

智能车道路场景数字字符识别技术

白睿, 徐友春, 李永乐*, 李炯, 谢枫

陆军军事交通学院, 天津 300161

摘要 针对道路场景中数字字符高噪声、多视角和难以定位识别的问题, 提出了一种稳健的道路场景数字字符定位识别算法。采用基于色彩空间和边缘增强的最大稳定极值区域(MSER)算法来提取候选区域, 设计了几何约束滤波器, 并与笔画宽度变换(SWT)联合滤除非字符区域, 得到字符定位结果。对 Lenet-5 中的收敛函数和池化窗进行改进, 将定位后的字符区域归一化输入网络中, 得到最终的字符识别结果。实验结果表明, 本文算法的字符召回率达到 90.0%, 综合性能值达到 0.89, 字符识别率达到 88.6%, 优于同类算法性能。

关键词 机器视觉; 数字字符识别; 色彩空间; 最大稳定极值区域; 笔画宽度变换; 卷积神经网络

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.191506

Digital Character Recognition Technique for Intelligent Vehicles in Road Scenes

Bai Rui, Xu Youchun, Li Yongle*, Li Jiong, Xie Feng

Army Military Transportation University, Tianjin 300161, China

Abstract To address the problems of large noise, multi-view, and difficult to locate and identify digital characters in road scenes, a robust method for recognizing digital characters in road scenes is proposed. According to this method, the maximally stable extremum region algorithm based on the color space and enhanced edge is used first to obtain candidate regions. Then, a geometrically constrained filter is designed and combined with the stroke width transform to filter non-character regions. The convergence function and pooling window of Lenet-5 are improved, and the localized character regions are normalized and input into the network to obtain the final recognition results. According to the experimental results, the recall rate of the proposed method is 90.0%, the comprehensive performance value is 0.89, and the character recognition rate is 88.6%. These results are higher than those of the existing algorithms.

Key words machine vision; digital character recognition; color space; maximally stable extremum region; stroke width transformation; convolution neural network

OCIS codes 150.1135; 150.5758; 150.1708

1 引言

道路环境中的数字字符可以为智能车提供限速、限高、停车位置编号等重要信息, 是智能车环境感知的重要部分, 例如现在非常成熟的车牌识别系统^[1-3]。在自然环境下, 特别是在智能车行驶过程中, 字符往往包含非常复杂的噪声, 识别过程经常会受到光照变化、字符粘连、对比度低、部分遮挡、字体以及视角多样化等因素的影响^[4-5]。

目前场景字符定位的方法大致可归纳为 3 种^[6]: 基于滑动窗口(sliding window)的方法、基于深度学习的方法和基于连通域分析(CCA)的方法。

滑动窗口法主要是通过设计多尺度的滑动窗口以及对应的分类器来实现字符定位。例如文献[7-9]设计了不同的滑动窗口并结合分类器得到字符区域, 但需要人为设定滑动窗口的尺度和步长, 且难以解决字符倾斜程度较大时的定位问题。

深度学习的方法可以达到较为理想的定位效

收稿日期: 2019-03-15; 修回日期: 2019-04-22; 录用日期: 2019-06-05

基金项目: 国家自然科学基金(91220301)、国家重点基础研究计划(2016YFB0100903)

* E-mail: Rui_bai_berry@126.com

果,而且对环境适应性较为稳健。方定邦等^[10]、马永杰等^[11]利用卷积神经网络(CNN)实现了对字符的定位与识别,但是对图形处理器(GPU)等计算资源占用率较高,同时难以满足智能车对实时性的要求,因此该方法实用性不强。

目前,使用较为广泛的连通域分析方法是最大稳定极值区域(MSER)和笔画宽度变换(SWT)^[12]。MSER具有稳定性和仿射不变性的特征,对不同字体字符、视角字符较为稳健^[13]。但MSER对低对比度图像的定位效果较差,而且它只利用了图像的灰度信息,忽略了图像的颜色信息^[14]。Mosleh等^[15]利用Bandlet改进SWT进行字符定位,但算法误检率较高。同时也有不少研究人员将两种方法结合在一起,取得了更好的效果。张国和等^[16]将MSER与SWT结合进行字符定位,但字符漏检率较高,对个别场景具有局限性。

对于字符识别方法,研究人员主要通过Gabor变换、局部二值模式(LBP)和梯度方向直方图^[17-18](HOG)等人工特征提取方法进行提取,然后将提取特征输入到支持向量机(SVM)、后向传输(BP)神经

网络等分类器中进行分类识别。K近邻(KNN)算法在MINIST数据集中达到95.4%的识别率^[19],但是对于受强噪声干扰的数字字符,其识别率并不理想。Lenet-5卷积神经网络^[20](CNN)在数字字符识别领域表现优异,它避免了复杂的人工特征提取过程,识别过程高效,但存在收敛速度慢、噪声适应性弱等问题。

本文提出基于色彩空间与边缘增强的MSER来实现字符区域提取,并对Lenet-5中的收敛函数和池化窗进行改进,克服原有算法存在的问题,以提高对道路环境数字字符的定位识别效果。

2 道路场景数字字符定位

采用基于色彩空间与边缘增强的MSER提取图像的字符候选区域,设计了适用于道路场景字符的几何约束滤波器,初步排除非字符连通域,通过SWT进一步剔除非字符区域,得到最终的字符区域。将定位后的字符区域归一化输入到改进的Lenet-5中得到字符识别结果,最后通过实验验证了算法的性能。本文算法流程如图1所示。

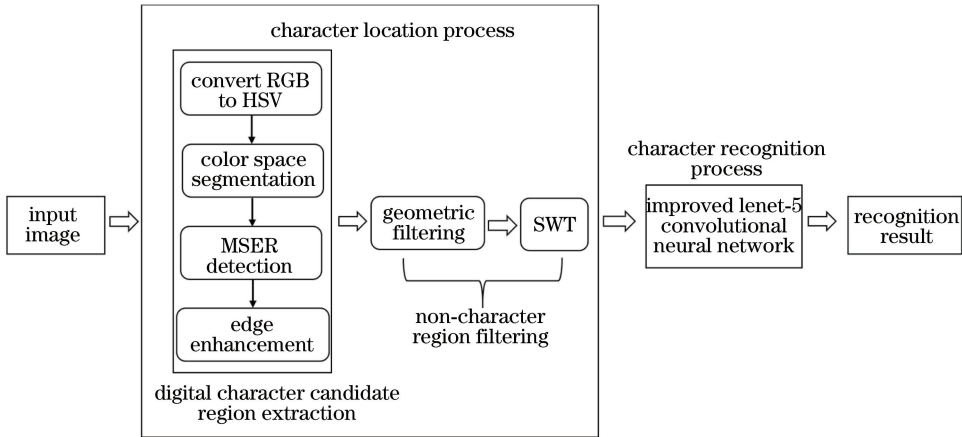


图1 本文算法流程

Fig. 1 Flowchart of proposed algorithm

2.1 基于色彩空间与边缘增强的MSER字符候选区域提取

Neumann等^[21]将最大稳定极值区域用于字符区域提取。MSER的原理是若灰度图像 I 满足映射关系 $D \subset Z^2 \rightarrow S$,其中 S 为全序结构, D 为连通分量, Z 为连通分量的一对像素点。定义邻域关系 $A \subset D \times D$, Q 是 D 的一个连续的子区域,令 $Q_1, \dots, Q_{i-1}, Q_i, \dots$ 为嵌套的极值区域序列,在动态二值化过程中,区域变化范围为 $(Q_{i-\Delta}, Q_{i+\Delta})$,其中 Δ 为阈值变化量。设 $q(i)$ 为连通域变化率, i 为二值化的阈值,存在 Q_i 满足

$$q(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{Q_i}, i \in [0, 255], \quad (1)$$

当 $q(i)$ 取得最小值时,连通区域 Q_i 为MSER。

该方法主要思想是根据字符像素点在纹理上的相似性或者空间上的临近性,排除背景像素点,从而得到字符候选区域,具体做法是依次将 $0 \sim 255$ 作为阈值对灰度图像进行二值化,将二值化过程中变化很小的区域作为MSER输出。

针对MSER对模糊或者低对比度区域的提取能力较差的问题,在原有算法的基础上,进行以下两点改进:1)由灰度图像改为在色彩空间对极值区域

进行提取,即在六角锥体模型(HSV 空间)内对 H、S、V 三个通道进行 MSER 提取,HSV 空间色彩与亮度分离,且更适应道路场景的光照变化,如图 2 所示。在字符对比度低的情况下,传统方法中对灰度图像进行 MSER 提取无法得到对应的字符区域,而对 S 通道进行 MSER 提取则可得到对应的字符区域。大量实验结果表明,通过 MSER 对色彩空间的利用,有效避免了 MSER 漏检的现象,提高了字符的召回率;2)利用 Canny 算子^[22]得到原始图像对应

的边缘图像,同时剔除了 Canny 边缘外的 MSER 像素点。利用 Sobel 算子^[23]得到边缘的梯度方向,根据梯度方向对边缘像素点进行扩展生长,目的是进一步增强字符轮廓,突出字符候选区域,便于进行后续筛选过程,上述过程如图 3 所示。

通过对色彩空间的提取与边缘信息的增强,在字符与背景对比度不高的情况下也能有效提取字符候选区域,避免了漏检的现象,在字符对比度不高的情况下,仍然能取得较好的检测效果。



图 2 本文提取方法与传统 MSER 提取方法对比。(a)某停车位数字编号原始图;(b)传统 MSER 提取效果图;(c)对 S 通道进行 MSER 提取的效果图

Fig. 2 Comparison between proposed extraction method and traditional MSER extraction method. (a) Original diagram of a parking space number; (b) traditional MSER extraction effect; (c) MSER extraction effect for S channel



图 3 字符边缘增强过程。(a)字符候选区域;(b)字符候选域的边缘提取图;(c)字符的边缘增强图

Fig. 3 Character edge enhancement process. (a) Character candidate regions; (b) edge extraction graph of character candidate regions; (c) edge enhancement graph of characters

2.2 几何滤波

基于数字字符连通域属性,通过大量实验,设计了适用于道路场景数字字符的几何约束滤波器,主要包括连通域面积 (Area)、连通域离心率 (Eccentricity)、连通域与其凸包面积之比 (Solidity)、宽高比 (Ratio) 4 个部分,如表 1 所示。通过几何滤波,可以剔除明显非字符区域,例如一些非常小的噪声点或者非常大的区域,包含大量孔洞的非字符连通域等,同时在宽高比中选择了较为保守的阈值,避免字符“1”被误判。

表 1 几何约束滤波器参数

Table 1 Geometric constraint filter parameters

Parameter	Area	Eccentricity	Solidity	Ratio
Threshold	[75,600]	[0.1,0.995]	[0,0.4]	[0.3,7]

2.3 基于笔画宽度变换的字符区域定位

通过几何滤波滤除部分明显非字符区域之后,需进一步滤除非字符区域。Epshtein 等^[24]充分利用字符的笔画特征实现了字符提取,而且能够实现对自然场景中字符的提取,取得了良好的效果。其原理是利用 Canny 算子对图像进行边缘提取,利用 Sobel 算子计算其边缘的梯度方向,得到图像的梯度图。基于笔画边缘像素点 P 及其梯度方向 R ,找到与其梯度方向相反的另一笔画边缘像素点 p ,计算两个笔画边缘点的欧氏距离 L ,即该边缘点的笔画宽度值 S_w 。通过求取字符连通域上所有边缘点的笔画宽度值完成该字符的笔画宽度变换。图 4 为某场景中字符“6”的边缘提取图与笔画宽度图。

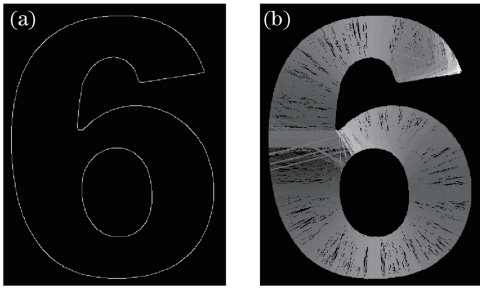


图4 字符“6”的边缘提取图与笔画宽度图。

(a)字符“6”的边缘提取图;(b)字符“6”的笔画宽度图

Fig. 4 Edge extraction and stroke width maps of character “6”. (a) Edge extraction of character “6”; (b) stroke width of character “6”

笔画宽度变换剔除非字符区域的思想是通过计算每个连通域边缘像素点的笔画宽度值,利用笔画连通图启发式规则^[22]滤除非字符连通域,最终达到字符定位的目的。

本研究通过实验得到的启发式规则为:1)当前像素点与邻域点的笔画宽度值之比 K 满足 $K < 5$; 2)字符连通域笔画宽度值的方差 $\epsilon < \overline{S_{w_i}}$, 其中 $\overline{S_{w_i}}$ 为连通域笔画宽度值的平均值。通过自建数据集中不同尺度字符的实验得到了参数 K 的阈值, 阈值范围涵盖了不同尺度的数字字符。

直接利用 SWT 对原始图像进行处理容易受到复杂背景的干扰,因此只将算法用于经过几何滤波后的字符候选域上,以提高算法效率。

利用笔画宽度变换得到了最终的数字字符的闭合连通域,通过 OpenCV 库函数可以进一步求取字符闭合连通域的最小外接矩形,进而得到图像中每个字符的位置坐标。若将图像左上角点作为图像坐标系的坐标原点,那么字符位置坐标可以用该字符连通域最小外接矩形的左上角坐标和右下角坐标来表示,即 (X_{i1}, Y_{i1}) 与 (X_{i2}, Y_{i2}) , 根据字符定位结果的纵横坐标值,也可以进一步确定字符的语义先后顺序。

3 基于改进 Lenet-5 数字字符识别

3.1 网络结构

Lenet-5、AlexNet、VGGNet 等网络均在目标检测与识别中取得了很好的效果,但字符特征与其他目标特征相比较为简单,无需过于繁琐的特征提取过程。同时为满足算法的实时性,采用 Lenet-5 卷积神经网络的基本结构,并在其基础上作了改进,使其满足实际的应用需求。主要从以下 2 个方面进行改进:1)将激活函数 sigmoid 改为 ReLu (Rectified

Linear Units)函数,减少了训练时间,提高了模型的收敛速度和准确率;2)将原有的 2 个平均池化窗改为 2 个最大池化窗,以突出字符的局部特征,有效减少噪声干扰。

Lenet-5 原有输入尺度为 $32 \text{ pixel} \times 32 \text{ pixel}$,通过实验测试了在 MINIST 数据集中不同输入尺度下的识别率,如表 2 所示。由表 2 可知,对原有输入尺度不同程度的改变都会导致识别准确率的下降,因此本文保持原有输入尺度不变。

表 2 不同输入尺度下识别率下降值的实验结果

Table 2 Experimental results of recognition rate drop-out values under different input scales

Input scale / (pixel×pixel)	18×18	24×24	28×28	36×36	48×48
Drop value /%	8.3	1.8	0.6	2.4	2.9

网络结构如图 5 所示,共包含 7 层(不含输入层)。从功能上主要包括 3 层:卷积层、池化层、全连接层。其中卷积层(conv)用于特征提取;池化层(pool)通过抽样降低训练参数;全连接层(FC)类似于多层感知器,用于模式识别;C 层是由卷积神经元组成的网络层;S 层是由次抽样神经元组成的网络层。

卷积层的作用是增强输入图像的特征,并降低噪声,提取输入图像的不同特征。主要通过不同种类的可训练的卷积核来对图像进行卷积运算,将上一层节点的值乘以对应的连接参数,之后加上偏置项输入到激活函数,得到卷积层节点的值,并通过卷积核在输入图像上的移动得到特征图的值。卷积层的实现为

$$x = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), \quad (2)$$

式中: x_i^{l-1} 为输入特征图; x 为输出特征图; k_{ij}^l 为卷积核; b_j^l 为偏置项; f 为激活函数,即 ReLu 激活函数。

池化层基于图像局部相关性的原理,利用池化窗对图像进行子抽样,减少数据处理量的同时保留有用信息,降低网络训练参数及模型的过拟合程度。沿用目标网络中 $2 \text{ pixel} \times 2 \text{ pixel}$ 的池化窗,即将上一层的 4 个输入节点作为下一层 1 个节点的输入,且输入域不重叠,即每次滑动 2 pixel。不同的是,对每个下抽样节点的 4 个输入节点进行最大值池化,之后将得到的结果乘以权重加上一个偏置参数作为激活函数的输入,激活函数的输出即为下一层节点的值,池化层实现的一般形式为

$$x_j^{(l)} = f[\beta_j^{(l)} \text{down}(x_j^{(l-1)}) + b_j^{(l)}], \quad (3)$$

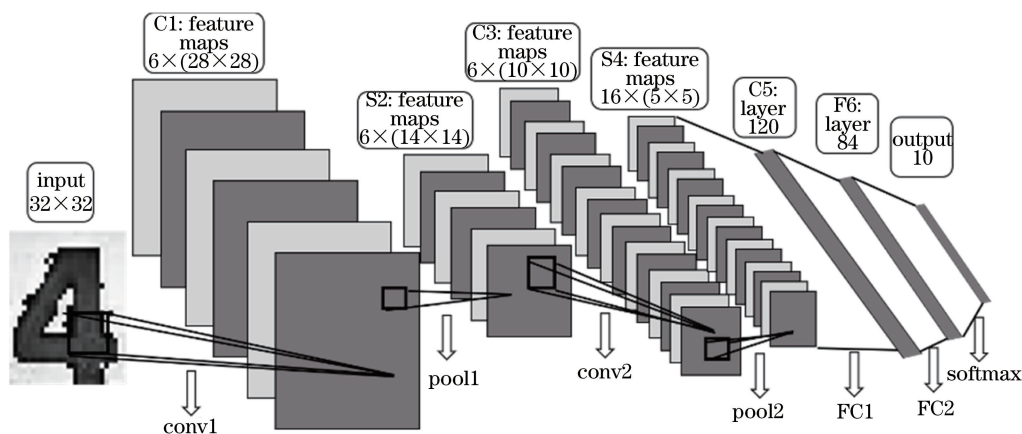


图 5 Lenet-5 网络示意图

Fig. 5 Diagram of the Lenet-5 network

式中:down()表示池化函数,这里指的是最大值池化; $x_j^{(l-1)}$ 和 $x_j^{(l)}$ 分别为由第 l 个池化窗抽样形成的第 j 个特征图; $\beta_j^{(l)}$ 和 $b_j^{(l)}$ 分别为特征图对应的乘性偏置和加性偏置。

全连接层整体结构可以看成是一个多层感知器,神经元采用全连接的方式,输出层的单元采用了径向欧氏距离函数(ERBF),计算输入向量和参数向量的欧氏距离,实现的一般形式为

$$y_n = \sum_m (x_m - w_{mn})^2 \quad (4)$$

式中: x_m 是上一节点的输入; y_n 是函数输出的第 n 个特征向量; w_{mn} 为全连接的权值; m 为前向神经元的编号。

3.2 数据集的建立

利用自建数据集对网络进行训练与测试,将SVHN、CTW data、ICDAR2013、Chars74K等公开数据集以及车载相机采集的部分道路场景中的数字样本进行人工粗分割得到单个数字字符,并将其归一化为 $32 \text{ pixel} \times 32 \text{ pixel}$ 的灰度图像,生成数字字符数据集,最终获得了18006个数字字符。为使训练得到的模型适应实际的道路场景,对数据集不作任何处理,使得到的数据集包含不同倾斜程度以及不同噪声程度的字符,以提高训练模型的泛化能力。图6为数据集中字符的示意图,表3为各类字符的数量,对于个别数量较少的字符,额外增加了字符的数量,使得用于训练的10类数字字符数量较为均衡。

表 3 数据集中字符数量

Table 3 Number of characters in the dataset

Character	0	1	2	3	4	5	6	7	8	9
Number	1811	1786	1837	1826	1778	1839	1745	1805	1798	1781



图 6 数据集中字符示意图。(a)数据集分割前示意图;(b)数据集粗分割后示意图

Fig. 6 Diagrams of characters in the dataset.
(a) Diagram before dataset segmentation;
(b) diagram after rough dataset segmentation

3.3 网络训练

从数据集每个类别的字符中随机抽取1500个样本组成训练集,同时,在每个字符剩余样本中随机选取

200个样本作为该字符的测试集,并将训练集与测试集的数字字符转化为caffe框架所支持的lmdb格式,同时为训练集添加标签,输入到上述的Lenet-5网络中进行训练。训练过程中train loss、test accuracy与迭代次数的关系曲线如图7所示,训练损失不断下降,测试样本损失不断减小,Lenet-5网络处于不断学习状态,经过约700次迭代后,训练样本损失趋于0,最终测试样本的准确率达到95.8%。

4 实验结果与分析

在实验中,算法基于Ubuntu 16.04,结合Caffe-Linux框架和OpenCV库函数实现,主要在Intel Core i5 6700 2.30GHz CPU上测试运行,在NVIDIA GTX960 GPU上进行网络训练,通过车载

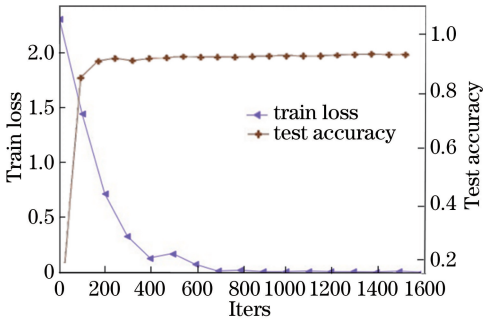


图7 训练损失和测试精度随迭代次数变化的示意图
Fig. 7 Train loss and test accuracy vary with number of iterations

相机采集道路场景图像,如图8所示。实验主要通过与相关文献算法进行对比分析以及公开的评价方法来原因算法的综合性能。

4.1 数字字符定位实验与分析

将本文所述字符定位方法以及相关文献中的多种字符定位方法进行对比实验。从公开数据集以及



图8 实验平台
Fig. 8 Experimental platform

实车采集的道路场景图片中随机选择200张图片作为测试集,其中包括限速牌、限高牌、停车位等多个道路场景在内的482个数字字符。为避免初始输入图像大小对数字字符定位性能的影响,将初始输入图像大小归一化为600 pixel×400 pixel。图9为利用本文方法得到的数字字符定位结果示意图,每幅道路场景图片均给出了经过滤除非字符区域后得到的数字字符连通域,如图9(a)所示中的二值图,最终的定位结果以字符连通域的最小外接矩形的形式给出。

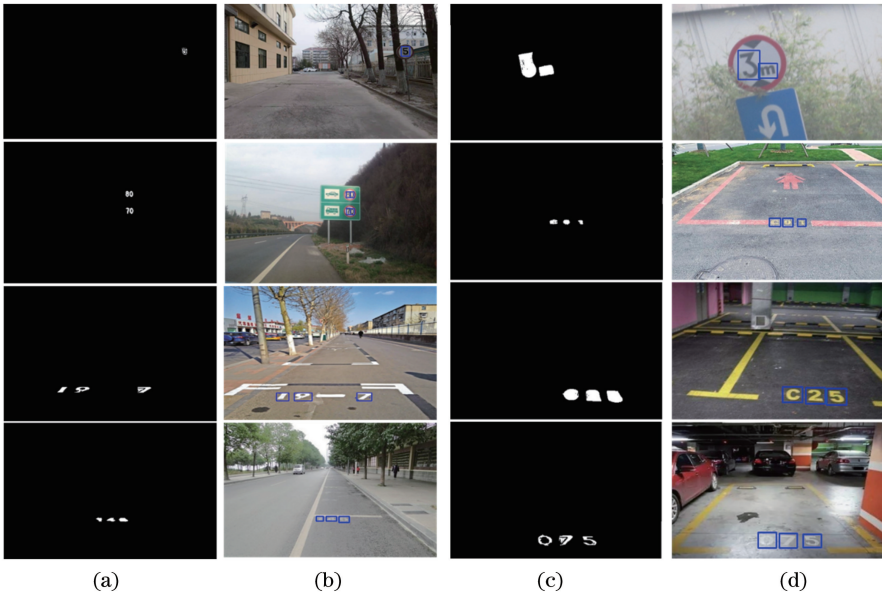


图9 字符定位结果示意图。(a)字符连通域;(b)道路环境数字字符定位结果;
(c)字符连通域;(d)道路环境数字字符定位结果

Fig. 9 Diagrams of character positioning results. (a) Character connected regions; (b) character location results in road scenes; (c) character connected regions; (d) character location results in road scenes

采用文档分析与识别国际会议(ICDAR)^[25]中对文本定位的评估方法,利用准确率 H 、召回率 R 、 F 值三个指标对字符定位方法进行对比和评价,其指标含义为

$$H = \frac{C}{C + E}, \quad (5)$$

$$R = \frac{C}{C + T}, \quad (6)$$

$$F = \frac{1}{\alpha/A + (1 - \alpha)/R}, \quad (7)$$

式中: C 为图像中检测到的字符区域的个数; E 为检测到的非字符区域的个数; T 为未检测到的字符区域的个数; α 为准确率 H 与召回率 R 之间的权重,本文选取 α 值为0.5。

根据上述三个性能指标对不同定位方法进行对比分析,结果如表4所示。

表4 不同算法字符定位性能对比

Table 4 Character location performance comparison of different algorithms

Method	<i>C</i>	<i>E</i>	<i>T</i>	<i>H</i>	<i>R</i>	<i>F</i>
Neumann ^[21]	399	70	83	0.85	0.82	0.83
Epshtein ^[24]	398	66	84	0.86	0.82	0.84
Lee ^[7]	404	63	78	0.86	0.83	0.84
Zhang ^[16]	411	54	71	0.88	0.85	0.86
Chen ^[13]	417	76	65	0.84	0.86	0.85
Sung ^[5]	420	60	62	0.84	0.87	0.87
Huang ^[9]	422	51	60	0.89	0.89	0.88
Ours	438	50	44	0.89	0.90	0.89

将测试集输入到算法中进行测试,分别得到测试集的 *C*、*E* 和 *T* 值以及评价指标 *H*、*R* 和 *F* 的统计值。从表 4 中数据可以看到,本文所提算法在测试集取得了最高的准确率和召回率,达到了最优的 *F* 值。准确率和召回率均优于同类算法的原因:基于色彩空间和边缘增强的 MSER 将道路场景中低对比度的字符提取出来了,通过几何滤波以及基于实验启发式规则的 SWT 对非字符区域进行了有效剔除,同时也有效避免了字符误判。

从图 9 的实验结果也可以看出,本文所提算法能够对不同道路环境数字字符实现稳健的定位效果,例如一些倾斜程度较大,受光照影响强烈以及部分遮挡等比较复杂的字符场景,定位后的字符二值连通域也基本能反映原有字符形状和结构,有利于后续识别。但在个别复杂场景下,出现定位得到的字符二值连通域部分缺失甚至完全无法定位字符的现象,得不到正确的识别结果,所以对数字

字符连通域的准确定位是得到准确识别结果的前提和重要环节。

4.2 数字字符识别实验与分析

在实验过程中采用了不同的数字字符识别方法进行测试,将定位得到的字符最小外接矩形作为感兴趣区域 (ROI),并将其归一化为 $32 \text{ pixel} \times 32 \text{ pixel}$ 的图片,从中随机选取 300 个数字字符作为实验测试样本,其中每个数字字符测试样本量为 30。采用不同识别方法对测试样本进行测试,并与真实值进行比较,得到了测试样本在不同算法中的识别准确率,用 *G* 来表示,如表 5 所示,其中 *N* 为测试样本中识别成功的样本数量。

表5 不同算法字符识别效果对比

Table 5 Comparison of character recognition effects of different algorithms

Method	<i>N</i>	<i>G</i> / %
KNN	241	80.1
HOG+SVM	237	79.4
BP Neural Network	230	76.8
Ours	265	88.6

图 10 为将图 3 中字符‘1’、‘4’、‘6’输入到改进后 Lenet-5 的识别结果,其中横坐标 0~9 对应 10 个数字字符,纵坐标为网络输出的结果值,代表与每个字符的相似度。从图中可以看到,字符‘1’与字符‘4’作为输入得到了准确的识别结果,但将字符‘6’作为输入得到了错误的识别结果 0,主要原因是字符倾斜程度太大,对比度较低,造成在该视角下字符的部分边缘产生粘连,因而最终产生了错误的识别结果。

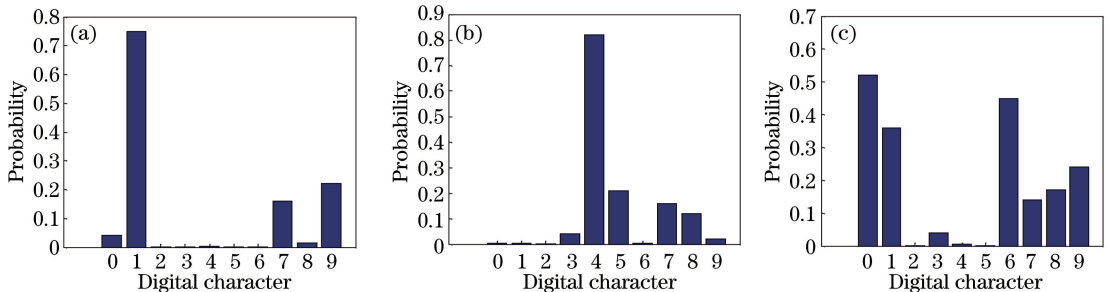


图 10 字符的识别结果。(a)字符‘1’的识别结果;(b)字符‘4’的识别结果;(c)字符‘6’的识别结果

Fig. 10 Character recognition results. (a) Recognition result of character ‘1’; (b) recognition result of character ‘4’; (c) recognition result of character ‘6’

5 结 论

针对智能车对复杂道路环境中数字字符识别,提出一种基于色彩空间和边缘增强的最大稳定极值区域的字符提取方法。设计了一种用于滤除非字符

区域的几何约束滤波器,并与笔画宽度变换联合滤除非字符区域,同时将字符定位与识别整合在一个系统中,实现了道路场景数字字符的同时定位与识别。通过实验验证了所提方法的先进性和有效性。本研究对汉字字符以及英文字符也有一定的定位效

果,但受限于字符数据量等因素未得到良好的识别效果,下一步主要针对某些光照强烈、字符对比度低的情况下开展研究,以提高算法的适应性。

参 考 文 献

- [1] Li L J, Socher R, Li F F. Towards total scene understanding: classification, annotation and segmentation in an automatic framework[C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 2036-2043.
- [2] Du S, Ibrahim M, Shehata M, *et al.* Automatic license plate recognition (ALPR): a state-of-the-art review [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(2): 311-325.
- [3] Wen Y, Lu Y, Yan J Q, *et al.* An algorithm for license plate recognition applied to intelligent transportation system [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(3): 830-845.
- [4] Lü P F, Wang S G. Intelligent target recognition method of unmanned aerial vehicle combat platform [J]. Laser & Optoelectronics Progress, 2019, 56(7): 071001.
吕攀飞, 王曙光. 无人机作战平台的智能目标识别方法[J]. 激光与光电子学进展, 2019, 56(7): 071001.
- [5] Sung M C, Jun B, Cho H, *et al.* Scene text detection with robust character candidate extraction method[C] // 2015 13th International Conference on Document Analysis and Recognition (ICDAR), August 23-26, 2015, Tunis, Tunisia. New York: IEEE, 2015: 15616211.
- [6] Dong Y B. Research on text location and character recognition of scene image[D]. Wuhan: Huazhong University of Science and Technology, 2016: 3-23.
董杨博. 场景图像文本定位与字符识别方法研究[D]. 武汉: 华中科技大学, 2016: 3-23.
- [7] Lee J J, Lee P H, Lee S W, *et al.* AdaBoost for text detection in natural scene [C] // 2011 International Conference on Document Analysis and Recognition, September 18-21, 2011, Beijing, China. New York: IEEE, 2011: 429-434.
- [8] Ye J, Huang L L, Hao X L. Neural network based text detection in videos using local binary patterns[C] // 2009 Chinese Conference on Pattern Recognition, November 4-6, 2009, Nanjing, China. New York: IEEE, 2009: 11001628.
- [9] Huang R, Shivakumara P, Feng Y K, *et al.* Scene character detection and recognition with cooperative multiple-hypothesis framework [J]. IEICE Transactions on Information and Systems, 2013, E96.D(10): 2235-2244.
- [10] Fang D B, Feng G, Cao H Y, *et al.* Handwritten formula symbol recognition based on multi-feature convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(7): 072001.
方定邦, 冯桂, 曹海燕, 等. 基于多特征卷积神经网络的手写公式符号识别[J]. 激光与光电子学进展, 2019, 56(7): 072001.
- [11] Ma Y J, Li X Y, Song X F. Traffic sign recognition based on improved deep convolution neural network [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121009.
马永杰, 李雪燕, 宋晓凤. 基于改进深度卷积神经网络的交通标志识别[J]. 激光与光电子学进展, 2018, 55(12): 121009.
- [12] Zhang X N, Gao X B, Tian C N. Text detection in natural scene images based on color prior guided MSER[J]. Neurocomputing, 2018, 307: 61-71.
- [13] Chen H Z, Tsai S S, Schroth G, *et al.* Robust text detection in natural images with edge-enhanced maximally stable extremal regions [C] // 2011 18th IEEE International Conference on Image Processing, September 11-14, 2011, Brussels, Belgium. New York: IEEE, 2011: 2609-2612.
- [14] Fan Y H, Deng D X, Yan J. Natural scene text detection based on maximally stable extremal region in color space[J]. Journal of Computer Applications, 2018, 38(1): 264-269, 294.
范一华, 邓德祥, 颜佳. 基于色彩空间的最大稳定极值区域的自然场景文本检测[J]. 计算机应用, 2018, 38(1): 264-269, 294.
- [15] Mosleh A, Bouguila N, Hamza A B. Automatic inpainting scheme for video text detection and removal [J]. IEEE Transactions on Image Processing, 2013, 22(11): 4460-4472.
- [16] Zhang G H, Huang K, Zhang B, *et al.* A natural scene text extraction method based on the maximum stable extremal region and stroke width transform [J]. Journal of Xi'an Jiaotong University, 2017, 51(1): 135-140.
张国和, 黄凯, 张斌, 等. 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[J]. 西安交通大学学报, 2017, 51(1): 135-140.
- [17] Wang M. Under the natural scene character area

- location and recognition [D]. Harbin: Harbin Engineering University, 2016: 22-53.
- 王猛. 自然场景下字符区域的定位与识别[D]. 哈尔滨: 哈尔滨工程大学, 2016: 22-53.
- [18] Jiang J Q. Research on the algorithms of text character processing in natural scene[D]. Chengdu: University of Electronic Science and Technology, 2017: 19-24.
- 姜继琦. 自然场景中字符处理方法研究[D]. 成都: 电子科技大学, 2017: 19-24.
- [19] Wu X Y, Wang S H, Zhang Y D. Survey on theory and application of K-nearest-neighbors algorithm[J]. Computer Engineering and Applications, 2017, 53(21): 1-7.
- 毋雪雁, 王水花, 张煜东. K最近邻算法理论与应用综述[J]. 计算机工程与应用, 2017, 53(21): 1-7.
- [20] Zhao Z H, Yang S P, Ma Z Q. License plate character recognition based on convolutional neural network LeNet-5[J]. Journal of System Simulation, 2010, 22(3): 638-641.
- 赵志宏, 杨绍普, 马增强. 基于卷积神经网络 LeNet-5 的车牌字符识别研究[J]. 系统仿真学报, 2010, 22(3): 638-641.
- [21] Neumann L, Matas J. A method for text localization and recognition in real-world images [M] // Kimmel R, Klette R, Sugimoto A. Computer vision-ACCV 2010. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2011, 6494: 770-783.
- [22] Canny J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8(6): 679-698.
- [23] Kittler J. On the accuracy of the Sobel edge detector[J]. Image and Vision Computing, 1983, 1(1): 37-42.
- [24] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform[C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 2963-2970.
- [25] Lucas S M, Panaretos A, Sosa L, *et al.* ICDAR 2003 robust reading competitions [C] // Seventh International Conference on Document Analysis and Recognition, August 6, 2003, Edinburgh, UK. New York: IEEE, 2003: 682-687.