基于边缘检测和特征融合的自然场景文本定位

王梦迪 张友梅 常发亮
(山东大学控制科学与工程学院 济南 250061)

摘 要 文本定位作为文本识别的基础和前提,对图像深层信息的理解至关重要。针对自然场景下的文本定位受光照、复杂背景等因素影响较大的问题,提出了一种基于多方向边缘检测和自适应特征融合的自然场景文本定位方法。该方法首先将自然场景图像进行三通道八方向的边缘检测;然后通过启发式规则对得到的边缘图像进行过滤从而提取出备选文本域,进而对备选文本域进行自适应权值的HOG-LBP特征提取与融合;最后采用支持向量机进行特征分类学习,实现文本定位。实验结果表明,该方法能准确定位自然场景图片的文本区域,对光照和复杂背景具有较强的鲁棒性。

关键词 自然场景,文本定位,边缘检测,特征融合
中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.056

Text Localization Based on Edge Detection and Features Fusion in Natural Scene

WANG Meng-di ZHANG You-mei CHANG Fa-liang
(College of Control Science and Engineering, Shandong University, Jinan 250061, China)

Abstract As the basis and premise of text recognition, text localization has an important influence on the analysis of images. Since the text localization in natural scene can be effected by illumination and the complex backgrounds significantly, we proposed a text localization method based on edge detection and features fusion. The method began with edge detection from three channels and eight directions, and then we filtered the detected edge images with heuristic rules to extract candidate text regions. On top of that, the HOG-LBP features were extracted and fused by adaptive weights. Finally, we applied support vector machine (SVM) to classify the candidate regions and realized text localization. Experimental results indicate that the proposed method can locate the text region accurately in natural scene images while reducing the influence of illumination and complex backgrounds effectively.

Keywords Natural scene, Text localization, Edge detection, Feature fusion

1 引言

随着数字信息化的发展和互联网的普及,图像的传输量呈现爆炸式增长,其中所含的信息量非常巨大,快速而准确地理解图像与视频中所包含的信息十分重要[1]。图像中往往蕴含着丰富的文本信息,提取这些文本信息对图像深层信息的获取和理解具有重要价值。文本定位作为文本信息提取的前提,其关键是提取有效的文本特征,通过适当的分类算法实现准确定位。目前文本定位算法大致分为3类:基于边缘检测的文本定位[2-3]、基于纹理特性的文本定位[4-5]和基于连通域的文本定位[6-7]。

基于边缘检测的文本定位利用了文本区域包含较多边缘的特性,此种定位方法操作简单,运算速度快。Lyu等[8]提出的定位视频中文本的算法通过改进的Sobel算子进行边缘提取,而后通过局部自适应阈值法将边缘图像转换成二值图像,最后通过投影分析来定位文本区域。Srivastav等[9]提出了一种基于笔画宽度和最近邻约束的文本定位算法,通过改进的Canny算子得到边缘图像,利用自适应阈值和启发式规则生成备选文本域,最后利用笔画宽度、背景颜色等特征实现文本定位。

基于纹理特征的文本定位利用了文本字符方向一致且呈现一定的纹理特征的特性,对字符大小和背景复杂度具有较强的鲁棒性,但纹理分析计算量较大,运算速度慢。Ye等[10]提出了一种基于局部二值模式(Local Binary Patterns, LBP)的文本定位算法。该算法首先根据改进的LBP算法得到的直方图对图像进行分块,然后利用概率神经网络(Probabilistic Neural Network, PNN)对特征向量进行分类,最后通过启发性规则实现精确定位。Mao等[11]提出了一种基于小波变换的文本定位算法,该算法首先对图像进行Haar小波分解,通过计算局部能量差异得到局部能量差异图,然后通过连通域分析和启发式规则实现精确定位。

基于连通域的文本定位利用了文本颜色的一致性,该方

到稿日期:2016-08-26 返修日期:2016-11-04 本文受国家自然科学基金项目(61673244),高等学校博士学科点专项科研基金资助课题(20130131110038)资助。

王梦迪(1993—),女,硕士生,主要研究方向为模式识别、计算机视觉,E-mail:wang_mengdi@yeah.net;**张友梅**(1991—),女,博士生,主要研究方向为模式识别、计算机视觉;**常发亮**(1965—),男,教授,博士生导师,主要研究方向为模式识别、计算机视觉。

法对背景复杂的图像有较高的敏感度，不适用于自然图像文本定位。Luca 等[12]提出了一种单个字符定位算法，即首先利用文本与背景具有较强对比度的特性，采用 Niblack 算法将图像像素分为文本像素、背景像素和其他像素，然后利用特征描述子进行连通域分析，得到备选文本域，最后利用级联分类器实现字符定位。

针对自然场景文本定位受光照、背景复杂度因素影响较大的问题，本文提出了一种基于多方向边缘检测和自适应特征融合的自然场景文本定位方法。该算法的流程如图 1 所示，首先将自然场景图像进行三通道八方向的边缘检测；然后通过启发式规则对得到的边缘图像进行过滤从而提取出备选文本域；最后对备选文本域进行自适应权值的 HOG-LBP 特征提取与融合，进而采用支持向量机（Support Vector Machine，SVM）进行特征分类学习，实现文本定位。

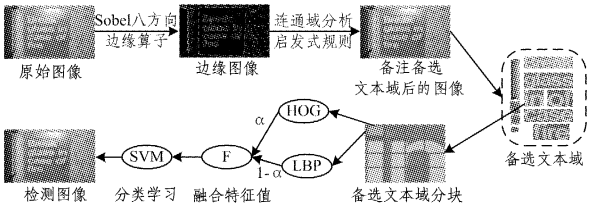


图 1　本文算法流程图

2　边缘与备选文本域的提取

图像中包含着丰富的边缘信息。从视觉角度来看，在观察一幅图像时，最先获取的是图像轮廓，即图像的边缘信息。图像中文本与背景的对比度较大，边缘检测算子能有效地获取字符的边缘，且文本区域具有较高的边缘密度。同时，图像中的文本具有不同于非文本的纹理特性[13]，例如，文本域具有一定的宽高比，而对于噪声区域，其形状具有不规则性；文本间相邻字符的距离紧密，字符间距与尺寸之间满足一定的比例，而非文本却不满足这一条件。利用文本域的这些特性，可将图像中的非文本域剔除，从而得到备选文本域。

边缘与备选文本域提取的具体流程如图 2 所示。

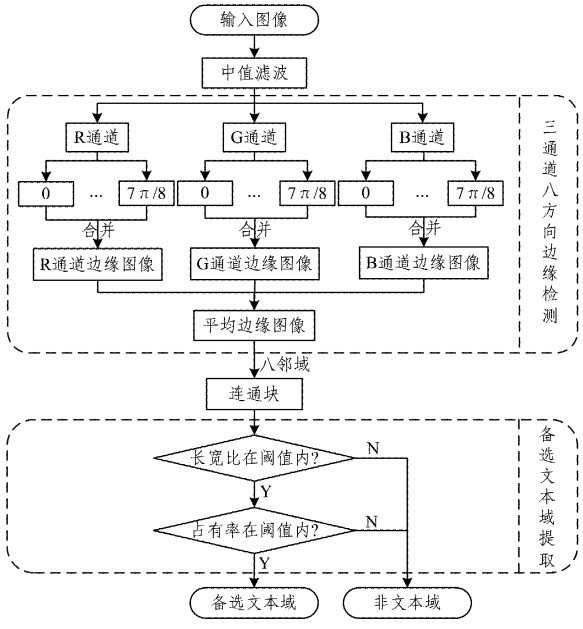


图 2　边缘与备选文本域提取流程图

给定一幅文本图像，首先通过中值滤波初步排除噪声等干扰因素；然后通过三通道八方向边缘检测得到更加清晰、规整的图像；最后运用八邻域联通法构建连通块，并根据连通块的长宽比等特性剔除非文本域，从而得到备选文本域。

2.1　基于多方向 Sobel 算子的边缘检测

当图像具有明显的边缘信息时，Sobel 算子可以获取较为准确的边缘方向信息；而 Canny 算子不仅可以获取清晰的图像边缘信息，而且获取的边缘的连续性较好。但 Canny 算子相较于 Sobel 算子会引入更多的背景噪声，从而影响文本定位的精确性[14]，因此本文采用 Sobel 算子进行边缘提取。

首先将原始图像进行 5 * 5 窗口的中值滤波和归一化处理。为了在光照不均情况下保留更多的边缘信息，分别在图像的 R，G，B 空间 3 个通道进行边缘检测。考虑到文本笔画的方向性，为了更好地保留文本的边缘信息，本文对 R，G，B 3 个通道都采用 0°，22.5°，45°，67.5°，90°，112.5°，135°，157.5° 8 个方向的 Sobel 算子[15]进行边缘检测，8 个方向的 Sobel 算子模板如图 3 所示。

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & -2 & -4 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 4 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) $Sobel_{0°}$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -4 & -2 & -0 \\ -1 & -4 & 0 & 4 & 1 \\ 0 & 2 & 4 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) $Sobel_{22.5°}$

$$\begin{bmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & -2 & -4 & 0 & 1 \\ 0 & -4 & 0 & 4 & 0 \\ -1 & 0 & 4 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(c) $Sobel_{45°}$

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -2 & -4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & 4 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

(d) $Sobel_{67.5°}$

$$\begin{bmatrix} 0 & -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{bmatrix}$$

(e) $Sobel_{90°}$

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & -4 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

(f) $Sobel_{112.5°}$

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & -4 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

(g) $Sobel_{135°}$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ -1 & -4 & 0 & 4 & 1 \\ 0 & -2 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(h) $Sobel_{157.5°}$

图 3　八方向 Sobel 算子模板

将 8 个方向的边缘图像进行合并，得到该通道上的边缘图像；然后将 3 个通道的边缘图像求和，得到较为完整的边缘图像，其组合方式如下：

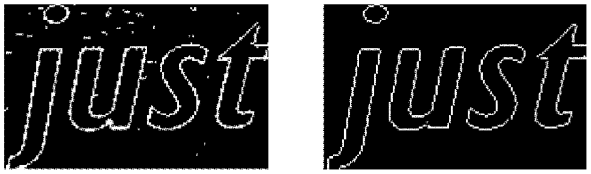
$$I_i=\sqrt{\frac{1}{8}\sum_{\theta=0}^{7\pi/8} I_{\theta Ci}^2} \tag{1}$$

$$I=\frac{1}{3}(I_R+I_G+I_B) \tag{2}$$

其中，$I_{\theta Ci}$ 表示 θ 方向在 i 通道上的边缘图像，此处 i=R，G，B，且 θ=0，π/8，π/4，3π/8，π/2，5π/8，3π/4，7π/8，I_i 为各方向在 i 通道上的平均边缘图像。

图 4 为传统 Sobel 算子与八方向 Sobel 算子边缘检测对

比图。可以看出，与传统 Sobel 算子边缘图像相比，八方向 Sobel 算子边缘图像中文字部分的边缘更加清晰、规整，且滤除了一部分不符合笔画特征的背景噪声，更利于后续的文本标注过程。



(a)传统 Sobel 算子边缘图像 (b)八方向 Sobel 算子边缘图像

图 4 两种 Sobel 算子检测图像的结果对比

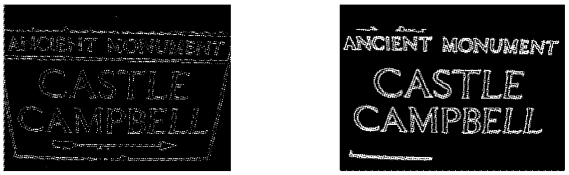
2.2 备选文本域的提取

采用八邻域联通法对边缘检测部分得到的平均边缘图像进行连通域处理，将连通区域标注成连通块；随后运用启发式规则对连通块进行备选文本域的提取。本文采用区域形状大小、长宽比与区域占有率 3 种启发式规则来过滤非文本域。区域形状大小包含连通块的面积、宽度和长度；长宽比 r 为连通区域的宽度 w 与长度 l 之比；占有率 p 为连通块面积 a_1 与连通块最小外接矩形面积 a_2 之比。经多次实验，我们得到区分文本域与非文本域的最佳阈值，将满足式(3)和式(4)的区域视为文本域，将超出阈值范围的区域视为非文本域，从而将其剔除。

$$0.125 \leqslant r = \frac{w}{l} \leqslant 8 \tag{3}$$

$$0.1 < p = \frac{a_1}{a_2} < 0.9 \tag{4}$$

图 5 分别给出了启发式规则过滤前与后的图像。对比图 5(a)、图 5(b)可以看出，经过启发式规则的过滤，图中一些明显不符合文本特征的噪声像素(如路牌的边框、箭头)被剔除，文本区域得以保留。



(a)启发式规则过滤前的图像 (b)启发式规则过滤后的图像

图 5 启发式规则过滤前、后的图像

在此基础上，对图像进行形态学处理，将文本区域聚集成文本行：首先对图像进行形态学膨胀操作，然后对其进行开运算使边缘轮廓平滑并过滤噪声。图 6 示出了备选文本域标记结果。

图 6 备选文本域标记结果

3 自适应 HOG-LBP 特征融合

方向梯度直方图(Histogram of Oriented Gradient, HOG)最早是由法国研究人员 Dalal 于 2005 年提出的用于数字图像处理的一种特征描述算子[16]，具体是图像局部像素区域的梯度方向直方图。LBP 特征是一种用来描述图像局部纹理特征的算子[17]，它首先由 T. Ojala 等[18]于 1994 年提出，用于提取图像局部纹理特征。文本由于由笔画组成，相较于非文本具有其独特的 HOG 信息，因此可通过提取图像中的 HOG 值来进行文本定位，但 HOG 特征对噪声边缘敏感度高，可利用 LBP 有效地过滤噪声边缘。

传统的 HOG-LBP 特征融合方法[19]分别提取这两种特征并直接进行堆叠，其特征维数是 HOG 和 LBP 特征维数相加之和，这种方法增加了信息的复杂度，加大了计算量。为了在保持检测性能的前提下尽可能降低特征维数，本文采用一种自适应权值的 HOG-LBP 特征融合方法，首先将备选文本域标准化为 128 * 96，然后将其分成 4 * 4(即 16)个小块，如图 7 所示。

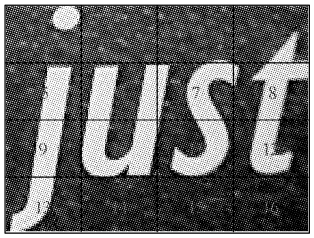


图 7 分块示意图

分别计算每一个小块的梯度值，然后根据该分块内的梯度直方图确定该分块中 HOG 与 LBP 的权重系数。其融合特征值的计算公式如下：

$$F_w(i) = (1-\alpha(i)) * F_{LBP}(i) + \alpha(i) * F_{HOG}(i) \tag{5}$$

其中，$F_w(i)$为第 i 块的加权融合特征值；$F_{HOG}(i)$，$F_{LBP}(i)$分别为第 i 块的 HOG 值和 LBP 值；$\alpha(i)$为加权系数，其计算公式如下：

$$\alpha(i) = \frac{gradient(i)}{\max\{gradient(1), \cdots, gradient(16)\}} \tag{6}$$

其中，$gradient(i)$为第 i 个分块内的梯度值。自适应特征融合后的特征维数为 178，而传统的 HOG-LBP 特征融合的特征维数为 347(HOG 特征 288 维，LBP 特征 59 维)。采用该特征提取方法可以有效地减少特征维数，缩减计算量。另外，在 HOG 值较大的分块内增加 LBP 值的权重可有效滤除备选文本域内的背景噪声，从而提高文本定位的精确度。

4 仿真实验结果

仿真实验平台为 Matlab R2012b，计算机配置为 64 位 Windows 系统，处理器 Inter(R) Core(TM) i3-2100 CPU，主频 3.10GHz，内存 4.00GB。训练集采用 ICDAR 数据库中的训练样本库，其中正例样本(文本图片)419 张，负例样本(背景图片)100 张。如图 8 所示，正例样本为含有文本的样本，而负例样本为不含文本的背景样本，正、负样本都被归一化为同样的尺寸大小。利用 SVM 分别对正样本图像和负样本图像进行训练，提取图像中 178 维自适应融合特征，将正样本图像标定为 1，存入正样本数据文件中；将负样本标定为 −1，存入负样本数据文件中。将正、负样本数据载入学习程序中用

于训练 SVM，生成学习模板。本文测试集采用 ICDAR 数据库中的测试样本库，含有 419 张样本，载入生成的学习模板进行图像的测试。

(a)正例样本

(b)负例样本

图 8　训练集的正、负样本

采用 ICDAR 文本定位竞赛算法评价标准[20]对实验结果进行评估，将准确率 P(Precision Rate)、召回率 R(Recall Rate)以及综合评价值 f 作为评价标准。其计算方法如下：

$$P=\frac{\sum_{r_E\in E} m(r_E,T)}{|E|},R=\frac{\sum_{r_T\in T} m(r_T,E)}{|T|} \tag{7}$$

$$m(r,R)=\max(m_p(r,r')|r'\in R) \tag{8}$$

$$m_p(r_1,r_2)=\frac{a(r_1)\bigcap a(r_2)}{a(r_1)\bigcup a(r_2)} \tag{9}$$

其中，$a(r)$为面积，E 为检测到的文本域集合，T 为实际文本域集合，$|R|$为集合内元素的个数。如图 9 所示，阴影部分为 $a(r_1)\bigcap a(r_2)$，虚线部分为 $a(r_1)\bigcup a(r_2)$，通过面积匹配得到的部分为 $m_p(r_1,r_2)$。矩形框(r_1,r_2)的匹配面积为阴影部分面积除以虚线部分面积。若两个矩形框完全重合，则 $m_p(r_1,r_2)$为 1；若两个矩形框不相交，则 $m_p(r_1,r_2)$为 0。针对估计集的每一个矩形框，在目标集中找到一个与其匹配面积最大的矩形框，得到 $m(r,R)$。

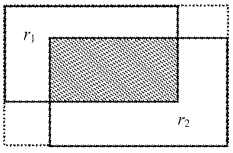


图 9　区域匹配图

综合评价值 f 的计算方法如下：

$$f=\frac{1}{\frac{a}{P}+\frac{1-a}{R}} \tag{10}$$

其中，a 代表 P 与 R 两者相关的权重，一般取 $a=0.5$。

图 10 示出了本算法仿真实验结果与 ICDAR 文本定位竞赛中其他参赛算法[20-21]和传统 HOG-LBP 融合算法[19]的比较结果。从比较结果可以看出，本算法在准确率、召回率和综合评价值 3 个指标上均有提高。

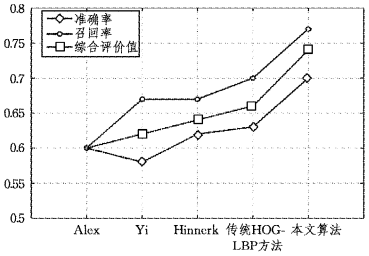


图 10　算法评价值比较结果

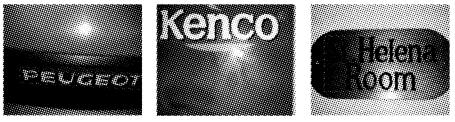
相较于传统边缘检测算法，本文算法采用的多通道多方向边缘检测在运算量和复杂度上势必有一定程度的增加，但在随后的特征融合过程中，文本算法采用的是自适应权值的特征融合算法，其相较于传统的特征融合算法减少了特征维数的计算量，从而解决了边缘检测运算量增大的问题。为了验证这个问题，将本文算法与采用了传统边缘检测和特征融合的算法进行了对比实验，分别测量两种方法在边缘检测环节的平均用时和总过程平均用时，实验结果如表 1 所列。在边缘检测环节，相比传统算法，本文算法处理每幅图像的平均用时增加了 0.041s。但由于特征融合过程降低了数据维度，减小了后续计算复杂度，在文本定位总用时上，本文算法比传统算法减少了 0.7s。由此表明，本文所提出的文本定位方法在提高准确性的同时也提高了定位效率。

表 1　两种方法的平均用时比较/s

算法	边缘检测用时	总用时
传统算法	0.033	2.1
本文算法	0.072	1.4

图 11 示出了测试中的正确检测结果。前 3 幅为复杂背景图像的文本定位，后 3 幅为光照不均图像的文本定位。由图 11 可以看出，本算法对复杂背景图像、光照不均的图像均能准确地进行定位。

(a)复杂背景图像



(b)光照不均图像

图 11　正确检测结果

图 12 示出了测试中的错误检测结果，在文本与背景具有较低的对比度或边缘信息过于复杂的情况下，本文算法不能有效地进行文本定位。

(a)文字与背景对比度较低

(b)边缘信息过于复杂

图 12　错误检测结果

结束语　本文提出一种基于多方向边缘检测和自适应特

(下转第 314 页)

[28] NESTEROV Y. Gradient Methods for Minimizing Composite Objective Function [R]. Universite Catholique de Louvain, 2007.

[29] BECKER S, BOBIN S J, CANDES E J. NESTA: A Fast and Accurate First-Order Method for Sparse Recovery [J]. SIAM Journal on Imaging Sciences, 2011, 4(1): 1-39.

[30] XIAO C, SALERNO M, YANG Y, et al. Motion-Compensated Compressed Sensing for Dynamic Contrast-Enhanced MRI Using Regional Spatiotemporal Sparsity and Region Tracking: Block LOw-rank Sparsity with Motion-guidance (BLOSM) [J]. Magnetic Resonance in Medicine, 2014, 72(4): 1028-1038.

(上接第 303 页)

征融合的自然场景文本定位方法。该方法采用三通道八方向的边缘检测方法提取边缘图像的 HOG 和 LBP 特征，并进行自适应权值的特征融合，使其在保持检测性能的基础上降低了特征维数，解决了多通道边缘检测环节所带来的计算量增加的问题。本文算法在 ICDAR 数据集上进行了训练及测试，实验结果表明，该方法的准确率为 70%，召回率为 77%，综合评价值为 74%。相较于其他竞赛算法和传统的 HOG-LBP 特征融合算法，该方法在定位效果上有较大提升，且有效地解决了自然场景下文本检测受光照、背景复杂度因素影响较大的问题。但是本文算法对文本与背景对比度较低和边缘信息过于复杂的情况的鲁棒性较低，针对该问题可做进一步研究。

参 考 文 献

[1] YE Q X, DOERMANN D. Text Detection and Recognition in Imagery: A Survey[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(7): 1480-1500.

[2] YANG H J, QUEHL B, SACK H. Text detection in video images using adaptive edge detection and Stroke Width verification [C]// 19th International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, 2012: 9-12.

[3] RAJESHABABA M, ANITHA T. Detect and separate localization text in various complicated-colour image[C]// International Conference on Circuits, Power and Computing Technologies. IEEE, 2013: 866-872.

[4] MORADI M, MOZAFFARI S. Hybrid approach for Farsi/Arabic text detection and localisation in video frames[J]. Iet Image Processing, 2013, 7(2): 154-164.

[5] YI C C, TIAN Y L. Text string detection from natural scenes by structure-based partition and grouping[J]. IEEE Transactions on Image Processing, 2011, 20(9): 2594-2605.

[6] FENG Y Y, SONG Y H, ZHANG Y L. Scene text localization using extremal regions and Corner-HOG feature[C]// IEEE International Conference on Robotics and Biomimetics. IEEE, 2015: 881-886.

[7] BHARDWAJ D, PANKAJAKSHAN V. Image Overlay Text Detection Based on JPEG Truncation Error Analysis[J]. IEEE Signal Processing Letters, 2016, 23(8): 1027-1031.

[8] LYU M R, SONG J C, CAI M. A comprehensive method for multilingual video text detection, localization, and extraction[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2): 243-255.

[9] SRIVASTAV A, KUMAR J. Text detection in scene images using stroke width and nearest-neighbor constraints[C]// 2008 IEEE Region 10 Conference. IEEE, 2008: 1-5.

[10] YE J, HUANG L L, HAO X L. Neural network based text detection in videos using local binary patterns[C]// Chinese Conference on Pattern Recognition(CCPR). IEEE, 2009: 1-5.

[11] MAO W G, CHUNG F L, LAM K K M, et al. Hybrid Chinese/English text detection in images and video frames[C]// 16th International Conference on Pattern Recognition. IEEE, 2002, 16(3): 1015-1018.

[12] ZINI L, DESTRERO A, ODONE F. A classification architecture based on connected components for text detection in unconstrained environments[C]// 6th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2009: 176-181.

[13] SU F, XU H L. Robust seed-based stroke width transform for text detection in natural images[C]// 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015: 916-920.

[14] KUTTY S B, SAAIDIN S, YUNUS P N A M, et al. Evaluation of canny and sobel operator for logo edge detection[C]// International Symposium on Technology Management and Emerging Technologies. IEEE, 2014: 153-156.

[15] KAUR B, GARG A. Mathematical morphological edge detection for remote sensing images[C]// 3rd International Conference on Electronics Computer Technology (ICECT). IEEE, 2011: 324-327.

[16] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2005: 886-893.

[17] HUANG F F. Research on face recognition based on LBP operator[D]. Chongqing: Chongqing University, 2009. (in Chinese)
黄非非. 基于 LBP 的人脸识别研究[D]. 重庆：重庆大学，2009.

[18] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[19] LIU Y Y, YU F Q, CHEN Y. Text location in image based on connected-component and statistical features[J]. Computer Engineering and Applications, 2016, 52(5): 165-68. (in Chinese)
刘亚亚，于凤芹，陈莹. 基于连通区域和统计特征的图像文本定位[J]. 计算机工程与应用，2016，52(5)：165-168 .

[20] LUCAS S M. ICDAR 2005 text locating competition results[C]// 8th International Conference on Document Analysis and Recognition(ICDAR'05). IEEE, 2005: 80-84.

[21] YI C C, TIAN Y L. Text detection in natural scene images by stroke gabor words[C]// International Conference on Document Analysis and Recognition(ICDAR). IEEE, 2011: 177-181.