

DOI: 10.7652/xjtuxb201701021

最大稳定极值区域与笔画宽度变换的自然场景文本提取方法

张国和, 黄凯, 张斌, 符欢欢, 赵季中

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 针对从背景复杂、视角多变、语言形式多样的场景图像中难以准确提取文本信息的问题,提出了一种基于最大稳定极值区域(MSER)和笔画宽度变换(SWT)场景文本提取方法。该方法结合MSER、SWT算法的优点,采用MSER算法的准确检测文字区域,建立文本候选区域,利用SWT算法计算文本候选区域笔画宽度得到候选文本区域的笔画宽度;根据笔画宽度图,利用连通域标记建立笔画宽度连通图,然后根据笔画宽度连通图,建立笔画连通图的启发性规则,删除非文本候选区域,并根据文本的几何特征分析及局部自适应窗口最大类间方差(Otsu)分割,有效提取出自然场景图像中的文本,文本提取的准确率、召回率及综合性能分别为0.74、0.64及0.68。仿真实验结果表明,在文本视角多变,字符大小、尺寸、字体各异的复杂条件下,所提方法具有较好的鲁棒性,适用于多语言和多字体混合的场景文本提取。

关键词: 自然场景图像;文本提取;最大极值稳定区域;笔画宽度变换

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 0253-987X(2017)01-0135-06

A Natural Scene Text Extraction Method Based on the Maximum Stable Extremal Region and Stroke Width Transform

ZHANG Guohe, HUANG Kai, ZHANG Bin, FU Huanhuan, ZHAO Jizhong

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To extract text information effectively from natural scene image with complex background, multi-orientation perspective and multilingual languages, a scenario text extraction method based on maximum stable extremal region (MSER) and stroke width transform (SWT) is presented. The method combines the merits of MSER and SWT algorithms. It establishes text candidate regions by utilizing MSER algorithm to detect text regions, and SWT algorithm is used to calculate the text stroke width of candidate region to get its stroke width. According to the stroke width graph, the stroke-connected graph is established by using connected component labeling. Then the heuristic rules of stroke-connected graph are established to remove non-text candidate regions according to the stroke-connected graph. By using the geometrical feature analysis and the local adaptive window Otsu segmentation, the text in natural scene images can be extracted effectively. The text extraction accuracy, recall rate and comprehensive performance of this method are 0.74, 0.64 and 0.68, respectively. Simulation experiment shows that the method can achieve good robustness for complex background with multi-orientation perspective, various

收稿日期: 2016-03-28。 作者简介: 张国和(1981—),男,副教授,博士生导师;张斌(通信作者),男,讲师。 基金项目: 国家自然科学基金资助项目(61306111)。

网络出版时间: 2016-10-20

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20161020.1044.006.html>

characters and font sizes, and it is suitable for variety of languages and fonts.

Keywords: natural scene image; text extraction; maximum stable extremal region; stroke width transform

近几年,随着各种智能设备和便携式手持设备的广泛应用,人们可以随时随地使用手机或者数码相机拍摄自己感兴趣的场景图像。提取的场景文本通过互联网实时搜索,可以提供地理位置查询、旅游导航和语言翻译等多样化的便捷服务和帮助。

针对复杂场景的文本提取问题,国内外学者进行了大量的研究和实验,提出了许多自然场景文本的提取方法。Zhang 等提出一种基于聚类 and 马尔科夫随机场(MRF)模型的场景文字提取方法^[1],该方法利用色彩、边缘及纹理信息,选用基于 HSL 颜色空间的改进模糊均值颜色聚类方法来提取候选文字区域,但该方法准确率较低。Meng 等提出了一种两步边缘分析(边缘合并和边缘分类)的自然场景文本提取方法^[2],该方法首先通过图像边缘检测将图像分成若干部分,然后将邻近相似笔画宽度和颜色的边缘合并,准确率较高,但该方法采用支持向量机(SVM)分类器进行文字特征合并,属于监督学习范畴。Yi 等提出一种辅助失明人士阅读自然场景文字的方法^[3],有效解决了失明人士阅读困难的问题。Neumann 等提出了基于学习的文本提取方法^[4-5],这些方法虽然能够适应多种场景变化且提取准确性高,但使用机器学习需要训练大量的样本。

现有的方法大多采用文本的色彩、纹理和边缘等特征,在光照不均匀和边缘轮廓模糊的情况下不可避免地产生文字区域的漏检。Matas 等提出的 MSER 算法^[6]可很好地解决这一问题,MSER 检测子具有旋转、尺度和仿射不变性、鲁棒性和准确性等特性,在视角多变、尺度变化的情况下能有效检测文本,检测效率高。

自然场景文本提取算法中对多语种场景文本提取的研究相对较少,在自然场景文本中很难对单个字符的大小、颜色、位置和方向等先验知识进行判定。针对此问题,Epshtein 等提出了笔画宽度变换(SWT)^[7],成功地完成了多语种场景文本提取任务,该方法巧妙利用文本字符显著的笔画特征,从而很好适应自然场景文本环境下的文本提取。自然场景文本提取过程中通常会对图像进行二值分割,然而场景文本的灰度变化不一,采用固定灰度阈值对文本进行二值分割的效果不佳。因此,Otsu 提出最大类间方差(Otsu)自适应阈值文本分割算法^[8],该

方法通过计算图像前景与背景的最大类间方差得到最佳分割阈值,从而完成图像的二值分割。

本文提出一种基于 MSER 检测和 SWT 的自然场景文本提取方法。该方法以自然场景图像文本为研究对象,首先分析文本区域的仿射不变性特征,利用 MSER 进行文本区域检测,产生一系列嵌套的最大极值稳定区域。为解决 MSER 重复检测的问题,采用区域清理和区域像素拓展进行了优化,合并嵌套的极值区域,建立独立的文本候选区域。然后,以笔画宽度作为文本提取的关键特征,实现对文本候选区域的准确筛选。最后,利用笔画宽度连通图启发性规则和 Otsu 阈值分割完成文本提取。

1 算法流程

本文算法主要结合 MSER 与 SWT 完成自然场景文本的检测和提取,算法提取准确度高,综合性能好。具体算法流程如图 1 所示,该算法分为建立文本候选区域和判定文本区域两部分。

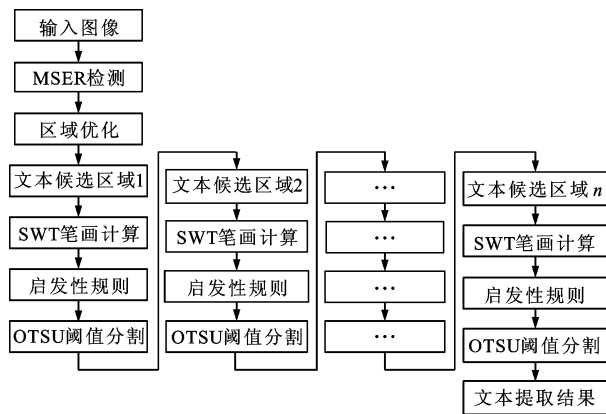


图1 整体算法流程

1.1 文本候选区域的建立

针对文本候选区域的建立而言,最难解决的问题是如何在视角多变、尺度变化的情况下有效检测文本,建立文本候选区。为了保证在文本检测的环节不存在漏检现象,本文选用仿射不变性 MSER 算子进行文本检测,MSER 数学定义如下。

图像 I 是一种映射, $I: D \subset \mathbb{Z}^2 \rightarrow S$, 且满足条件: S 是全序, 且 $S = \{0, 1, \dots, 255\}$; 邻域关系 $A \subset D \times D$, 如果 $\sum_{i=1}^d |p_i - q_i| \leq 1$, 那么 $p, q \in D$ 就是相邻

关系,记为 pAq 。

区域 Q 是 D 的一个连续子集,对于任意 $p, q \in Q$,都存在一个连通路程 $p, a_1, \dots, q_n, q$,使得 $pAa_1, \dots, a_iAa_{i+1}, \dots, a_nAq$ 。区域 Q 的边界 $\partial Q = \{q \in D \setminus Q: \exists p \in Q: qAp\}$, ∂Q 与 Q 至少一个像素相邻,但不属于 Q 的像素集合。极值区域 $Q \subset D$: 对于所有的 $p \in Q, q \in \partial Q$,满足 $I(p) > I(q)$ (最大灰度区域) 或者 $I(p) < I(q)$ (最小灰度区域) 的区域。

最大稳定极值区域 MSER: 设 $Q_1, \dots, Q_{i-1}, \dots$ 为极值嵌套区域 ($Q_i \subset Q_{i+1}$) 的一个序列,如果稳定性方程 $q(i) = |Q_{i+\Delta}/Q_{i-\Delta}| / |Q_i|$ 在 i^* 处存在局部极小值,那么极值区域 Q_i 就是 MSER。其中 i^* 为嵌套极值区域的某层, $|\cdot|$ 为集合的势, $\Delta \in S$ 为参数。

MSER 的原理是使用不同的灰度阈值对图像进行二值化,在比较宽的灰度阈值范围内保持形状稳定的区域。经典 MSER 会对稳定的单个字符区域重复检测,产生一系列嵌套的最大稳定极值区域,如果仅仅利用 MSER 检测到的文本区域作为文本的候选区域,那么同一文字区域会产生多个候选文本区域,从而增加了计算复杂度,降低了文本提取速度,如图 2 所示。



图 2 MSER 文本区域检测

本文采用区域清理和区域像素拓展对 MSER 检测区域进行优化,尽可能地排除非文本区域,加快后述的文本区域判定速度。分析文本特征可知,文字的尺寸在一幅图像中有一定范围,设定文字最小尺寸为 $T_{\min}$,最大尺寸为 $T_{\max}$,用区域像素个数表示该文字区域面积,区域清理过程为:若一个区域的像素个数小于 $T_{\min}$,则删除该区域;若一个区域的像素个数大于 $T_{\max}$,则删除该区域;求解 MSER 检测椭圆区域的长宽比,根据文字区域特征限定其范围 $[1/8, 8]$,据此删除明显的非文字区域。

区域像素扩展如图 3 所示,文本中单个字符之间具有一定的水平和垂直间距,将每个 MSER 合并距离相近的重叠区域,利用像素拓展可以合并 MSER 重复检测区域,同时确保一些非完整连接文字的完整性。根据文字特征对最大稳定极值区域做像素拓展,像素拓展过程为:对每个最大稳定极值区域

建立矩形边界框 (Bounding Box),圈定极值区域范围;对矩形边界进行水平和垂直方向上的像素拓展,通过实验分析,像素拓展范围一般限定在 $[2, 7]$;将有重叠的区域合并得到文本候选区域。非文字区域的清理以及重叠区域的合并有利于加快文本区域判定的处理速度。



图 3 区域像素拓展合并重叠区域

1.2 文字区域的判定

候选区域提取完成后,需要进一步判定是否为文字区域。本文将采用文本笔画宽度特征来实现文本区域的判定,通过计算一幅图像中每个文字边缘像素点的笔画宽度,生成一幅尺寸大小与原图一致的笔画宽度图,像素点原值将被该点所对应的笔画宽度值代替,计算文字笔画宽度图的详细流程如图 4 所示。利用 Canny 算子^[9]具有精细边缘定位和边缘响应次数最少的优良特性,对文本图像进行边缘提取,边缘提取结果如图 5 所示。利用 Sobel 算子^[10]计算文本图像的边缘梯度方向,得到梯度方向图,结果如图 6 所示。笔画宽度变换计算过程如图 7 所示,从笔画边缘上的一点 p 沿着梯度方向 r 出发,找到与 r 相反的另一笔画边缘点 q ,那么 p 与 q 点之间直线的距离就是其所包含的像素个数 w ,即 p 点的笔画宽度值。逐个求取边缘点的笔画宽度后会生成一幅尺寸大小与原图一致的笔画宽度图。

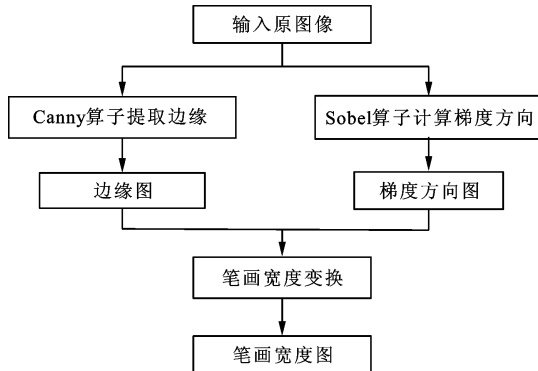


图 4 笔画宽度计算流程图

基于笔画宽度的连通域标记算法^[11]思想为:在经典标记算法基础上扫描笔画宽度图,如果当前像素点的值与其邻域点的笔画宽度值之比小于 3,则将标记为同一个区域,从而建立笔画连通图。利用启发式规则对笔画连通图进行约束,剔除非文本区



图5 文本边缘提取

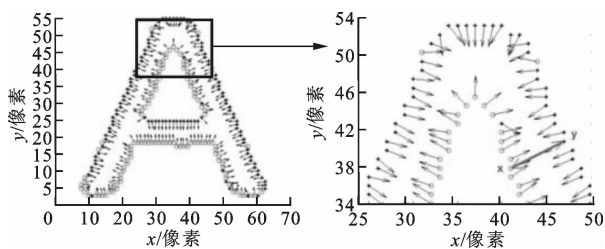


图6 笔画边缘梯度方向图

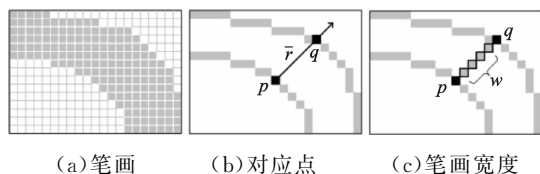


图7 笔画宽度计算

域。笔画连通图启发式规则为:计算每个连通区域笔画宽度值变化的方差,删除方差过大的连通区域,方差阈值设定为笔画宽度平均值之比的一半;连通区域长宽比要求在 $0.1 \sim 10$ 之间,若不在该范围内的区域则判定为非文本区域,如此可剔除掉不符合要求的区域,排除长宽比较大的区域;删除面积过大或过小的连通区域。

基于上述规则,文字区域的判定步骤为:①采用 SWT 计算文本候选区域正向和反向的笔画宽度,得到笔画宽度图 $SWTMap1$ 、 $SWTMap2$;②利用笔画宽度的连通域标记算法对笔画宽度图进行连通域标记,得到笔画宽度连通图 $SWTLabel1$ 、 $SWTLabel2$;③为连通图建立候选文本尺寸大小相同的索引矩阵 $Letter1$ 、 $Letter2$,初始化为 1,利用启发性规则,对不符合条件的区域在索引矩阵中置 0,剔除不符合条件的区域;④对索引矩阵进行求和,结果为 S_1 、 S_2 ,如果 $S_1 = 0$ 或 $S_2 = 0$,表明该候选区域是非文字区域。

通过笔画宽度连通域分析和启发式规则,可准确判定文本区域,剔除非文本区域,从而保留文本区域。对判定后的文本区域,保留的区域中文本与背景在亮度方面有明显的差异,并且候选区域相对于整幅图像来说尺寸较小,文本能量比较集中,前景与背景存在明显的类间方差。因此,选择 Otsu 算法对筛选的文本区域进行二值分割,完成场景文本的

提取。利用 SWT 对整幅图像做像素级处理时,是在笔画宽度计算过程中沿着梯度方向查找符合条件的边缘点,算法复杂度高。本文仅利用 SWT 对文本候选区域进行处理,加速了文本区域的判定速度。

2 实验及分析

本文首先对自然场景图片进行 MSER 文字区域检测和区域优化实验,确保在文本检测过程中不存在文本漏检现象,实验结果如图 8 所示。检测实验结果表明,本文所给出的 MSER 检测方法可以对复杂背景中视角多变,混合多语种,字体、尺寸大小有差异的自然场景图像建立文本候选区域,鲁棒性较好。在建立文本候选区域过程中,对于受到强烈光照等环境因素干扰、有异物遮挡及有严重模糊失真变形等情况的场景文本,本文算法具有一定的局限性,可能剔除一些文本区域。



图8 建立文本候选区域

本文对文本区域提取速度进行了实验,实验平台采用 Intel Core i5-2400 CPU,主频为 3.1 GHz,处理图像大小为 512×512 。实验结果表明,SWT 算法复杂度较高,利用 SWT 对整幅图像的处理速度较慢,处理时间长。本文利用区域拓展合并 MSER 重复检测区域,采用 SWT 处理 MSERs 文本候选区域,并没有直接应用 SWT 处理整幅图像,该方法节省了 30% 以上的时间,加速了场景文本提取的速度。

本文对自然场景文本中混合语言场景文本和非印刷场景文本进行了相关的实验,部分混合语言场景文本提取结果如图 9 所示。实验结果表明,本文算法能完成多语种场景文本的提取任务,通用性较强,非印刷场景文本部分实验结果如图 10 所示。实验结果表明,本文算法对非印刷字体有很好的鲁棒性,将笔画宽度作为文字区域的判定条件,对于文字笔画边缘轮廓清晰、同一水平线上文本连续、字符笔画宽度相对均匀的非印刷字体具有很好的提取效

果,对于尺寸较大的单个非印刷字体提取效果不佳。



图 9 混合语言场景文本提取结果



图 10 不同字体场景文本提取结果

为了评价本文算法,选用 ICDAR 图像库进行算法验证,图像库包含 250 张自然场景图像,这些自然场景图片具有文本视角多变、背景纹理丰富的特点。自然场景图像内容广泛,包括常见的广告牌、杂志封面、标志牌、街景商店名等。本文提出算法的测试结果与现有的场景文本提取算法进行比较。ICDAR 图像库部分实验结果如图 11 所示。



图 11 ICDAR 图像库部分实验结果

对于各种不同的检测提取文本方法来说,一种客观有效的评估标准是十分必要的。本文采用国际会议 ICDAR 所提出的评估方法^[12],通过计算文本区域提取的准确率 p 和召回率 r 来衡量算法的优劣,即

$$p = \frac{C}{E} \tag{1}$$

$$r = \frac{C}{T} \tag{2}$$

式中: E 为算法提取出来的区域数量; T 为图像中人工找出的文本区域数量; C 为 E 、 T 的交集。为方便算法性能对比,给出其综合性能

$$f = \frac{1}{\alpha/p + (1 - \alpha)/r} \tag{3}$$

式中: α 为 p 、 r 的权重因子,用来设定这两个参数对文本提取效果的影响程度,一般取 $\alpha = 0.5$ 。

本文算法结果与其他方法的性能比较如表 1 所示。由表 1 可知,本文方法提取准确率高、综合性能好。

表 1 自然场景文本提取性能对比

方法	p	r	f
文献[1]	0.59	0.53	0.56
文献[2]	0.78	0.63	0.70
文献[4]	0.73	0.64	0.68
文献[5]	0.83	0.67	0.74
文献[13]	0.68	0.67	0.67
文献[14]	0.73	0.60	0.66
本文	0.74	0.64	0.68

3 结 论

本文研究了一种基于 MSER 文本检测和笔画宽连通域标记的自然场景文本提取方法。该方法利用仿射不变性 MSER 检测子检测文本区域,将区域优化与笔画宽度分析方法有效结合,采用 SWT 连通域分析、启发式规则和局部 Otsu 完成文本提取,属于综合性场景文本提取方法。仿真结果表明,所提出的场景文本提取算法具有较高的精确度、提取率,对文本颜色、字体、尺寸以及语言种类具有较好的鲁棒性,但本文算法对受到强烈光照等环境因素干扰和有严重扭曲变形等场景文本具有一定的局限性,下一步将对场景文本光照不均和几何变形的矫正等算法进行研究。

参考文献:

[1] ZHANG T Z, ZHAO Y M. Scene text extraction method based on clustering and MRF model [J]. Computer Engineering, 2011, 37(21): 176-181.
[2] MENG Q, SONG Y, ZHANG Y, et al. Text detection in natural scene with edge analysis [C]// IEEE International Conference on Image Processing. Piscat-

- away, NJ, USA: IEEE, 2013: 4151-4155.
- [3] YI C, TIAN Y. Assistive text reading from natural scene for blind persons [M]. Berlin, Germany: Springer, 2015: 219-241.
 - [4] NEUMANN L, MATAS J. Real-time scene text localization and recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2012: 3538-3545.
 - [5] QU Y, LIN L, LIAO W, et al. Text localization with hierarchical multiple feature learning [M]. Berlin, Germany: Springer, 2015: 634-643.
 - [6] MATAS J, CHUM O, URBAN M, et al. Robust wide-baseline stereo from maximally stable extremal regions [J]. Image & Vision Computing, 2004, 22(10): 761-767.
 - [7] EPSHTEIN B, OFEK E, WEXLER Y. Detecting text in natural scenes with stroke width transform [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2010, 119(5): 2963-2970.
 - [8] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
 - [9] CANNY J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.
 - [10] KITTLER J. On the accuracy of the Sobel edge detector [J]. Image and Vision Computing, 1983, 1(1): 37-42.
 - [11] WU K, OTOO E, SUZUKI K. Optimizing two-pass connected-component labeling algorithms [J]. Formal Pattern Analysis & Applications, 2009, 12(2): 117-135.
 - [12] LUCAS S M, PANARETOS A, SOSA L, et al. IC-DAR 2003 robust reading competitions [C]// International Conference on Document Analysis and Recognition. Piscataway, NJ, USA: IEEE, 2003: 682-687.
 - [13] PAN Y F, HOU X, LIU C L. A hybrid approach to detect and localize texts in natural scene images [J]. IEEE Transactions on Image Processing, 2011, 20(3): 800-813.
 - [14] CHEN H, TSAI S S, SCHROTH G, et al. Robust text detection in natural images with edge-enhanced maximally stable extremal regions [C]// 18th IEEE International Conference on Image Processing. Piscataway, NJ, USA: IEEE, 2011: 2609-2612.

[本刊相关文献链接]

- 张胜杰, 查宇飞, 李运强, 等. 一种利用局部结构信息的加权哈希图像检索算法. 2016, 50(10): 78-85. [doi: 10.7652/xjtuxb201610012]
- 周奇才, 吴青龙, 熊肖磊, 等. 桁架结构拓扑及截面尺寸优化设计方法. 2016, 50(9): 1-9. [doi: 10.7652/xjtuxb201609001]
- 黄晓冬, 孙亮, 刘胜蓝. 一种判别极端学习的相关反馈图像检索方法. 2016, 50(8): 96-102. [doi: 10.7652/xjtuxb201608016]
- 张凤娜, 张天奎, 王东明, 等. 复杂快中子源编码成像高效模拟方法研究. 2016, 50(7): 118-124. [doi: 10.7652/xjtuxb201607018]
- 张泰, 张为, 刘艳艳. 周界视频监控中人员翻越行为检测算法. 2016, 50(6): 47-53. [doi: 10.7652/xjtuxb201606008]
- 郝维娜, 令锋超, 刘志刚, 等. 轴承滚珠面型误差激光干涉测量系统的研究. 2016, 50(6): 83-89. [doi: 10.7652/xjtuxb201606013]
- 瞿磊, 王忠, 李瑞娜, 等. 甲醇/生物柴油燃烧颗粒的结构特征研究. 2016, 50(5): 81-86. [doi: 10.7652/xjtuxb201605012]
- 王富平, 水鹏朗. 采用多尺度方向微分比率的角度点检测算法. 2016, 50(4): 68-75. [doi: 10.7652/xjtuxb201604011]
- 姚翠萍, 周湘连, 王晶, 等. 纳米金标记细胞的荧光寿命成像及其三维重建. 2016, 50(4): 153-158. [doi: 10.7652/xjtuxb201604023]
- 朱爱斌, 何大勇, 邹超, 等. 刀具磨损图像视差图的非标定方法. 2016, 50(3): 8-15. [doi: 10.7652/xjtuxb201603002]
- 张鹏, 陈湘军, 阮雅端, 等. 采用稀疏 SIFT 特征的车辆识别方法. 2015, 49(12): 137-143. [doi: 10.7652/xjtuxb201512022]
- 崔汪莉, 卫军胡, 纪鹏, 等. 基于加权局部梯度直方图的头部三维姿态估计. 2015, 49(11): 71-76. [doi: 10.7652/xjtuxb201511012]
- 周远, 周玉生, 刘权, 等. 一种适用于图像拼接的 DSIFT 算法研究. 2015, 49(9): 84-90. [doi: 10.7652/xjtuxb201509015]
- 毛彦斌, 张选平, 杨晓刚. 伪 DNA 密码图像加密算法研究. 2015, 49(9): 91-98. [doi: 10.7652/xjtuxb201509016]
- 刘凯, 张立民, 孙永威, 等. 利用深度玻尔兹曼机与典型相关分析的自动图像标注算法. 2015, 49(6): 33-38. [doi: 10.7652/xjtuxb201506006]
- 吴一全, 孟天亮, 吴诗姘. 人工蜂群优化的非下采样 Shearlet 域引导滤波图像增强. 2015, 49(6): 39-45. [doi: 10.7652/xjtuxb201506007]
- 朱爱斌, 胡浩强, 何大勇, 等. 采用频域融合方法的砂轮刀具磨损三维重构技术. 2015, 49(5): 82-86. [doi: 10.7652/xjtuxb201505013]

(编辑 赵炜)