残片图像的方向。

2)图像标准化。提取图像中的字符,确定标准字符尺寸,将所有残片的字符与标准字符进行比较,计算出缩放比例 $Scale_i$ 。将所有图像按照 $Scale_i$ 进行调整,确保缀合残片上的文字大小相同,消除因拍摄角度、拍摄高度和拍摄设备等原因造成图像比例不一致对缀合结果的影响。在此过程中,为了有效利用碎片的文本特征,需要进行残片字符分割和图像尺寸调整。

针对字符分割,本文采用投影法,即通过统计投影的像素数目,作为字符分割的依据。

(1)进行图像二值化处理。

$$dst(x, y) = \begin{cases} 0 & src(x, y) \geq T \\ 255 & src(x, y) < T \end{cases} \quad (1)$$

式中, dst 为阈值分割结果图像,与原始图像具有相同的大小和类型, $src(x, y)$ 表示源图像在 (x, y) 坐标处的像素值, T 为进行图像分割的阈值。

(2)在投影之前进行形态学调整,做膨胀操作。

A 和 B 是 Z^2 中的集合, B 对 A 的膨胀定义为

$$A \oplus B = \left\{ z \mid \left(\hat{B} \right)_z \cap A \neq \emptyset \right\} \quad (2)$$

式中, A 是被膨胀的集合(图像物体), B 是一个结构元, \hat{B} 是 B 的反射。式(2)表示 B 的反射进行平移与 A 的交集不为空。

(3)进行水平投影和垂直投影。

(4)分割字符。分割结果如图3所示。



图3 字符分割结果

Fig. 3 Character segmentation results

理想状态下,分割框应为一字一框,但通过观察图3可知,存在过度分割和粘连的现象。考虑中文结构具有的复杂性以及手写字体的相对不规则性,对分割框的宽高比 $Ratio$ 做出限制。由于汉字具有方块字的特征,且本文算法不论是图像比例缩放还是字距提取,均只需将文本按字符切分,因此宽高比限制设为 $0.9 \leq R \leq 1.1$ 。改进后的字符分割如图4所示,可以看出,过度分割、粘连现象明显减少。

字符分割后,得到分割框序列,具体为

$$Res = \left\{ (h_1, w_1), (h_2, w_2), \dots, (h_n, w_n) \right\}$$

式中, h_k 和 w_k 为分割框的高和宽,统计 h_k 和 w_k 的分布,选取最大值作为参照数据 h 和 w 。

依据标准字符尺寸,计算缩放比例 S_i ,调整残片图像尺寸。残片图像的缩放比例为

$$S_i = \frac{H}{(h + w)/2} \quad (3)$$

式中, H 为选取的标准字符大小的边长像素数。若 $S_i > 1$,表示需对图像进行上采样调整;若 $S_i < 1$,表示

需对图像进行下采样调整。

标准化后的残片如图5所示。



图4 筛选后的字符分割结果

Fig. 4 Improved character segmentation results



图5 带有比例尺的标准化残片

Fig. 5 Standardized fragments with scale

3)建立残片 i 的缀合向量 A_i 。针对每个残片 i ,提取外轮廓 W_i 及点集合 V_i ,计算残片面积 A_i 等。缀合向量 A_i 包括质心、外轮廓坐标点集合、局部坐标系和残片面积等属性。

1.2 基于碴口特征的局部匹配

完成图像预处理后,进行碎片局部匹配。基于碴口特征的局部匹配包括碴口特征提取、特征描述子构建和碴口匹配结果评估等,其核心是碴口特征提取。

建立残片 i 的外轮廓坐标点集合 V_i ,进行特征提取,得到碴口特征描述符 $S_i = (\theta_k, L_k, \delta_k)$,其中 θ_k 表示旋转角序列, L_k 表示边长序列, δ_k 表示倾斜角度序列。具体步骤如下:

1)通过相邻的3个点 V_{k-1}, V_k, V_{k+1} ,得到两个向量 $P_k:(V_k \rightarrow V_{k-1})$ 和 $Q_k:(V_k \rightarrow V_{k+1})$,则 P_k 和 Q_k 的夹角 θ_k 的余弦值为

$$\cos(\theta_k) = \frac{P_k \cdot Q_k}{\|P_k\| \times \|Q_k\|} \quad (4)$$

式中,“ \cdot ”表示向量 P_k 和 Q_k 进行点积运算。

2)定义相邻两点 $V_{k-1}(x_{k-1}, y_{k-1})$ 和 $V_k(x_k, y_k)$ 的欧几里得距离 L_k 为

$$L_k = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2} \quad (5)$$

在局部坐标系中,设沿坐标轴 X 方向的单位向量为 x_0 ,定义相邻两点 V_{k-1} 和 V_k 之间线段倾斜角的向量 δ_k 为

$$\cos \delta_k = \frac{P_k \cdot x_0}{\|P_k\| \times \|x_0\|} \quad (6)$$

3)综合式(4)—(6),得到3个序列,即旋转角序列 θ_k 、边长序列 L_k 和倾斜角序列 δ_k ,其中, $k = 1, 2, \dots, n$,从而完成碴口特征描述符构建。

完成描述符构建后,可依据相似度准则在多个碎片中寻找可以进行匹配的碎片,选取合适的阈值进行评估,完成局部匹配。

1.3 多层特征融合

敦煌遗书具有物理上、结构和语义上的多层特征,不同层次特征具有异构性,又存在一致性和互补性,因此可以利用多层特征共同描述研究对象。多层特征匹配包括在物理层设置基于杰卡德相关系数的局部灰度相似度量,在结构层设置基于Freeman编码(刘勇奎,2001)的几何轮廓匹配,在语义层设置基于灰度波动的列间距一致性度量等。

1.3.1 基于杰卡德相关系数的局部灰度相似度量

颜色是物质的一种物理性质,常用做全局特征。但通过深入考量敦煌遗书碎片图像,发现碎片纸张的颜色因污渍等原因存在不均匀现象,甚至边缘与内部颜色也不尽相同,如图6所示,在这种情况下,通过全局特征判断两张碎片的匹配度,并不能取得非常理想的效果。因此,提出基于杰卡德相关系数的局部灰度相似度量,通过提取碎片边缘附近的灰度值,判断匹配情况。

算法步骤如下:

输入:两张碎片的灰度图。

输出:在物理层两张碎片是否匹配。

1)经过形态学操作去噪,利用拉普拉斯算子提



图6 颜色不均的残片示例

Fig. 6 Example of uneven color fragments

取当前两张碎片轮廓,得到碎片边缘上的点 $(X1_i, Y1_i)$ 和 $(X2_i, Y2_i)$ 。

2)通过点 $(X1_i, Y1_i)$ 和 $(X2_i, Y2_i)$ 检索到图像对应位置的灰度值向量,分别为 A 和 B 。

3)计算杰卡德相关系数 $J(A, B)$,具体为

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

式中, $|A \cap B|$ 和 $|A \cup B|$ 得到的都是集合运算后的具体值。杰卡德相关系数的值可以衡量两张碎片在物理层的相似度,值越大说明相似度越高。

4)匹配判断。具体计算为

$$BP_y = \begin{cases} \text{True} & J(A, B) > P_y T \\ \text{False} & \text{其他} \end{cases} \quad (8)$$

式中, BP_y 为物理层匹配结果布尔值, $P_y T$ 表示设置的评估阈值。

1.3.2 基于Freeman编码的几何轮廓匹配

链码具有平移不变性,但对于描述碎片轮廓信息,还应具有旋转不变形,因此选择了一阶差分链码。

通过边缘检测,获取碎片图像边缘上像素点的坐标序列。具体为

$$C = \{C_1(x_1, y_1), C_2(x_2, y_2), \dots, C_i(x_i, y_i), \dots, C_n(x_n, y_n)\}$$

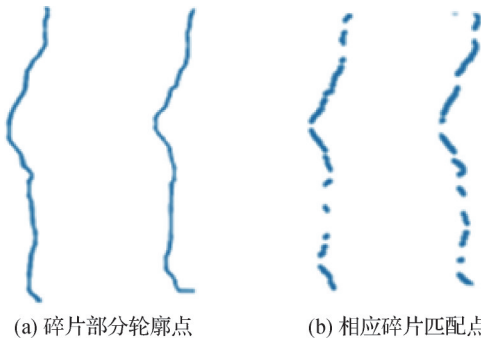
然后,将坐标序列转化成差分链码序列。

相似性评估采用最长公共子序列方法。古籍碎片存在边角磨损,丢失了部分轮廓信息,最长公共子序列方法鲁棒性好,可削弱缺失部分带来的影响,如图7所示,分别是两个碎片的部分轮廓及其最长公共子序列匹配结果。最长公共子序列方法关注相似

部分的总长度,可设为 F_c 。匹配判断计算式为

$$BSt = \begin{cases} \text{True} & F_c > StT \\ \text{False} & \text{其他} \end{cases} \quad (9)$$

式中, BSt 为结构层匹配结果布尔值, StT 表示设置的评估阈值。



(a) 碎片部分轮廓点 (b) 相应碎片匹配点

图7 最长公共子序列匹配结果

Fig. 7 Matching results of longest common subsequence

((a) partial contour points of fragments;
(b) match points for corresponding fragments)

1.3.3 基于灰度波动的列间距一致性度量

本文称文档中列与列之间的间距为列间距。古籍的列间距比较显著且在同一篇古籍中较为稳定,通过投影法提取古籍碎片图像中列间距。

对进行二值化后的古籍图像(图8)进行垂直投影,投影效果如图9所示,投影值为0的区域对应每列的间隔,取平均值,设为 Col 。匹配判断计算式为

$$BSe = \begin{cases} \text{True} & Col > SeT \\ \text{False} & \text{其他} \end{cases} \quad (10)$$

式中, BSe 为语义层匹配结果布尔值, SeT 表示设置的评估阈值。

1.4 基于多层特征融合的评估算法

多层特征融合评估缀合的目标是进一步确认匹

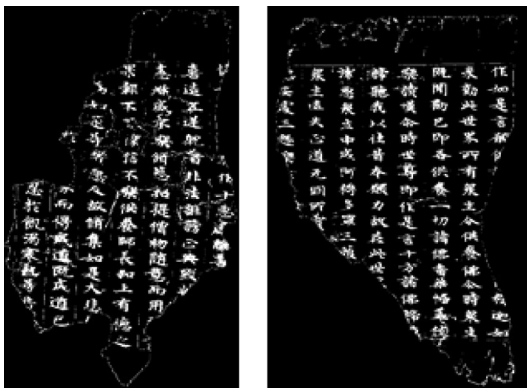


图8 二值化古籍图像

Fig. 8 Binarized ancient manuscript images



图9 垂直投影结果

Fig. 9 Vertical projection results

配结果,并确定需拼接残片的位置和方向。

首先进行初始化,包括:

1)建立缀合平面 Z ,以黑底纯色图像作为缀合结果背景,建立全局缀合坐标系。

2)初始化汉诺塔结构,将所有待缀合残片按照面积从大到小顺序排列,置入堆栈 $S1$,栈顶残片面积最大,栈顶元素设为 SF ,即待匹配的小碎片。另设2个堆栈,分别记为 $S2$ 和 $S3$,完成汉诺塔结构。

3)取 $S1$ 栈顶残片放入全局缀合平面 Z ,令其质心与全局坐标系原点 O 重合,局部坐标系方向与全局坐标系方向一致。

设缀合平面已完成缀合的残片整体为 BF , BF 采用全局坐标系。缀合算法的步骤如下:

输入:残片集合 $F = \{F_1, F_2, \dots, F_i, \dots, F_n\}$

输出:所有缀合平面的匹配结果集合和可视化结果。

1)获取 $S1$ 栈顶残片 SF ,并从栈顶弹出此元素。

2)将残片 SF 与 BF 基于磕口特征进行局部匹配,记录局部匹配位置。若匹配成功,进入步骤4),否则将 SF 置 $S2$ 中。

3)如果 $S1$ 不空,转到步骤1);否则,转到步骤7)。

4)将残片 SF 与 BF 进行多层特征匹配,并进行匹配评估,若评估结果为成功,则转到步骤5);否则,将 SF 置 $S2$ 中,转到步骤3)。

5)确定 SF 在缀合平面内的方向和位置,将 SF 与 BF 进行缀合,更新 BF 。

6)如果 $S1$ 不空,则转到步骤1)。

7)处理 $S2$ 。若 $S2$ 空,则转步骤8);否则,将 $S2$ 中元素依次出栈,进入 $S3$,再从 $S3$ 依次出栈进入 $S1$ 。完成一组缀合结果,初始化,得到新的 BF ,转到步骤3)。

8)结束。

对于步骤4)进行的多层特征匹配和评估,核心是多层特征的提取和融合。在特征融合和评估过程中,3层特征单独计算,但又同时为整体残片 BF 和当前残片 SF 进行缀合评估。在融合过程中,考虑到

特征重要性和量纲影响,对每层结果进行布尔逻辑的与操作,将3层特征的重要性放在同等位置,并且设置准入标准。最终评估式为

$$BA = BPy \& BSt \& BSe \quad (11)$$

式中, Bpy , BSt 和 BSe 分别为物理层、结构层和语义层的评估结果布尔值, BA 表示总体评价结果。

在工作栈S1上的碎片匹配方式为

$$BA_{ij} = BPy_{ij} \& BSt_{ij} \& BSe_{ij} \quad (12)$$

式中, i 和 j 都代表碎片的索引值,且 $i \neq j$ 。设 i 是整体碎片 BF 的索引值,则 j 为当前残片 SF 的索引值,即从工作栈S1中取出的栈顶元素。

在步骤5)中需要确定 SF 在缀合平面内的位置。根据在步骤2)中记录的局部匹配点位置 $SF(SFX_1, SFY_1)$ 和 $BF(BFX_1, BFY_1)$,计算在 SF 中质心 (SFX_0, SFY_0) 与匹配点的相对位置变化,记为 $(SFX_0 - SFX_1, SFY_0 - SFY_1)$ 。在全局缀合坐标系中,当前残片的质心位置应为 $(BFX_1 + SFX_0 - SFX_1, BFY_1 + SFY_0 - SFY_1)$ 。

2 实验和结果

2.1 环境说明与实验设置

实验算法在Pycharm2021上建立,软件环境为Windows 10操作系统,处理器为Intel(R) Core(TM) i5-8250U,所用开发语言为Python3.6,实验的计算主要依靠CPU完成。物理层设置阈值标准为0.5,结构层设置自适应阈值为碎片边缘向量长度的25%,语义层设置阈值为1。

2.2 敦煌遗书残片缀合数据集

由于敦煌遗书的收藏存在多、散、对外公布标准不统一等现象,单一数据来源很难满足需要。为了能够量化评价并分析本文算法,本文构建了独立的敦煌遗书残片缀合数据集。数据来源包括国际敦煌项目(International DunHuang Project, IDP)网站、国家图书馆中华古籍资源库、《英藏敦煌文献》(四川人民出版社1990年出版)、《国家图书馆藏敦煌遗书》(北京图书馆出版社2005年出版)、《俄藏敦煌文献》(上海古籍出版社2001年出版)等。敦煌遗书残片缀合数据集主要分为两部分,一是根据已有的缀合研究相关专业文献整理的可缀合残片组,共计11组,31张;另一部分是未发现可匹配项的孤片

225张,作为实验干扰项;数据集总计256张。

2.3 实验结果

2.3.1 基于碴口特征的局部匹配及评估

现有方法中,边角特征一般包含旋转角序列 θ_k 以及边长序列 L_k ,在对碎片进行这两项特征提取并粗匹配后,可分别在两个碎片上得到一个可匹配的轮廓子集。但有时可能出现误匹配情况,如图10所示,S9040和S9069号碎片是正确的匹配组合,但基于边角特征碴口匹配结果显示的匹配段(图中方框内线段)却是错误的。

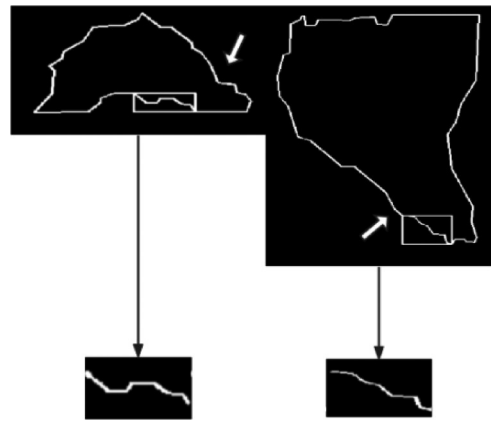


图10 边角特征碴口匹配的误匹配结果

Fig. 10 Error matching result with edge and corner feature

古籍残片轮廓信息较为丰富,只考虑边长和角度,可能出现轮廓信息近似而导致的误匹配情况。为减少这种误匹配现象,本文在碴口特征中加入倾斜角序列特征,实验结果如图11所示,S9040和S9069号碎片在加入倾斜角序列特征后,成功找到正确匹配段(图中方框内线段)。

2.3.2 基于分层模型的缀合评估

轮廓信息相似度较高的碎片可能不止一个,因此在碴口匹配完成后,需要进行进一步的匹配和评估。用匹配度量 $Match_Deg$ 来表示每个匹配段的匹配程度(房然然,2015),按其值由小到大排列,值越小,代表两个碎片轮廓信息越匹配,表1为S8983号碎片的碴口匹配结果。

由表1可知,经碴口特征判断,与S8983最匹配的碎片为S9081。但真实匹配碎片应为排名第2的S8961,因此可知仅凭借碴口特征得出的匹配碎片是误匹配结果。图12展示了S8983(左)与S9081(右)的碴口匹配结果。通过图12也可看出,S8983与S9081方框内匹配段的边角走势确有相似之处,但

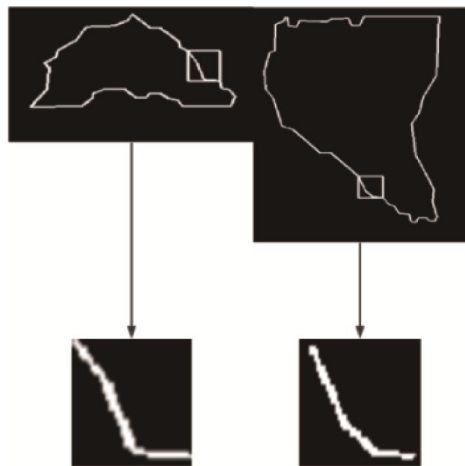


图 11 加入倾斜角序列特征的碴口匹配结果
Fig. 11 Matching result by adding the oblique sequence features

表 1 S8983 残片碴口匹配结果

Table 1 S8983 fragment matching results

排名	碎片编号	Match_Deg
1	S9081	0.712 9
2	S8961	0.735 7
3	BD9428	3.125 7
4	Or15012	6.219 3
5	Or9912	12.332 5

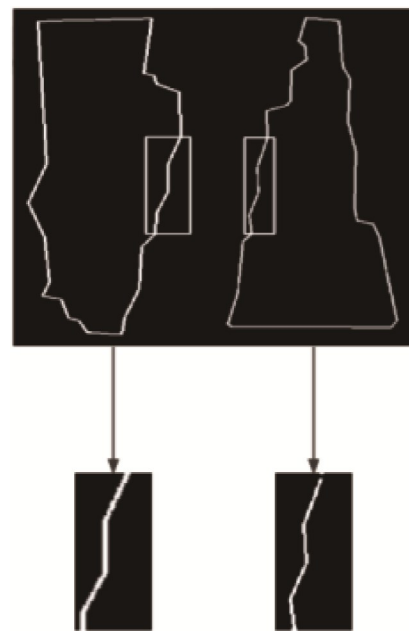


图 13 S8961(左)与S8983(右)碴口匹配结果
Fig. 13 Matching result of fragment S8961(left) and S8983(right)

此本文加入了融合多层特征的进一步匹配和评估,通过增加特征筛选提高准确率。以列字距一致性特征的加入为例,表 2 是表 1 中 5 个碎片的列字距提取结果。

通过字距结果可以看出,S8983 碎片和 S9081 碎片的列字距差较大,不符合匹配要求;而列字距差同样较小的碎片还包括 Or9912,但该碎片在碴口匹配阶段表现不佳。因此通过综合碴口匹配结果和列字距一致性结果,能够筛选出唯一匹配碎片为 S8961,符合实际情况,说明了多层特征融合的重要性。

表 2 列字距提取结果

Table 2 Results of column space

碎片编号	列字距	S8983 碎片的列字距差
S8961	16	1
S9081	30	15
BD9428	26	11
Or15012	27	12
Or9912	17	2

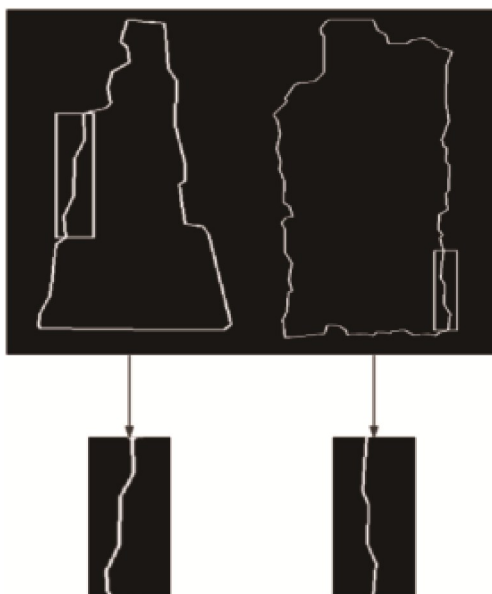


图 12 S8983(左)与S9081(右)碴口匹配结果
Fig. 12 Matching result of fragment S8983(left) and S9081(right)

图 13 的 S8961 才是正确匹配。

可见,碴口特征匹配并不能有效排除误匹配,因

2.3.3 实验结果分析

本文采用 PA (partial accuracy) 和 CA (complete accuracy) 作为评估指标。PA 即不完整缀合准确率,包含组内元素未找全的情况;CA 即完整缀合准确率,指

组内碎片都找全且无杂片混入。

经过实验,多特征融合算法成组碎片输出匹配结果如表3所示。可以看出,以组为单位计算CA(完整缀合准确率),能够完整找到8个成组和218张(组)孤片,共输出236组结果,CA为95.76%。以张为单位计算PA(不完整缀合准确率),可完整找到8个组,共计23张,不完整找到2个成组,共计4张,找出孤片仍为218张,碎片总片数为256片,则计算PA为95.70%。算法最终输出部分缀合结果和孤片如图14所示。

表3 输出序列结果
Table 3 Output sequence results

匹配序列编号	序列内容
1	'6_2', '6_1'
2	'13_1', '13_2', '13_3'
3	'3_2', '3_1'
4	'15_1', '15_2', '15_3', '15_4', '15_5', '15_6'
5	'16_1'
6	'16_2'
7	'28_3', '28_1'
8	'28_2'
9	'19_1', '19_2', '19_3', '19_4'
10	'26_2', '26_1'
11	'20_1'
12	'25_2', '25_1'
13	'20_1', '20_2'
14	'22_1', '22_2'

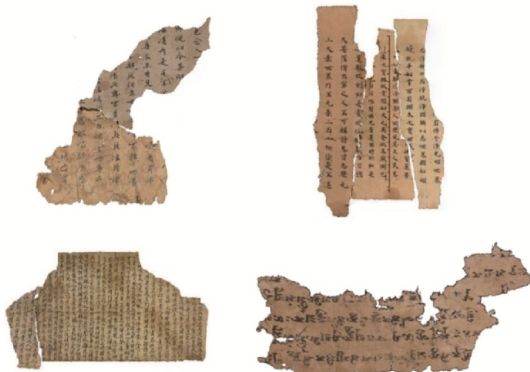


图14 缀合孤片结果

Fig. 14 Results of reconstruction

2.3.4 消融实验

为验证各个层次特征的有效性及其对最终缀合结果的影响程度,在固定其他条件的情况下,分别在单个层次特征、两两组合层次特征和3个层次特征组合下进行实验。结果如表4所示,其中,*Py*,*St*,*Se*分别表示物理层、结构层和语义层。

从表4可以看出,从特征单独使用的结果上看,物理层和结构层特征鲜明,有效性明显高于语义层特征,原因可能是古代对书写字距有明确规定,简单的列间距特征区分性不强;从两层特征融合使用来看,缀合成功率得到了提升,语义层特征虽然没有得到表现,成为其他层的附属,但也没有导致漏匹配的提升;从3层特征融合使用的实验结果看,语义层的加入得到了最优的结果,说明语义层特征的使用在降低误匹配上是有效的。

实验结果表明,模型中的每一层特征对缀合成功均有着正面影响;融合使用全部特征时,充分利用了多层特征的一致性和互补性,缀合效果最佳。说明了多特征融合的必要性以及算法设计的合理性。

表4 不同层组合的缀合性能对比

Table 4 Comparison of reconstruction properties of different layer combinations

组合	CA	PA	%
<i>Se</i>	57.48	29.30	
<i>St</i>	91.24	77.34	
<i>Py</i>	92.14	87.89	
<i>Py+Se</i>	92.14	87.89	
<i>St+Se</i>	91.24	77.34	
<i>Py+St</i>	94.47	92.19	
<i>Py+St+Se</i>	95.76	95.70	

注:加粗字体表示各列最优结果。

2.4 与其他算法的对比

考虑到实际应用场景,在进行不同算法对比时,本文的评估指标除PA和CA之外,还加上了耗时指标。算法对比结果如表5所示。其中,算法1是基于轮廓的二维非规则碎片拼接(房然然,2015),提出了一种新的基于轮廓的二维碎片的拼接过程,使用一种适用于多边形特征的、改进的局部匹配方法寻找空间邻接碎片对,并度量匹配程度,通过生成全局路径并将碎片依

照路径拼接;实现了多碎片的拼接;算法2是基于角序列的二维碎片轮廓匹配算法(周丰和黄晓鸣,2007),是一种以角序列为特征的碎片轮廓匹配算法,可以更加有效地表示尖角信息;算法3是仅采用倾斜角序列特征的碴口匹配算法,用于验证进行多层特征融合的有效性和必要性。

由表5可知,本文算法在两种准确率评判指标上均优于对比算法,在PA上相比其他3个算法分别提高了20.62%、63.44%和23.43%,在CA上相比其他3个算法分别提高了39.85%、68.09%和23.33%。但在运行耗时上,本文算法要比其他算法更长。

表5 算法对比结果

Table 5 Comparison of different algorithms

算法	PA/%	CA/%	耗时/s
算法1	55.08	55.51	5 169.784
算法2	32.26	27.27	5 556.383
算法3	72.27	72.03	5 063.228
本文	95.70	95.76	7 829.484

注:加粗字体表示各列最优结果。

2.5 性能分析

算法1采用多边形逼近法提取多边形顶点表示轮廓,提升了模型处理不规则碎片的能力。但古籍残片由于年久腐蚀存在噪声,会影响多边形顶点的分布,继而影响匹配结果,导致准确率较低。

算法2通过向量积求得轮廓线的角度序列表示。特征值的选取及计算直接影响匹配的结果,在碎片数据大且复杂的情况下不适用,而且古籍残片的缀合并不一定是严丝合缝的,角度存在误差是正常现象,所以需要其他层次特征的融合进行辅助。

算法3较算法1和算法2有优势,PA和CA显著提升,倾斜角序列特征的引入有效提高了准确率。本文方法与算法3对比,证明了特征融合的必要性和分层模型的有效性。

3 结论

本文提出了在碴口特征的基础上,进行多层特征融合评估的古籍残片缀合方法,并以敦煌遗书残片为实例进行了算法实现和模型评估,准确率为

95%,有效提升了古籍残片缀合的效率。

与其他算法相比,本文模型在缀合准确率上表现更优。古籍残片多数存在损坏、腐蚀、破碎、污渍,甚至部分丢失的情况,匹配任务复杂,在仅依赖单一特征的情况下,准确率较低。而分层模型不仅能够提取古籍碎片多层次的特征,提升碎片的描述水平,还能更好地完成碎片缀合评估,减少缀合过程中的漏判和错判,进一步提升碎片缀合的准确性。

但是,本文仍有如下不足:1)受限专家缀合的真实数据难以获取,且目前对敦煌古籍残片自动缀合的关注和研究较少,仍属冷门方向,没有适用的公开数据集,验证数据集的规模和覆盖范围仍有提升空间。2)算法借助多层特征信息帮助筛选,增添了处理步骤。同时,汉诺塔结构没有引入并行优化,造成耗时较长。3)研究尚处探索阶段,实验中得到的缀合结果,除了有标准答案的分组,其他结果因碍于专业限制,不能给出确切解释,需向文献专家寻求进一步的帮助。

因此,未来的研究方向重点聚焦在以下方面:1)不断收集和整理数据,持续扩充古籍残片数据集规模,使之成为敦煌遗书残片缀合研究的公开数据集;2)引入并行算法,优化全局匹配策略,减少运行时间,提升算法的整体运行效率;3)古籍文献中含有大量的语义信息可辅助提升缀合精度与效率,未来可探索基于残字、残词、残句和字形进行更深入的基于语义的缀合研究。

参考文献(References)

- Chen L P. 2014. Piecing together and research of the Dunhuang manuscript *Datang Tianxia Jun Xingshi Zupu* focusing on S. 5861. *Dunhuang Research*, (1): 78-86 (陈丽萍. 2014. 敦煌本《大唐天下郡姓氏族谱》的缀合与研究——以S.5861为中心. *敦煌研究*, (1): 78-86) [DOI: 10.13584/j.cnki.issn1000-4106.2014.01.011]
- da Gama Leitão H C and Stolfi J. 2002. A multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9): 1239-1251 [DOI: 10.1109/TPAMI.2002.1033215]
- De Smet P, De Bock J and Corluy E. 2003. Computer vision techniques for semi-automatic reconstruction of ripped-up documents//*Proceedings Volume 5108, Visual Information Processing XII*. Orlando, USA: SPIE: 189-197 [DOI: 10.1117/12.501078]
- Fang G C. 2014. A Dunhuang documents collected in the national library of China. *Dunhuang Research*, (3): 123-131 (方广钊. 2014. 中

- 国国家图书馆藏敦煌遗书. 敦煌研究, (3): 123-131 [DOI: 10.13584/j.cnki.issn1000-4106.2014.03.016]
- Fang R R. 2015. Research on Automatic Reassembly of 2D Fragments. Jinan: Shandong Normal University (房然然. 2015. 2维碎片自动拼接技术研究. 济南: 山东师范大学)
- Jing S X and Chen L. 2017. A preliminary reconstruction of the Mahāparinirvāna-sūtra from Dunhuang manuscripts in the British Library. *Dunhuang Research*, (3): 92-107 (景盛轩, 陈琳. 2017. 英藏敦煌《大般涅槃经》残卷初步缀合. 敦煌研究, (3): 92-107) [DOI: 10.13584/j.cnki.issn1000-4106.2017.03.011]
- Kong W X and Kimia B B. 2001. On solving 2D and 3D puzzles using curve matching//Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, USA: IEEE: II [DOI: 10.1109/CVPR.2001.991015]
- Lin Y B, Wang C, Cheng J, Chen B L, Jia F K, Chen Z G and Li J. 2015. Line segment extraction for large scale unorganized point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102: 172-183 [DOI: 10.1016/j.isprsjprs.2014.12.027]
- Liu J G, Wu Z P, Liu S Q and Yin S M. 2002. A merging algorithm for images based on segmentation of feature regions. *Journal of Xidian University*, 29(6): 768-771 (刘金根, 吴志鹏, 刘上乾, 殷世民. 2002. 一种基于特征区域分割的图像拼接算法. 西安电子科技大学学报(自然科学版), 29(6): 768-771) [DOI: 10.3969/j.issn.1001-2400.2002.06.016]
- Liu Q J, Chen P and Wang Z Y. 2016. Algorithm design on scraps of paper splicing based on text feature. *Research and Exploration in Laboratory*, 35(11): 110-113 (刘秋菊, 陈平, 王仲英. 2016. 一种基于文字特征的碎片拼接算法设计. 实验室研究与探索, 35(11): 110-113) [DOI: 10.3969/j.issn.1006-7167.2016.11.027]
- Liu X G, Jia Z Y, Liu W J and Han M. 2016. Study of document recovery method based on gray-level matrix. *Application Research of Computers*, 33(12): 3901-3903, 3908 (刘宪国, 贾子钰, 刘万军, 韩敏. 2016. 一种基于灰度值矩阵的文档复原方法研究. 计算机应用研究, 33(12): 3901-3903, 3908) [DOI: 10.3969/j.issn.1001-3695.2016.12.085]
- Liu Y K. 2001. Research on the compression algorithm for Freeman chain code. *Chinese Journal of Computers*, 24(12): 1294-1298 (刘勇奎. 2001. Freeman链码压缩算法的研究. 计算机学报, 24(12): 1294-1298) [DOI: 10.3321/j.issn:0254-4164.2001.12.009]
- Lu Y H. 2017. A method of fragment splicing restoration under minimal human intervention. *China Computer and Communication*, (14): 51-54 (陆羽翊. 2017. 一种在极小人工干预下的碎片拼接复原方法. 信息与电脑, (14): 51-54) [DOI: 10.3969/j.issn.1003-9767.2017.14.023]
- Luo Z Z. 2012. Semi-auto stitching of scrapped paper based on character characteristic. *Computer Engineering and Applications*, 48(5): 207-210 (罗智中. 2012. 基于文字特征的文档碎片半自动拼接. 计算机工程与应用, 48(5): 207-210) [DOI: 10.3778/j.issn.1002-8331.2012.05.060]
- McBride J C and Kimia B B. 2003. Archaeological fragment reconstruction using curve-matching//Proceedings of 2003 Conference on Computer Vision and Pattern Recognition Workshop. Madison, USA: IEEE: 3 [DOI: 10.1109/CVPRW.2003.10008]
- Papaodysseus C, Panagopoulos T, Exarhos M, Triantafillou C, Fragoulis D and Doumas C. 2002. Contour-shape based reconstruction of fragmented, 1600 BC wall paintings. *IEEE Transactions on Signal Processing*, 50(6): 1277-1288 [DOI: 10.1109/TSP.2002.1003053]
- Tao B, Yu Z W and Zheng X X. 1997. Automatic image mosaic. *Chinese Journal of Biomedical Engineering*, (4): 29-35 (陶波, 于志伟, 郑筱祥. 1997. 图像的自动拼接. 中国生物医学工程学报, (4): 29-35)
- Tsamoura E and Pitas I. 2010. Automatic color based reassembly of fragmented images and paintings. *IEEE Transactions on Image Processing*, 19(3): 680-690 [DOI: 10.1109/TIP.2009.2035840]
- Tu C P, Chai Y H, Li G L and Liu J F. 2011. An image mosaic method based on refined corner points feature matching. *Research and Exploration in Laboratory*, 30(10): 40-43 (涂春萍, 柴亚辉, 李广丽, 刘觉夫. 2011. 一种基于Harris角点特征精确匹配的图像拼接方法. 实验室研究与探索, 30(10): 40-43) [DOI: 10.3969/j.issn.1006-7167.2011.10.012]
- Wang L, Mo Y L and Qi F H. 1996. An improved method of edge detection based on Canny's theory. *China Journal of Image and Graphics*, 1(3): 191-195 (王磊, 莫玉龙, 戚飞虎. 1996. 基于Canny理论的边缘提取改善方法. 中国图象图形学报, 1(3): 191-195) [DOI: 10.11834/jig.19960360]
- Yuan J, Zhou M Q, Geng G H and Zhang Y H. 2018. Automatic reassembly of fractured fragments using Morse topological features. *Acta Automatica Sinica*, 44(8): 1486-1495 (袁洁, 周明全, 耿国华, 张雨禾. 2018. 基于Morse-Smale拓扑特征的文物碎片拼接算法. 自动化学报, 44(8): 1486-1495) [DOI: 10.16383/j.aas.2017.c160778]
- Zhang K and Li X. 2014. A graph-based optimization algorithm for fragmented image reassembly. *Graphical Models*, 76(5): 484-495 [DOI: 10.1016/j.gmod.2014.03.001]
- Zhang X Y. 2015. The patching-up and study on the fragments of Xinpusajing, Quanshanjing and Jiuzhuzhongshengkunanjing in Dunhuang manuscripts. *Fudan Journal (Social Sciences)*, 57(6): 12-20 (张小艳. 2015. 敦煌本《新菩萨经》、《劝善经》、《救诸众生苦难经》残卷缀合研究. 复旦学报(社会科学版), 57(6): 12-20) [DOI: 10.3969/j.issn.0257-0289.2015.06.002]
- Zhang Y Q and Luo M J. 2016. Key factors of patching up fragmentary Dunhuang Buddhist scriptures. *Journal of Zhejiang University (Humanities and Social Sciences)*, 46(3): 5-20 (张涌泉, 罗慕君. 2016. 敦煌佛经残卷缀合释例. 浙江大学学报(人文社会科学版), 46(3): 5-20) [DOI: 10.3785/j.issn.1008-942X.CN33-6000/C.2016.01.252]
- Zhang Y Q, Luo M J and Zhu R X. 2021. Exploring the Dunhuang Sutra cave riddle. *Social Sciences in China*, (3): 180-203, 208 (张涌

- 泉, 罗慕君, 朱若溪. 2021. 敦煌藏经洞之谜发覆. 中国社会科学, (3): 180-203, 208]
- Zhang Y Q and Zhu R X. 2015. A study on the patching-up of the fragments of Golden Light Sutra in the Russian collections of Dunhuang manuscripts. Fudan Journal (Social Sciences), 57(6): 1-11, 20 (张涌泉, 朱若溪. 2015. 俄藏《金光明经》敦煌残卷缀合研究. 复旦学报(社会科学版), 57(6): 1-11, 20) [DOI: 10.3969/j.issn.0257-0289.2015.06.001]
- Zhao F Q, Zhou M Q and Geng G H. 2017. Fracture surface matching method of rigid blocks. Journal of Image and Graphics, 22(1): 86-95 (赵夫群, 周明全, 耿国华. 2017. 刚体碎块的断裂面匹配. 中国图象图形学报, 22(1): 86-95) [DOI: 10.11834/jig.20170110]
- Zhong J Q and Wang R S. 2001. Image registration using edge-based methods. Computer Engineering and Science, 23(6): 25-29 (钟家强, 王润生. 2001. 基于边缘的图象配准改进算法. 计算机工程与科学, 23(6): 25-29) [DOI: 10.3969/j.issn.1007-130X.2001.06.008]
- Zhou F and Huang X M. 2007. Algorithm of 2D contours matching based on angle-sequence. Science Technology and Engineering, 7(15): 3757-3760 (周丰, 黄晓鸣. 2007. 基于角序列的2维碎片轮廓匹配算法. 科学技术与工程, 7(15): 3757-3760) [DOI: 10.3969/j.issn.1671-1815.2007.15.027]
- Zhou P, Tan Y and Xu S S. 2002. A new method of image registration based on corner detection. Journal of University of Science and Technology of China, 32(4): 455-461 (周鹏, 谭勇, 徐守时. 2002. 基于角点检测图像配准的一种新算法. 中国科学技术大学学报, 32(4): 455-461) [DOI: 10.3969/j.issn.0253-2778.2002.04.013]
- Zhu Y J, Zhou L S and Liu Y. 2007. Matching algorithm of two-dimensional irregular fragments. Computer Engineering, 33(24): 7-9 (朱延娟, 周来水, 刘毅. 2007. 2维非规则碎片匹配的算法. 计算机工程, 33(24): 7-9) [DOI: 10.3969/j.issn.1000-3428.2007.24.003]

作者简介

郑玉彤, 女, 副教授, 主要研究方向为数字图像处理、模式识别和计算机视觉。E-mail: 9900572@muc.edu.cn

翁彧, 通信作者, 男, 教授, 主要研究方向为人工智能、大数据和机器学习。E-mail: wengyu@muc.edu.cn

李雪龙, 男, 硕士研究生, 主要研究方向为数字图像处理。E-mail: 21302029@muc.edu.cn

殷梓轩, 男, 硕士研究生, 主要研究方向为数字图像处理。E-mail: 626857685@qq.com

高歌, 女, 博士研究生, 主要研究方向为计算语言学与数字人文。E-mail: 21400169@muc.edu.cn