



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
04
VOL.25

ISSN1006-8961
CN11-3758/TB



海岛岸线自动分割 P768



第25卷第4期 (总第288期)

2020年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@radi.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

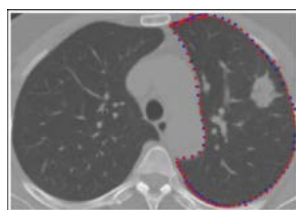
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



自适应在线判别外观学习的
分层关联多目标跟踪
(第0708页)



低秩和主动形状模型的肺区
分割(第0759页)



结合曲率滤波的HTM算法去
除遥感影像云雾(第0791页)

综述

深度学习目标检测方法综述

赵永强, 饶元, 董世鹏, 张君毅 0629

图像处理和编码

结合矩阵分解与差分隐私的人脸图像发布

张啸剑, 付聪聪, 孟小峰 0655

图像分析和识别

多尺度深度特征融合的变化检测

樊玮, 周末, 黄睿 0669

面向类内差距表情的深度学习识别

陈亮, 吴攀, 刘韵婷 0679

带后缀印刷体维吾尔文映射关系检索

伊克萨尼·普尔凯提, 阿布都萨拉木·达吾提, 艾斯卡尔·艾木都拉 0688

结合混沌鸟群算法的阴极铜板表面缺陷检测

王卓, 张长胜, 李伟, 钱俊兵, 唐都作, 蔡兵, 常以涛 0697

自适应在线判别外观学习的分层关联多目标跟踪

方岚, 于凤芹 0708

图像理解和计算机视觉

多方向Laplacian能量和与tetrolet变换的图像融合

沈瑜, 陈小鹏, 杨倩 0721

双重对偶生成对抗网络的跨年龄素描—照片转换

吴柳玮, 孙锐, 阚俊松, 高隽 0732

医学图像处理

残差密集相对平均CGAN的脑部图像配准

王丽芳, 张程程, 秦品乐, 蔺素珍, 高媛, 窦杰亮 0745

低秩和主动形状模型的肺区分割

孙申申, 田丹, 吴微, 康雁, 赵宏 0759

遥感图像处理

改进Deeplab网络的遥感影像海岸线分割

王振华, 钟元芾, 何婉雯, 曲念毅, 徐利智, 张文苹, 刘智翔 0768

卷积神经网络提取风云影像土壤湿度

韩颖娟, 周顺武, 王文文, 张学艺, 张承明, 栗佩琪 0779

结合曲率滤波的HTM算法去除遥感影像云雾

杨珂珂, 贾渊, 沈川 0791

ChinaMM 2019会议专栏

利用稳健非负矩阵分解实现无监督高光谱解混

宋晓瑞, 吴玲达, 孟祥利 0801

关键点深度特征驱动人脸表情识别

王善敏, 帅惠, 刘青山 0813

融合区域和边缘特征的水平集水下图像分割

孙杨, 陈哲, 王慧斌, 张振, 沈洁 0824

深度时空能量特征表示下的人体行为识别

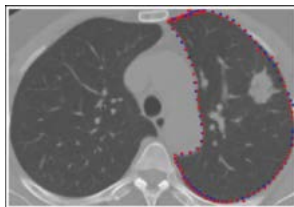
巢新, 侯振杰, 李兴, 梁久祯, 宦娟, 刘浩昱 0836

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association(P0708)



Lung segmentation by active shape model approach based on low rank theory(P0759)



Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering(P0791)

Review

Survey on deep learning object detection

Zhao Yongqiang, Rao Yuan, Dong Shipeng, Zhang Junyi 0629

Image Processing and Coding

Private facial image publication through matrix decomposition

Zhang Xiaojian, Fu Congcong, Meng Xiaofeng 0655

Image Analysis and Recognition

Multiscale deep features fusion for change detection

Fan Wei, Zhou Mo, Huang Rui 0669

Depth learning recognition method for intra-class gap expression

Chen Liang, Wu Pan, Liu Yunting 0679

Mapping relationship retrieval of Uyghur-printed suffix word

Eksan Firkat, Abdusalam Dawut, Askar Hamdulla 0688

Cathodic copper plate surface defect detection based on bird swarm algorithm with chaotic theory

Wang Zhuo, Zhang Changsheng, Li Wei, Qian Junbing, Tang Duzuo, Cai Bing, Chang Yitao 0697

Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association

Fang Lan, Yu Fengqin 0708

Image Understanding and Computer Vision

Image fusion of multidirectional sum modified Laplacian and tetrolet transform

Shen Yu, Chen Xiaopeng, Yang Qian 0721

Double dual generative adversarial networks for cross-age sketch-to-photo translation

Wu Liuwei, Sun Rui, Kan Junsong, Gao Jun 0732

Medical Image Processing

Image registration method with residual dense relativistic average CGAN

Wang Lifang, Zhang Chengcheng, Qin Pinle, Lin Suzhen, Gao Yuan, Dou Jieliang 0745

Lung segmentation by active shape model approach based on low rank theory

Sun Shenshen, Tian Dan, Wu Wei, Kang Yan, Zhao Hong 0759

Remote Sensing Image Processing

Island shoreline segmentation in remote sensing image based on improved Deeplab network

Wang Zhenhua, Zhong Yuanfu, He Wanwen, Qu Nianyi, Xu Lizhi, Zhang Wenping, Liu Zhixiang 0768

Extracting soil moisture from Fengyun satellite images using a convolutional neural network

Han Yingjuan, Zhou Shunwu, Wang Wenwen, Zhang Xueyi, Zhang Chengming, Li Peiqi 0779

Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering

Yang Keke, Jia Yuan, Shen Chuan 0791

Column of ChinaMM 2019

Unsupervised hyperspectral unmixing based on robust non-negative matrix factorization

Song Xiaorui, Wu Lingda, Meng Xiangli 0801

Facial expression recognition based on deep facial landmark features

Wang Shanmin, Shuai Hui, Liu Qingshan 0813

Level set method combining region and edge features for segmenting underwater images

Sun Yang, Chen Zhe, Wang Huibin, Zhang Zhen, Shen Jie 0824

Action recognition under depth spatial-temporal energy feature representation

Chao Xin, Hou Zhenjie, Li Xing, Liang Jiuzhen, Huan Juan, Liu Haoyu 0836

中图法分类号: TP391.1 文献标识码: A 文章编号: 1006-8961(2020)04-0688-09

论文引用格式: Eksan F, Abdusalam D and Askar H. 2020. Mapping relationship retrieval of Uyghur-printed suffix word. Journal of Image and Graphics, 25(04):0688-0696(伊克萨尼·普尔凯提, 阿布都萨拉木·达吾提, 艾斯卡尔·艾木都拉. 2020. 带后缀印刷体维吾尔文映射关系检索. 中国图象图形学报, 25(04):0688-0696) [DOI:10.11834/jig.190244]

带后缀印刷体维吾尔文映射关系检索

伊克萨尼·普尔凯提¹, 阿布都萨拉木·达吾提¹, 艾斯卡尔·艾木都拉²

1. 新疆大学软件学院, 乌鲁木齐 830046; 2. 新疆大学信息科学与工程学院, 乌鲁木齐 830046

摘要: **目的** 维吾尔文属于黏着性语言,其组成方式是在词干上添加词缀来实现不同的语义,在添加词缀的过程中词干的尾部会发生一定的形态变化,而且词干添加词缀的时候也可能发生弱化、脱落、增音等音变现象导致进一步的形态变化,所以利用目前的图像文字检索(word spotting)技术只能检索到某一具体的维吾尔文词汇,却不能以某一词干为检索词,检索出其对应的带后缀的词语。为此,提出了基于映射关系的带后缀印刷体维吾尔文词语检索技术。**方法** 首先利用局部特征对维吾尔文词图像进行特征提取,其次将获得的特征用快速最近邻搜索(fast library for approximate nearest neighbors, FLANN)双向匹配来获得特征匹配集,最后将特征匹配集进行单应性变换和透视变换到待检索维吾尔文词图像上,把特征匹配集转化为空间关系,经过映射匹配对特征匹配集的空间关系进行后缀词检索,从而实现印刷体维吾尔文图像带后缀词检索的需求。**结果** 实验数据选取190幅维吾尔文印刷体文本图像中的17 648幅切割词图像,并对其中30幅词图像的167幅后缀词图像进行后缀检索,采用不同的局部特征算法进行后缀检索对比,结果表明,尺度不变特征变换(scale-invariant feature transform, SIFT)算法的后缀检索效果优于SURF(speeded up robust features)算法,精确率和召回率分别达到了94.23%和88.02%,在印刷体文档图像中,可以高效地检索到词干组成的后缀词,能够满足用户的不同检索需求,具有普适性。在弱化、脱落、增音和多种音变同时出现以及词干尾部发生变化的不同情况下进行后缀检索对比实验,实验结果表明在弱化和词干尾部变化而导致的形态变化中,检索效率最佳。**结论** 本文提出的基于映射关系进行后缀词图像检索的方法,是第一次对维吾尔文带后缀词检索方式的一种实现,利用匹配集之间的空间关系,对维吾尔文带后缀词图像实现了高效检索的目的。**关键词:** 维吾尔文;带后缀词检索;局部特征;单应矩阵;透视变换

Mapping relationship retrieval of Uyghur-printed suffix word

Eksan Firkat¹, Abdusalam Dawut¹, Askar Hamdulla²

1. School of Software, Xinjiang University, Urumqi 830046, China;

2. School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

Abstract: Objective Uyghur belongs to adhesive language, and the formation and meaning of words in Uyghur language depend on affix connection, which add affixes to the stems to achieve different semantic meaning. For example, in Uyghur “مەكتەپ” refers to school, and it could be added by a first person singular in Uyghur “مەن” as a suffix to form a new word “مەكتەپم”, which means my school. During adding suffixes, certain morphological changes will occur at the tail of the stems. In addition, phonetic change also happens at that process, such as weakling (some morphological changes occur at the tail of a stem), epenthesis (a few characters were added at the tail of the stem), and deletion (a few characters were

收稿日期:2019-06-02;修回日期:2019-09-14;预印本日期:2019-09-21

基金项目:国家自然科学基金项目(61662076)

Supported by: National Natural Science Foundation of China(61662076)

deleted at the tail of the stem). Multiple phonetic changes also appear simultaneously, causing further morphological change. All the semorphological changes form the word different from the stem. Therefore, using the current word spotting technique, only a specific Uyghur vocabulary can be retrieved, and a certain stem cannot search its corresponding suffixed words. In addition, traditional word spotting approaches only aim at the number of the matching sets and noton the spatial relationship of the matching sets. Therefore, some drawbacks for word spotting technique occur. This study proposes Uyghur-printed suffix word retrieval based on mapping relationship that take advantage of the spatial relationship of the matching sets to retrieve the corresponding suffix words of the stems. **Method** The process of the proposed approach is described as follows. First, the segmentation algorithm segments printed the Uyghur document images to word image corps. Then, the local features of the Uyghur word image are extracted. To compare the efficiency of the different local features, scale-invariant feature transform (SIFT) and speeded up robust features (SURF) have been adopted. The experiment result shows that SIFT feature has better performance than SURF because SIFT can obtain more feature points than SURF, and the distribution of SIFT feature points is more diverse than SURF, which is very helpful for further retrieval steps. However, SURF is more efficient than SIFT considering the time efficiency. Then, Brute-Force matching and fast library for approximate nearest neighbor (FLANN) bilateral feature matching have been adopted as matching algorithm. The experiment result shows that FLANN bilateral feature matching has better performance than Brute-Force matching because FLANN bilateral feature matching can filter more mismatch pairs than Brute-Force matching. In addition, the correctness of the feature matching set is very important to the following suffix word retrieval, and the accuracy of FLANN bilateral feature matching is very outstanding. Finally, the feature matching sets are subjected to homography transformation and perspective transformation to the Uyghur word image for the final retrieval steps. After homography transformation and perspective transformation, a quadrilateral is built. If this quadrilateral belongs to rectangle and the right part of the acquired rectangle simply match with the outline of the query word image, then the retrieved word belongs to corresponding suffixed words of the query stem. Meanwhile, if this quadrilateral does not belong to rectangle and does not match with the outline of the query word image, then the retrieved word does not belong to the corresponding suffixed words of the query stem but to the mismatch. This result indicates that the proposed method not only can retrieve corresponding suffixed words of the query word but also can filter the mismatch word. In other words, the feature matching sets are transformed into a spatial relationship and are further determined whether the retrieved word belongs to suffix word or mismatched word according to the spatial relationship. The spatial relationship of the feature matching set is searched for the suffix word, thereby implementing the printed Uyghur Text image suffix word retrieval. **Result** The experimental data selected 17 648 segmented word images in 190 Uyghur print text images and 30 word images, which have 167 corresponding suffix words considered as the search terms. In the experiment, we used different local feature algorithms to suffix retrieval. The comparison results show that the SIFT algorithm's suffix retrieval effect is better than SURF algorithm, and its accuracy and recall rate reach 94.23% and 88.02%, respectively. In addition, we carried out comparative experiments in different situations, such as weakling, epenthesis, deletion, and replacement. Moreover, multiple phonetic changes appear simultaneously and change in the tail of the word stem. In those five different situations, the retrieval result of changes in the tail of the word stem was the best, and its accuracy and recall rate reach 98.9% and 96.07%, respectively. The main reason for this result is that the changes in the tail of the word stem do not make very obvious formation changes of the stem word, which can be very helpful for the suffix word retrieval. However, in different cases, multiple phonetic changes appear simultaneously that the accuracy and recall rate of the retrieval reach 66.6% and 22.2%. The reason for such low performance is that multiple phonetic changes simultaneously change the formation of the stem part in the suffix word. In addition, several characters in the stem part have been replaced by other characters, which are very difficult to retrieve its corresponding suffix word by the original stem word. However, this kind of situation only takes few percentages of the whole morphological changes. **Conclusion** Therefore, the proposed algorithm meets the different retrieval needs of users. The method of suffix word retrieval based on mapping relationship is also the first implementation of Uyghur suffix word retrieval method. The Uyghur suffix image is efficiently retrieved by the spatial relationship between the matching sets.

Key words: Uyghur language; suffix word retrieval; local feature; homograph; perspective transform

0 引言

随着信息化时代交互信息的频繁传递,搜索引擎应运而生。印刷体文字图像搜索是搜索引擎技术的一个分支(Jameson, 2004)。一般来说,印刷体文字图像检索技术分为两种:1)传统的光学字符识别(optical character recognition, OCR)技术,将印刷体文字图像中的信息转化为计算机可读内容后再从中进行文字图像检索。2)关键词(word spotting)识别技术,主要是通过模式匹配的方式直接从图像文档中对关键词进行检索(Wei 等, 2015),不需要将印刷体文字图像转化为可读格式。

传统的OCR技术因易受文档图像质量及字体等因素影响,检索成功率不是很高,越来越多的学者开始青睐于word spotting技术(温子潇等, 2018)。作为OCR的替代方法,word spotting技术已广泛应用于文档图像检索领域(Singhai 和 Shandilya, 2010)。如Lee等人(2012)利用尺度不变特征变换(scale-invariant feature transform, SIFT)方法对无分段单词进行定位,Almazán等人(2013)描述了一种将单词图像和文本字符串都嵌入到公共矢量中的子空间的方法,Ghosh 和 Valveny(2015)提出了一种基于字符串词定位的无分段查询方法,Hussain等人(2016)提出了一种基于关键字的乌尔都语文档图

像信息检索系统,胡海青等人(2013)基于改进SIFT方法在文字图像匹配中进行了应用,杨娜娜等人(2014)利用SIFT特征对维吾尔文字符进行检索等,这些方法在不同程度上促进了图像检索的应用(Sfikas 等, 2015)。

维吾尔文属于黏着性语言,其组成方式是在词干上添加词缀来实现不同的语义,在添加词缀的过程中词干的尾部会发生一定的形态变化,而且词干添加词缀的时候也可能会发生弱化、脱落、脱音等音变现象,其具体音变现象如表1所示。

从表1可以看出,词干和词缀的连接有非常丰富的形态变化,虽然这些形态变化可以表达不同的语义,但对由某一词干构成的词进行检索时却造成了非常大的困难。在英文中虽然也存在检索基于某个词干构成的词的情况,但英文的拉丁文书写方式,非常巧妙地解决了基于某个词干检索时对其构成的词的检索情况,如change(变化)→changeable(可变化的),service(服务)→serviceable(有用的),只需要考虑对词干进行匹配,就可以进行检索。但维吾尔文却因为上述的音变现象以及自身的词干尾部的形态变化,导致词干本身发生一定的形态变化时,直接对词干进行匹配,这种解决方案是不可行的。比如以“مەكتەپ”(学校)为词干,检索由词缀构成的词“مەكتەپىم”(我的学校)时,发生了弱化这种音变现象,导致词干发生了明显的形态变化。

表1 维吾尔语的音变现象

Table 1 Phenomenon of morphological sandhi in Uyghur

音变现象	实例	说明
弱化	مەكتەپ (学校) + مە (第一人称单数) = مەكتەپىم (我的学校)	其中“ە”弱化成“ى”
增音	سېگىل (抓饭,词干) + ۈم (第一人称单数,词缀) = يولۇپۇم (我的抓饭)	词干后增加了一个字母“ي”
脱落	كىل (妹妹,词干) + مە (第三人称单数,词缀) = كىلمە (我的妹妹)	有时出现多个字母同时脱落的情况
多种现象同时出现	كىل (来) + تۇ + پەن (系助动词) = پەن كىل (听说他来了)	词干中的“ە”弱化成“ى”,而“ى”“ۇ”“ۇ”脱落

表2是以“مەكتەپ”(学校)为词干的后缀词,可以明显看出直接用词干进行完全匹配是不可行的,所以当对基于某个词干和其词缀构成的词(下文称“带后缀词”)进行检索时,会对检索造成一定的困难。

为解决以上难点,本文提出了基于映射关系的带后缀印刷体维吾尔文词语检索技术,利用图像文字

检索(word spotting)将获得的匹配特征集转化为空间映射关系,根据空间映射关系对带后缀词进行检索。

1 系统框架

本文通过对印刷体文档图像进行切割,获得词

图像数据集,提取查询词图像和待检索词图像数据集的局部特征,通过特征匹配对待检索词进行匹配,获得匹配特征集,最后通过映射匹配查询由检索词构成的词图像,具体系统框架如图 1 所示。

表 2 带后缀词举例

Table 2 Illustration of stem word with different suffix

汉语	维吾尔语
学校	مەكتەپ
我的学校	مەكتەپم
在学校	مەكتەپتە
学校的	مەكتەپنىڭ
往学校	مەكتەپكە
直到学校	مەكتەپكىچە
从学校	مەكتەپتىن

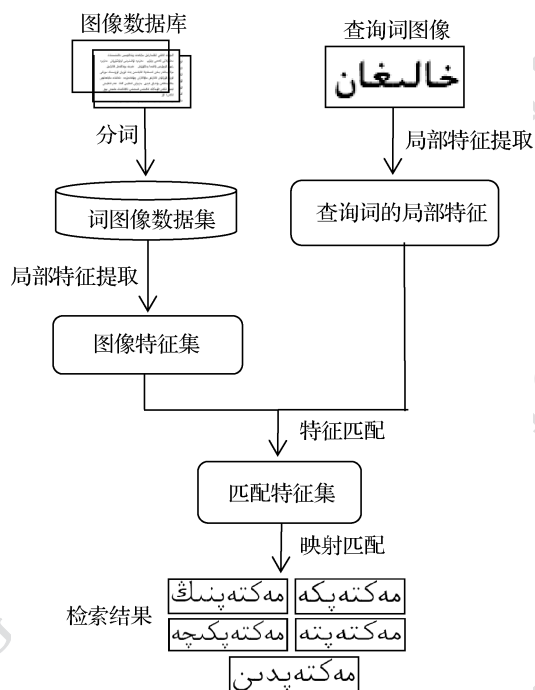


图 1 检索系统框架

Fig. 1 Framework of retrieval system

2 特征提取与特征匹配

由于局部特征存在本身的优势,已经成功应用于涉及图像处理的不同场景,与其他特征相比,局部特征包含更多的细节信息(吴永国等,2012),本文利用两种具有代表性的特征 SIFT 和 SURF(speeded

up robust features) 算法对图像进行特征提取,进行特征匹配,得到匹配特征集用于后续检索。

2.1 SIFT 特征算法

SIFT 算法由 Lowe(1999) 提出并完善的(Lowe, 2004)的一种检测和描述图像局部特征的算法,具有非常好的稳定性。

SIFT 算法的稳定性是因为其特征点通过不同尺度特征提取,在提取过程中考虑了图像的旋转不变形,过滤了由于干扰因素而产生变化的特征点,因此,特征点不受其他干扰因素影响(Aldana-Murillo 等,2015),算法的具体步骤如下:

1) 尺度空间极值检测。通过高斯尺度核对原图像进行卷积运算获得不同尺度情况下的尺度空间,用于进行特征点计算,具体为

$$L(x, y, \sigma) = G(x, y, k\sigma) * I(x, y) \quad (1)$$

式中, $G(x, y, k\sigma)$ 为尺度可变高斯函数, $I(x, y)$ 为原始图像, k 为尺度变换量。获得高斯尺度空间 $L(x, y, \sigma)$ 后,对其进行差分操作以获取尺度不变的高斯差分尺度空间,即

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) *$$

$$I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2)$$

式中, $D(x, y, \sigma)$ 为获得的高斯差分尺度空间,最后经过采用相应的降采样方法与尺度空间相结合构建图像的高斯差分金字塔结构。

2) 关键点定位。获得高斯差分尺度空间后,对每个监测点的邻点像素以及上下两层的像素点进行比较,从而确保检测点的准确性。

3) 方向确定。对检测到的特征点,进行邻域梯度方向的直方图统计,确定其主方向。需要注意的是,有些特征点除主方向外,还存在辅方向,它对于后续匹配的稳定性至关重要。

4) 关键点描述。特征点的描述是在特征点周围取 16×16 像素的区域,然后将该区域划分为多个 4×4 像素的区域,在区域内计算其方向梯度方向,得到 128 维特征描述子。

2.2 SURF 特征算法

SURF 算法是一种稳健的局部特征点检测和描述算法,是对 SIFT 算法的改进,主要改进为以特征向量的描述方式提取兴趣点,有助于提高整个算法的速度。首先构建 Hessian 矩阵,生成所有的兴趣点,用于特征的提取。构建 Hessian 矩阵的目的是为了生成图像稳定的边缘点,加快检测速度。Hes-

sian 矩阵是描述函数的局部曲率的一个由多元函数的 2 阶偏导数构成的方阵,数学表达式为

$$\mathbf{H}(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{yx}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (3)$$

式中, $\mathbf{H}(x, \sigma)$ 为 Hessian 矩阵, $L_{xx}(x, \sigma)$ 和 $L_{yy}(x, \sigma)$ 分别为 X 和 Y 方向的 2 阶偏导数卷积, $L_{xy}(x, \sigma)$ 和 $L_{yx}(x, \sigma)$ 为混合偏导数卷积。

SURF 算法通过改变盒子滤波器的模板大小建立尺度图像金字塔。通过 Hessian 矩阵获得兴趣点后,比较与领域范围内的其他点的亮度或暗度,从而获得关键点的位置。将关键点与周围的点以及尺度空间的点进行比较,获得一个初步的关键点,经过过滤后获得稳定的特征点。对得到的特征点经过 Haar 小波响应确定其主方向,得到一个 SURF 特征向量 (Aliya 等, 2019)。

2.3 暴力匹配

暴力匹配是一种通用的模式匹配算法,在进行图像匹配时,通过找出两幅图像中相似度最高的两个特征向量为匹配对象,匹配过程为检索图像中的特征点与待检索图像中所有特征点依次进行匹配,在匹配过程中得到最佳匹配结果。

2.4 FLANN 双向匹配

快速最近邻搜索 (fast library for approximate nearest neighbors, FLANN) 匹配算法由 Muja 和 Lower (2009) 提出,用于在高维空间内通过欧氏距离来寻找与查询点最邻近的点。本文通过 FLANN 匹配算法,将查询词图像的特征点集 $\mathbf{A} = \{x_1, x_2, \dots, x_N\}$ 和待检索词图像的特征点集 $\mathbf{B} = \{y_1, y_2, \dots, y_N\}$ 进行匹配,找到特征点 x_i 的最近邻 y_{j1} 和 y_{j2} , 得到 x_i 到 y_{j1} 和 y_{j2} 的欧氏距离 D_1 和 D_2 。并设定可调阈值 ε , 将 D_1/D_2 与阈值进行比较,判断是否为匹配点,如果匹配,将匹配点放入到匹配集 $\langle x_i, y_j \rangle$, 匹配规则为

$$D(x, y) = \|\mathbf{A} - \mathbf{B}\| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2} \quad (4)$$

匹配结果为

$$\begin{cases} \text{若 } D_1/D_2 < \varepsilon & y_{j1} \text{ 为匹配点} \\ \text{其他} & y_{j1} \text{ 为非匹配点} \end{cases} \quad (5)$$

得到匹配集 $\{x_i^A, y_j^B\}$ 后,对待检索词特征集进行同样的匹配,得到特征集 $\{y_j^B, x_i^{A'}\}$ 。比较 x_i^A 和 $x_i^{A'}$ 的大小,相等时则两点完全匹配,得到最终特征

点匹配集。这种匹配方式可以进一步有效去除误匹配对象。FLANN 双向匹配在词干和带后缀词的双向匹配效果如图 2 所示。



图 2 FLANN 双向匹配效果图

Fig. 2 Result of FLANN bidirectional matching

3 后缀检索算法

获得匹配特征集后,进行单因性变换和透视变换,来将查询词匹配特征集的空间关系映射到待检索词图像中,根据两者的空间关系实现对带后缀词的检索,这种方法不仅可以解决词干添加词缀时词干尾部发生形态变化和词缀添加时发生的音位变化问题,同时可以自动将误匹配的待检索词自动过滤。后缀检索的流程如图 3 所示。

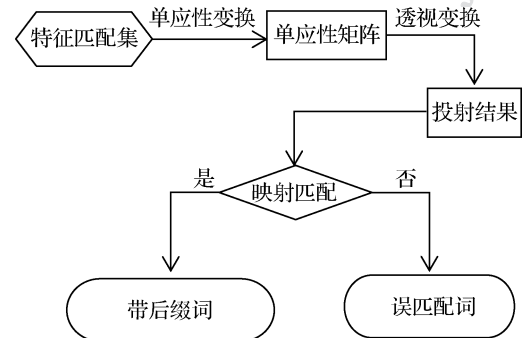


图 3 后缀检索流程图

Fig. 3 Flow chart of suffix retrieval

3.1 单应性变换

单应性变换简单来说就是两个平面之间的一种映射关系的线性变换。在特征点配对中,就是从一平面上的特征点到另一平面上特征点的线性变换,其数学原理为

$$\begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & h_8 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (6)$$

式中, $h_0, h_1, \dots, h_7$ 为变换参数, $\mathbf{H}$ 为变换矩阵,是 8 个自由度矩阵, $\mathbf{H}$ 的计算需要通过两幅图像的对特征点的坐标进行求解。匹配到的特征点越多,

得到的 H 越准确。在特征匹配过程中所匹配的特征集由于是局部特征的关系, 会导致所匹配的特征集实际上并不含有相同词干, 所以不同的特征算法以及特征点匹配算法会对后期的后缀检索有一定影响。通过求解图像之间的单应性矩阵 H 后, 需要进行一步操作去进行后缀检索。

3.2 透视变换

得到映射矩阵 H 后, 将查询词的特征集通过透视变换投射到待检索词图像中, 透视变换是将图像投影到一个新的平面里, 称为投影映射, 其相对于仿射变换来说, 具有更大的灵活性, 是将一个四边形区域映射到另一个四边形区域。通用的变换公式为

$$\begin{bmatrix} x' & y' & w' \end{bmatrix} = \begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (7)$$

式中, u, v 为原始图像里的坐标, 对应得到变换后的坐标为 x, y , 其中, $x = x'/w', y = y'/w'$ 。而变换

矩阵 $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ 可以拆为 4 部分, 其中,

$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ 为线性变换, $\begin{bmatrix} a_{31} & a_{32} \end{bmatrix}$ 用于平移,

$\begin{bmatrix} a_{13} & a_{23} \end{bmatrix}^T$ 表示透视变换。

3.3 映射匹配

经过单应性变换和透视变换后, 词干的局部特征匹配集会以四边形 Ra 的方式映射到待查询词的图像中。经实验发现, 如果待检索词是带后缀词, 那么 Ra 的右半部分将会与待查询词图像的边界重合, 即 Ra 的右半部分会是一个标准的矩形。本文通过判断 Ra 与待检索词的边界关系来判断是否为后缀匹配。当 Ra 的右上角和右下角与待检索词图像的右上角和右下角重合, 且 Ra 的长度小于待检索词图像的长度时, 则判断待检索词是带后缀词。由此也解决了添加词缀后, 因词干尾部发生形态变化和音变现象而难以进行检索的问题。当 Ra 为一个不规则四边形, 或者投射到待检索词边界的 Ra 超过其边界时, 则可以判定为误匹配词。这样就把传统的匹配问题利用空间关系来进一步进行检索, 其中映射匹配流程如图 4 所示, 其中红色矩形即 Ra 。

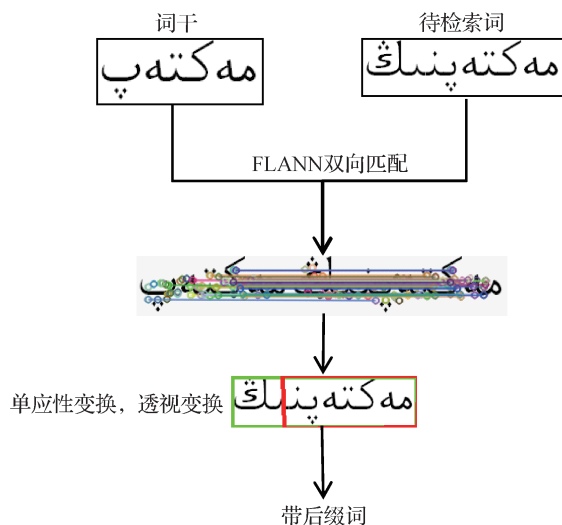


图 4 映射匹配流程图

Fig. 4 Flow diagram of homograph matching

4 实验结果分析

4.1 实验数据

实验用的图像数据集是通过网络爬虫方式从中央人民广播电台中国维吾尔文广播网 (<http://www.uycnr.com/>) 获得的维吾尔文印刷体文本图像, 共 190 幅, 由 17 648 个词构成。

实验在 Windows10 操作系统用 Visual Studio 2013 + opencv 编程开发。

4.2 实验结果与分析

在 SIFT 和 SURF 两种特征算法上进行后缀检索对比实验。由于各个特征算法的特征点提取方式不同, 所以在特征提取过程中得到的特征集的数量和分布方式也不同。对一幅文字图像进行特征提取实验, 效果如图 5 所示。

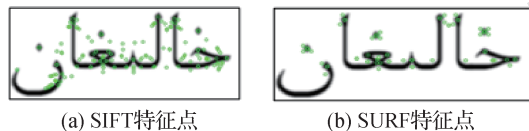


图 5 局部特征提取图

Fig. 5 Feature extraction of different local feature

((a) SIFT; (b) SURF)

从图 5 可以看出, 不同的特征算法的特征点数量和特征分布方式都有明显差异, 在实验中使用对数据集中心切割好的图像进行特征提取, 计算平均特征数目和平均特征提取时间, 统计结果如表 3 所示。

表3 不同局部特征算法的特征点个数和时间对比
Table 3 Comparison of the number and time of feature points of different local feature algorithms

特征算法	平均特征数目	平均特征提取时间/s
SIFT	76	0.07
SURF	63	0.017

经实验发现, SIFT 特征算法得到的特征点最多, 检测时间也最多。同时发现 SIFT 和其他特征算法相比特征分布点比较广泛。SURF 特征算法的特征点检测时间最短, 但在特征提取方面, 特征点数量与 SIFT 相比, 效果不是很好, SIFT 特征算法相对而言更加精确。在特征匹配阶段, 从数据集中选取 20 幅词图像及其对应的后缀图像进行预匹配, 在不同特征条件下进行 FLANN 双向匹配和暴力匹配, 发现 FLANN 双向匹配的阈值设置为 0.8 时最佳, 在此情况下对结果进行统计, 如表 4 所示。

表4 不同匹配算法的性能对比
Table 4 Comparison of the different matching algorithms

性能	FLANN 双向匹配		暴力匹配	
	SIFT	SURF	SIFT	SURF
平均关键点数目	68	58	68	58
平均匹配点数目	59	52	65	56
正确匹配点数目	56	47	49	43
正确匹配率/%	94.90	90.40	75.40	76.70
错误匹配点数目	3	5	16	13
错误匹配率/%	5.10	9.60	24.60	23.30

从表 4 可以看出, 暴力匹配到的点数多于 FLANN 双向匹配, 但正确匹配率低于 FLANN 双向匹配。因为暴力匹配时, 虽然检测图像的每个特征点匹配到待检测图像的特征点, 但因为匹配的特征点为局部特征点, 再加上维吾尔文书写方式原因, 导致暴力匹配在匹配过程中会出现误匹配对象。这些误匹配对象对后期的后缀检索的正确性会产生较大影响, 因为后缀检索对特征匹配集的正确率要求非常高, 如果出现过多的误匹配点, 会导致后缀检索失效。而 FLANN 双向匹配可以进一步有效去除这些误匹配对象, 所以暴力匹配的正确匹配率要小于 FLANN 双向匹配。因此本文以 FLANN 双向匹配为特征匹配算法。

本文扩大数据集, 在 190 幅印刷体维吾尔文本图像数据集所含的 17 648 幅切割词图像中, 对 30 幅词图像中的 167 个词干和词缀构成的词进行后缀检索, 以“مەكتەپ” (学校) 为查询词为例, 检索到的带后缀词的结果如表 5 所示, 在词干检索结果表的映射匹配中, 红色四边形为查询词映射结果, 绿色矩形为带后缀词本身边界, 从两个四边形的空间关系可以判断后缀检索的有效性, 即当检索词构成的红色映射四边形和带后缀词的绿色边界矩形满足映射匹配条件时, 可以判定待检索词是带后缀词。

表5 检索结果表
Table 5 Results of retrieval word

后缀词	特征匹配结果	映射匹配	匹配结果
مەكتەپلەر			带后缀词
مەكتەپتە			带后缀词
مەكتەپنىڭ			带后缀词
مەكتەپكە			带后缀词
مەكتەپدەن			带后缀词

对不同特征算法进行的后缀检索实验, 效果也不相同。将 SIFT 和 SURF 两种特征算法进行的后缀检索实验的召回率和精确率进行对比, 结果图 6 所示。

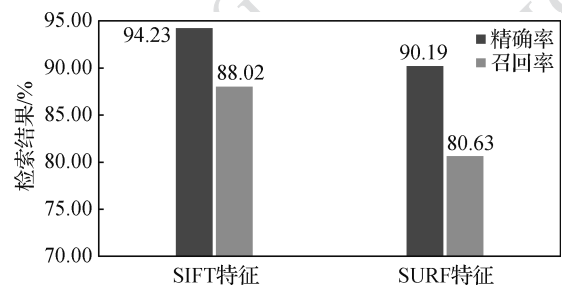


图6 不同特征算法在后缀检索中的结果
Fig. 6 Results of different feature extractor under suffix retrieval

从图 6 可以看出, 基于映射关系的词干检索技术具有非常好的效果, 两种特征算法进行后缀检索时, 精确率和召回率存在明显差异。SIFT 算法的后缀检索效果优于 SURF 算法, 主要是因为维吾尔文字图像的局部特征点中 SIFT 特征点的数量比 SURF 多, 同时在特征匹配过程中 SIFT 特征算法的匹配率高于 SURF, 因此选择 SIFT 特征算法为本文的局部特征检索算法。

为了验证本文方法在不同音变和词干尾部发生变化情况下的检索性能,首先对167幅带后缀词弱化、脱落、增音、多种音变同时出现、词干尾部发生变化等5种类型构成的词图进行人工标注,各类型数量如表6所示。

表6 不同构词结果
Table 6 Results of different morphological formation

构词类型	数量
词干尾部变化	102
弱化	23
脱落	17
增音	16
多种现象同时出现	9

从表6可以看出,词干尾部变化情况较多,多种现象同时出现的音变现象最少。在得到标注结果后,对不同构词情况进行后缀检索,结果如图7所示。

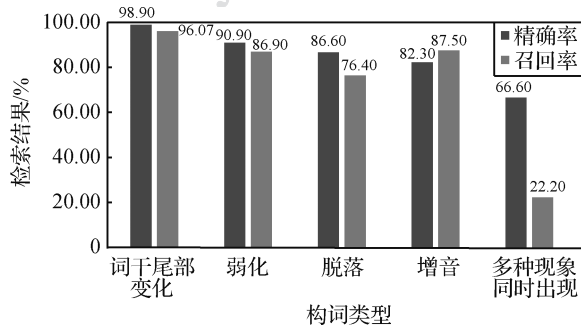


图7 不同构词类型在后缀检索中的结果

Fig. 7 Results of different word-formation methods in suffix retrieval

由图7可以看出,本文方法在词干尾部发生变化时的检索效率最佳,其原因为词干尾部变化程度对整体后缀检索的影响非常小。在弱化和增音音变现象的检索效果较好,也表明本文方法对词干发生一定程度的形态变化时的检索具有鲁棒性。但本文方法对脱落和多种音变现象同时出现的情况的效果有待提高,原因是这两种音变现象导致词干发生很大程度的形态变化,一是在脱落音变现象中多个字母同时脱落导致了词干发生非常大的形态变化,二是在多种音变现象同时出现的情况中,词干发生了非常大的形态变化导致检索失败。比如,由“كىل”

(来)构成的带后缀词“كىلىپتىمىز” (我们来了)中词干发生了非常大的形态变化,导致本文方法效果不是很好。

5 结 论

针对目前的 word spotting 技术只能检索到某一具体的维吾尔文词汇,不能以某一词干为检索词检索出其对应的带后缀的词语的问题,本文提出了基于映射关系的带后缀印刷体维吾尔文词语检索技术,即利用映射关系将匹配好的局部特征集转化为空间关系,并据其达到对带后缀词的检索目的。这是第一次在图像检索层面对带后缀词的词语实现检索,该技术将局部特征的匹配特征集从传统的基于匹配率的思维方式转化为基于匹配特征集在空间关系的思维方式,进一步优化了检索效率,对其他黏着性语言的印刷体文字图像的带后缀词检索也具有一定启发性。但是本文方法对词干添加词缀时发生脱音以及多种音变现象同时发生等导致词干发生较大形态变化的带后缀词的检索的效果不是很好,这将是下一步的研究重点。

参考文献 (References)

- Aldana-Murillo N G, Hayet J B and Becerra H M. 2015. Evaluation of local descriptors for vision-based localization of humanoid robots// Proceedings of the 7th Mexican Conference on Pattern Recognition. Mexico City: Springer: 179-189 [DOI: 10. 1007/978-3-319-19264-2_18]
- Aliya B, Nurbiya Y, Hornisa M, Alimjan A and Kurban U. 2019. Complex Uyghur document image matching and retrieval based on modified SURF feature. CAAI Transactions on Intelligent Systems, 14(2): 296-305 (阿丽亚·巴吐尔, 努尔毕亚·亚地卡尔, 吾尔尼沙·买买提, 阿力木江·艾沙, 库尔班·吾布力. 2019. 改进 SURF 特征的维吾尔文复杂文档图像匹配检索. 智能系统学报, 14(2): 296-305) [DOI: 10. 11992/tis. 201709014]
- Almazán J, Gordo A, Fornés A and Valveny E. 2013. Handwritten word spotting with corrected attributes//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney: IEEE: 1017-1024 [DOI: 10. 1109/ICCV. 2013. 130]
- Ghosh S K and Valveny E. 2015. Query by string word spotting based on character bi-gram indexing//Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE: 881-885 [DOI: 10. 1109/ICDAR. 2015. 7333888]
- Hu H Q, Tan J L, Zhu Y T, Gong G C and Liu J G. 2013. Application

- of improved SIFT algorithm in text image matching. *Computer Engineering*, 39(1): 239-243 (胡海青, 谭建龙, 朱亚涛, 龚国成, 刘金刚. 2013. 改进 SIFT 算法在文字图像匹配中的应用. *计算机工程*, 39(1): 239-243) [DOI: 10.3969/j.issn.1000-3428.2013.01.052]
- Hussain R, Khan H A, Siddiqi I, Khurshid K and Masood A. 2016. Keyword based information retrieval system for Urdu document images//*Proceedings of the 11th International Conference on Signal-Image Technology and Internet-Based Systems*. Bangkok: IEEE: 27-33 [DOI: 10.1109/SITIS.2015.16]
- Jameson M. 2004. Promises and challenges of digital libraries and document image analysis: a humanist's perspective//*Proceedings of the 1st International Workshop on Document Image Analysis for Libraries*. Palo Alto: IEEE: 54-61 [DOI: 10.1109/DIAL.2004.1263237]
- Lee D R, Hong W and Oh I S. 2012. Segmentation-free word spotting using SIFT//2012 IEEE Southwest Symposium on Image Analysis and Interpretation. Santa Fe, USA: IEEE: 65-68 [DOI: 10.1109/SSIAI.2012.6202454]
- Lowe D G. 1999. Object recognition from local scale-invariant features//*Proceedings of the 7th IEEE International Conference on Computer Vision*. Kerkyra: IEEE: 1150-1157 [DOI: 10.1109/ICCV.1999.790410]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110 [DOI: 10.1023/b:visi.0000029664.99615.94]
- Muja M and Lowe D G. 2009. Fast approximate nearest neighbors with automatic algorithm configuration//*Proceedings of 2009 International Conference on Computer Theory and Applications*. Lisboa, Portugal: [s. n.]: 331-340 [DOI: 10.5220/0001787803310340]
- Sfikas G, Giotis A P, Louloudis G and Gatos B. 2015. Using attributes for word spotting and recognition in polytonic Greek documents//*Proceedings of the 13th International Conference on Document Analysis and Recognition*. Tunis: IEEE: 686-690 [DOI: 10.1109/ICDAR.2015.7333849]
- Singhai N and Shandilya S K. 2010. A survey on: content based image retrieval systems. *International Journal of Computer Applications*, 4(2): 22-26 [DOI: 10.5120/802-1139]
- Wei H X, Gao G L and Su X D. 2015. A multiple instances approach to improving keyword spotting on historical Mongolian document images//*Proceedings of the 13th International Conference on Document Analysis and Recognition*. Tunis: IEEE: 121-125 [DOI: 10.1109/ICDAR.2015.7333738]
- Wen Z X, Bao F L, Gao G L, Wang Y H and Su X D. 2018. Design and implementation of Mongolian information retrieval system. *Journal of Chinese Information Processing*, 32(7): 44-51, 57 (温子潇, 包飞龙, 高光来, 王勇和, 苏向东. 2018. 蒙古文信息检索系统的设计与实现. *中文信息学报*, 32(7): 44-51, 57) [DOI: 10.3969/j.issn.1003-0077.2018.07.006]
- Wu Y G, Li D Q, Luo H B, Wei Y W and Xia R B. 2012. Image retrieval method combining global description and local description. *Computer Engineering and Design*, 33(2): 634-638 (吴永国, 李德强, 罗海波, 魏永旺, 夏仁波. 2012. 融合全局和局部描述的图像检索方法. *计算机工程与设计*, 33(2): 634-638) [DOI: 10.3969/j.issn.1000-7024.2012.02.043]
- Yang N N, Halidan A and Yiliyaer D. 2014. Uyghur character recognition method based on SIFT image registration. *Transducer and Microsystem Technologies*, 33(3): 40-43 (杨娜娜, 哈力旦·阿布都热依木, 伊力亚尔·达吾提. 2014. 基于 SIFT 图像配准的维吾尔语文字识别方法. *传感器与微系统*, 33(3): 40-43) [DOI: 10.3969/j.issn.1000-9787.2014.03.012]

作者简介



伊克萨尼·普尔凯提, 1996 年生, 男, 硕士研究生, 主要研究方向为场景文本检测与识别、强化学习、图像处理。

E-mail: eksan_xju@qq.com



艾斯卡尔·艾木都拉, 通信作者, 男, 教授, 主要研究方向为语音识别与合成、模式识别与图像处理、自然语言处理、信息检索与内容安全。

E-mail: askar@xju.edu.cn

阿布都萨拉木·达吾提, 男, 副教授, 主要研究方向为教育信息化、教育智能化和现代化。

E-mail: abdulalam@xju.edu.cn