



动态云台摄像机无人机检测与跟踪算法

谢家阳, 王行健, 史治国, 吴均峰, 陈积明, 陈潜, 王滨

引用本文:

谢家阳, 王行健, 史治国, 等. 动态云台摄像机无人机检测与跟踪算法[J]. *智能系统学报*, 2021, 16(5): 858–869.

XIE Jiayang, WANG Xingjian, SHI Zhiguo, et al. Drone detection and tracking in dynamic pan-tilt-zoom cameras[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(5): 858–869.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202103032>

您可能感兴趣的其他文章

多特征融合的异视角目标关联算法

Target association from different perspectives based on multi-feature fusion

智能系统学报. 2020, 15(5): 847–855 <https://dx.doi.org/10.11992/tis.202006037>

区域损失函数的孪生网络目标跟踪

Regional loss function based siamese network for object tracking

智能系统学报. 2020, 15(4): 722–731 <https://dx.doi.org/10.11992/tis.201910005>

面向环境探测的多智能体自组织目标搜索算法

Self-organizing target search algorithm of multi-agent system for environment detection

智能系统学报. 2020, 15(2): 289–295 <https://dx.doi.org/10.11992/tis.201908023>

仿猛禽视顶盖信息中转整合的加油目标跟踪

Aerial refueling target tracking using a falcon visual tectum information integrating like method

智能系统学报. 2019, 14(6): 1084–1091 <https://dx.doi.org/10.11992/tis.201909005>

一种自适应模板更新的判别式KCF跟踪方法

Adaptive template update of discriminant KCF for visual tracking

智能系统学报. 2019, 14(1): 121–126 <https://dx.doi.org/10.11992/tis.201806038>

高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function

智能系统学报. 2018, 13(3): 388–394 <https://dx.doi.org/10.11992/tis.201612040>

微信公众平台



关注微信公众号, 获取更多资讯信息

吴文俊人工智能技术发明奖一等奖

成果名称：低空飞行器智能监测关键技术及应用

获奖人：史治国、陈积明、吴均峰、严鹏、陈剑锋、孙优贤

完成单位：浙江大学、上海无线电设备研究所、浙江欧软低空防务技术有限公司



史治国

浙江大学信息与电子工程学院教授、博导、院长助理，浙江大学工程师学院副院长，IET Fellow, IEEE Senior Member, 中国电子学会高级会员，IEEE Network 编委，IET Communications 编委，Journal of The Franklin Institute 编委，浙江省151人才工程第一层次培养人员，2001年于浙江大学竺可桢学院/信息与电子工程学系获工学学士学位，2006年于浙江大学电子科学与技术专业获工学博士学位，2009年12月晋升副教授，2014年12月晋升教授。2011年9月至2013年11月在加拿大滑铁卢大学宽带通信研究组(BBCR)进行合作访问研究。2012年获浙江省科学技术二等奖，2015年获教育部科技进步一等奖，2016年获日本JSPS Invitation Fellowship。已授权发明专利70余项，发表学术论文100余篇，论文已被引用3400余次，作为项目负责人承担了国家重点研发项目课题、国家自然科学基金、浙江省重点研发计划项目等多项课题。主要研究兴趣包括信号处理及其定位应用。

史治国博士注重将最新的研究成果用于本科生教学，推动浙江大学信息与电子工程学院电子信息类主干课程贯穿式创新实践“个人实验室”研制与推广，取得了良好的教学效果，史治国博士2017年获得浙江大学唐立新教学名师奖，2020年获得浙江大学永平教学贡献奖。

团队简介

低空小型飞行器防控不仅仅是城市安全和隐私保护的问题，更是关系到国家安全与国家形象。研究团队面向低空小型无人机监测难题，针对基于视频、射频、音频等多种技术手段进行低空飞行器智能监测的主要技术挑战，课题组经过近十年的研发，发明了一系列低空飞行器监测技术，并应用于实际智能监测系统。团队在史治国和陈积明教授的领导下，在多源融合低空小型飞行器监测方面取得了一系列理论和关键技术成果，主要有：（1）提出了一种基于关键帧搜索目标匹配的分布式主从相机协同智能检测方法，并设计了基于神经网络的双流目标检测识别算法；（2）提出了基于信号时频特征和频谱特征时空融合学习的检测方法及基于张量阵列信号处理的目标角度估计方法；（3）提出了基于谐波能量包络特征与注意力机制的多特征融合残差神经网络检测方法，发明了基于贝叶斯推断和多谐波信号关联性估计的声源定位技术；（4）发明了基于关联性分析的异构传感器融合检测算法，设计研发了低功耗高速边缘视频智能处理硬件系统和基于Jacobi并行加速的射频定位硬件电路。成果在国家重大活动、关键基础设置以及城市公共安全保障方面得到了广泛的应用，并出口到阿联酋、沙特等国家，经济和社会效益显著。

DOI: 10.11992/tis.202103032

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210830.1334.006.html>

动态云台摄像机无人机检测与跟踪算法

谢家阳¹, 王行健¹, 史治国¹, 吴均峰¹, 陈积明¹, 陈潜², 王滨³

(1. 浙江大学信息学部, 浙江杭州 310027; 2. 上海无线电设备研究所, 上海 200090; 3. 杭州海康威视数字技术股份有限公司, 浙江杭州 310052)

摘要: 为应对小型无人机的黑飞、滥飞对个人隐私、公共安全造成的威胁, 本文采用高清云台摄像机定点巡航的方式对近地动态复杂背景中的无人机进行检测与跟踪, 并提出了一种适用于动态云台摄像机的闭环无人机检测与跟踪算法, 包含检测与跟踪两种模式。在检测模式下, 本文设计了一种基于运动背景补偿的运动目标检测算法来提取分类候选区域, 然后利用基于神经网络结构搜索得到的轻量级卷积神经网络对候选区域进行分类识别, 可在不缩小高清视频图像的条件下实现无人机检测; 在跟踪模式下, 本文提出了一种结合卡尔曼滤波的局部搜索区域重定位策略改进了核相关滤波跟踪算法, 使之在高清云台伺服追踪过程中仍能对目标进行快速稳定的跟踪; 为将检测模式与跟踪模式结合在闭环框架中, 本文还提出了一种基于检测概率和跟踪响应图状态的自适应检测与跟踪切换机制。实验表明, 本文算法可应用于定点巡航状态的高清云台摄像机, 实现近地复杂动态背景中无人机的实时准确检测、识别与快速跟踪。

关键词: 反无人机; 运动目标检测; 背景运动补偿; 目标识别; 目标跟踪; 深度学习; 机器视觉; 卷积神经网络
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2021)05-0858-12

中文引用格式: 谢家阳, 王行健, 史治国, 等. 动态云台摄像机无人机检测与跟踪算法 [J]. 智能系统学报, 2021, 16(5): 858-869.

英文引用格式: XIE Jiayang, WANG Xingjian, SHI Zhiguo, et al. Drone detection and tracking in dynamic pan-tilt-zoom cameras[J]. CAAI transactions on intelligent systems, 2021, 16(5): 858-869.

Drone detection and tracking in dynamic pan-tilt-zoom cameras

XIE Jiayang¹, WANG Xingjian¹, SHI Zhiguo¹, WU Junfeng¹,
CHEN Jiming¹, CHEN Qian², WANG Bin³

(1. Faculty of Information Technology, Zhejiang University, Hangzhou 310027, China; 2. Shanghai Radio Equipment Research Institute, Shanghai 200090, China; 3. Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou 310052, China)

Abstract: To cope with the threat to personal privacy and public safety caused by illegal uses of small unmanned aerial vehicles, we propose a closed-loop UAV detection and tracking algorithm suitable for dynamic PTZ camera, which uses the fixed-point cruise mode of high-definition PTZ camera to detect and track unmanned aerial vehicles in the near-earth dynamic complex background, including two modes of detection and tracking. In the detection mode, this paper designs a moving target detection algorithm based on moving background compensation to extract the classification candidate areas, and then uses the lightweight convolutional neural network based on neural network structure search to classify and identify the candidate areas, which can realize UAV detection without reducing the high-definition video image; In the tracking mode, this paper proposes a local search area relocation strategy combined with Kalman filter to improve the tracking algorithm of kernel correlation filter, so that it can still track the target quickly and stably in the servo tracking process of HD PTZ; In order to combine detection mode and tracking mode in a closed-loop framework, this paper also proposes an adaptive detection and tracking switching mechanism based on detection probability and tracking response graph state. Experiments show that the proposed algorithm can be applied to high-definition PTZ cameras in a fixed-point cruise state, and realize real-time accurate detection, recognition and fast tracking of UAV in complex dynamic background near the ground.

Keywords: anti-UAV; moving target detection; background motion compensation; object recognition; object tracking; deep learning; computer vision; convolutional neural network

在其市场快速增长的同时,无人机的“黑飞”与“滥飞”也对个人隐私及公共安全产生了严重的威胁和挑战^[2]。无人机防控不仅是城市安全和隐私保护的问题,更是关系到国家安全与形象的大事,因此迫切需要研发一整套完整有效的技术手段以实现对低空飞行器的监测和防控。

Shi等^[3]对反无人机监控手段进行详细的介绍和对比,反无人机监测主要有雷达、射频、音频及视觉探测4种方式,4种方式各有优劣。雷达探测通过无人机反射的回波测量无人机的速度、方位,但雷达成本高、辐射大、易受地面杂波影响,不适用于城市环境;射频探测利用无人机与遥控器之间的通信或图传信号对无人机进行检测与定位,但射频天线成本较高,检测受无人机通信协议限制;音频探测利用麦克风阵列对无人机旋翼发出的声音进行检测与定位,但作用距离较短,且易受环境噪声和多径效应影响。视觉检测利用摄像机获取监控区域的视频,利用图像特征进行无人机检测与跟踪。由于视觉监测方式成本低、普适性强、检测距离较远且能直观显示检测结果,已成为反无人机监测领域研究的热点。

目标检测是计算机视觉最基础的任务之一,自计算机视觉诞生以来,一直是学术界研究的重点领域。视觉目标检测算法可分为传统目标检测算法和深度学习目标检测算法。传统目标检测算法通常使用滑动窗口生成大量候选区域,然后在提取其人工设计特征后使用机器学习算法对目标进行检测与识别,但这些方法检测效率低下,且对复杂背景中形态多样、姿态多变的检测效果不佳^[4-5]。

随着图形处理器(graphics processing unit, GPU)并行计算技术的快速发展与深度学习算法的兴起,基于深度学习的目标检测算法发展迅速,其主要可以分为一阶段和二阶段目标检测算法。一阶段深度学习目标检测将图像中的所有位置均视作潜在目标,通过神经网络直接产生目标的类别与位置,检测速较快,但精度相对于二阶段目标检测算法略有不如,代表算法有基于anchor机制的YOLO(you only look once)系列算法^[6]、SSD(single shot multibox detector)算法^[7]和基于特征点的CornerNet^[8]、CenterNet^[9]。除了基于深度卷积神经网络架构的目标检测算法外,基于Transformer的目标检测算法也是一阶段目标检测算法的重要分支^[10]。二阶段深度学习算法将目标检测划分为候选区域生成、候选区域分类与边界框调整两个阶段,代表算法为RCNN(region-based convolutional neural networks)系列目标检测算法^[11]。候选区域生成的目的在于快速准确地找出图像中目标可能存在的区域,除了最暴力的多尺度滑动

窗口法,研究者们还提出了基于图像分割的超像素聚类候选区域生成方法,如基于图结构的图像分割算法^[12]和选择性搜索算法^[13],此类方法消除规则网格、固定形状与尺度对候选区域的限制。窗口评分候选区域生成方法如Objectness^[14]、EdgeBoxes^[15]等是另一类重要候选区域生成方法,此类算法对每个滑动矩形窗口是否包含物体进行评分,并只能返回矩形候选框而非目标轮廓,速度相较于超像素聚类方法更快。除非窗口采样比较密集,窗口评分候选区域生成法的位置精度都较低。Faster RCNN^[16]首次将基于锚框的候选区域生成网络(region proposal network, RPN)引入并将其嵌入到Fast RCNN的检测网络中,实现了整个检测框架端到端的优化。如何实现快速准确且适用于尺度目标检测的候选区域生成方法仍是目标检测研究领域的重要研究方向之一。第二阶段对候选区域进行分类与边界框调整一般通过卷积神经网络实现,经典的卷积神经网络模型有AlexNet、VGG Net、GoogleNet、ResNet、MobileNet、ResNeX和ShuffleNet等^[17],为兼顾神经网络模型的可训练性、推理速度与精度,对于特征任务,现在通常采用神经网络结构搜索的方法以获取最优的网络模型^[18]。在检测到无人机后,需要控制云台对快速飞行的无人机进行持续伺服追踪,使无人机一直保持在摄像机的视角中^[19]。在云台伺服追踪过程中,仍需对视频中的无人机进行检测,为云台的控制算法提供反馈信息。在确认视频图像中目标的类别与位置后,仍持续地对后续的每一帧高分辨率图像进行全局检测是对计算资源的浪费,为快速高效地对画面中位置已知的无人机进行跟踪,本文采用基于局部搜索的图像单目标跟踪算法对后续视频中的无人机进行跟踪定位。基于局部检测的相关滤波目标跟踪算法是单目标跟踪算法的经典方法之一,此类算法利用循环卷积和快速傅里叶变换高效地训练分类器,并可以快速对局部区域内的跟踪目标进行检测与定位,其代表算法有最小均方误差和输出跟踪算法^[20]、核相关滤波跟踪算法^[21]、尺度自适应跟踪算法^[22]等。在刚引入深度学习时,只是使用卷积神经网络特征替换方向梯度直方图(histogram of oriented gradient, HOG)^[5]等手工特征,使相关滤波器能获得更鲁棒的跟踪能力,如C-COT(continuous convolution operators tracking)^[23]、ECO(efficient convolution operators)^[24]等。完全端到端的深度学习跟踪算法肇始于2016年牛津大学Luca Bertinetto等^[25]提出的Siamese-fc孪生网络目标跟踪,在此基础上Siam RPN^[26]、Siam Mask^[27]等深度学习目标跟踪相继提出。虽然深度学习目标跟踪算法的性能

更佳,其检测速度比相关滤波目标跟踪算法慢得多,但是这种可实现快速处理的相关滤波目标跟踪算法在后续帧的局部搜索区域会随运转的摄像机发生偏移,导致跟踪失败。

为实现全方位、高复杂、远距离环境下的对无人机的实时准确检测、识别与跟踪,本文通过高清云台摄像机对监测范围进行持续动态扫描,并提出了一种反无人机视觉闭环检测与跟踪框架。本文的主要贡献如下:

1) 提出了一种适用于云台摄像机定点巡航的二阶段目标检测算法,此算法首先对由于摄像机运动产生的视频背景运动进行补偿,然后使用三帧差分法生成运动目标候选区域,并使用由神经网络结构搜索得到的深度卷积神经网络进行分类识别,可在不缩小高清视频图像的情况下实现对复杂动态背景中无人机的快速准确检测;

2) 提出了一种基于背景补偿和卡尔曼滤波的

局部搜索区域重定位策略改进了核相关跟踪算法,解决了由于高清云台摄像机伺服追踪引局部搜索区域偏移问题,在视频背景运动的情况下也能对快速飞行的无人机进行准确、稳定地跟踪;

3) 提出了一种自适应切换机制将检测模块和跟踪模块结合成一个闭环系统,通过深度卷积神经网络的分类结果自动为跟踪模块提供初始化跟踪框,并使用相关滤波跟踪响应图的平均峰值能量比和最大峰值能量来表征跟踪效果的优劣,在跟踪失败时自动切换至检测模式。

1 算法基本框架

本文提出了一种适用于云台摄像机动态扫描状态的自适应闭环无人机检测与跟踪算法,包含检测与跟踪两种模式,并通过一种自适应切换机制将检测模式与跟踪模式有机结合在闭环框架中,算法的框架图如图 1 所示。

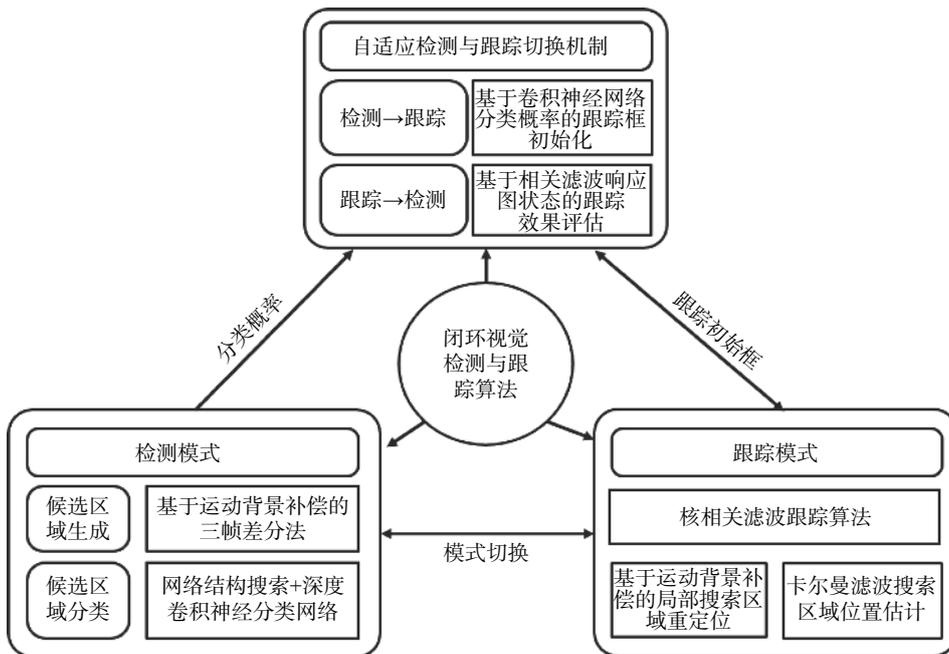


图 1 算法框架

Fig. 1 Algorithm framework

首先本文采用高清云台摄像机对中近程近地动态复杂背景进行定点巡航扫描,此时摄像机处于检测模式,使用基于运动背景补偿三帧差分法提出候选目标并通过一个轻量级的深度卷积神经网络对候选目标进行分类;根据自适应检测-跟踪切换机制,当目标属于无人机的概率大于预设阈值时,摄像机切换到跟踪模式,采用引入了局部搜索区域重定位策略的相关滤波跟踪算法对无人机进行跟踪;自适应检测-跟踪机制将根据相关滤波响应图的状态对跟踪效果进行评估,若跟踪效果较好,云台摄像机将根据跟踪结果对无

人机进行持续地伺服追踪;若跟踪失败,云台摄像机则重新切换到定点巡航检测模式。算法的可视化流程如图 2 所示。

2 检测模式算法设计

云台摄像机定点巡航状态下的检测模式采用如图 2 所示的两阶段目标检测框架。准确且快速的候选区域生成对实时目标检测至关重要。入侵的无人机基本都是处于飞行状态,故可采用运动目标检测算法提取候选区域,然后再使用深度卷

积神经网络对其进行分类识别。但定点巡航状态的云台摄像机采集的视频背景是动态变化的, 无法使用传统的帧间差分法或背景减除法进行运动目标检测。光流法可检测运动背景中的运动目

标, 但其计算量太大, 无法实现对高清视频的实时处理。为实现摄像机巡航状态下运动目标检测, 本文提出了一种基于运动背景补偿的三帧差分法对疑似运动目标进行检测。

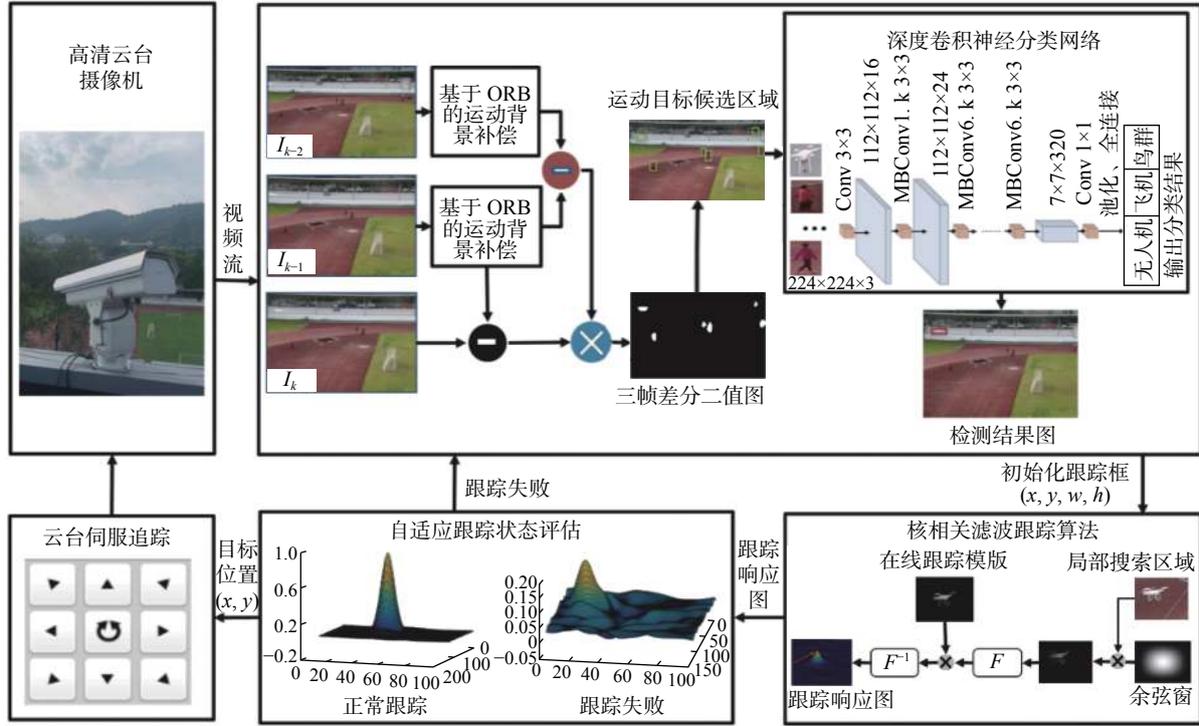


图 2 算法可视化流程

Fig. 2 Visualization of the algorithm flowchart

2.1 运动背景补偿

为消除定点巡航时旋转的云台摄像机产生的自运动效应, 本文提出了一种运动背景补偿算法对前后视频帧进行背景对齐。在计算机视觉中, 平面的单应性被定义为从一个平面到另一个平面的投影映射。由于远距离的摄像机成像场景可被近似成一个平面, 并且无人机等运动目标仅占视频画面的极小部分, 在云台摄像机旋转时, 连续两帧视频背景之间由于摄像机转动产生的运动可由一个 3×3 的单应性矩阵进行刻画^[28]。本文中用 I_k 指代第 k 帧视频图像, 根据单目摄像机对极几何和射影变换, 图像 I_k 中坐标位置为 $[x_i^k, y_i^k]$ 的背景特征点在云台摄像机经过微小旋转后拍摄的图像 I_l 中的坐标 $[x_i^l, y_i^l]$ 为

$$\begin{bmatrix} x_i^l \\ y_i^l \\ 1 \end{bmatrix} = \mathbf{H}_{kl} \begin{bmatrix} x_i^k \\ y_i^k \\ 1 \end{bmatrix}$$

式中: $\mathbf{H}_{kl} \in \mathbf{R}^{3 \times 3}$ 为两帧中对应特征点坐标转化的单应性矩阵; $i = 1, 2, \dots, N$, N 为图像中特征点的数量。首先选取云台摄像机旋转过程中的邻近的两帧视频图像 I_l 和 I_k , 分别提取两帧图像中的 ORB(oriented FAST and rotated BRIEF)^[29] 特征点, 然后利用快速最近邻逼近搜索对两帧图像中的

ORB 特征点进行初步匹配, 由于视频帧率较高, 拍摄连续几帧视频时摄像机转动的角度较小, 则两帧中对应的特征点相距较近, 故可以设置阈值滤除汉明距离过大的错误特征点对, 以提高匹配的准确率。为进一步减弱错误匹配的特征点对和前景特征点对对单应性矩阵估计的影响, 本文采用随机抽样一致性算法^[30]对特征点对进行采样, 然后使用最小二乘法对单应性矩阵进行估计, 最后对视频图像 I_l 上的每一个像素点都进行坐标变换, 即可得到背景与 I_k 对齐的图像 I_{kl} 。运动背景补偿前后的运动目标检测效果图对比如图 3 所示。

2.2 运动目标检测

为消除复杂背景对目标检测的影响和快速提取目标候选区域, 本节将对运动目标候选区域生成算法进行介绍。由于背景减除法^[31]需要对背景进行建模与持续维护, 不适用于快速运动的背景, 而帧间差分法^[32]直接将相邻视频帧作为背景, 背景不积累, 算法简单、计算量小, 比较适合本文场景。帧间差分法对相邻帧图像对应位置像素作差值计算, 由于前后帧的背景已通过运动背景补偿进行配准, 若相邻帧图像对应位置的像素差小于预设阈值, 则可认为此处为背景像素; 反之, 若相邻帧图像对应位置的像素差值大于某阈

值, 则标记为前景, 即视频中运动目标所在区域。但无人机往往飞行较快, 导致帧间差分法受到“鬼影”干扰, 即无人机在前后帧的所在区域都会被误认为运动目标候选区域, 造成误检。为了消除“鬼影”现象, 本文采用了三帧差分法对运动背景补偿后的连续三帧视频图像 $I_{k(k-1)}$ 、 I_k 、 $I_{k(k+1)}$ 进行处理。具体地, 先分别对 $I_{k(k-1)}$ 、 I_k 和 I_k 、 $I_{k(k+1)}$ 应用帧间差分法得到两张差分二值图 D_k^1 和 D_k^2 , 然后对 D_k^1 和 D_k^2 做与运算 (\cap), 得到三帧差分二值图 D_k , 其计算过程为

$$D_k^1(x,y) = \begin{cases} 1, & I_k(x,y) - I_{k(k-1)}(x,y) \geq \theta \\ 0, & I_k(x,y) - I_{k(k-1)}(x,y) < \theta \end{cases}$$

$$D_k^2(x,y) = \begin{cases} 1, & I_k(x,y) - I_{k(k+1)}(x,y) \geq \theta \\ 0, & I_k(x,y) - I_{k(k+1)}(x,y) < \theta \end{cases}$$

$$D_k(x,y) = D_k^1(x,y) \cap D_k^2(x,y)$$

式中 θ 为二值化阈值 (在本文中设置为 25)。由于运动目标内部颜色相近, 会导致检测出的运动目标区域内部出现空洞, 另外差分二值图的与操作 (\cap) 会加剧空洞的产生, 为得到更为准确、完整的运动目标候选区域, 在进行与操作前, 先对差分二值图 D_k^1 和 D_k^2 做了形态学闭处理^[33], 减少内部空洞, 将距离较近的前景区域进行连接整合, 在得到三帧差分图 D_k 后, 对其中的前景区域进行了最小凸包检测^[34], 获得更加完整的运动目标候选区域。帧间差分法、三帧差分法及它们经后处理的效果对比如图 4 所示。

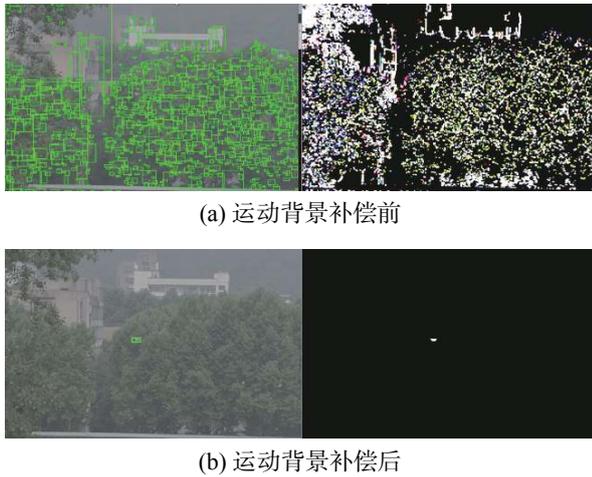


图 3 运动背景补偿前后的运动目标检测效果对比
Fig. 3 Comparison of the moving target detection results before and after background motion compensation

2.3 深度卷积神经网络分类器

在生成运动目标候选区域后, 本文设计了一个深度卷积神经网络对候选区域进行分类, 判断其是否为无人机。为提高检测速度, 可以先根据无人机形状的先验信息将区域外接矩形纵横比小于 1 或大于 3 的候选区域剔除。为兼顾分类的实时性和准确度, 本文以运算量和准确度为优

化目标, 结合复合模型扩张方法与神经网络结构搜索寻找面向分类任务最优的网络深度、宽度与分辨率^[18], 设计了一个高效的轻量级深度卷积神经网络 EfficientNet-B0, 其主干网络的结构如图 2 和表 1 所示。

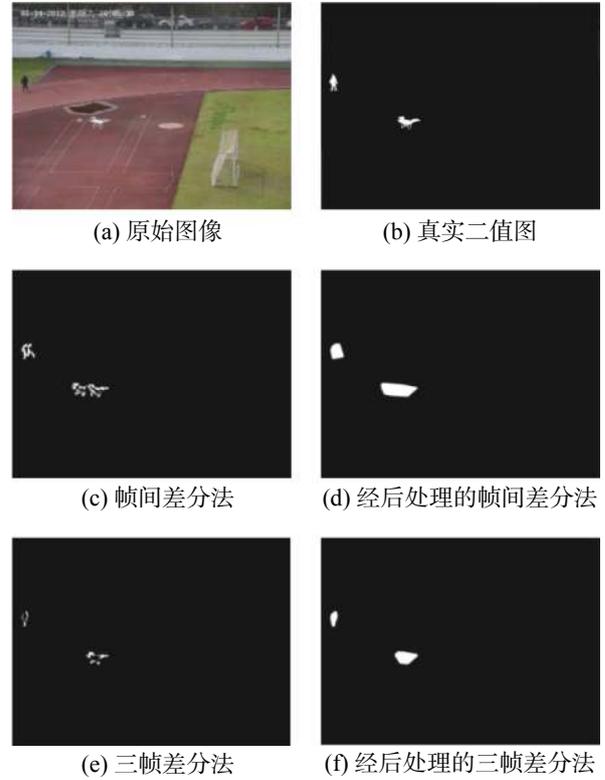


图 4 不同运动目标区域生成算法效果对比
Fig. 4 Comparisons of different moving region proposal generation methods

表 1 EfficientNet-B0 的网络结构
Table 1 Architecture of EfficientNet-B0

阶段	模块Operator	尺寸(w×h×c)	层数
1	Conv3×3	224×224×32	1
2	MBCConv1,k3×3	112×112×16	1
3	MBCConv6,k3×3	112×112×24	2
4	MBCConv6,k5×5	56×56×40	2
5	MBCConv6,k3×3	28×28×80	3
6	MBCConv6,k5×5	14×14×112	3
7	MBCConv6,k5×5	14×14×192	4
8	MBCConv6,k3×3	7×7×320	1
9	Conv1×1&Pooling&FC	7×7×1280	1

首先将输入图像缩放为 224×224, 经过 3×3×32 的卷积核 7 个串联的 MBCConv(mobile inverted residual bottleneck convolution) 模块来循环扩张和压缩特征通道, 以获取更好的学习能力, 增加网络的深度。其中, MBCConv 模块通过 1×1 卷积核将特征进行通道扩展, 经过 3×3 和 5×5 的深度可分卷积核进行通道特征提取后, 由 1×1 卷积核进

行特征通道压缩,输出经尺寸压缩及通道扩张后的特征图。此外在 MBConv 模块中,还引入通道注意力机制 (squeeze-and-excitation block, SE)^[35] 和神经元随机失活机制^[36],选择性加强包含有用信息的特征,降低过拟合风险。7层 MBConv 模块输出的 $7 \times 7 \times 320$ 特征图经过 1×1 卷积核的通道压缩、池化层的空间压缩后,将特征张量扁平化为一维向量,输入全连接层利用 sigmoid 函数输出一维概率向量。该预测向量给出目标属于各个类别的概率,并选择最大值所在的类别作为分类结果,当候选目标属于无人机的概率大于预设的检测-跟踪切换阈值时,使用该目标的检测框初始化跟踪算法。

3 跟踪模式算法设计

当检测到无人机后,算法将切换到跟踪模式,在此模式下,结合局部搜索区域重定位策略的核相关滤波跟踪算法将对经检测算法初始化的无人机在画面中的位置进行持续跟踪,同时为使无人机保持在视频画面中,云台控制算法以目标在视频画面中离视频中心的距离与方位为反馈量,采用比例-积分-微分 (proportion integration differential, PID) 控制算法控制云台转动对无人机进行伺服跟踪^[19]。

3.1 核相关滤波跟踪算法原理

由 Henriques 等^[21]提出的核相关滤波跟踪算法是一种判别式局部区域跟踪算法,在当前帧训练一个局部区域目标检测器,使用该目标检测器在下一帧的邻近区域对目标进行搜索,并使用新的检测结果更新目标检测器,检测速度可达每秒几百帧。为避免密集采样带来的重复特征提取和复杂训练过程的高耗时,核相关滤波跟踪器以检测器提供的初始检测框为正样本,提取其 fHOG^[37]特征图 $p \in \mathbf{R}^{m \times n \times l}$,其中 m 、 n 、 l 分别为 fHOG 特征图的长、宽与通道数,通过循环移位对所提取的 fHOG 特征图进行近似密集采样生成样本集合 $S = \{(\mathbf{p}_i, g_i) | i = 1, 2, \dots, (m \times n)\}$,其中 \mathbf{p}_i 为合成样本, $g_i \in [0, 1]$ 为其对应标签,所有的 g_i 根据空间排布可生成一张对应于循环移位大小的高斯分布图 $G \in \mathbf{R}^{m \times n}$,离中心越近,标签值越大。此局部目标检测器可建模成一个岭回归优化问题^[38]:

$$\min_w \sum_{i=1}^{m \times n} (\phi(\mathbf{p}_i) \mathbf{w} - g_i)^2 + \lambda \|\mathbf{w}\|^2$$

式中: $\phi(\cdot)$ 函数可将低维空间的样本映射到高维空间; λ 为正则化因子。根据表示定理^[39],此优化问题的最优解可表示为 $\mathbf{w} = \sum_i \alpha_i \phi(\mathbf{p}_i)$,则优化问

题转化为求最优的线性组合系数 α ,由于训练样本均是由循环位移产生的,利用循环矩阵的性质^[40],将分类器训练过程转化到频域中计算,将复杂费时的矩阵求逆操作转化为简单快速的元素点乘,最优的线性组合系数 α 为

$$\alpha = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(G)}{\mathcal{F}(k^{pp}) + \lambda} \right)$$

$$k^{pp} = \exp \left(-\frac{1}{\sigma^2} (\|p\|^2 + \|p'\|^2 - 2\mathcal{F}^{-1} \left(\sum \mathcal{F}(p) \circledast \mathcal{F}(p) \right)) \right)$$

式中: \mathcal{F} 和 \mathcal{F}^{-1} 分别代表离散傅里叶变换和反向离散傅里叶变换; $*$ 为复共轭算子; \circledast 表示元素点乘; \oslash 代表元素点除; σ 为高斯函数的方差; k^{pp} 是初始检测框 fHOG 特征图 p 的高斯核函数。

在原核相关滤波算法中,该检测器是在上一帧目标所在位置的中心点的矩形区域内进行搜索,得到跟踪响应图 $\Omega = \mathcal{F}^{-1} (\mathcal{F}(k^{pq}) \circledast \mathcal{F}(\alpha))$,其中 k^{pq} 为初始检测框与搜索框对应的 fHOG 特征图 p 、 q 的高斯核函数。对应于跟踪响应图 Ω 峰值的坐标即为被跟踪目标的新位置。核相关滤波跟踪算法的可视化流程图如图 2 所示。

3.2 局部搜索区域重定位

由于云台摄像机在追踪无人机过程中自身的转动,即使无人机在物理空间中是静止的,其在视频不同帧中的图像坐标也将出现变化,另外当无人机飞行速度较快时,它可能会飞出局部搜索框,为了更准确地跟踪无人机,在上一帧的相同位置对无人机进行跟踪是不合理的,为实现在动态追踪的云台摄像机中对快速运动的无人机进行跟踪,本文对无人机在下一帧中的搜索位置进行预测和补偿,以实现搜索区域重定位,使核相关滤波跟踪算法在云台摄像机转动导致背景运动时仍能对快速飞行的无人机进行稳定的跟踪。

为消除云台转动引起的局部搜索区域偏移,本文将通过运动背景补偿实现。与 2.1 小节所述方法相同,首先通过在跟踪前后两帧图像 I_{k-1} 与 I_k 中提取 ORB 特征点并配对,为避免被跟踪目标的特征点对补偿结果的影响,去除在原跟踪框内的特征点,得到剩余特征点对,其中 $[x_i^{k-1}, y_i^{k-1}]$ 和 $[x_i^k, y_i^k]$ 分别是 I_{k-1} 与 I_k 中对应特征点的图像坐标,其中 $i = 1, 2, \dots, N$, N 为特征点对的数量。这些特征点的位移可被视作画面整体的偏移,则在新的一帧中,搜索框中心的水平补偿偏移 Δx^k 与竖直补偿偏移 Δy^k 分别为

$$\Delta x^k = \frac{1}{N} \sum_{i=1}^N (x_i^k - x_i^{k-1})$$

$$\Delta y^k = \frac{1}{N} \sum_{i=1}^N (y_i^k - y_i^{k-1})$$

在补偿搜索区域位置由于云台转动引起的自运动偏移后,为进一步提高核相关滤波算法对快速飞行无人机的跟踪性能,本文将通过卡尔曼滤波^[41]对新一帧中的局部搜索区域位置进行估计。假设无人机在第 k 帧时的图像坐标 Z^k 为 $[x^k, y^k]$,而第 $k-1$ 帧-第 k 帧由于云台引起的自运动像素偏移为 $[\Delta x^k, \Delta y^k]$,则经云台自运动偏移修正后的目标坐标 Z_m^k 为 $[x_m^k, y_m^k] = [x^k - \Delta x^k, y^k - \Delta y^k]$,无人机的实际图像速度 $[v_x^k, v_y^k] = [x_m^k - x_m^{k-1}, y_m^k - y_m^{k-1}]$,由以上定义,视频中无人机的运动模型为

$$\begin{aligned} X_m^k &= AX^{k-1} + w^k \\ Z_m^k &= BX_m^k + v^k \end{aligned}$$

式中: $A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ 分别为

为模型的状态转移矩阵和观测矩阵; $X_m^k = [x_m^k, y_m^k, v_x^k, v_y^k]$ 为第 k 帧经背景运动偏移修正后的运动状态向量;过程噪声 $w^k \sim N(0, Q)$ 和观测噪声 $v^k \sim N(0, R)$ 都是均值为 0 的高斯噪声; Q 和 R 分别是对应的噪声协方差矩阵。则基于卡尔曼滤波的局部搜索区域位置预测过程为

$$\begin{aligned} \hat{P}_k &= AP_{k-1}A^T + Q \\ K_k &= \hat{P}_k B^T (A \hat{P}_k A^T + R)^{-1} \\ \hat{X}_m^k &= A \hat{X}_m^{k-1} \\ \hat{X}_m^k &= \hat{X}_m^k + K_k (Z_m^k - B \hat{X}_m^k) \\ P_k &= (I - K_k B) \hat{P}_k \end{aligned}$$

式中: P_k 与 \hat{P}_k 分别为第 k 帧的预测误差协方差矩阵和估计误差协方差矩阵; \hat{X}_m^k 和 \hat{X}_m^k 分别代表目标的预测运动向量和估计运动向量。在经过卡尔曼滤波和运动背景补偿后,第 k 帧局部搜索区域位置被重定位至 $[\hat{x}_m^k + \Delta x^k, \hat{y}_m^k + \Delta y^k]$ 。

图 5 给出了 2 种不同情况下局部搜索区域重定的跟踪效果示意图,图 5(a) 为无人机运动较慢情形,是否进行局部搜索区域重定位在这种条件下对目标跟踪效果影响不大,但是在图 5(b) 所示无人机运动速度较快或者云台旋转较快时,无人机在第 k 帧中将处于未经重定位的黄色局部搜索区域外,而在经过局部搜索区域重定位后,无人机在第 k 帧中将仍处于蓝色局部搜索区域内。



(a) 无人机运动或云台转动较慢场景



(b) 无人机运动或云台转动较快场景

图 5 局部搜索区域重定位跟踪效果示意

Fig. 5 Illustration of tracking algorithms with and without local search area relocation

4 自适应检测跟踪切换机制

为实现检测模式与跟踪模式的自适应切换,本文设计了一种自适应切换机制,将前文提到的检测模式与跟踪模式整合到一个闭环框架中。如前文所述,检测模式的检测结果可为跟踪算法提供精确可靠的初始框,具体地,当深度卷积神经网络分类结果中,属于无人机的概率超过某阈值的目标检测框即跟踪算法的初始跟踪框。

在对无人机进行跟踪的过程中,遮挡、光线变化、模糊、目标变形等因素均会造成跟踪失败,一旦跟踪失败,算法会自动切换至检测模式,为此本文根据核相关滤波跟踪响应图提出了一种自动跟踪效果评估机制。如图 2 所示,无人机在搜索框中的位置是由响应图 Ω 中峰值对应的位置决定的,跟踪响应图一定程度上也可以反映跟踪器的跟踪效果,如图 6 所示。在正常跟踪时,响应图一般只有一个较大的尖峰,其余部分都较为平坦;在目标被遮挡或消失导致跟踪失败时,其响应图可能出现多个尖峰,峰值均较小,且各处凹凸不平,为量化描述跟踪响应图的状态,本文提出使用平均峰值能量比 ρ 来表征跟踪效果的优劣:

$$\begin{aligned} \rho &= \tau^2 / \sum_{i=1}^m \sum_{j=1}^n \Omega(i, j)^2 \\ \tau &= \max_{i=1, j=1}^{m, n} (\Omega(i, j)) \end{aligned}$$

式中: τ 为响应图的最大峰值。如图 6 所示,在正常跟踪时,响应图的平均峰值能量比 ρ 可达 152.6,而跟踪失败时, ρ 仅为 8.4。图 7 为某次跟踪平均峰值能量比 ρ 在不同帧的变化曲线,在开始时 ρ 在较高值附近上下波动,此时认为跟踪正常,继续跟踪;当 ρ 迅速减小,本文设定当最新视频帧中平均峰值能量比 ρ 小于其历史平均值的 1/2 时,判定跟踪失败,立即切换至检测模式。

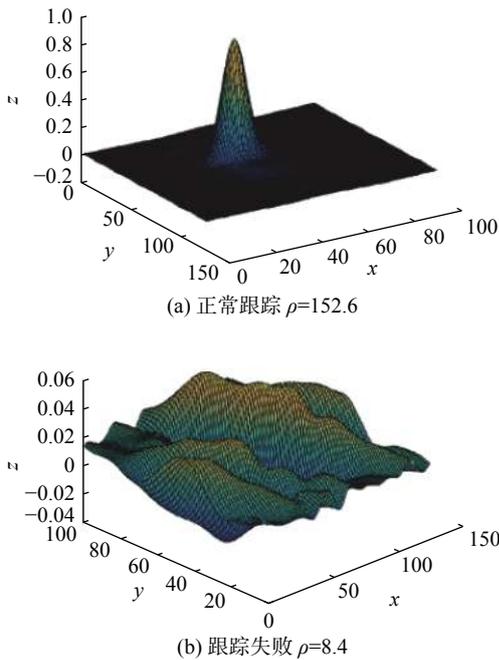


图 6 正常跟踪和跟踪失败时的响应图对比
Fig. 6 Response maps of good and bad tracking

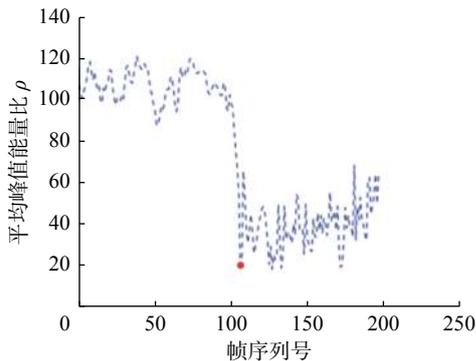


图 7 平均峰值能量比 ρ 变化曲线
Fig. 7 Changing curve of the mean peak energy ratio

5 实验与分析

为验证本文提出的反无人机视觉闭环检测与跟踪算法的有效性, 本文对算法的不同模式和整体性能都做了详尽的测试和评估。

5.1 运动目标候选区域生成算法

为测试检测模式中的基于运动背景补偿的运动目标检测算法, 本文采用开源数据集 changeDetection^[42] 中 PTZ 摄像机目标下的 4 个图像序列作为测试集, 使用准确率 P 、召回率 r 和 F_1 分数对检测结果进行评估, 其具体定义为

$$P = \frac{TP}{TP + FP}, r = \frac{TP}{TP + FN}, F_1 = \frac{2pr}{p+r}$$

式中: TP 为检测正确的前景像素点个数; FP 为将背景检测成前景的像素个数 (误检); FN 为将前景检测为背景的像素个数 (漏检)。本文提出的算法

与已有算法 EFIC^[43]、WeSamBE^[44] 在此数据集上的比较结果如表 2 所示。

表 2 运动目标检测算法性能比较
Table 2 Comparisons of moving target detection

算法	准确率	召回率	F_1 分数
EFIC	0.918	0.528	0.671
WeSamBE	0.814	0.313	0.452
本文算法	0.720	0.701	0.710

由表 2 可知, 本文提出的算法虽然在准确率上相较于 EFIC、WeSamBE 较低, 但其召回率与 F_1 参数均是 3 种算法中最高的, 在候选区域生成阶段应尽可能发现可能的目标, 故相较于准确率, 候选区域生成算法更关注召回率指标。

5.2 深度卷积神经分类网络

为评估本文中的深度卷积神经分类网络, 本文构建了一个无人机分类数据集, 包含复杂环境下的无人机图像作为正样本, 共 600 张, 另包含了飞机、鸟类、行人、树木等负样本图像, 共 200 张, 如图 8 所示, 数据集中训练集和测试集的比例为 7:3。



图 8 无人机分类数据集
Fig. 8 Drone classification dataset

为在达到高精度的同时兼顾运算速度, 实验通过网络结构搜索, 共生成了 8 种不同尺寸的网络 (EfficientNet- $b_0 \sim b_7$), 其参数量依次增加, 前向计算速度逐渐降低。由于在本文面向的场景中, 分类类别数量较少, 如果盲目应用参数量较大的网络, 反而会提升过拟合风险并降低分类精度。为实现精度与速度的平衡, 实验在所构成的分类数据集上依次评测所有生成的 EfficientNet 系列网络, 使用单块 Tesla T4 GPU, 实验结果如图 9 所示。结果表明, EfficientNet- b_0 的分类准确率和速度可达到 97.2% 和 258 FPS, 在达到无人机辨别任务精度要求的同时, 满足检测系统的速度要求, 综合性能最好。

5.3 核相关滤波跟踪算法实验结果

为评估局部搜索区域重定位核相关滤波跟踪算法, 本文构建了一个无人机检测与跟踪测试数据集, 包含 24 段云台摄像机拍摄视频片段, 图像大小均为 2 048×1 536, 帧率为 24 FPS, 包含不同天气、背景和距离下不同种类、尺度和姿态的无人机。本文利用距离精度 (precision) 和成功率

(success rate) 为指标来量化算法的跟踪效果, 定义中心位置误差 (center location error, CLE) 为跟踪框与真实框中心点之间的距离, 定义跟踪框与真实框相交的面积与其共同占据的面积之比为交并比 (intersection over union, IOU)。距离精度为 CLE 小于某阈值的帧数占总帧数之比, 而跟踪成功率则定义为 IOU 大于某阈值的帧数占总帧数之比, 在不同阈值下的距离精度和成功率可由距离精度图和成功率图表示。

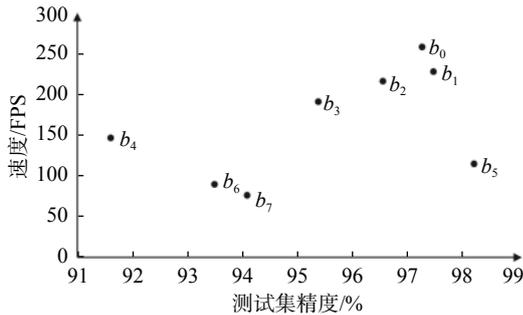
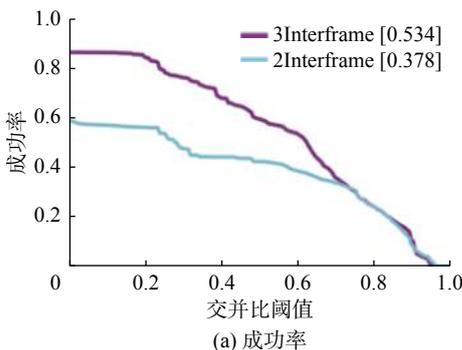


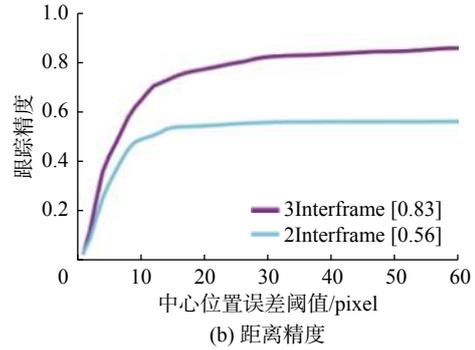
图 9 EfficientNet 系列网络的性能对比评估

Fig. 9 Performance assessment of EfficientNet networks

由于影响闭环检测与跟踪算法的因素很多, 为详尽地探究不同因素对跟踪性能的影响, 本文进行了一系列消融性实验。在控制了其他因素相同的条件下, 首先本文对比了基于帧间差分法和基于三帧差分法检测器初始化跟踪器对跟踪效果的影响, 其跟踪效果如图 10 所示。由图 10 可知, 基于三帧差分的检测算法的跟踪框初始化方法可取得更好的跟踪效果; 然后对比了改进前后的核相关滤波跟踪对跟踪效果, 如图 11 所示, 使用了局部搜索区域重定位策略的核相关滤波跟踪算法跟踪效果更佳; 最后验证了自适应检测跟踪切换机制对目标跟踪的影响, 与之作为对比, 比较算法每隔固定时间 (实验设置为 30 帧) 对跟踪器进行初始化, 两种检测-跟踪切换机制的跟踪效果如图 12 所示。由图 12 可知自适应检测-跟踪切换机制下的目标跟踪效果优于固定周期检测-跟踪切换机制。



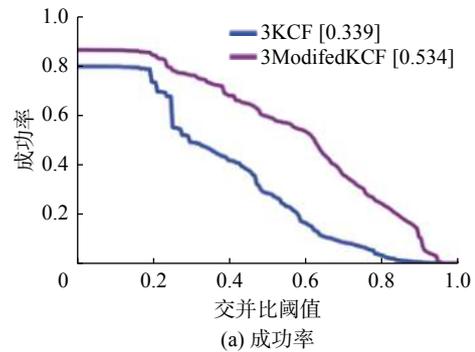
(a) 成功率



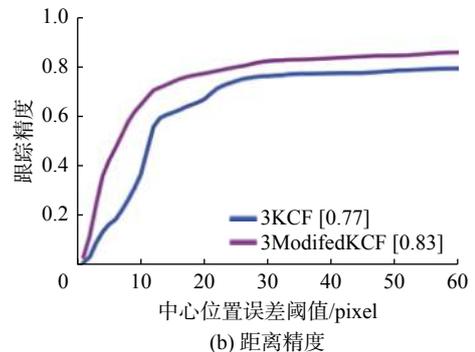
(b) 距离精度

图 10 不同初始化方法的成功率和距离精度

Fig. 10 Success plot and precision plot of the tracker with different initialization method



(a) 成功率



(b) 距离精度

图 11 核相关滤波跟踪改进前后的成功率和距离精度

Fig. 11 Success plot and precision plot of the tracker with and without searching area relocation

5.4 算法实现细节与实验测试

在实际测试中, 本文以 CPU 型号为 Intel core i7-7700、内存 8 GB、GPU 为 11 GB 的 Nvidia GeForce GTX 1080ti 计算机为中央处理器, 使用 4 路型号为 HIKVISION DS2DF7330IW 的云台摄像机对周边区域进行定点巡航扫描, 视频图像分辨率为 2 048×1 536, 帧率为 24 FPS, 支持水平 360°、竖直 -2°~90° 旋转和 30 倍光学变焦。所有算法由 C++和 python 混合实现, 图 13 为由 C#编写的人机交互界面, 其中左侧为视频图像显示框, 右侧则包括云台控制按钮、检测报警灯、识别目标与分类结果框。为验证算法的优越性, 本文对比 Yolo v3^[6]、Faster RCNN^[16] 与本文所提出的算法在

5.3 节提出的无人机检测与跟踪数据集进行了对比实验, 并采用平均正确率 (average precision, AP) 和帧率 (frame per second, FPS) 对算法进行评估, 其比较结果如表 3 所示。其中在使用 Yolo v3 进行检测时, 需将视频图像缩小至 608×608 。由表 3 可知, 本文提出的算法对于无人机检测与跟踪数据集的 AP 为 94.90%, FPS 可达 33.1, 在检测效果与其他 2 种经典目标检测算法相媲美的条件下, 检测速度远远优于其它两种算法。本文做了大量的实际测试, 测试结果显示本文提出的算法及其系统实现可对各种复杂场景下的无人机检测与跟踪, 实验测试检测与跟踪效果图如图 14 所示, 检测距离最远达 1 000 米。

表 3 不同检测算法性能比较
Table 3 Comparisons of different detection algorithm

检测算法	AP/%	FPS
Yolo v3	93.87	10.5
Faster RCNN	94.34	7.6
本文算法	94.90	33

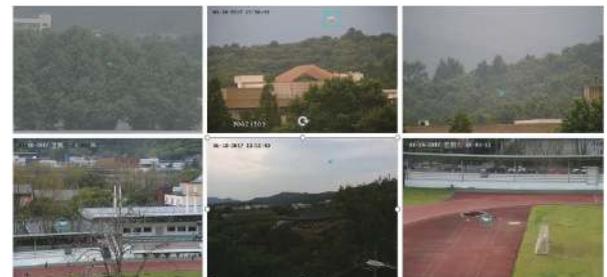


图 14 实际实验场景检测与跟踪效果
Fig. 14 Detection and tracking results of field tests

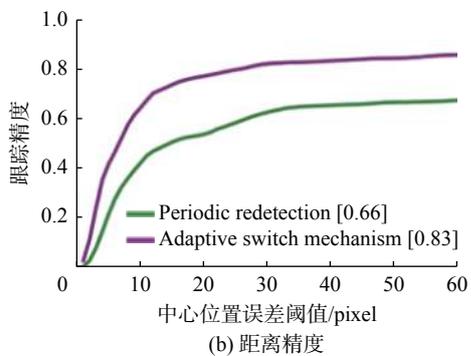
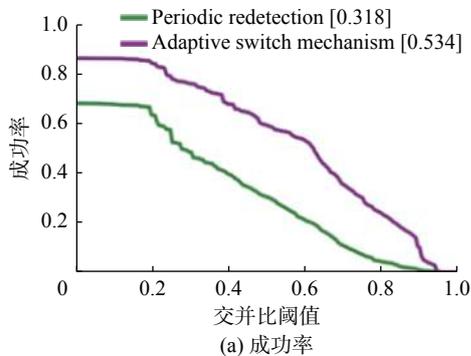


图 12 基于自适应切换机制与固定周期切换机制的跟踪成功率和距离精度

Fig. 12 Success plot and precision plot of the tracker with adaptive switching and periodic redetection mechanisms



图 13 系统人机图像交互界面
Fig. 13 Graphic user interface of the system

6 结束语

本文提出了一个适用于高清云台摄像机定点巡航扫描状态的无人机闭环检测与跟踪算法, 此算法通过一种自适应检测-跟踪切换机制将检测模式和跟踪模式有机的结合在了一个闭环框架中, 可实现快速实时、准确、稳定的无人机检测、识别与跟踪。与以往工作不同, 本文在目标检测模块中提出了一个快速准确、无需对高清视频图像进行尺度缩小的二阶段目标检测框架; 并用局部搜索区域重定位策略改进了核相关跟踪算法; 最后提出了一种自适应切换机制实现了检测模块和跟踪模块的自动切换。大量实验和实际测试也证明了算法不同模式与自适应闭环框架的有效性。

参考文献:

- [1] SUN Yuyi, CHEN Jiming, HE Shibo, et al. High-confidence gateway planning and performance evaluation of a hybrid LoRa network[J]. IEEE internet of things journal, 2021, 8(2): 1071-1081.
- [2] FLOREANO D, WOOD R J. Science, technology and the future of small autonomous drones[J]. Nature, 2015, 521(7553): 460-466.
- [3] SHI Xiufang, YANG Chaoqun, XIE Weige, et al. Anti-drone system with multiple surveillance technologies: architecture, implementation, and challenges[J]. IEEE communications magazine, 2018, 56(4): 68-74.
- [4] XIE Jiayang, YU Jin, WU Junfeng, et al. Adaptive switching spatial-temporal fusion detection for remote flying drones[J]. IEEE transactions on vehicular technology, 2020, 69(7): 6964-6976.

- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005: 886–893.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
- [7] LIU Wei, ANGUÉLOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 21–37.
- [8] LAW H, DENG Jia. CornerNet: detecting objects as paired keypoints[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 765–781.
- [9] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. CenterNet: keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South), 2019: 6568–6577.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK, 2020: 213–229.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 580–587.
- [12] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph-based image segmentation[J]. International journal of computer vision, 2004, 59(2): 167–181.
- [13] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154–171.
- [14] ENDRES I, HOIEM D. Category independent object proposals[C]//Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece, 2010: 575–588.
- [15] ZITNICK C L, DOLLÁR P. Edge boxes: locating object proposals from edges[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland, 2014: 391–405.
- [16] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Quebec, Canada, 2015: 91–99.
- [17] CANZIANI A, PASZKE A, CULURCIELLO E. An analysis of deep neural network models for practical applications[J]. arXiv preprint arXiv: 1605.07678, 2016.
- [18] TAN Mingxing, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA, 2019: 6105–6114.
- [19] ZHANG Lei, XU Ke, YU Shiqi, et al. An effective approach for active tracking with a PTZ camera[C]//Proceedings of IEEE International Conference on Robotics and Biomimetics. Tianjin, China, 2010: 1768–1773.
- [20] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2544–2550.
- [21] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(3): 583–596.
- [22] DANELLJAN M, HÄGER G, KHAN F S, et al. Discriminative scale space tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(8): 1561–1575.
- [23] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 472–488.
- [24] DANELLJAN M, BHAT G, SHAHBAZ KHAN F, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6931–6939.
- [25] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 850–865.
- [26] LI Bo, YAN Junjie, WU Wei, et al. High performance visual tracking with Siamese region proposal network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8971–8980.
- [27] WANG Qiang, ZHANG Li, BERTINETTO L, et al. Fast online object tracking and segmentation: a unifying approach[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 1328–1338.
- [28] DÉNIZ O, BUENO G, BERMEJO E, et al. Fast and accurate global motion compensation[J]. Pattern recognition, 2011, 44(12): 2887–2901.
- [29] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of International Conference on Computer Vision.

- Barcelona, Spain, 2011: 2564–2571.
- [30] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381–395.
- [31] BRUTZER S, HÖFERLIN B, HEIDEMANN G. Evaluation of background subtraction techniques for video surveillance[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 1937–1944.
- [32] CHENG Yuanhang, WANG Jing. A motion image detection method based on the inter-frame difference method[J]. Applied mechanics and materials, 2014, 490: 1283–1286.
- [33] SERRA J, VINCENT L. An overview of morphological filtering[J]. Circuits, systems and signal processing, 1992, 11(1): 47–108.
- [34] GRAHAM R L, YAO F F. Finding the convex hull of a simple polygon[J]. Journal of algorithms, 1983, 4(4): 324–331.
- [35] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7132–7141.
- [36] GHIASI G, LIN T Y, LE Q V. DropBlock: a regularization method for convolutional networks[C]//Proceedings of 33rd Conference on Neural Information Processing Systems. Montréal, Canada, 2018: 10750–10760.
- [37] FELZENSZWALB P F, GIRSHICK R B, MC-ALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 32(9): 1627–1645.
- [38] HOERL A E, KANNARD R W, BALDWIN K F. Ridge regression: some simulations[J]. Communications in statistics, 1975, 4(2): 105–123.
- [39] SCHÖLKOPF B, SMOLA A J. Learning with kernels: support vector machines, regularization, optimization, and beyond[M]. Cambridge: MIT Press, 2002.
- [40] GRAY R M. Toeplitz and circulant matrices: a review[J]. Foundations and trends® in communications and information theory, 2006, 2(3): 155–239.
- [41] KIM Y, BANG H. Introduction to Kalman filter and its applications[M]. GOVAERS F. Introduction and Implementations of the Kalman Filter. IntechOpen, 2019.
- [42] GOYETTE N, JODOIN P M, PORIKLI F, et al. Changedetection.net: a new change detection benchmark dataset[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, USA, 2012: 1–8.
- [43] ALLEBOSCH G, DEBOEVERIE F, VEELAERT P, et al. EFIC: edge based foreground background segmentation and interior classification for dynamic camera viewpoints[C]//Proceedings of the 16th International Conference on Advanced Concepts for Intelligent Vision Systems. Catania, Italy, 2015: 130–141.
- [44] JIANG Shengqin, LU Xiaobo. WeSamBE: a weight-sample-based method for background subtraction[J]. IEEE transactions on circuits and systems for video technology, 2018, 28(9): 2105–2115.

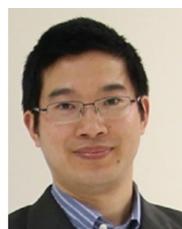
作者简介:



谢家阳, 博士研究生, 主要研究方向为计算机视觉、反无人机系统、深度学习、目标检测与跟踪。



王行健, 硕士研究生, 主要研究方向为计算机视觉、深度学习、目标检测与跟踪。



史治国, 教授、博士生导师, 主要研究方向为研究方向为信号处理及定位应用, 物联网。IET Fellow、IEEE Network 编委、IET Communications 编委、Journal of The Franklin Institute 编委, 2020 年获中国人工智能学会技术发明一等奖, 2015 年获教育部科技进步一等奖, 2012 年获浙江省科学技术二等奖。主持国家重点研发项目、国家自然科学基金项目、浙江省重点研发计划项目等多项, 获授权发明专利 80 余项。发表学术论文 100 余篇。