

结合 $Y^d U^a V^a$ 颜色模型和改进 MobileNetV3 的视频烟雾检测方法*

刘 通,程江华,华宏虎,罗笑冰,程 榜
(国防科技大学 电子科学学院,湖南 长沙 410073)

摘要:为降低视频烟雾检测中的虚警率和提升检测效率,提出 $Y^d U^a V^a$ 颜色模型,该模型可以表征烟雾的空间域分布特性和时间域变化特性。利用该颜色模型快速筛选出疑似烟雾图像块,降低虚警率和提升运算效率。提出改进的 MobileNetV3 网络结构,用于提取图像深度特征并对疑似烟雾图像块进行分类识别,检测视频中是否存在烟雾。视频烟雾检测仿真结果表明:该方法准确率和检测帧率高,虚警率低。

关键词:烟雾检测;深度学习;颜色模型;轻量级网络;MobileNet

中图分类号:TP391 **文献标志码:**A **文章编号:**1001-2486(2021)05-080-06

Video smoke detection method combining $Y^d U^a V^a$ color model and improved MobileNetV3

LIU Tong, CHENG Jianghua, HUA Honghu, LUO Xiaobing, CHENG Bang

(College of Electronic Science, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to reduce the false alarm rate and improve the detection efficiency for video smoke detection, the $Y^d U^a V^a$ color model was proposed, which can characterize the spatial distribution and temporal variation of smoke. By using this color model to quickly screen the suspected smoke image blocks, the false alarm rate was reduced and the computing efficiency was improved. An improved MobileNetV3 network structure was proposed, which is aimed to extract deep features of images and to classify the suspected smoke image blocks so as to detect whether there is smoke in a video. The simulation results of video smoke detection show that this method has high accuracy, high detection frame rate, and low false alarm rate.

Keywords: smoke detection; deep learning; color model; lightweight network; MobileNet

快速检测烟雾有利于及早发现火灾,降低火灾危害。传统的烟雾传感器需要烟进入传感器且浓度达到一定程度才能检测到烟雾,难以在室外开放空间使用。基于视觉的烟雾检测技术不受空间限制、覆盖区域大、成本低,是室外烟雾检测研究的主要方向^[1]。就烟雾目标的视觉表现而言,颜色^[2]、纹理^[3]和运动^[4]特征是常用的烟雾检测特征,基于这些传统特征检测烟雾目标通常具有复杂度低、检测效率高的优点,但往往检测正确率偏低,因为烟雾的颜色、纹理特征并不显著,运动特征受自然界的其他物体运动干扰较大,自然界中存在许多与烟雾类似的目标,如云等,导致烟雾目标的可靠检测非常困难。采用深度学习提取图像或者视频中的烟雾目标的深度特征,相比传统特征而言不受人为选择影响,特征泛化能力更强,在大数据集下性能通常优于传统特征,是近些年烟雾检测的主流方法。常用的卷积神经网络

(Convolutional Neural Networks, CNN)、区域卷积神经网络(Region-CNN, R-CNN)等网络模型都可以用于烟雾目标的检测^[5],也有专门针对烟雾特性优化的深度网络模型,如深度归一化卷积神经网络(Deep Normalization and Convolution Neural Network, DNCNN)^[6]。该网络将传统 CNN 中的卷积层改进为批规范化卷积层,有效地解决了网络训练过程中梯度弥散和过拟合的问题,以此加速训练过程和提高检测效果;并且对训练样本进行了数据增强,从而解决正负样本不平衡和训练样本不足的问题。目前基于深度学习开展的烟雾检测研究主要是提取烟雾图像中的深度特征,没有充分利用烟雾运动特性的深层特征,在烟雾检测性能上还有待提升,尤其是虚警率还偏高。而且,采用深度网络通常运算效率偏低,难以满足烟雾视频检测应用对时效性的要求。

为进一步降低视频烟雾检测中的虚警率和提

* 收稿日期:2020-03-17

基金项目:国家自然科学基金资助项目(61303188);湖南省自然科学基金资助项目(2020JJ4670);基础加强计划技术领域基金资助项目(2019-JCJQ-JJ-209)

作者简介:刘通(1982—),男,河南南阳人,副研究员,博士,E-mail:liutong1129@126.com

升检测效率,本文提出一种结合 $Y^dU^aV^a$ 颜色模型和改进 MobileNetV3 的视频烟雾检测方法。首先提出 $Y^dU^aV^a$ 颜色模型,描述相邻帧烟雾的运动和颜色变化特征;然后在分块图像中利用 $Y^dU^aV^a$ 颜色模型粗筛选疑似烟雾图像块;最后结合烟雾快速检测要求改进 MobileNetV3 网络结构,实现视频中烟雾的快速可靠检测。

1 $Y^dU^aV^a$ 颜色模型

烟雾检测领域目前采用 CNN、DCNN 等深度网络来提升烟雾特征的显著性和稳健性,进而提升烟雾检测性能。通常是将 RGB 图像送入深度网络,在 R、G、B 三个通道上提取特征。然而,R、G、B 三个通道不能很好地展示烟雾的视觉特性。通过对烟雾的视觉特性进行分析发现:烟雾的亮度会随着颗粒的浓度以及成分的变化而大幅变化,可能很亮,也可能很暗,但烟雾的色度比较均匀且稳定,而且烟雾是运动的,存在向上和向四周的扩散运动。基于此,提出 $Y^dU^aV^a$ 模型,用于更好地描述烟雾的视觉特性。基本思路是:在 YUV 颜色模型上,计算 Y 空间上相邻帧图像的变化图像,用于反映烟雾的运动扩散特性;对 U 空间和 V 空间的图像进行均值滤波,用于反映烟雾的色度变化特性。具体地,对于视频中的第 k 帧图像,有:

$$Y_k^{(d)}(x,y) = |Y_k(x,y) - Y_{k-1}(x,y)| \quad (1)$$

$$U_k^{(a)}(x,y) = \frac{U_k(x,y) + U_{k-1}(x,y)}{2} \quad (2)$$

$$V_k^{(a)}(x,y) = \frac{V_k(x,y) + V_{k-1}(x,y)}{2} \quad (3)$$

图 1 展示了烟雾视频 wildfire_smoke_4.avi^[7] 中 RGB、YUV 和 $Y^dU^aV^a$ 三种模型在第 115 帧的对比情况。在 RGB 和 YUV 的各颜色分量中,烟雾与白云等颜色相近目标的差异不显著。在 $Y^dU^aV^a$

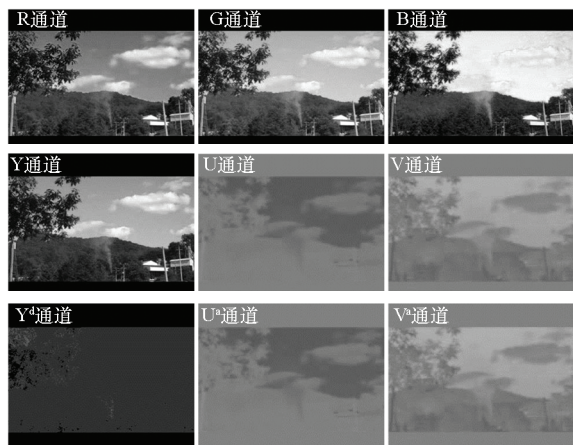


图 1 三种颜色模型对比

Fig. 1 Comparison of three color models

模型的三个分量中,运动的烟雾与类烟雾的白云有了较为明显的差异,而且烟雾与背景的颜色差异仍然得以体现。因此,采用 $Y^dU^aV^a$ 模型既利于表征烟雾的运动特性,又能描述烟雾的颜色差异,更适于提取烟雾的显著和稳健特征。

2 疑似烟雾图像块检测

运用深度网络检测图像中的烟雾目标通常采用以下两种方式:

一是按照网络模型的输入要求将视频图像进行变换,之后送入网络模型得到图像的分类结果,如 MobileNetV3 的输入图像为 224×224 的三通道 RGB 图像。在实际图像中,烟雾区域通常只占据整幅图像中的一小部分,此时烟雾目标并不显著,难以训练较好的模型。

二是采用目标检测的思想,主要包括两类方法。一是 two-stage 的方法,如 Fast R-CNN。此类方法先生成许多候选框,然后对每一个候选框进行分类和回归来完成目标分类工作,精度较高但效率偏低^[8]。二是 one-stage 的方法,如 SSD。此类方法通过一遍网络搜索即可得到目标的位置和类别信息,速度较快但仍难以满足实时视频检测的要求^[9]。

事实上,对于森林火灾监测预警等视频烟雾检测应用而言,其主要任务是快速判断实时视频中某时刻是否存在烟雾,而并不关心烟雾的面积、轮廓等详细信息。因此,为了实现视频中烟雾目标的快速检测,采用分块的思想,先将图像划分为互不重叠的子块,然后将各子块的图像送入深度网络模型进行分类。与前述的第一种方式相比,在各个图像块中,烟雾目标的显著性会得到提升,便于训练出分类性能好的深度网络模型。与前述的第二种方式相比,图像块的数量与多尺度候选框的数量相比是非常小的,从而可以提升整体的运算效率。进一步地,还将对送入深度网络模型的图像块进行滤波处理,依据烟雾的运动和颜色特性筛选出疑似烟雾的图像块,仅将疑似烟雾区域的图像块送入深度网络模型进行分类,这样又将提升视频烟雾的检测效率。

图像分块的思路具体是:首先将输入图像的尺寸调整为 560×448 ;然后按照从左到右、从上到下的顺序,将图像划分为互不重叠的 56×56 的图像子块,每幅图像共可划分为 80 个图像子块。需要解释的是,图像子块尺寸是为适应后续改进的 MobileNetV3 网络而定的,将后续网络的输入图像尺寸设计为 56×56 ,可以更多地继承原网络

的结构优势以及已训练好的模型资源。

对于每一个图像分块,依据烟雾的运动和颜色特性来检测疑似烟雾图像子块。具体是在 $Y^d U^a V^a$ 颜色模型上,统计每一个图像子块中疑似烟雾像素点的数量 N ,当 N 超过图像子块面积的 10% 时,就认定该图像子块是疑似烟雾图像子块。这里,疑似烟雾图像块的筛选条件比较宽松,目的是尽可能避免遗漏,后续会通过深度网络模型对疑似烟雾图像块进行细分类,滤除此类虚检的图像块。其中,疑似烟雾像素点是依据烟雾的运动和颜色特性来判定的,公式为:

$$g_k(x, y) = \begin{cases} 1, & Y_k^{(d)}(x, y) > T_1 \text{ 且} \\ & |U_k^{(a)}(x, y) - V_k^{(a)}(x, y)| < T_2 \\ 0, & \text{其他} \end{cases} \quad (4)$$

式中: $g_k(x, y) = 1$ 表示第 k 帧图像中的像素点 (x, y) 为疑似烟雾像素点,否则不是。 T_1 和 T_2 为经验阈值,通过观察小样本烟雾像素点在 $Y^d U^a V^a$ 颜色模型上的变化情况主观确定,为宽松起见,取 $T_1 = 10$ 、 $T_2 = 40$ 。

3 改进的 MobileNetV3 网络

为了满足视频烟雾检测对于烟雾检测效率的要求,以及前端嵌入式视频采集处理平台部署的要求,通常需要采用轻量级网络模型实现烟雾图像深度特征的提取及烟雾目标的识别。MobileNetV3 网络是谷歌新推出的轻量级网络,其中的 MobileNetV3 large 版本相对于 MobileNetV2 而言,在 COCO 上达到相同精度的前提下,速度可提升 25%^[10]。这对于视频烟雾检测的效率提升而言很有意义。因此,基于 MobileNetV3 large 版本设计视频烟雾检测的网络模型。主要从视频烟雾检测的实际出发,对 MobileNetV3 large 版本进行改进,基本思路是:

第一,为了提取显著性和稳健性更强的烟雾运动和颜色变化特征,对 MobileNetV3 网络的输入层进行改进,具体是采用 $Y^d U^a V^a$ 颜色模型替代 MobileNetV3 网络常用的 RGB 颜色模型。 $Y^d U^a V^a$ 颜色模型如第 1 节所述。

第二,由第 2 节所述,深度网络的目标是对疑似烟雾区域的图像进行分类,结合这一需求,对 MobileNetV3 large 版本的网络结构进行修改,修改后的网络结构如表 1 所示,具体修改内容包括:

1)在第一层,“Input”的尺寸改为 $56^2 \times 3$,以适应疑似烟雾区域的尺寸。同时,步长“s”改为

1,在这一层卷积运算过程中不降低特征尺度,但扩展了通道数。

2)合并了原始 MobileNetV3 large 版本网络结构的第二层和第三层,因为结合本文第一层的修改,原始的第二层和第三层存在较大的冗余,在修改后的结构中,第二层网络能够抽取相似的特征,而且由于删除了一层网络,运算效率得以进一步提升,且资源占用也相应减少。

3)在最后一层,根据应用需求,只需要判断疑似烟雾区域是否为真实的烟雾区域,这是个二分类问题,因此将“#output”改为 2。

表 1 改进的 MobileNetV3 网络
Tab. 1 Modified MobileNetV3 network

输入	操作	扩展 尺寸	输出 通道数	SE	NL	步 长
$56^2 \times 3$	conv2d		16	无	HS	1
$56^2 \times 16$	bneck, 3×3	64	24	无	RE	2
$56^2 \times 24$	bneck, 3×3	72	24	无	RE	1
$56^2 \times 24$	bneck, 5×5	72	40	有	RE	2
$28^2 \times 40$	bneck, 5×5	120	40	有	RE	1
$28^2 \times 40$	bneck, 5×5	120	40	有	RE	1
$28^2 \times 40$	bneck, 3×3	240	80	无	HS	2
$14^2 \times 80$	bneck, 3×3	200	80	无	HS	1
$14^2 \times 80$	bneck, 3×3	184	80	无	HS	1
$14^2 \times 80$	bneck, 3×3	184	80	无	HS	1
$14^2 \times 80$	bneck, 3×3	480	112	有	HS	1
$14^2 \times 112$	bneck, 3×3	672	112	有	HS	1
$14^2 \times 112$	bneck, 5×5	672	112	有	HS	1
$14^2 \times 112$	bneck, 5×5	672	160	有	HS	2
$7^2 \times 160$	bneck, 5×5	960	160	有	HS	1
$7^2 \times 160$	conv2d, 1×1		960	无	HS	1
$7^2 \times 960$	Pool, 7×7			无	HS	
$1^2 \times 960$	conv2d, 1×1 , NBN		1 280	无	HS	1
$1^2 \times 1 280$	conv2d, 1×1 , NBN		2	无		

注:conv2d 指二维卷积运算,bneck 指瓶颈运算,SE 表示块中是否有 Squeeze-And-Excite 结构,NL 表示采用的非线性激活函数的类型,HS 指 hard-swish,RE 指 ReLU, NBN 指不采用批标准化。

4 算法实现

综上所述,算法的主要实现步骤如下所示。

Step 1: 采用双线性插值方法,将视频的第 k 帧图像尺寸调整为 560×448 ;

Step 2: 依据式(1)~(3),计算第 k 帧图像的 $Y^dU^aV^a$ 颜色模型;

Step 3: 依据式(4),判别疑似烟雾像素点;

Step 4: 依据第2节所述,统计各个 56×56 的图像子块的疑似烟雾像素点数量,筛选出疑似烟雾图像块;

Step 5: 将疑似烟雾图像块送入第3节所述的改进 MobileNetV3 网络模型,开启多线程并行计算,只要有一个疑似烟雾图像块分类结果为烟雾,则判断第 k 帧图像为烟雾图像,继续下一帧检测。

5 实验与分析

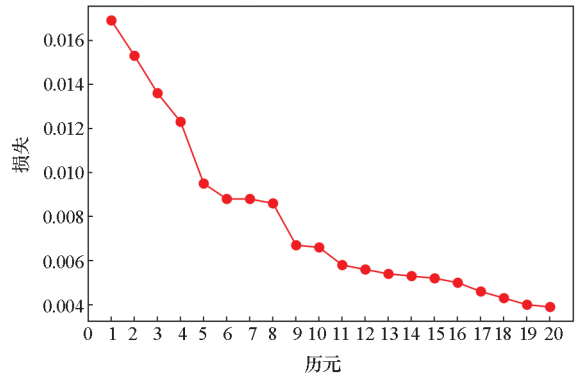
5.1 实验说明

选择烟雾检测领域公开的视频数据集进行实验仿真,包括 Yuan 公开的3段烟雾视频和3段非烟雾视频^[11],以及 CVPR 实验室公开的4段烟雾视频和10段非烟雾视频^[7]。其中,每段视频的前500帧用作测试。剩余的视频帧都用作训练。

视频烟雾检测的评价指标选用文献[6]所述的准确率(Accuracy Rate, AR)和虚警率(False Alarm Rate, FAR),同时增加检测帧率(detected frames per second, dfps)指标,用于评测算法的运算效率。实验平台环境为:Windows 7 系统,python 3.6.2、tensorflow 1.11.0 和 keras 2.2.4 软件平台,NVIDIA GeForce GTX1080Ti 显卡和 Intel Core i7-8700K CPU。

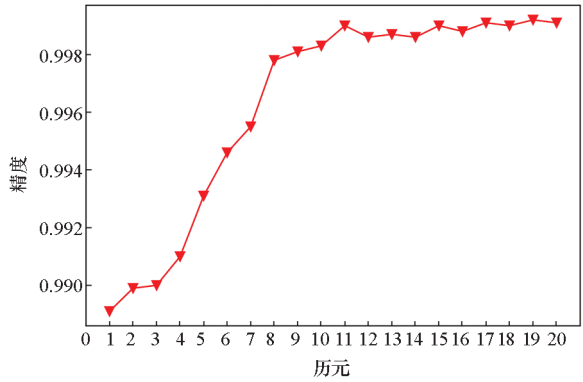
5.2 网络模型训练

改进的 MobileNetV3 网络模型的训练方法是:对于训练数据,先按照第4节所述的 Step 1 至 Step 4 生成视频中每帧图像(第一帧除外)的各个 56×56 图像子块的 $Y^dU^aV^a$ 图像,其中疑似烟雾图像块的 $Y^dU^aV^a$ 图像与其他图像块分开存放。然后分别从中人工筛选出烟雾图像块和非烟雾图像块的 $Y^dU^aV^a$ 图像,并为每一个图像块人工设定标签,烟雾图像块标签为1,非烟雾图像块标签为0,构建正负样本集。筛选的正样本集图像数量为12 182,负样本集图像数量为42 860。基于该数据集训练表1所示的改进 MobileNetV3 网络模型,训练过程的损失曲线和精度曲线如图2所示。同时,存储 $Y^dU^aV^a$ 图像对应的 RGB 图像,同样人工设定标签,构建正负样本集,作为后续对比实验中所对比方法的训练样本集。



(a) 损失曲线

(a) Loss curve



(b) 精度曲线

(b) Accuracy curve

图2 训练过程曲线

Fig.2 Training process curves

5.3 实验对比分析

5.3.1 不同输入图像和网络的对比测试

主要创新之处包括:1)提出 $Y^dU^aV^a$ 颜色模型,作为深度网络模型的输入;2)采用图像分块和疑似烟雾块筛选,降低虚警率和提高运算效率;3)改进 MobileNetV3 网络。下面对比这三个阶段所达到的性能指标,如表2所示。其中,实验1直接用原 RGB 图像作为 MobileNetV3 large 版本网络的输入(网络模型训练数据集为训练数据中的 RGB 图像);实验2先对 RGB 图像进行分块(分块方法与本文所述一致),然后再将图像块采用双线性插值方法归一化到 224×224 ,之后再作为 MobileNetV3 large 版本网络的输入(网络模型训练数据集为 $Y^dU^aV^a$ 图像对应的 RGB 图像构建的正负样本集);实验3采用分块后的 $Y^dU^aV^a$ 图像作为改进的 MobileNetV3 网络的输入;实验4是在实验3的基础上,对输入的 $Y^dU^aV^a$ 图像块进行疑似烟雾块的粗筛选,然后再送入改进的 MobileNetV3 网络进行细分类,详见第4节的算法实现。由表2可见,采用 RGB 图像作为输入时,

AR 偏低且 FAR 偏高,这主要是因为测试视频中烟雾目标较小,在整幅图像中特征不够显著。实验 2 对 RGB 进行分块之后,AR 提升较为明显,但因为参与分类的图像块大幅增加,检测帧率明显降低。实验 3 采用分块后的 $Y^dU^aV^a$ 图像作为改进 MobileNetV3 网络的输入,因为 $Y^dU^aV^a$ 图像相比 RGB 图像更能呈现烟雾的显著特征,虚警率有一定下降。同时由于改进 MobileNetV3 网络层数减少且输入层尺度下降,检测帧率也有一定提升。进一步地,实验 4 对输入的 $Y^dU^aV^a$ 图像块进行疑似烟雾块的粗筛选,降低了静止目标可能引起的虚警,从而降低了虚警率,而且检测帧率也得以提升。而且,由于式(4)中疑似烟雾像素点判决的参数 T_1 和 T_2 设置宽松,实验中经过疑似烟雾块粗筛选之后尽管存在部分烟雾占比很小的真实烟雾块被误判为非烟雾块,但每一帧烟雾图像中仍能检测出多个疑似烟雾图像块,而每帧图像中最终只要检测出一个真实的烟雾图像块就会将该帧图像标记为烟雾图像,因此粗筛选没有引起漏检,实验 4 的准确率指标在虚警率下降后得以提升。

表 2 不同输入图像和网络的实验结果

Tab.2 Results of different input images and network

组号	AR/%	FAR/%	dfps
实验 1	82.38	3.32	15
实验 2	90.46	1.94	3
实验 3	92.53	1.07	7
实验 4	97.89	0.49	21

5.3.2 不同方法性能对比

表 3 给出了本文方法与部分常用的烟雾检测方法的性能对比结果。其中,对比方法中使用的图像训练数据集是 5.2 节所述的存储 $Y^dU^aV^a$ 图像对应的 RGB 图像构建的正负样本集。由表 3 可见,采用深度学习的烟雾检测方法(DNCNN, MobileNetV3 + SSD 和本文方法)的准确率普遍高于传统方法的特征提取方法(HS'I, LBP + LBPV 和 optical flow),这主要是因为烟雾的颜色、纹理、运动特征都不是特别显著,人为选择最优的特征比较困难,难以单一鉴别能力强的特征。但是深度学习往往复杂度高,检测帧率一般较低,相对而言,使用轻量级网络(MobileNetV3 + SSD 和本文方法)可以提升运算效率,而本文方法在使用轻量级网络前先采用运动、颜色等快速筛选策略剔除了大量的非烟雾区域图像子块,进一步降

低了运算量。使用运动特征的烟雾检测方法(optical flow 和本文方法)的虚警率通常较低,这是因为场景中与烟雾颜色、纹理类似的物体比较多,容易引起虚警,而运动可以剔除大量静止物体,从而降低虚警率。综合分析,本文方法使用 $Y^dU^aV^a$ 颜色模型融合烟雾的空间域分布特性和时间域变化特性,快速剔除非烟雾区域图像块,同时改进 MobileNetV3 轻量级网络来快速提取鉴别能力更强的烟雾特征,视频烟雾检测的准确率最高,虚警率最低,而且检测帧率在深度学习方法中最高,尽管与传统特征提取方法相比检测效率偏低,但基本能满足视频烟雾检测对时效性的要求。

表 3 不同方法的实验结果

Tab.3 Results of different methods

方法	AR/%	FAR/%	dfps
HS'I ^[2]	51.83	17.78	137
LBP + LBPV ^[3]	69.89	11.62	57
optical flow ^[4]	79.93	1.81	27
DNCNN ^[6]	95.36	0.97	2
MobileNetV3 + SSD	93.17	1.32	13
本文方法	97.89	0.49	21

6 结论

利用计算机视觉技术检测视频中的烟雾是近些年火灾探测领域的热点研究方向。为降低视频烟雾检测中的虚警率和提升检测效率,提出 $Y^dU^aV^a$ 颜色模型,该模型可以表征烟雾的颜色和运动变化特性,有助于快速筛选出疑似烟雾图像块和提取鉴别能力强的特征。在此基础上,改进 MobileNetV3 网络模型,可以快速准确地对疑似烟雾图像块进行细分类。仿真实验证实本文方法准确率高,虚警率低,且检测帧率高,但检测帧率仍然不能充分地满足实时性的要求,后续还需进一步提升。

参考文献(References)

- [1] 史劲军,袁非牛,夏雪. 视频烟雾检测研究进展[J]. 中国图象图形学报, 2018, 23(3): 303-322.
SHI Jinting, YUAN Feiniu, XIA Xue. Video smoke detection: a literature survey [J]. Journal of Image and Graphics, 2018, 23(3): 303-322. (in Chinese)
- [2] KRSTINIĆ D, STIPANIČEV D, JAKOVČEVIĆ T. Histogram-based smoke segmentation in forest fire detection

- system [J]. Information Technology and Control, 2009, 38(3): 237 – 244.
- [3] RUSSO A U, DEB K, TISTA S C, et al. Smoke detection method based on LBP and SVM from surveillance camera [C] // Proceedings of International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018: 1 – 4.
- [4] KOPILOVIC I, VAGVOLGYI B, SZIRANYI T. Application of panoramic annular lens for motion analysis tasks; surveillance and smoke detection [C] // Proceedings of 15th International Conference on Pattern Recognition, 2000: 714 – 717.
- [5] FRIZZI S, KAABI R, BOUCHOUICHA M, et al. Convolutional neural network for video fire and smoke detection [C] // Proceedings of 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016: 877 – 882.
- [6] YIN Z J, WAN B Y, YUAN F N, et al. A deep normalization and convolutional neural network for image smoke detection [J]. IEEE Access, 2017, 5: 18429 – 18438.
- [7] CVPR Lab. Fire & Smoke Database [EB/OL]. (2014 – 8 – 24) [2020 – 02 – 02]. <https://cvpr.kmu.ac.kr/>.
- [8] GIRSHICK R. Fast R-CNN [C] // Proceedings of IEEE International Conference on Computer Vision (ICCV), 2016.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C] // Proceedings of European Conference on Computer Vision, 2016: 21 – 37.
- [10] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C] // Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1314 – 1324.
- [11] YUAN F N. Video smoke detection [EB/OL]. (2011 – 08 – 28) [2020 – 02 – 02]. <http://staff.ustc.edu.cn/~yfn/vsd.html>.