

基于改进的 ViBe 和 YOLO v3 算法的行人检测方法

李士骥, 李忠民, 李 威

(南昌航空大学 信息工程学院, 江西 南昌 330063)

摘要: 针对传统视觉背景提取 (visual background extractor, ViBe) 算法在进行行人检测时会产生鬼影的缺点, 本文提出了一种基于改进的 ViBe 和 YOLO v3 算法的行人检测方法。利用改进的 YOLO v3 算法 YOLO v3-SPP (spatial pyramid pooling) 对 ViBe 算法的初始化策略进行改进以消除鬼影。运用 YOLO v3-SPP 算法对首帧图像进行行人检测, 使用本文提出的行人消除方法将检测出的行人进行消除, 并将输出图像代替 ViBe 算法的首帧, 从而达到消除鬼影的目的。经过分析和实验验证, 结果表明该算法能够有效解决鬼影问题。

关键词: 图像处理; 行人检测; 视觉背景提取; 鬼影消除; YOLO v3 算法

中图分类号: TP391

文献标志码: A

文章编号: 1001-8891(2023)02-0137-06

Pedestrian Detection Method Based on Improved ViBe and YOLO v3 Algorithms

LI Shiji, LI Zhongmin, LI Wei

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: This paper presents an improved algorithm to solve the problem of ghosting in the visual background extractor (ViBe) algorithm for pedestrian detection. The initialization strategy of the ViBe algorithm is improved using the YOLO v3-spatial pyramid pooling (SPP) algorithm to eliminate ghosts. Thus, the YOLO v3-SPP algorithm detects pedestrians in the first frame, eliminates the detected pedestrians, and replaces the first frame of the ViBe algorithm with the output image to eliminate ghosts. The results of the analysis and verification show that the algorithm can effectively solve the ghost problem.

Key words: image processing, pedestrian detection, visual background extraction, ghost elimination, YOLO v3

0 引言

背景提取是一种目标检测算法, 其准确性、实时性和鲁棒性较好^[1]。ViBe 算法^[2]是背景提取算法中的一种像素级的前景检测算法, 这种算法检测迅速、检测精度优异。虽然这种算法的性能十分优异, 但其仍具有不足之处。比如, 算法所存在的鬼影问题就是影响算法性能的最主要原因之一。所谓鬼影, 是指在没有运动目标的区域中被错误检测出的前景区域, 也就是不真实的前景^[3]。

对于如何解决鬼影的问题, 许多学者提出的解决措施中, 主要包含两种处理方式: 一是对背景模型初始化作出改进。比如, 文献[4]使用多帧图像进行背景模型的初始化。文献[5]将最大类间方差法 (OTSU)^[6]引

入到帧间差分法中, 把前面几帧图片加以适当的处理, 组成一张背景样本图像, 以这幅图片为依据来创建初始的背景模型。文献[7]用视频的前 25 帧图像中的像素点及其它邻域的像素点中的像素值来初始化背景模型。此方法的背景模型初始化需要通过多帧图像来建立, 而这会大大降低背景模型的初始化速度。但同时, 此方法无法处理因背景突变而产生的鬼影现象。第二种处理方式是在前景检测的过程中对鬼影进行识别, 若识别出鬼影, 则将其视为背景。比如, 文献[8]比较了前景目标与相邻背景像素的直方图, 根据直方图的不同来辨别鬼影区域。文献[9]通过分析前景的像素值来进行鬼影的消除。然而这种算法对速度慢并且颜色单调的目标的检测敏感性不高, 容易将其误判为鬼影, 并且算法较为复杂, 无法确保实时性。

收稿日期: 2022-03-18; 修订日期: 2022-04-04.

作者简介: 李士骥 (1998-), 男, 硕士研究生, 主要从事图像处理方面的研究。E-mail: sdlishiji@163.com。

通信作者: 李忠民 (1975-), 男, 博士, 副教授, 硕士生导师, 主要从事图像处理与人工智能方面的研究。E-mail: zhongmli@163.com。

基金项目: 国家自然科学基金 (61263040), 江西省自然科学基金 (20202BABL202005), 南昌航空大学“三小”项目 (2022XG13)。

综上所述,当前消除鬼影的方法都具有一定的不足。为了解决鬼影问题,本文提出了一种基于改进的ViBe和YOLO v3算法的行人检测方法。算法采用YOLO v3-SPP对首帧中的行人进行检测,将检测出的行人消除后,将输出的图像作为ViBe算法的首帧,从而解决鬼影问题。

1 算法原理

1.1 ViBe 算法原理

ViBe算法为每一个像素点都创建了一个背景样本模型,通过比较当前帧与背景模型中相对应位置的像素值,并由预设的阈值将其定为背景点或前景点。一旦像素被确认为背景点,则随机更新其背景模型里的一个样本值。同时基于空间一致性原则,将其邻域像素背景模型中的一个样本值进行随机更新。ViBe算法主要由3部分组成,即背景建模、前景检测和背景模型的更新。

1) 背景建模

背景模型是背景样本的集合。在背景模型的初始化过程中,从一帧图像每个像素的八邻域中随机选择 N 个像素的灰度值存放到背景模型样本中。背景模型也可以描述为:

$$M(x) = \{v_1, v_2, \dots, v_N\} \quad (1)$$

式中: $M(x)$ 为背景样本; v_i 表示背景模型样本中的像素值; N 表示样本的大小。

2) 前景检测

将当前的像素值与背景模型中同一位置像素的样本作比较来判定当前像素点是否为前景。如图1所示, $v(x)$ 为当前帧图像中的像素 x 的像素值; C_1 与 C_2 组成了二维的色彩空间(C_1, C_2); $S_R(v(x))$ 是以 $v(x)$ 作为圆心的圆形区域, R 是此区域的半径; $v(x)$ 与背景样本模型之间的相似程度由 $S_R(v(x))$ 所包含的 $M(x)$ 的样本数 M 来描述。当 M 大于或者等于设定的阈值时,判定 $v(x)$ 为背景;当 M 小于设定的阈值时,判定 $v(x)$ 为前景。

3) 背景模型的更新

本文中,ViBe算法的背景模型更新方法使用了随机更新策略。设定抽样因子为 λ 。假设第二步前景检测过程后,算法判定当前像素点为背景像素点,那么它会有 $1/\lambda$ 的概率更新其在背景模型中的样本值,并将随机替换样本模型中的一个样本。每一个样本经过时间 Δt 仍被保存的概率可以描述为:

$$p(t, t + \Delta t) = e^{-\ln(\frac{N}{N-1})\Delta t} \quad (2)$$

该过程将有助于确保样本生命周期的平滑。基于图像空间的一致性原则,该算法可以通过相同的方式与概率随机选取八邻域中的一个像素来更新其背景模型。

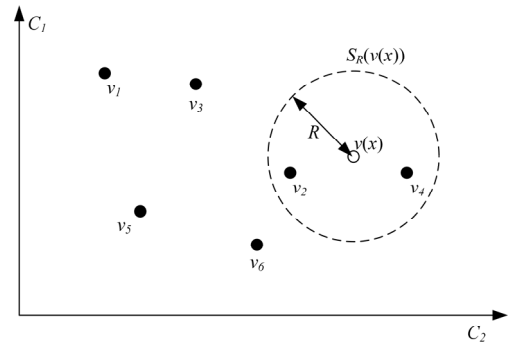


图1 当前帧像素值与样本值在二维色彩空间的比较

Fig. 1 Comparison of current frame pixel value and sample value in two-dimensional color space

1.2 YOLO v3-SPP 介绍

YOLO v3借鉴了在Faster R-CNN中提到的锚框(Anchor box)机制^[10]。为了增强检测的精度,采用 K 均值聚类算法得出边框的尺寸,从而获得合适的锚框^[11]。同时,采用特征金字塔网络(feature pyramid networks, FPN)来实现多尺度预测^[12],检测时可以应对不同尺寸的物体,因此能够大大提高对精细物体的识别能力。YOLO v3结合了锚框和特征金字塔网络的思想,每个网格可以预测3个预测框,并选择具有最优交并比(IoU)的锚框作处理后作为输出的预测框。

YOLO v3使用Darknet53网络为主干,舍弃了最大池化层,采用步长为2的卷积层作为代替。

由于YOLO v3算法在行人遮挡检测上的漏检率较高,不利于对图像进行特定的处理,所以对网络结构进行改进以增加算法的准确率与检测精度。

YOLO v3-SPP是提出的YOLO v3算法的增强网络。与YOLO v3不同的是,它在网络的YOLO层前增加了一个SPP层,并为了避免过拟合而进行多尺度训练,从而能够增加检测精度(在COCO公共数据集上)。

SPP层的基本构造如图2所示。利用 5×5 , 9×9 与 13×13 这3个最大池化层对特征图进行池化,再连接起池化后的特征图,并将其传输到下一层网络^[13]。

加入SPP层的YOLO v3算法(YOLO v3-SPP)的结构如图3所示。

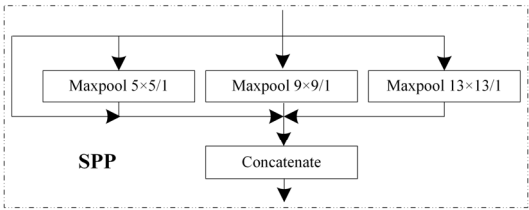


图 2 SPP 层结构

Fig. 2 SPP layer structure

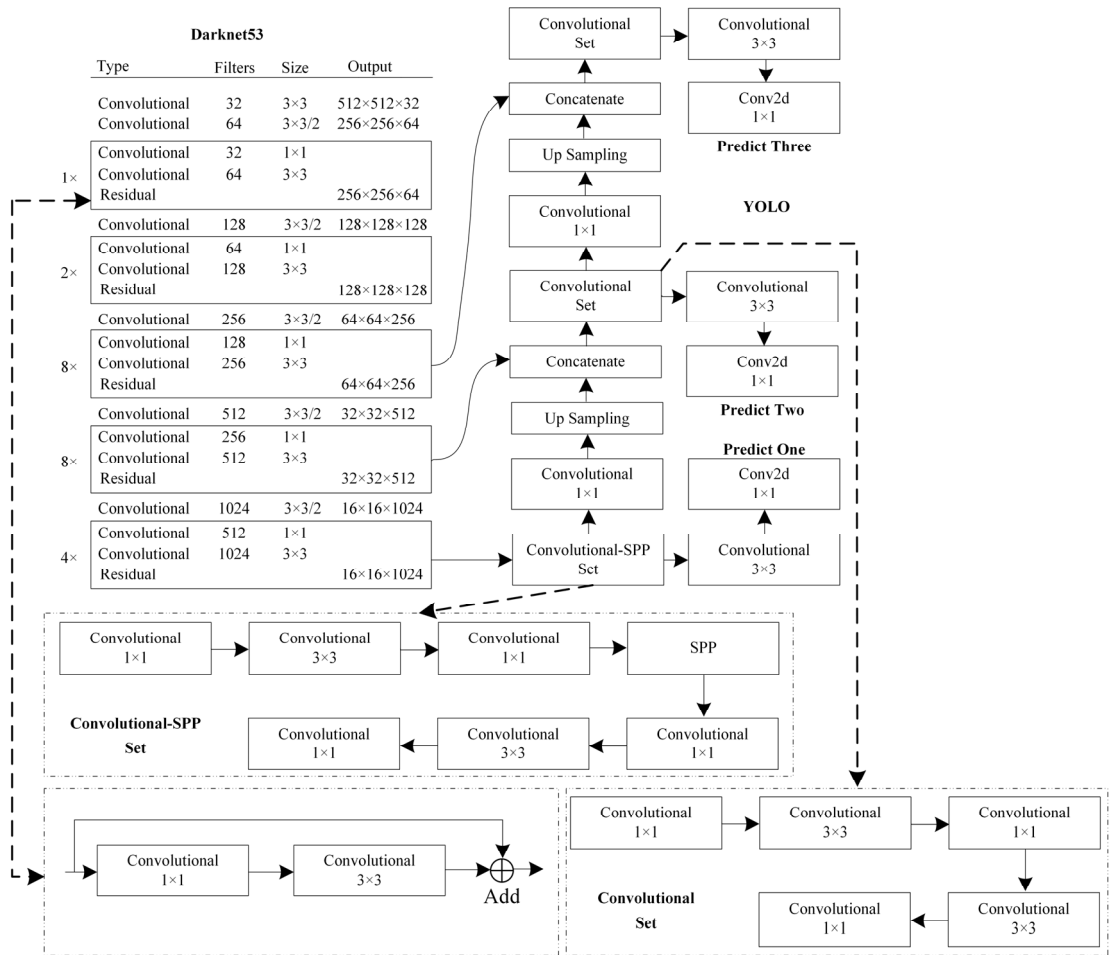


图 3 YOLO v3-SPP 结构

Fig. 3 YOLO v3-SPP structure

2 基于 YOLO v3 SPP 的改进 ViBe 算法

2.1 鬼影问题

ViBe 算法为了提升初始化的速度,通过只处理视频图像的首帧来对背景样本模型作初始化。但是这种措施存在着不足,当待检测的运动前景存在于首帧时,会使后续检测中出现鬼影。如图 4 所示。

2.2 提出的改进措施

因为鬼影产生的原因是第一帧中出现了运动目标,所以利用 YOLO v3-SPP 算法本身对静止行人目标的检测优势,在进行背景初始化之前,对首帧图像进行行人检测。检测出行人后,将检测出的行人进行消除,从而避免鬼影的产生。

基于 YOLO v3 SPP 的改进 ViBe 算法框图如图 5 所示。



(a) 原始图像 (b) ViBe 算法
(a) Original image (b) ViBe algorithm

图 4 鬼影问题展示

Fig. 4 Ghost problem display

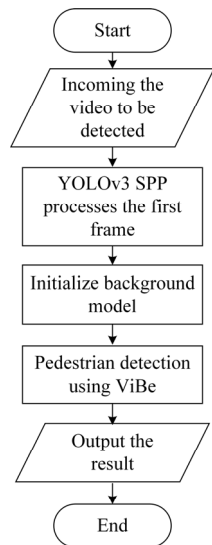


图5 基于 YOLO v3 SPP 的改进 ViBe 算法框图

Fig. 5 Block diagram of improved ViBe algorithm based on YOLO v3 SPP

本文使用 PASCAL VOC2012 数据集对 YOLOv3-SPP 网络进行训练后得到权重文件。在输入待检测的视频后，对首帧作处理。利用训练好的权重文件进行行人检测。整幅首帧图像可以分为两部分，即前景部分与背景部分。针对背景部分的像素点，与传统算法的背景初始化步骤一致。针对前景部分的像素点，采用以下的行人消除方法对行人进行消除。具体步骤如下：

1) 将每个检测出的行人，分为 6 个部分。如图 6 所示。

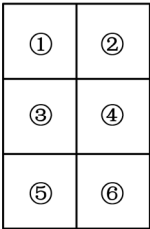


图6 分成六部分的行人

Fig. 6 Pedestrians divided into six parts

2) 如果检测出的行人不在图像的边界处，则用行人周围的像素来代替图 6 中各方框内的像素值。代替规则如下：

- ①和②分别用各自上方的像素值来代替。
- ③和④用各自两边的像素值来代替。
- ⑤和⑥分别用各自下方的像素值来代替。

3) 如果检测出的行人在图像的边界处，则基于 2) 中的规则，做出适当的调整。具体调整如图 7 所示。像素代替方法的总流程图如图 7 所示。

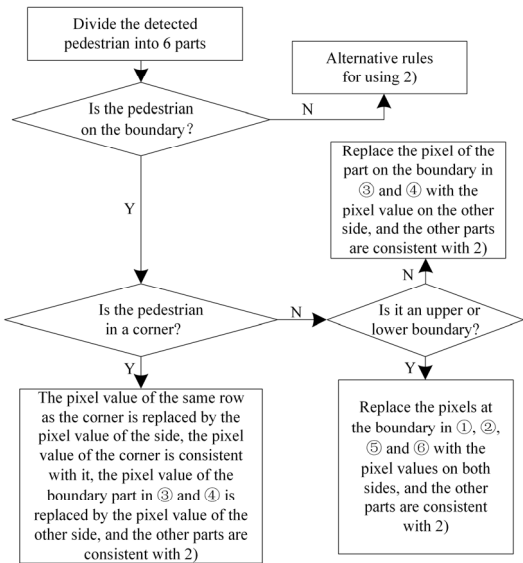


图7 像素代替方法的总流程

Fig. 7 General flow chart of pixel replacement method

用这种方法可以将首帧中的行人进行消除。行人消除的效果如图 8 所示。

用处理好的首帧图像进行初始化背景建模，再用 ViBe 检测行人，对鬼影的消除效果如图 9 所示。

可见，本文算法可以有效地解决鬼影问题，并改善算法的检测性能。

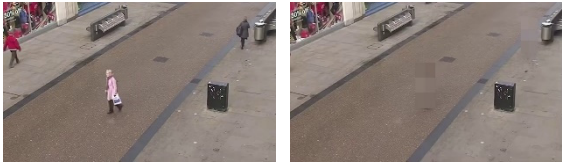


图8 行人消除的效果

Fig. 8 Effect of pedestrian elimination



(a) 原始图像
(a) Original image

(b) ViBe 算法
(b) ViBe algorithm



(c) 本文算法
(c) The algorithm in our paper
图 9 鬼影消除的效果

Fig. 9 Results of ghost elimination

3 实验结果分析与评价

3.1 实验结果分析

根据以上的算法描述, 选择 PETS2006^[14]测试视频来进行实验验证。实验环境和参数如下: Windows 10 系统, AMD Ryzen 5 2600X Six-Core Processor CPU, 3.60GHz 主频, 16.0G RAM, Python 3.7。背景模型建立参数 $N=20$, $R=20$ 。 N 为背景样本的容量, R 为颜色距离判定阈值。

本文将提出的算法和 ViBe 算法、三帧差分法^[15] (three-frame differential method, TFDM)、高斯混合模型^[16] (Gaussian mixture model, GMM) 算法进行对比测试。三帧差分法是一种对视频中连续 3 帧图像进行差分运算来检测运动目标的方法, 这种方法的优点是计算量小, 实时性较好。GMM 算法是有限个单一高斯分布的加权和, 能反映各个像素点不同的分布情况。

测试视频的检测结果如图 10 所示, 第 1 行分别是测试视频中第 534 帧、第 564 帧和第 583 帧的原始图像。第 2 行是传统 ViBe 算法对这 3 帧原始图像的检测结果, 可以清楚地看到用这种方法检测后所产生的鬼影; 第 3 行是使用三帧差分法 (TFDM) 对这 3 帧原始图像的检测结果, 能够看出这种方法只能检测出运动物体的轮廓信息, 而当物体运动比较快时, 轮廓的部分信息也会丢失。第 4 行是高斯混合模型 (GMM) 算法对 3 帧原始图像的检测结果, 可以看到, 这种方法检测出的前景会出现部分缺失和空洞的现象。第 5 行是本文所提出的算法, 可以看出, 本文所提出的算法可以有效地解决鬼影问题, 并且检测出的运动目标完整无空洞。

3.2 实验结果评价

采用文献[17]中提出的方法对本文提出的算法进行定量评价。使用精确率 (Precision)、召回率 (Recall) 以及 F1 分数 (F1-score) 等指标, 可以描述为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

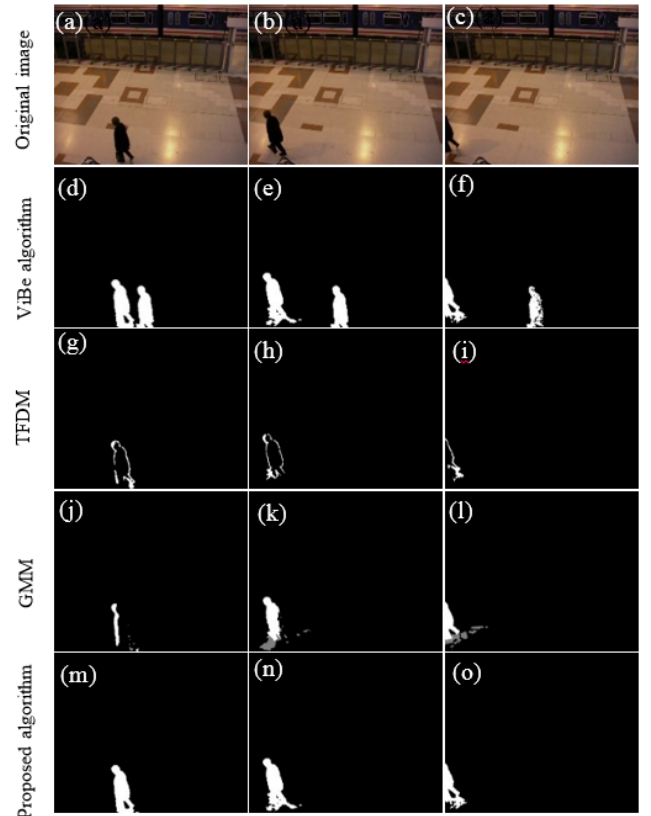


图 10 PETS2006 视频检测效果: (a)(b)(c) 原始图像; (d)(e)(f) ViBe 算法; (g)(h)(i) 三帧差分法; (j)(k)(l) 高斯混合模型算法; (m)(n)(o) 本文算法

Fig.10 PETS2006 video detection effect: (a)(b)(c) Original image; (d)(e)(f) ViBe algorithm; (g)(h)(i) Three-frame difference method; (j)(k)(l) Gaussian mixture model algorithm; (m)(n)(o) The algorithm in our paper

$$\text{Pecall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

式中: TP 与 TN 分别为检测正确的前景像素与背景像素的数量; FP 与 FN 分别检测错误的前景像素与背景像素的数量。一般来说, 精确率与召回率越高, 查准率和查全率就越高, 算法性能就越好; F1 分数则为前两者的调和平均, F1 分数的值越大, 算法的综合性能越强。表 1 为对实验结果做出的对比评价, 其中 frame per second 指的是每秒处理的帧数, 每秒处理的帧数越多, 算法的实时性能就越强。可见, 虽然三帧差分法的实时性较强, 但其精确率明显低于本文提出的算法, 而本文提出的算法在精确率、召回率和 F1 分数上均高于其他 3 种算法, 综合性能最强。

表 1 实验结果评价
Table 1 Evaluation of experimental results

Algorithm	Precision	Recall	F1-score	Frame per second
ViBe algorithm	0.7186	0.4535	0.5560	5.2629
Three-frame differential method	0.5354	0.6422	0.5839	9.9624
Gaussian mixture model algorithm	0.6552	0.3554	0.4617	4.1905
Proposed algorithm	0.8056	0.7070	0.7531	5.3419

4 总结

本文首先对背景提取技术和 ViBe 算法的优点与缺点做了总体上的描述,并对一些当前的改进 ViBe 算法作出了介绍,然后详细介绍了 ViBe 算法的原理与 YOLO v3-SPP 的结构,并在此基础上提出了一种基于 YOLO v3-SPP 的改进 ViBe 算法以解决传统 ViBe 算法中的鬼影问题。改进算法利用 YOLO v3-SPP 对静止行人的检测能力来改进 ViBe 算法本身的背景初始化策略,对检测出的行人进行了消除,从而解决了传统 ViBe 算法由于首帧出现行人而产生的鬼影问题。本文选择在 PETS2006 数据集下对所提出算法的性能与传统算法、三帧差分法和高斯混合模型算法进行比较测试。经比较后的实验测试结果表明,本文算法能够有效地解决传统 ViBe 算法的鬼影问题,且检测效果得到提升。

参考文献:

[1] Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos[J]. *Computer Vision and Image Understanding*, 2014, **122**(5): 4-21.

[2] Barnich O, Van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences[J]. *IEEE Transactions on Image Processing*, 2011, **20**(6): 1709-1724.

[3] 闫硕, 陈科山. 基于双背景模型的鬼影抑制方法研究[J]. *计算机应用与软件*, 2016, **33**(5): 162-165.

YAN Shuo, CHEN Keshan. Study on ghost suppression method based on double background model[J]. *Computer Applications and Software*, 2016, **33**(5): 162-165.

[4] 郭冲宇, 杜慧敏. 视觉背景提取算法关键技术研究[J]. *计算机应用研究*, 2017, **34**(5): 1548-1552,1589.

GUO Chongyu, DU Huimin. Key technologies of visual background extraction algorithm[J]. *Computer Application Research*, 2017, **34**(5): 1548-1552,1589.

[5] 陈树, 丁保阔. 一种基于改进视觉背景提取算法的前景检测[J]. *计算机工程与科学*, 2018, **40**(4): 673-680.

CHEN Shu, DING Baokuo. A foreground detection based on improved visual background extraction algorithm[J]. *Computer Engineering and*

Science, 2018, **40**(4): 673-680.

[6] HAN X, GAO Y, LU Z, et al. Research on moving object detection algorithm based on improved three frame difference method and optical flow[C]//*Fifth International Conference on Instrumentation & Measurement*, 2015: 580 - 584.

[7] 杨丹, 戴芳. 运动目标检测的 ViBe 算法改进[J]. *中国图象图形学报*, 2018, **23**(12): 1813-1828.

YANG Dan, DAI Fang. Improvement of ViBe algorithm for moving target detection[J]. *China Journal of Image Graphics*, 2018, **23**(12): 1813-1828.

[8] 方岚, 于凤芹. 去除鬼影及阴影的视觉背景提取运动目标检测算法[J]. *激光与光电子学进展*, 2019, **56**(13): 131002.

FANG Lan, YU Fengqin. Moving object detection algorithm based on removed ghost and shadow visual background extractor[J]. *Laser & Optoelectronics Progress*, 2019, **56**(13): 131002.

[9] ZHAO D W, TAN J J, YANG W B, et al. An improved vibe algorithm for fast suppression of ghosts and static objects[C]//*IEEE International Conference on Mechatronics and Automation(ICMA)*, 2018: 889 - 893.

[10] Girshick R. Fast R-CNN[C]//*IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[11] Redmon J Farhadi A. YOLO v3 anincremental improvement[EB/OL]. [2018-04-08] [2022-02-08], <https://arxiv.org/abs/1804.02767>.

[12] HUANG G, LIU Z, Weinberger K Q, et al. Densely connected convolutional networks[C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261-2269.

[13] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1904-1916.

[14] PETS2006. <http://www.cvg.reading.ac.uk/PETS2006/data.html>[DB/OL]. 2006.

[15] 赵建. 基于三帧差法的运动目标检测方法研究[D]. 西安: 西安电子科技大学, 2013.

ZHAO Jian. The Research of Moving Target Detection Method Based on Three-frame Difference[D]. Xi'an: Xidian University, 2013.

[16] STAUFFER C, GRIM SON W. Learning patterns of activityusing real-time tracking[J]. *IEEE Translations on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 747-757.

[17] Goyette N, Jodoin P M, Porikli F, et al. Change detection.net: a new change detection benchmark dataset[C]//*Computer Vision and Pattern Recognition Workshops*, 2012: 1-8.