计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2019.20180543
56(7): 1525-1533, 2019

基于城市监控的自然场景图像的中文文本提取方法

肖 珂[1] 戴 舜[1] 何云华[1] 孙利民[2]
[1](北方工业大学信息学院 北京 100144)
[2](中国科学院信息工程研究所 北京 100093)
(zehan_xiao@163.com)

Chinese Text Extraction Method of Natural Scene Images Based on City Monitoring

Xiao Ke[1], Dai Shun[1], He Yunhua[1], and Sun Limin[2]
[1](*School of Information Science and Technology, North China University of Technology, Beijing* 100144)
[2](*Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093)

Abstract Efficient environment monitoring and information analysis in urban scenes has become one of primary tasks of smart cities. In smart cities, the recognition of text information in scene images, especially the extraction of Chinese text in scene images, is an intuitive and efficient method for analyzing scene information. However, the Chinese text extraction of the current scene images fails to achieve good results because of the uneven illumination and blurred images. In addition, the complexity of Chinese character structure is also an important factor affecting the Chinese text extraction. In order to solve this problem, this paper proposes an edge enhanced maximally stable extremal regions (MSER) detection method, which can extract the MSER under the conditions of illumination and blurring influence, and the non-MSER can be efficiently filtered by geometric feature constraints to obtain high quality candidate MSER. Then the proposed central aggregation is used to aggregate the candidate Chinese text field that has been divided into multiple MSER, so that the candidate region becomes a single candidate Chinese text component, and then these components are analyzed, and finally the correct Chinese text is selected by machine learning. Experiments show that the algorithm can extract Chinese text in natural scene images more effectively.

Key words text extraction; maximally stable extremal regions (MSER); Chinese aggregation; support vector machine (SVM); Internet of things (IoT)

摘 要 智慧城市的首要任务是城市场景监控及其信息分析,场景图像中文本信息的识别是一种直观且高效的场景信息分析手段,但目前场景图像的中文文本提取由于图像光照和模糊、中文字符结构复杂等因素,未能达到很好的效果.为解决这一问题,提出一种边缘增强的最大稳定极值区域(maximally stable extremal regions, MSER)检测方法,可在光照和模糊影响的条件下提取 MSER,通过几何特征约束条件高效地过滤明显的非 MSER,得到高质量的候选 MSER.之后使用提出的中心聚合方法对分割成多个 MSER 的候选中文文本域进行中文的聚合,使得候选区域成为单个候选的中文文本分量,再对这些分量进行分析,并运用机器学习选出正确的中文文本.实验结果表明:该算法能够更有效地提取出自然场景图像中的中文文本.

收稿日期:2018-08-02;**修回日期**:2018-12-13
基金项目:国家重点研发计划项目(2017YFB0802300);国家自然科学基金项目(61802005);北京市自然科学基金项目(4184085)
This work was supported by the National Key Research and Development Program of China (2017YFB0802300), the National Natural Science Foundation of China (61802005), and the Beijing Natural Science Foundation (4184085).
通信作者:何云华(heyunhua610@163.com)

关键词 文本提取;最大稳定极值区域;中文聚合;支持向量机;物联网

中图法分类号 TP311

目前传感器已被应用于各种环境的实时感知,感知数据的分析与利用逐渐改变着人们的生活方式,由此激起了各类物联网[1](Internet of things, IoT)场景应用,如智慧城市、智能医疗和国防军事等[2].随着城市化的进展,智慧城市在大数据基础上,通过物联网将现实城市与数据进行有效融合,自动和实时地感知现实世界中人与物体的各种状态和变化,为城市管理和公众提供各种智能化的服务.在智慧城市的推动过程中,视频图像的检测和识别成为一项关键的任务.视频图像检测和识别是基于内容的视觉媒体,对图像的颜色、纹理和布局等进行分析和检索,从中挖掘出规律性的内容,这样能方便城市电子警察对城市监控和管理.

针对城市应用环境,视频图像的检测与识别方案也存在一些问题,如捕获的照片模糊失真,无法用于城市管理.电子产品往往暴露在外,受外界环境影响较大,采集的图像会受到外界噪声、散射等因素影响导致处理效果不理想.本文针对智慧城市系统架构中图像处理模块,研究高效的自然场景文本提取算法,通过高效快速文本提取算法为智慧城市中场景检测和识别功能提供保障.

现有文本提取方法可以分为两大类:基于滑动窗口的方法和基于连通域的方法[3].1)基于滑动窗口的方法[4]通常利用固定大小的滑动窗口来搜索图像中的单个候选字符或候选字词,然后使用机器学习技术来分类和识别文本.尽管这样的方法对于噪声和模糊是鲁棒的,但是由于搜索空间大使得它们的速度偏慢.2)基于连通域的方法首先通过使用图像的局部属性(例如强度、颜色、笔画宽度)从图像中作为候选文本提取连通域,然后使用字符或文本行的属性作为特征,利用统计学或机器学习等来去除非文本连通域.该方法能够实现高鲁棒性和低计算量,且针对英文文本的检测在文档分析与识别国际会议(International Conference on Document Analysis and Recognition, ICDAR)的竞赛中已有了很好表现.但其应用到中文的文本提取,并不能达到处理英文时的良好效果.这是由于中文的单个字符并不具有英文那样单个连通域的形式,难以保证候选文本连通域的提取质量.再加上文本提取中的一些公开性问题,如光照不均和非文本的形状非常类似于文本字符等,针对中文的文本提取很难达到满意的效果.而已有的针对中文的提取算法在效率和提取能力上仍需提高.

针对上述问题,本文提出了一种基于边缘增强的最大稳定极值区域(maximally stable extremal regions, MSER)和支持向量机(support vector machine, SVM)结合的自然场景中文文本提取算法.首先,在考虑图像光照和模糊等因素的情况下,使用基于边缘增强的 MSER 检测方法,过滤和聚合候补 MSER 得到有效的中文文本域;再根据中文文本域的特征使用高效的机器学习算法将类似文本的结构剔除从而保障中文文本提取的准确性.

1 相关工作

近年来,对于自然场景的文本检测和提取的工作已备受关注,学者们也提出一些优秀方法值得参考.在基于滑动窗口的一些方法中,Huang 等人[5]提出了基于滑动窗口和 MSER 结合的文本提取方法,MSER 可以显著减少扫描的窗口数量,并增强对低质量文本的检测,最后使用卷积神经网络(convolutional neural network, CNN)分类出正确的文本;Gómez 等人[6]探讨了对象提议技术在场景文本理解中的适用性,提出了一种简单的文本特定的选择性搜索策略,搜索图像中的特定窗口,并通过凝聚聚类在层次结构中分组,对每个节点定义可能的语义假设,根据语义来检测场景图像中文本单词;Zhou 等人[7]提出了能够直接预测全图像中任意方位和矩形形状的文字或文字线管道的方法,通过设计高效的损失函数和神经网络结构,用单个神经网络消除不必要的中间步骤(例如候选聚合和单词分割).这些方法有效地利用滑动窗口的特性,得到不错的提取效果.

以连通域为基础的方法中,Minetto 等人[8]提出了一个结合自下而上和自上而下机制来检测文本框的综合策略,自下而上的部分是基于连通域分割和分组进行的,而自上而下的部分是通过基于框描述符的统计学习方法实现的,该部分主要贡献在于引入一个适用于文本框分析的新描述符——模糊方向梯度直方图,以此实现场景图像的文本提取;Rajan 等人[9]提出了一种基于分数泊松的增强模型

来提高拉普拉斯算子图像的质量,通过图像增强操作以获得目标和背景之间更好的对比度,增强图像有效避免拉普拉斯算子图像的噪声,实现了更高精度的文本检测和识别;Yao 等人[10]提出一种利用 2 级分类方案的文本提取方法,采用笔画宽度变化(stroke width transform, SWT),并根据文本的一些固有属性设计了对文本非常有效的 2 级分类方案,再以适度的训练来消除敏感的手动参数调整,在场景图像的文本提取方面取得了很好的效果.

以上这些算法虽然具有很好的效果,但它们的目标都是针对英文.而如果将该类方法用于中文的文本提取,难以达到他们处理英文时的优越性能,针对中文的提取方法,国内学者也做出一些不错的工作.例如张伟伟等人[11]通过剪枝策略对图像存在嵌套关系的连通域进行取舍,得到候选笔画区域,利用结构元参数对图像进行动态闭操作,以消除同一汉字笔画之间的间隙,得候选汉字区域,之后利用结构和角点规则过滤掉非汉字区域,并用颜色规则聚类得到候选文本区域;喻勃然等人[12]通过最大稳定极值算法提取区域,对于汉字笔画分离的问题,用形态学运算进行笔画融合,再根据汉字的特点,设计启发式规则过滤非文本区域,其中通过候选字符区域的椭圆拟合,引入椭圆的偏心率作为文本判别规则.但由于效率和图像噪声敏感等原因,这些算法无法满足物联网的环境,为了将文本提取算法实现在物联网这样的实时性平台上,本文提出了一种基于 MESR 和 SVM 结合的高效中文提取方法.

2　基于 MSER 和 SVM 的中文文本提取方法

在智慧城市概念中,有效监测和分析城市中各场景信息可加强对城市的管理,而场景中包含的一些文本信息可以极大地提高场景信息分析的效率.因此,本文研究针对自然场景下的高效中文文本提取算法,并解决现有中文文本算法因效率不足而无法应用于城市场景监测的问题.其中算法的流程如图 1 所示.首先,提出使用基于边缘增强的 MSER 检测算法,提取出图像的 MSER;以 MSER 为单位进行分析,并使用几何特征的约束,对所得到的 MSER 进行过滤;对于过滤后的 MSER 进行中文聚合,图像中的中文文本往往会被分割成多个 MSER,使分散的结构形成候选中文文本域;最后根据中文文本的特征,对文本进行 SVM 分类,得到正确的正确文本.

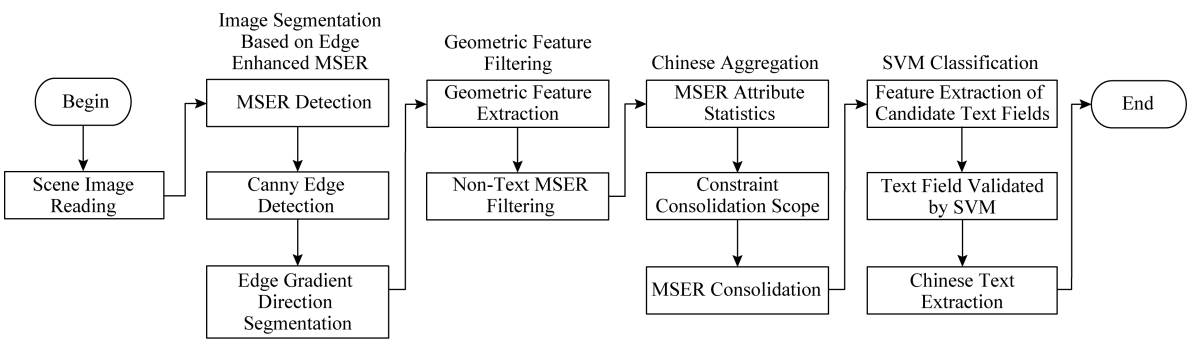


Fig. 1　Algorithm flowchart

图 1　算法流程图

2.1　边缘增强的 MSER 的图像分割

本文复现了 Matas 等人[13]提出的最大稳定极值区域检测算法,检测结果如图 2 所示.由图 2 可知,自然场景图像中存在大量的 MSER,这些区域中包含了大量的非文本区域,需要进一步的过滤.而在某些场景下部分文本区域并没有被判断为 MSER,这将直接影响后续的提取结果.

由于文本与其背景的灰度对比通常极为重要,并且可以假定每个文本具有均匀灰度或颜色,因此 MSER 是文本区域检测和提取的自然选择.虽然 MSER 被视为最好的区域检测器之一[14],但由于其对视点、比例和光照变化的鲁棒性,加上它对模糊图像的敏感,将 MSER 直接应用于有限分辨率的图像时,不能有效地检测或区分某些特殊的场景图像的文本区域.

针对多个文本被检测为单个 MSER 区域这类由图像模糊造成的现象,本文结合 Canny 边缘检测和 MSER 的提取特性,通过 Canny 边缘来增强极值区域的轮廓,然后沿着原始灰度图像计算出的梯度方向修剪 MSER,从而移除了由 Canny 边缘形成的边界外 MSER 像素.由于文本类型(亮或暗)在 MSER 检测阶段是已知的,因此可以调整梯度方向

以保证它们的指向背景.边缘增强的 MSER,提供了显著改进的文本表示,其中分开单独的连通区.不仅可以提高几何过滤器的性能,而且还可以增加在不同图像特殊条件下基于 MSER 的特征匹配的可重复性,这种边缘增强的 MSER 检测算法,结合边缘和 MSER 区域的优点,相比于传统的 MSER 算法,不仅能够提高检测算法对复杂场景的应用性,同时还可以减少背景的干扰,有利于后续对文本区域鉴别.图 2 显示了边缘增强的 MSER 图像分割的良好效果.

(a) Natural scene image (a) Image segmented by traditional MSER (c) Image segmented by edge enhanced MSER

Fig. 2 Comparison of results between MSER and edge enhanced MSER

图 2 MSER 与边缘增强的 MSER 实验结果对比

2.2 先验知识初步过滤

本文基于 MSER 的提取效果和中文字符的特点制定了一些高效的先验知识,作为约束条件进行初步的过滤:

1) 基于长短轴长度比的过滤.由于中文存在偏旁部首,而偏旁部首不像整个字符那样特征鲜明,所以适当放宽长轴与短轴的比例约束,将长短比大于 4∶1的 MSER 过滤掉.值得注意的是,中文字符有些特殊的偏旁部首,如“亻”、“一”和“刂”等结构不能满足先前的约束,为了防止这样的 MSER 被过滤,对这些 MSER 的过滤采用新的约束.经过研究发现,“亻”、“一”和“刂”等结构的共同特征是拟合椭圆方向都接近竖直或者水平方向,所以当 MSER 的椭圆拟合方向为水平和竖直时,长短比大于 8∶1 的 MSER 才会过滤.

2) 基于孔洞数过滤.中文字符中包含孔洞数的范围并不能轻易地约束,但是此时的 MSER 只是一个代表中文字符部分的连通区域,这样的区域通常没有过多的空洞数.在众多中文字符中,其所包含的单个偏旁部首结构,孔洞数量一般不超过 5 个,即 $MSER\_hole_i \leqslant 5$,所以该约束条件能够把孔洞数大于 5 的 MSER 都过滤.

3) 基于占空比过滤.中文字符的部分结构通常具有一定占空比,即像素面积与椭圆面积的比例,正确的结构其占空比通常不会太小也不会过大.因为字符偏旁部首的像素除了某些特殊“亻”和“一”等,其他都分布得相对松散.而由 MSER 通过整体形状拟合出椭圆,其面积必定不会比 MSER 像素组成面积小.所以将满足占空比小于 0.2 且大于 0.85 的 MSER 过滤掉.

2.3 中文聚合方法

中文字符不同于英文字符的一笔而就,它通常是由多个 MSER 组成,为了得到正确的中文文本,在验证之前需要将分散的多个 MSER 聚合成候补的文本区域.对此,本文提出了如算法 1 所示的基于文本中心聚合的方法,有 4 个步骤:

1) 统计 MSER 属性.得到每个 MSER 的矩形包围盒信息、质心坐标、平均颜色分量以及平均笔画宽度.由于中文字符被称为“方块字”,单个中文字符的最佳凸包通常是一个正方形.因此,矩形包围盒的使用能够更加有效地迎合中文字符的特点.

2) 约束合并范围.除了“一”等特殊的中文字符,单个完整的中文字符通常拥有相近的高度和宽度,并且在场景图像中,为了方便人们辨识文字,字体的各个结构会具有相似的笔画宽度和颜色.所以该聚合算法在 MSER 相互合并之前,先在二维空间中找出每个 MSER 能够实现合并的一些对象,即对每个待处理的 MSER 只考虑质心在距离约束范围内的 MSER 作为备选的合并结构,该距离约束范围是以待处理的 MSER 的质心为圆心、12 倍的平均笔画宽度为半径的圆圈.同时为了避免背景的类似结构误入,以 2 个 MSER 之间平均颜色比值(颜色分量的比值)和平均笔画宽度比值小于 1.2 作为约束.

3) 初步相交合并.由于中文的特性,无论是书写还是印刷体,为了不让汉字的偏旁部首,被误判成相邻汉字的一部分,相邻字体之间会有一定距离,而这个距离会比字体的部首结构之间距离大很多.这种距离的差距对字体的结构合并成一个完整的字体很重要,因为它有利于将正确的笔画结构归并到字

体中,从而得到完整的中文文本区域.在合并判断时,本文将所有情况分为 2 种:相交和相邻.遍历合并范围内的 MSER,首先判断 2 个 MSER 的包围盒之间是否相交:

$$intersect=\left(\frac{R\_h_i+R\_h_j}{\Delta h}\geqslant 1\right)\vee\left(\frac{R\_w_i+R\_w_j}{\Delta w}\geqslant 1\right),\tag{1}$$

$$\Delta h=\max(|R\_b_i-R\_t_j|,|R\_b_j-R\_t_i|),\tag{2}$$

$$\Delta w=\max(|R\_r_i-R\_l_j|,|R\_r_j-R\_l_i|),\tag{3}$$

其中,R_w 和 R_h 表示 MSER 的宽和高;R_t,R_l,R_b,R_r 分别表示连通域包围盒的左上角和右下角的横纵坐标,$intersect$ 代表 2 个 MSER 是否相交.如果 $intersect$ 值为真就进行合并操作,将已经合并的 MSER 标记.在第 1 次相交合并时,被处理的 MSER 有可能被扩大,造成与原本未相交的 MSER 开始出现相交.因此遍历完合并范围内的 MSER 后,对未被标记的 MSER 再次进行相交判断并标记.

4) 相邻合并.此时如果合并范围内的 MSER 仍未被完全标记,则进行相邻合并,当 2 个 MSER 满足:

$$\max(|R\_w_i+R\_w_j-\Delta w|,\ |R\_h_i+R\_h_j-\Delta h|)<\frac{\lambda}{\kappa}T,\tag{4}$$

$$\max(\Delta w,\Delta h)<\lambda T,\tag{5}$$

$$T=\frac{1}{N}\sum_{i=1}^{N}(meanSW(MSER_i)),\tag{6}$$

其中,N 表示约束范围内连通域的总个数,经实验证明 κ 和 λ 设置为 4 和 10 时效果最佳.通过限制合并集合的宽度、高度以及宽高比例,避免邻近的包含完整字符的 MSER 被合并.

算法 1. 中文文本中心聚合.

输入:过滤后最大稳定极值区域 C_{MSER};

输出:候选中文文本域 TC.

```
for c∈C_MSER do
  Fature←swt,color,size,pos←c;
end for
for f_i,f_j∈Fature do
  if f_j.pos∈Range(f_i)
    R←{C_MSER|∈f_j C_MSER};
  end if
end for
for c_i,c_j∈R do
  if similarColor(c_i,c_j)
    if similarSWT(c_i,c_j)
      if intersect(c_i,c_j)
        c_i←c_i∪c_j;
      else
        if adjacentLimit(c_i,c_j)
          c_i←c_i∪c_j;
        end if
      end if
    end if
  end if
end for
TC←{c_1,…,c_i,…,c_j,…}.
```

2.4　基于支持向量机的文本分类

经过中文聚合后,形成了大量候选中文字符区域,在中文聚合前,初步过滤伪 MSER 仍然会存在许多类似的文本结构,所以需要经过再次分类.本文选取了一些针对中文字符的特征,作为 SVM 的特征向量进行训练与分类.

1) 面积比例特征

$$f\_areaRatio=\frac{Area(CC)}{Area(Pic)},\tag{7}$$

其中,$Area(CC)$代表候选文本区域面积,$Area(Pic)$表示图像面积.

2) 长度比例特征

$$f\_lenRatio=\frac{\max(w,h)}{\max(PicW,PicH)},\tag{8}$$

其中,w,h 分别表示候选文本区域的宽和高,而 $PicW$ 和 $PicH$ 分别表示图像的宽和高.

3) 长宽比特征

$$f\_aspectRatio=\max\left(\frac{w}{h},\frac{h}{w}\right).\tag{9}$$

4) 边缘对比度特征

$$f\_EdgeContrast=\frac{Border(CC)\cap(Canny(CC)\cup Sobel(Pic))}{Border(CC)},\tag{10}$$

其中,$Canny(Picture)$和 $Sobel(Picture)$分别表示图像的归一化 Canny 和 Sobel 边缘检测;$Border(CC)$表示候选文本区域的边界框包含的像素.

5) 形状规则特征

$$\begin{cases}f\_contourRoughness=\dfrac{Area(CC-open(imfill(CC),2\times 2))}{Area(CC)},\\ f\_holes=imholes(CC),\end{cases}\tag{11}$$

其中,$imfill(CC)$表示填充候选文本区域;$open(\cdot)$表

示进行开运算;$imholes(CC)$表示统计候选文本区域中的孔洞数.

6) 笔画宽度特征

$$f\_stroke=\frac{varSW(CC)}{meanSW(CC)},\tag{12}$$

其中,$varSW(CC)$表示候选文本区域的笔画宽度方差,$meanSW(CC)$表示候选文本区域的笔画宽度均值.

7) 空间相干性面积比特征

$$f\_ratio\_S=\frac{Area(imdilate(CC,5\times5))}{Area(Pic)},\tag{13}$$

其中,$imdilate(\cdot,strel)$代表结构元素 $strel$ 的形态膨胀操作.

3 实验与结果分析

本文算法的实验平台为戴尔台式计算机,其CPU为Intel core i7的处理器,运行内存为8 GB,操作系统为64位的Windows 7系统.

3.1 数据集

公开的实验数据集对文本提取的研究责任重大,当研究人员使用公开数据集进行算法评估时,算法的性能体现才更具说服力.对于中文文本的提取,目前没有公开且权威的自然场景图像数据集.虽然有西安电子科技大学建立的中文图像数据集,却只在校园内研究使用.为了更好地评定本文的研究,根据ICDAR数据集的图像组成规则,建立了针对中文文本提取的图像库,具体建立方法如下:

1) 数量组成.220幅训练样本的图像和180幅测试图像.

2) 图像分辨率范围.ICDAR竞赛图像库的图像分辨率范围是860×640至1 600×1 200,本文采集的图像其分辨率从860×640至2 048×1 536.

3) 难度比例.根据图像中文本提取的难度,将图像分为难、中和易3个等级.ICDAR竞赛图像库中图像难度比例约为3∶1∶1,因此自建的中文图像库也遵循着这一难度比例.

4) 图像文本内容.ICDAR图像库中文本内容包括路边标志牌文本、服饰标签文本、图书封面文本、车辆车牌号、宣传字画文本、包装袋封皮文本和建筑物名称等,自建库也同样包含这些内容.

5) 字符组成.ICDAR图像库的图像中只包含英文文本,而自建库是针对中文的,因此主要由大量的中文文本、少量的阿拉伯数字和英文文本组成.

3.2 评价标准

在ICDAR比赛出现之后,学术界对文本检测、提取和识别的评价标准都迎合了ICDAR比赛中使用的评价方法.根据ICDAR评估协议,算法的性能是通过 f 值评定的,它是通过精确率和召回率调和平均值测定的.2个矩形之间的匹配度 m 被定义为交点面积与包含2个矩形的最小边界矩形的比值.由每种算法估计的矩形集称为估计值,而在ICDAR数据集中提供的基准矩形集称为目标.对于每个矩形,找到具有最大值的匹配.因此,1组矩形 R 中矩形 r 的最佳匹配是

$$m(r,R)=\max\{m(r,r_g)\mid r_g\in R\}.\tag{14}$$

然后,精确率和召回率的含义是

$$precision=\frac{\sum_{r_e\in E}m(r_e,T)}{|E|},\tag{15}$$

$$recall=\frac{\sum_{r_t\in T}m(r_t,E)}{|T|},\tag{16}$$

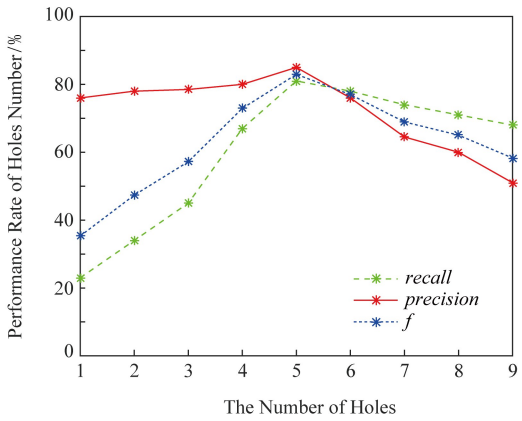
其中,E 和 T 分别是目标矩形和估计矩形的集合.f 是算法性能的单一度量,是精确率和召回率的组合指数.提取结果的精确率和召回率的相对权重由1个参数 α 控制,其被设置为0.5,得到相等权重的精确率和召回率:

$$f=\frac{1}{\frac{\alpha}{precision}+\frac{1-\alpha}{recall}}.\tag{17}$$

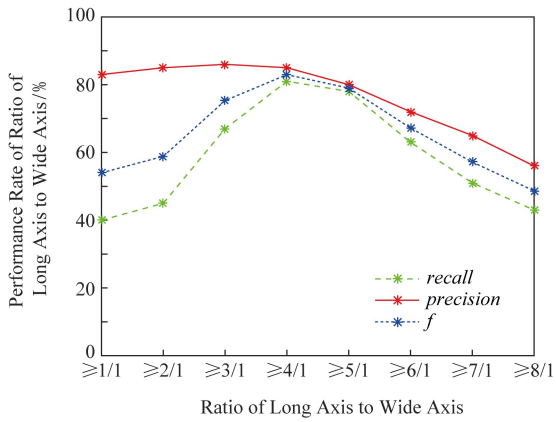
3.3 实验结果对比

本文对图像集中每张图像的平均处理时长为0.86 s,满足物联网的实时响应要求.在本文算法利用先验知识初步过滤处理过程中,选取合适的约束条件值进行初步的过滤,可以为后续的处理过程提供有力支撑.从图3可以看出,当过滤条件选取不恰当时,会对召回率的影响巨大,从而间接地影响了 f 值.

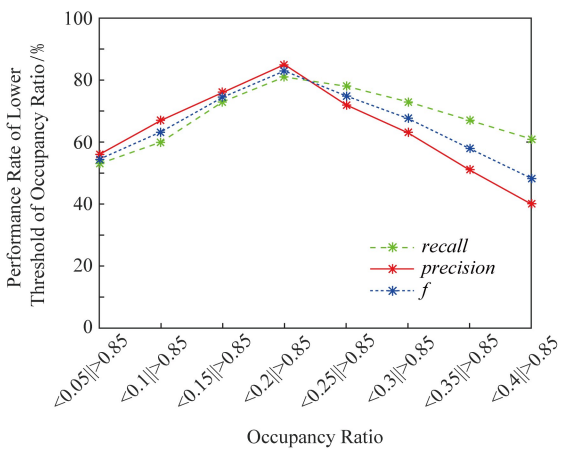
中文聚合步骤对于最终中文文本的提取至关重要,因此我们对其中涉及的关键参数 λ 的取值进行了实验.图4显示了不同的 λ 取值对算法性能 f 的影响,能够看出:当 $\lambda=10$ 时精确率和召回率都达到了峰值,算法具有最佳的性能;而当 λ 的取值逐渐增大,会造成聚合过度,即将2个独立中文文本被错误地合并成1个字符,这个错误聚合的文本无法通过SVM的验证,影响了精确率和召回率;相对地,当 λ 的取值逐渐变小,会逐渐地使得分散的笔画无法被有效地聚合为1个完整中文文本,同样会影响精确率和召回率,导致 f 值偏低.



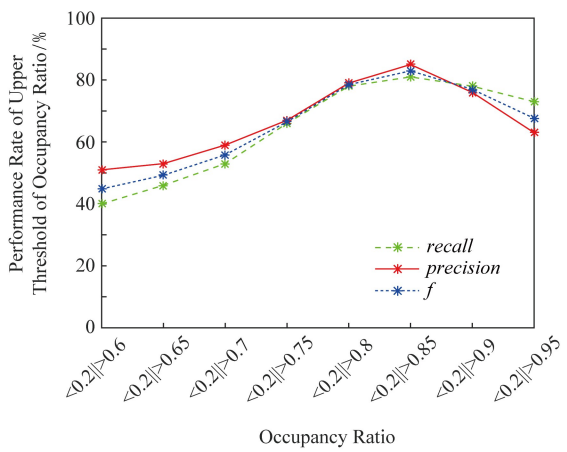
(a) Comparison of the number of holes



(b) Comparison of the ratio of long axis to wide axis



(c) Comparison of lower threshold selection of occupancy ratio



(d) Comparison of upper threshold selection of occupancy ratio

Fig. 3　Comparison of prior knowledge parameters

图 3　先验知识参数取值比较

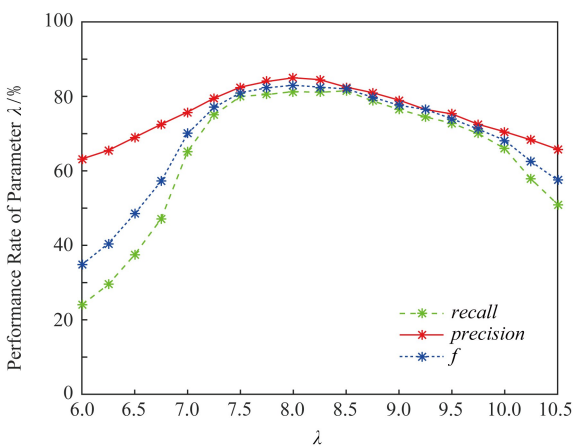


Fig. 4　Comparison of Chinese aggregate parameter λ

图 4　中文聚合参数 λ 取值比较

SVM 分类时本文使用多项式核函数,由于特征维数低,样本数远超过特征维数.分类的情况如表 1 所示,训练样本为 220 幅训练图像中提取出候选中文字符区域,而测试样本为 180 幅测试图像中提取的所有候选中文字符区域,可以看出 SVM 的分类效果明显优于 KNN 算法.

Table 1　Performance Comparison of Several Chinese Text Positioning Classification Methods

表 1　2 种中文文本定位分类方法的性能比较

Classification	Number of Training Samples	Number of Testing Samples	*precision*	*recall*	*f*
SVM	5 292	4 436	0.85	0.81	0.83
KNN	5 292	4 436	0.81	0.78	0.79

将本文的算法与近几年提到的中文文本提取算法进行对比,结果如表 2 所示.由表 2 可知,本文的算法在自建库上具有较好的提取效果,相较前人的中文文本提取算法,由于本文的算法对光照不均图像和模糊图像具有更好的处理能力,并对自然场景中复杂背景图像具有更稳定的提取效果,所以精确率和召回率有一定程度的提高.

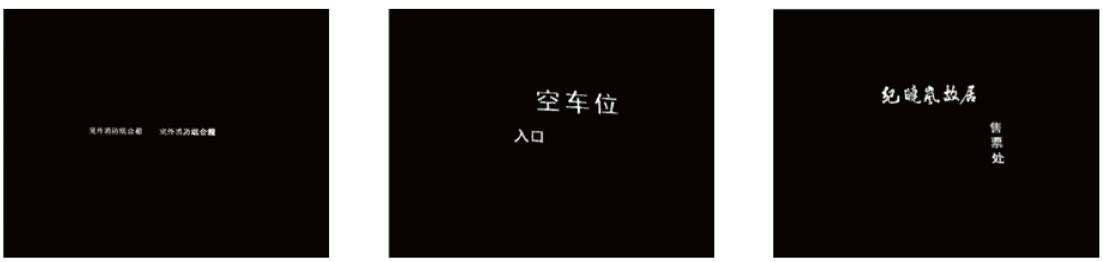
Table 2 Performance Comparison of Several Chinese Text Positioning Extraction Methods

表 2 5 种中文文本定位提取方法的性能比较

Chinese Text Extraction Algorithm	*precision*	*recall*	*f*
Edge-enhanced MSER Extraction Algorithm	0.85	0.81	0.83
Ref [11]	0.74	0.71	0.72
Ref [12]	0.75	0.78	0.76
Ref [15]	0.72	0.80	0.75
Ref [16]	0.78	0.81	0.80

图 5 所示的为本文算法中文文本提取的效果，图 5(a)为原始自然场景图像，而图 5(b)为提取后的二值图像.从图 5 中可以看出，自然场景图像里的中文文本都被很好地提取出.

(a) The original image



(b) Algorithm processed image

Fig. 5 Algorithm experimental result display

图 5 算法实验结果展示

4 总结与展望

本文在迎合物联网与图像处理结合的思想上，针对智慧城市中应用提取文本信息来加速对城市场景的监测，研究了针对自然场景下的高效中文文本提取算法，并解决了现有中文文本算法因效率不足而无法应用于城市场景监测的问题.提出了在自然场景图像下基于边缘增强的最大稳定极值区域中文文本提取方法，首先得到候选 MSER，并使用字符的长短轴和面积、空洞数目等约束条件高效地过滤明显的非 MSER，对候选文本进行初步验证.经过初步过滤后，运用中心聚合的方法，使得 MSER 聚合成各个候选文本区域，最后通过 SVM 验证得到文本.通过对算法性能的测试和评估，结果表明，本文提出的算法具有较高的精确率和召回率，解决了现有的在自然场景图像下针对中文文本提取效率不足的问题，且较少的处理时间也满足了智慧城市架构下对城市场景分析和识别的实效性.

参 考 文 献

[1] Gómez-Torres E, Luján-Mora S. An approach of context-aware mobile applications for Internet of things [C] //Proc of the 2nd Int Conf on Information Systems and Computer Science. Piscataway, NJ: IEEE, 2017: 41-48

[2] Sun Yimin. Research and simulation of large data differentiation classification technology under the Internet of things [C] //Proc of the 3rd Int Conf on Intelligent Transportation, Big Data & Smart City. Piscataway, NJ: IEEE, 2018: 191-194

[3] Wiwatcharakoses C, Patanukhom K. MSER based text localization for multi-language using double-threshold scheme [C] //Proc of the 1st Int Conf on Industrial Networks and Intelligent Systems. Piscataway, NJ: IEEE, 2015: 62-71

[4] Soni R, Kumar B, Chand S. Text detection and localization in natural scene images using MSER and fast guided filter [C] //Proc of the 4th Int Conf on Image Information Processing. Piscataway, NJ: IEEE, 2017: 351-356

[5] Huang Weilin, Qiao Yu, Tang Xiaoou. Robust scene text detection with convolution neural network induced MSER trees [C] //Proc of the 13th European Conf on Computer Vision. Berlin: Springer, 2014: 497-511

[6] Gómez L, Karatzas D. Object proposals for text extraction in the wild [C] //Proc of the 13th Int Conf on Document Analysis and Recognition. Piscataway, NJ: IEEE, 2015: 1786-1812

[7] Zhou Xinyu, Yao Cong, Wen He, et al. EAST: An efficient and accurate scene text detector [C] //Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2642-2651

[8] Minetto R, Thome N, Cord M, et al. Text detection and recognition in urban scenes [C] //Proc of the 2nd IEEE Int Conf on Computer Vision Workshops. Piscataway, NJ: IEEE, 2016: 227-234

[9] Rajan V, Raj S. Text detection and character extraction in natural scene images using fractional poisson model [C] // Proc of the 1st Int Conf on Computing Methodologies and Communication. Piscataway, NJ: IEEE, 2017: 1136-1141

[10] Yao Cong, Bai Xiang, Liu Wenyu, et al. Detecting texts of arbitrary orientations in natural images [C] //Proc of the 25th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 1083-1090

[11] Zhang Weiwei, Tang Guangming, Sun Yifeng, et al. Chinese scene text localization algorithm based on character features [J]. Journal of Information Engineering University, 2014, 15(6): 729-736 (in Chinese)
(张伟伟, 汤光明, 孙怡峰, 等. 一种针对汉字特点的场景图像中文文本定位算法[J]. 信息工程大学学报, 2014, 15(6): 729-736)

[12] Yu Boran, Wan Hongjie. Chinese text localization in natural scene based on heuristic rules and SVM [J]. Electronic Design Engineering, 2016, 24(24): 161-164 (in Chinese)
(喻勃然, 万洪杰. 基于启发式规则和 SVM 的自然场景中文文本定位[J]. 电子设计工程, 2016, 24(24): 161-164)

[13] 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

[14] Gao Shilin, Ji Lixin, Li Shaomei, et al. Fast scene-text localization algorithm based on MESR's fitting ellipse [J]. Computer Engineering and Design, 2015, 3(36): 693-698 (in Chinese)
(高士林, 吉立新, 李绍梅, 等. 基于 MSER 拟合椭圆的快速场景文本定位算法[J]. 计算机工程与设计, 2015, 3(36): 693-698)

[15] Li Chuang, Ding Xiaoqing, Wu Youshou. An algorithm for text location in images based on histogram features and AdaBoost [J]. Journal of Image and Graphics, 2006, 11(3): 325-331 (in Chinese)
(李闯, 丁晓青, 吴佑寿. 一种基于直方图特征和 AdaBoost 的图像中的文字定位算法[J]. 中国图像图形学报, 2006, 11(3): 325-331)

[16] Liu Xiaopei, Lu Zhaoyang, Li Jing. Complex scene text location method based on WTLBP and SVM [J]. Journal of Xidian University: Natural Science, 2012, 39(4): 103-108 (in Chinese)
(刘晓佩, 卢朝阳, 李静. 结合 WTLBP 特征和 SVM 的复杂场景文本定位方法[J]. 西安电子科技大学学报: 自然科学版, 2012, 39(4): 103-108)

Xiao Ke, born in 1980. PhD, professor. Member of IEEE. His main research interests include wireless comm-unications, physical layer security, the Internet of things, and embedded systems.

Dai Shun, born in 1994. Master. His main research interests include graphic processing and the Internet of things.

He Yunhua, born in 1987. PhD. Student member of IEEE. His main research interests include security and privacy in cyber-physical systems, bitcoin based incentive mechanism, security and privacy in vehicle ad hoc networks.

Sun Limin, born in 1966. PhD, professor, PhD supervisor. Member of IEEE. His main research interest include wireless sensor networks and vehicular ad hoc network.