ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 1,Jan. 2025,pp. 86－101
DOI:10. 16337/j. 1004-9037. 2025. 01. 007

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

低代价高动态大视场低慢小飞行器检测与跟踪

常宇轩，杨　文，吴金建

(西安电子科技大学人工智能学院,西安 710126)

摘　要：低空经济的兴起使小型无人机(Unmanned aerial vehicles，UAVs)在物流、测绘与娱乐等领域获得广泛应用,但由此带来的安全风险也日渐凸显。低慢小(Low-slow-small，LSS)目标检测在国家安全、空域监管以及无人机防御等领域具有重要意义,能够有效应对小型低空飞行目标带来的潜在威胁。针对现有传感器在低成本、复杂光照与大视场要求上存在的不足,本文提出了一种基于事件相机与RGB相机协同的LSS目标检测系统。首先,借助事件相机高速成像及大动态范围配合智能控制转台进行“扫视”,并通过基于事件的检测算法完成目标初步定位;随后,利用信息协同模块融合双模态数据以提升检测精度;最后,依托RGB相机的高分辨率与动态变焦特性实现“凝视”,并结合提出的图像识别算法进行目标精细化识别与跟踪。在复杂光照与宽视场条件下,该系统兼顾了低代价与高性能,为LSS目标检测提供了有效的新路径。

关键词：目标检测与追踪;低慢小目标;事件相机;RGB相机;跨模态协同

中图分类号：TP391.41　　**文献标志码**：A

Low-Cost, High-Dynamic, Large Field-of-View Detection and Tracking of Low-Slow-Small Aerial Vehicles

CHANG Yuxuan, YANG Wen, WU Jinjian

(School of Artificial Intelligence, Xidian University, Xi'an 710126, China)

Abstract: The rise of the low-altitude economy has led to the widespread adoption of small unmanned aerial vehicles (UAVs) in logistics, surveying, and entertainment, yet the associated security risks have grown increasingly prominent. Detecting low-slow-small (LSS) targets is therefore crucial in domains such as national security, airspace regulation, and UAV defense, as it effectively mitigates potential threats posed by small and low-altitude flying objects. In response to limitations of current sensors with regard to cost-effectiveness, operation under complex lighting conditions, and broad field-of-view coverage, this paper proposes an LSS target detection system that leverages both an event camera and an RGB camera. First, the high-speed imaging and wide dynamic range of event camera are employed for an initial "sweep", and an event-based detection algorithm provides preliminary target localization. Next, an information collaboration module fuses the two-modal data to enhance detection accuracy. Finally, the high-resolution and dynamic zoom features of RGB camera enable a "gaze" mode, combined with a dedicated image recognition algorithm for fine-grained target identification and tracking. Under complex

基金项目:国家自然科学基金(62402366)。
收稿日期:2024-12-31;**修订日期**:2025-01-23

lighting and wide field-of-view scenarios, this system achieves both low cost and high performance, offering an effective new approach to LSS target detection.

Key words: target detection and tracking; low-slow-small (LSS) targets; event camera; RGB camera; cross-modal collaboration

引　言

在低空经济蓬勃发展的背景下,小型无人机(Unmanned aerial vehicles, UAVs)在物流、测绘、娱乐等领域迅速普及,也给关键基础设施和公共活动区域带来潜在的安全隐患[1]。低慢小目标检测是应对现代复杂安全环境的关键技术之一,其核心在于解决低慢小(Low-slow-small, LSS)目标在复杂背景下难以被传统方法感知和识别的难题。低慢小目标具有的物理与成像特性使其难以被有效检测,可能对机场、关键基础设施和公共活动区域等重要场所存在潜在威胁,快速有效地检测和跟踪它们对于预防安全事故至关重要[2]。

低慢小目标体积小、速度慢、信号弱、特征不明显,导致检测难度大。同时复杂光照如空域的强光、逆光及高动态光照,以及建筑和树木等复杂背景会导致目标与环境对比度低,干扰目标识别。目前主流的探测手段是基于雷达系统,但雷达探测成本高、隐蔽性差,同时存在探测"盲区",且对低速或悬停目标捕捉能力不足。低慢小目标电磁反射面积小,使雷达接收到的反射信号微弱,容易导致目标丢失或漏检。光学探测系统作为补充手段来弥补雷达探测的短板。然而,传统光学相机成像慢、动态范围和视场角受限,复杂光照下难以全方位搜索和捕捉目标;红外探测依赖目标热辐射信号,仅呈现温差分布,但低慢小目标体积小、红外辐射弱、受环境影响大,易与背景对比度低而导致漏检。为此,本文基于自主研制的动态范围大、成像速度快的新型仿生动态成像相机(事件相机),研究低代价、高动态、大视场的低空飞行器监测技术,弥补传统光学探测系统的不足,力图在雷达"盲区"的大动态环境中增强对弱小无人机与鸟类以及在"盲区"之外对低速或悬停目标的检测精度。

事件相机是一种启发于生物视觉的新型动态视觉传感器,打破传统相机基于能量积分的成像范式,采用能量差分的新体制成像范式,只捕捉动态场景。事件相机具有成像动态范围大(~120 dB)、成像速度快(~10 μs)、数据冗余小、动态目标捕捉能力强以及静态背景自动滤除等特点。事件相机弥补了传统成像系统在高速和极端光照环境下成像能力不足的问题,可在复杂光照条件下有效捕捉运动目标并实时生成信息。相比于雷达,事件相机不受电磁干扰影响,对新型复合材料鲁棒性高,在扫描搜索状态下,对于运动或静止目标都能有效发现。因此,事件相机在低慢小目标检测领域有巨大的潜力。

综上,面向在复杂光照下及大视场远距离视野范围内的低慢小目标低代价检测识别需求,本文构建了基于事件相机+传统RGB相机的多传感器协同探测方法,依照"先动态差分成像快速扫视发现,后传统成像聚焦凝视识别"的设计思想,充分挖掘动态差分成像对弱小运动目标捕捉能力强的特性。首先,利用事件相机成像速度快的特点,利用"速度换空间"实现大视场的快速扫描,开发基于多尺度特征融合和时序特征聚合的事件目标检测方法,捕捉目标位置信息;然后,设计跨模态空间信息和语义信息融合方法,协助RGB相机进行目标捕捉,进一步调整焦距利用其空间分辨率特点更清晰地捕捉目标细节以进行精确凝视识别。本文所提多传感器协同工作方式可以充分发挥事件相机在动态场景中的快速响应能力和可见光相机在静态场景中的视觉清晰度,从而提升系统的检测精度和鲁棒性,适用于低空飞行目标的监测与识别。

1 相关工作

1.1 国内外研究现状

目前,国内外针对低慢小无人机目标的几种主流检测方式[3]有雷达探测、可见光检测以及红外探测,这几种检测方法各有优劣。

现有的低慢小检测方法主要是基于雷达的探测[4-5],其探测长距离与大范围表现出色并且可以全天候探测。雷达设备通过发射特定频率的电磁波,遇到目标物体时电磁波反射形成回波,处理解析得到目标物体信息,包括距离、速度、形状和尺寸等,可以精准地捕捉远距离目标,为大面积监控提供支撑。目前针对低慢小目标的雷达检测方法有变换域检测方法和微多普勒分析方法等[6]。但存在近距离“盲区”,信号互相干扰致目标信息难获取,且对悬停无人机探测能力弱,限定了其在特定场景的应用。同时,低慢小目标因其体积小,在雷达探测时电磁反射面积小,使雷达接收到的反射信号微弱,易造成目标丢失或漏检。此外,雷达设备的成本较高、配置复杂。针对这些问题,目前主流方案是采用光学探测解决,这是由于其不受电磁信号近场干扰影响,可以在中近程距离时获得高分辨目标图像。光学探测虽能够在一定程度上弥补雷达的短板,但仍存在一些问题。光学探测中的可见光探测以高分辨率和颜色还原能力著称,可以为低慢小目标探测提供外观信息。深度学习算法是其主要实现方式,由于其提取复杂抽象特征能力强,并且具有较强的泛化能力,因此已经被应用于包括低慢小无人机检测在内的许多领域。Hu等[7]对YOLOv3进行改进,以适应低慢小目标的检测,通过4个尺度特征图,预测物体的标记框,获得更多目标特征信息。Nalamati等[8]在远程无人机监控视频中,评估了多种目标检测模型Faster R-CNN[9]和SSD[10]。但可见光相机动态范围小,在强光及逆光等复杂光照条件下容易过曝或欠曝,降低成像质量。同时可见光相机视场角小,对于低慢小目标的检测需要全方位快速扫视,而其成像速度慢导致在快速旋转时图片帧模糊,无法检测到目标。此外,可见光相机前背景等权重成像,在复杂背景环境中,低慢小目标与背景颜色或纹理相近时易漏检。

光学探测的另一种方式——红外探测,也被应用于低慢小目标探测。Wang等[11]提出了一种基于多尺度融合滤波的红外图像目标探测算法。在此基础上,Zhang等[12]提出了一种基于四元数组离散余弦变换的时空增强红外小目标检测算法。红外探测同样存在一些问题,红外成像设备通过探测目标的热辐射信号生成图像,显示的是目标与背景的温差分布,而非物体的颜色或纹理等细节,无法提供丰富的外观信息。由于低慢小目标的体积较小、红外辐射特征不强,容易受到环境影响,尤其是在温差较小环境下,目标与背景信号的对比度较低,容易淹没于背景,造成漏检。

事件相机区别于传统光学相机,通过像素的变化触发事件,采用差分成像,仅对运动物体响应,能增强目标与背景的区分度,可以解决上述传统光学探测目标与背景对比度低的问题。其动态范围大,对光强变化鲁棒性强,改善了可见光相机复杂光照下过曝和欠曝的问题;并且成像速度快,解决了可见光相机在转台快速旋转时成像模糊的问题。但单路事件相机产生稀疏的事件流缺乏丰富的目标外观信息,难以直接目标识别获取细节特征。

综上,本文提出了基于事件相机与可见光相机协同检测的方案。表1展示了目前几种主流方案与本文方案的对比,可以看出本文方案的优势。利用光学探测弥补了雷达探测方法近距离“盲区”以及“盲区”外目标悬停的问题。利用事件相机成像速度快,解决了可见光相机快速扫视时帧模糊问题;利用事件相机异步触发事件,差分成像,解决了对比度低、目标淹没于背景漏检的问题。

表 1 本文方案与几种主流方案对比

Table 1 Comparison between the proposed scheme and several mainstream schemes

方案	近距离探测	环境适应性	目标信息获取	成本与配置
雷达探测	盲区,难测近距小目标	全天候	位置、速度等	成本高、配置复杂
可见光探测	可探测	光照、背景影响大	外观细节	成本低、配置简单
事件探测	可探测	背景适应好	目标位置信息	成本低、配置简单
红外探测	可探测	容易受环境温度影响	目标位置信息	成本低、配置简单
事件+可见光	能有效探测	有效应对复杂环境	兼备运动与外观细节	成本低、配置简单

1.2 事件相机成像原理

事件相机,也称动态视觉传感器(Dynamic vision sensor,DVS),是一种基于生物视觉原理的新型传感器[13-14],与传统基于帧的相机不同,它不捕捉图像的全局亮度信息,而是通过检测像素亮度的变化事件来获取动态场景的信息。这种工作原理使得事件相机具有高时间分辨率(通常在微秒级)、低功耗以及高动态范围等显著优势,特别适合动态场景和高对比度环境的感知任务。

DVS的核心是像素级独立工作。每个像素通过感知亮度变化来触发事件,而非捕获亮度的绝对值。式(1)表明,当像素的对数亮度发生变化,且变化量超过设定阈值时,该像素会生成一个事件e,由四元组(x_e, y_e, p_e, t_e)组成。其中,x_e、y_e分别为事件发生位置的横坐标和纵坐标,t_e表示事件发生的时间戳,p_e表示事件的极性,反映了光强变化的趋势。光强变大表明该事件为ON事件(即正极性事件),光强变小表明该事件为OFF事件(即负极性事件)。

$$\begin{cases}\Delta \ln I(t)=\ln I(u,t+\Delta t)-\ln I(u,t)\geqslant C_{\text{th}}\\ \Delta \ln I(t)=\ln I(u,t+\Delta t)-\ln I(u,t)\leqslant -C_{\text{th}}\end{cases} \tag{1}$$

式中:$u=(x,y)$代表像素坐标位置;I代表光强;t为上一事件的发生时刻;C_{th}为事件的触发阈值。当光强变化超过阈值C_{th}的时候,产生事件信号。每个DVS像素连续独立地重复这个过程,对像素光强的相对变化进行编码,从而产生异步事件流。

2 基于DVS的低慢小数据集

2.1 数据采集

数据采集过程中,利用事件相机和传统相机同时采集3种不同类型不同大小的无人机目标,并分别在无背景、树木背景及高楼背景下用DVS和RGB相机录制数据。同时在构建数据集时,无人机与镜头的相对距离不同,使相机画面中的无人机呈现出不同的尺度,这些不同尺度的样本有助于模型在不同观测条件下进行泛化。最终构建出了一个综合性强的多尺度、多背景低慢小目标数据集,旨在为无人机目标检测算法提供更具挑战性和多样性的训练数据。

2.2 事件表征方法

事件相机的输出是稀疏的异步事件流数据,只有在像素亮度发生变化时记录事件。因此本文通过在固定时间窗口Δt内累积事件,有效增强了图像中的信息密度,填补稀疏数据带来的空缺。同时这种在时间窗口累积事件的方法能引入时间维度,在目标检测任务中更好地捕捉目标的特征。

首先按固定时间窗口Δt划分原始事件流,之后按照式(2)将每个时间窗口内发生的事件$e_i(x_i, y_i, p_i, t_i)$进行累积得到一张事件图像I[15],即

$$I(x,y,\Delta t)=1_{(x_i=x \text{ and } y_i=y)} \tag{2}$$

式(2)表明在Δt时间内,若像素x_i、y_i处发生了事件,像素值设为1,否则为0。通过这种事件表征的方法[16],将原始事件流转化为一段包含时序关系的事件图像序列。

2.3 数据标注

本文对每张事件图像以及RGB图像进行人工标注。事件图像中一些运动目标形状比较相似,因此在标注前应观看单张图像前后的事件图像序列来确定目标位置,建立事件间的时空关联性,并对每张图像进行正确标注,每条标注包括目标的边界框及类别信息。图1展示了部分事件图像及RGB图像的标注可视化结果。

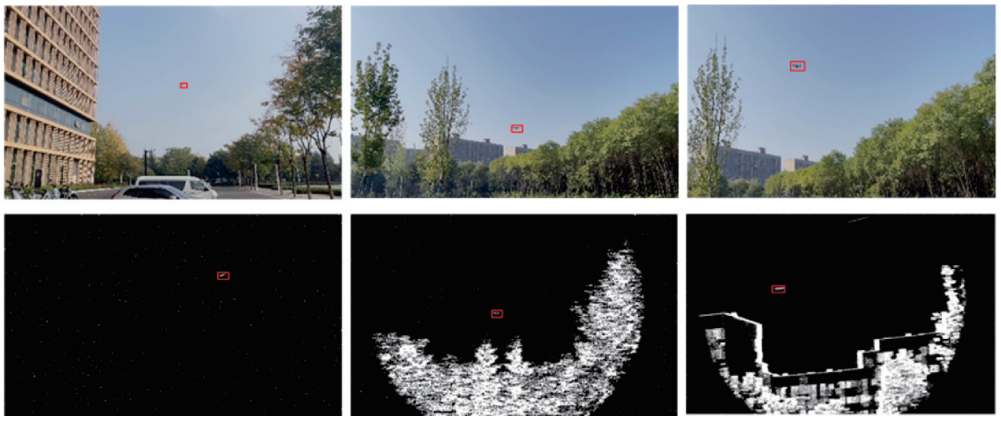

图1 数据集标注可视化结果

Fig.1 Dataset annotation visualization results

2.4 数据集统计特性分析

本文构建的基于DVS的低慢小数据集共包含10段不同无人机尺度、不同背景下的事件流,并按照时间窗口$\Delta t=20$ ms将事件流转换为事件图像序列。其中,3段为树木背景下的事件流,3段为高楼背景下的事件流,4段为无背景下的事件流,并且每个事件流中不同尺度大小的无人机目标均匀分配,总共得到8 123张图片并进行数据标注。数据集中无背景、树木背景、高楼背景的事件图像总数,以及各背景下训练集和测试集的事件数量如表2所示。

表2 基于DVS的低慢小数据集汇总

Table 2 Summary of LSS dataset based on DVS

背景类型	事件图像总数	训练集数量	测试集数量
无背景	3 248	2 598	650
树木	2 598	2 078	520
高楼	2 277	1 821	456
合计	8 123	6 497	1 626

3 面向DVS的低慢小目标检测

3.1 系统概述

本文首先搭建基于“事件相机+传统相机”的低慢小目标检测系统。如图2所示,低慢小目标检测系统是一个自动化旋转扫描平台,具有高稳定性和高可控性等特点。本系统在广泛场景下有效运行,适应不同的探测距离、光照与天气条件,并对周围环境具有较强的适应性。具体表现为:

(1)探测距离性能。本文方案以大疆mini2无人机为实验对象,配备120 mm焦距的事件相

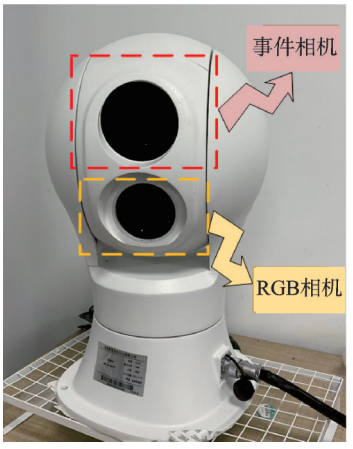

图2 双模协同低慢小目标检测平台

Fig.2 Dual-mode cooperative LSS target detection platform

机镜头、变焦可见光相机与可全方位灵活扫描的转速可调智能云台。本文方案对低慢小目标的探测距离为10～350 m。近距离时事件相机快速捕捉动态目标,触发云台并联动可见光成像;远距离时可见光相机利用变焦功能,配合云台,确保清晰成像与稳定跟踪。

(2)光照与天气适应性。事件相机基于异步触发机制对光照变化具有较强鲁棒性,能在强光(50 000 lux)、弱光(10 000 lux)、甚至低光(3 000 lux)环境下稳定捕获运动目标。相比之下,可见光相机在复杂光照条件下表现较差。本文方案通过多传感器协同,将事件相机中目标信息传递给可见光相机,克服复杂光照干扰,保证在不同光照及天气下对低慢小目标检测与跟踪性能良好。

(3)周围环境适应性。在不同环境下,如城市建筑、高楼、树木等复杂背景以及无背景的天空,本文方案能有效适应。低慢小目标纹理特征弱,与背景的纹理对比度低,传统方法可能导致漏检,利用事件相机大视场范围与仅对运动物体敏感的特性可以从复杂背景中筛选出运动目标与可见光相机共享信息,利用可见光高图像分辨能力,实现对低慢小目标的准确跟踪识别。

为应对1.1节所述挑战,本文还提出了一种融合事件相机扫视检测、传统相机凝视识别与跨传感器信息协同的多模态处理框架,整体结构如图3所示。具体而言,首先充分利用事件相机在成像速度与大动态范围方面的优势进行目标定位;随后,通过空间信息融合,引导RGB相机聚焦目标的大致区域;最后,将两种模态的语义信息结合起来,从而提升低慢小目标的检测精度,并在复杂环境下仍能保持出色的性能。

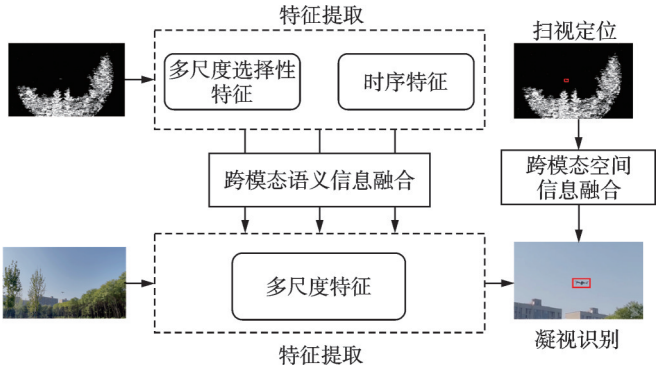

图3 基于事件相机+RGB相机系统的低慢小目标检测与识别框架

Fig.3 LSS target detection and recognition framework based on event camera +RGB camera system

3.2 基于事件的低慢小目标扫视检测算法

对于事件数据,本文设计了一种新的基于事件的检测算法。图4展示了算法的主要网络框架部分,负责提升事件数据的时空特征表达,主要包含以下几个部分:事件表征模块、多尺度选择性特征提取模块以及时序特征提取模块,将提取的特征送入检测头[17]得到预测结果。图中X表示将事件表征后的事件图像通过多尺度选择性特征提取(Multi-scale selective feature extraction, MSFE)模块提取得到的高层和低层信息融合特征;隐藏状态H存储了时间步及之前累积的特征信息,使得特征具有多尺度和时序信息,确保模型可以利用历史信息进行目标检测。

3.2.1 事件表征模块

对于事件表征,本文将一段事件流在时间维度上按照$\Delta t=20$ ms分割成多个时间窗口,随后通过将时间窗口中的事件流转化为事件图像,使事件数据具备更易于网络处理的形式。同时这一过程也保留了事件的时序信息,结合下文提出的时序特征提取模块,有助于捕捉目标的动态变化。

3.2.2 多尺度选择性特征提取模块

本文提出了一种基于空间和通道重建卷积的多尺度特征提取模块,即MSFE模块。MSFE模块包含3个特征提取器,其中每个特征提取器由空间和通道重建卷积SCConv[18]、普通卷积Conv[10, 19]以及基于层次尺度的特征金字塔网络HS-FPN[20]组成。SCConv模块包含2个单元:一个名为空间重构单元,一个名为通道重构单元。其中空间重构单元通过分离-重构方法来减少空间冗余,通道重构单元则使用分割-转换-融合方法来减少通道冗余。这2个单元协同工作,以减少卷积神经网络(Convolutional

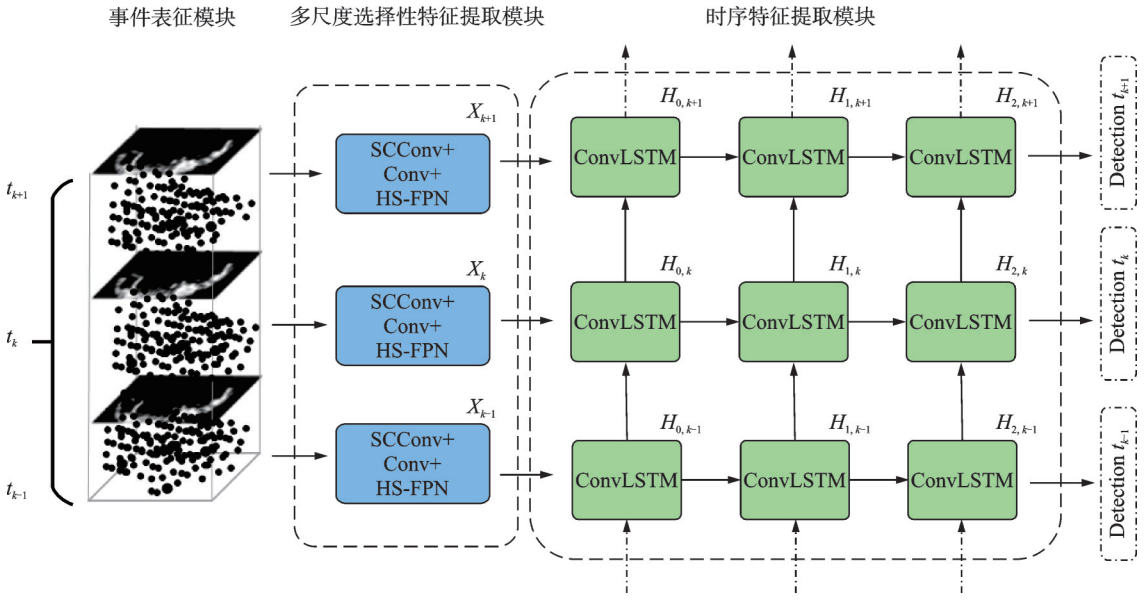

图4 基于事件的低慢小目标扫视检测算法框架

Fig.4 Event based saccade detection algorithm framework for LSS targets

neural network, CNN)中特征的冗余信息。

HS-FPN在传统特征金字塔网络的基础上进行了改进,主要由两个部分组成:特征选择模块和特征融合模块。最初,不同尺度的特征图在特征选择模块中经过筛选过程。随后,通过选择性特征融合(Selective feature fusion, SFF)机制,使用高级特征作为权重来过滤低尺度特征中包含的必要语义信息,将高层和低层信息协同集成。这种融合产生的特征具有丰富的语义内容,从而增强模型的检测能力。将事件表征后的特征 E_i 作为多尺度选择性特征提取模块的输入,之后通过融合不同尺度的特征图来实现对低慢小目标的有效提取,从而增强模型的表达。

3.2.3 时序特征提取模块

事件相机捕捉的是像素级别的光强变化,输出的数据是稀疏的事件流,包含了事件发生的时间戳、像素位置以及极性信息。这种数据形式天然具有时间连续性和动态特性,单纯依赖静态的空间信息难以充分表达其动态变化的本质。因此,对于低慢小目标的检测任务,仅靠某一帧的事件数据可能无法清晰判断目标的形态或运动方向,而通过建立相邻时间段内事件图像的关联,模型可以更准确地理解目标的运动轨迹、速度以及方向等关键属性,更有效地对小目标进行检测。同时,事件相机数据中可能存在由于环境噪声或感光器件特性引起的噪声事件。这些噪声事件往往是随机分布的,而真实事件在时间上具有连续性和规律性,通过时序分析[21-22]可以减少噪声对小目标检测的影响。

针对上述问题,本文设计了一个时序特征提取模块,目的是学习相邻事件图像之间的相似性来融合时序特征,从数据中提取时空信息[23-24]。如式(3)所示,每个时间步 t_k 的输入事件数据 E_k 通过MSFE模块提取高层和低层信息融合特征,即

$$X_k = \mathrm{MSFE}(E_k) \tag{3}$$

时间步 t_k 的特征 X_k 与上一时间步 t_{k-1} 的隐藏状态 $H_{l,k-1}$ 一起输入到ConvLSTM模块[25],生成当前时间步的隐藏状态,表达式为

$$H_{0,k} = \mathrm{ConvLSTM}(H_{0,k-1}, X_k) \tag{4}$$

$$H_{l,k}=\mathrm{ConvLSTM}(H_{l-1,k},H_{l,k-1}) \quad l\in[1,2] \tag{5}$$

式(5)中$H_{l,0}=0$,将得到的最后一层的隐藏状态$H_{2,k}$输入到目标检测头,生成当前时间步t_k的检测结果,表达式为

$$D_k=\mathrm{Detection}(H_{2,k}) \tag{6}$$

3.3 跨传感器信息协同算法

信息协同模块通过融合事件相机和可见光相机的数据,实现对低慢小目标的全面理解和高效感知。模块主要包含两个核心部分,分别是空间信息融合和语义信息融合,从物理空间的几何对齐与语义层面的信息互补两个角度增强多模态数据的协同性。

3.3.1 空间信息融合

空间信息融合模块旨在利用事件相机的高时间分辨率和运动目标定位能力,将事件流数据中的运动目标位置投影到可见光图像坐标系中,以作为兴趣区域。可见光相机通过对这些兴趣区域的深度分析(如高分辨率纹理检测)实现更加精准的识别。

实现空间信息融合的一个前提是事件相机和可见光相机的视场几何对齐。在本文的数据集中,事件由分辨率为1 280像素×800像素的CeleX相机记录,可见光图片由分辨率为1 920像素×1 080像素的RGB相机记录。由于两种相机具有不同的感知机制和分辨率,通过标定相机内参和外参,建立两者之间的空间映射关系。

事件相机的内参矩阵为$\boldsymbol{K}_1$,RGB相机的内参矩阵为$\boldsymbol{K}_2$,相机的内参矩阵$\boldsymbol{K}$包含了相机的焦距f_x和f_y,主点坐标c_x和c_y以及畸变系数等信息。旋转矩阵$\boldsymbol{R}$和平移向量$\boldsymbol{T}$用来描述两个相机的位置关系。外参矩阵$[\boldsymbol{R}|\boldsymbol{T}]$描述了从事件相机坐标系到RGB相机坐标系的转换。

事件相机图像为$I_1(x_1,y_1)$,RGB相机图像为$I_2(x_2,y_2)$,式(7)将图像坐标转化为归一化的相机坐标系,即

$$\boldsymbol{p}_1=\boldsymbol{K}_1^{-1}\begin{bmatrix}x_1\\y_1\\1\end{bmatrix},\quad \boldsymbol{p}_2=\boldsymbol{K}_2^{-1}\begin{bmatrix}x_2\\y_2\\1\end{bmatrix} \tag{7}$$

如果两个相机之间的外参矩阵为$[\boldsymbol{R}|\boldsymbol{T}]$,则第2个相机的归一化坐标$\boldsymbol{p}_2$可以通过第1个相机的坐标$\boldsymbol{p}_1$应用式(8)中的坐标变换得到,即

$$\boldsymbol{p}_2=\boldsymbol{R}\cdot\boldsymbol{p}_1+\boldsymbol{T} \tag{8}$$

最后,通过内参矩阵将归一化坐标映射回图像坐标系,即

$$\begin{bmatrix}x_2\\y_2\\1\end{bmatrix}=\boldsymbol{K}_2\cdot\boldsymbol{p}_2 \tag{9}$$

基于画面配准的基础,事件检测部分检测到某个低慢小目标时,可以根据其空间位置控制云台转动到指定位置,让目标位于RGB相机的中心位置,并初步推测出该目标在RGB相机视野中的大致位置坐标,进行粗检测的过程,从而引导RGB相机聚焦于该区域进行精细的目标识别。

3.3.2 语义信息融合

语义信息融合注重从高层次语义特征上实现数据的互补与增强。事件相机的数据包含丰富的时间动态信息,可用于预测目标的运动轨迹或分析动态行为,而可见光相机则提供纹理、颜色和类别等静态语义信息。Tomy等[26]提到了一种早期融合方法,将来自RGB相机和事件相机的输入简单连接起来,作为输入送到网络中。但这种早期融合方法存在一些缺陷,无法发挥各模态的不同优势。因此,本

文引入了通道注意力机制[27]来进行语义层次的跨模态融合，即交叉引用模块(Cross reference module，CRM)。

通道注意力能够通过动态学习每个模态特征通道的重要性，赋予关键信息更高的权重，忽略不相关或冗余特征。例如，事件数据可以在快速运动或低光场景下占主导地位，而在静态场景中，RGB数据的特征权重会更高。同时，通道注意力对环境变化具有较强的适应性，当某一个模态的数据质量下降时(RGB图像在低光或遮挡条件下不清晰)，机制会自动增强高质量模态的特征权重，提升任务的准确性和鲁棒性。

本文使用的CRM模块使单模态以全局角度从辅助模态中学习到更多的互补信息，图5给出了CRM模块的详细结构。

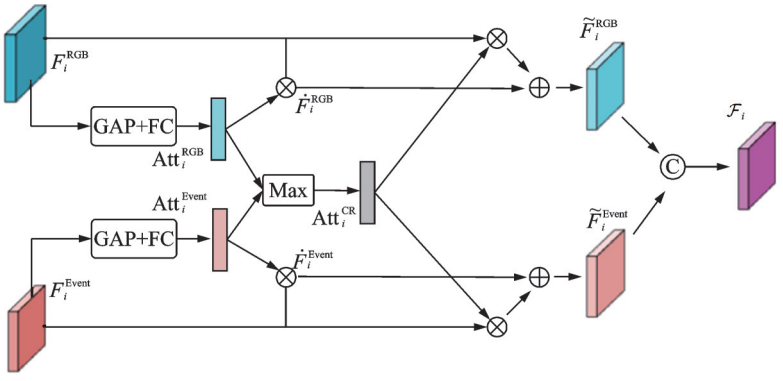

图5 基于通道注意力的交叉引用模块

Fig.5 Cross reference module based on channel attention

在CRM模块中，给定RGB图像和Event流第i个卷积块分别产生的两个输入特征F_i^{RGB}和F_i^{Event}，首先使用全局平均池化(Global average pooling，GAP)来获得RGB和Event中的全局统计信息。然后将这两个特征向量分别输入到一个全连接(Fully connected，FC)层和Softmax激活函数中，得到归一化的通道权重$\mathrm{Att}_i^{\mathrm{RGB}}$和$\mathrm{Att}_i^{\mathrm{Event}}$，如式(10)所示，分别反映了RGB和Event特征的重要性。之后用生成的权重对原始特征进行逐通道加权，这样能够明确关注重要部分，并抑制对场景理解不必要的部分。

$$\mathrm{Att}_i=\delta(\mathcal{W}_i\mathrm{AvgPooling}(F_i)+b_i) \tag{10}$$

式中：$\delta(\cdot)$表示Softmax激活函数；$\mathcal{W}_i$和b_i分别表示第i个特征的FC层的权重和偏置参数；AvgPooling$(\cdot)$表示全局平均池化操作；F_i为RGB和Event两个模态的第i个卷积块产生的输入特征。如式(11)所示，应用通道权重得到通道增强特征$\dot{F}_i$，即

$$\dot{F}_i=\mathrm{Att}_i\otimes F_i \tag{11}$$

此外，根据式(10)得到的通道权重$\mathrm{Att}_i^{\mathrm{RGB}}$和$\mathrm{Att}_i^{\mathrm{Event}}$由Max函数聚合，保留来自两个模态的有用特征通道，然后将其送到归一化操作$\mathcal{N}$，得到交叉引用通道权重$\mathrm{Att}_i^{\mathrm{CR}}$，即

$$\mathrm{Att}_i^{\mathrm{CR}}=\mathcal{N}\left(\mathrm{Max}(\mathrm{Att}_i^{\mathrm{RGB}},\mathrm{Att}_i^{\mathrm{Event}})\right) \tag{12}$$

基于式(12)中的融合通道权重$\mathrm{Att}_i^{\mathrm{CR}}$，并与增强特征$\dot{F}_i^{\mathrm{RGB}}$和$\dot{F}_i^{\mathrm{Event}}$相加，得到如式(13)所示的强化特征$\widetilde{F}_i^{\mathrm{RGB}}$和$\widetilde{F}_i^{\mathrm{Event}}$。式(14)将RGB和Event的增强特征进一步连接并馈送到1×1卷积层以生成跨模态融合特征$\mathcal{F}_i$，即

$$\widetilde{F}_i=\dot{F}_i+\mathrm{Att}_i^{\mathrm{CR}}\otimes F_i \tag{13}$$

$$\mathcal{F}_i = \mathrm{Conv}_{1\times 1}\left(\mathrm{Concat}\left(\widetilde{F}_i^{\mathrm{RGB}}, \widetilde{F}_i^{\mathrm{Event}}\right)\right) \tag{14}$$

3.4 基于图像的低慢小目标凝视识别算法

图像识别模型建立在RetinaNet网络[28]和ResNet-50主干网络[29]上。RetinaNet架构首先由骨干网络提取图像上各种尺度的特征,并将其传递给特征金字塔(Feature pyramid networks, FPNs),最后使用两个子网络进行分类和边界框回归。文献[26]中基于RetinaNet对两种模态的数据进行融合,两个模态经过主干网络特征提取后在不同的层生成多尺度特征图,来自两个模态分支的特征图在相同尺度上进行简单拼接,将拼接后的特征图送入FPN,再将每一层的特征送入到解码器Decoder中,预测所属类别和边界框回归参数。但简单融合方式未对两种模态的重要性进行区分,导致一些无关的或者低权重的信息引入,影响网络性能,同时它无法捕获RGB和事件特征之间的互补关系,缺乏对不同模态特征在任务中的重要性建模。因此本文引入3.3.2节中的语义融合模块CRM,利用通道注意力机制进行融合,提取关键特征。利用CRM融合模块来替换简单的串联拼接进行跨模态特征融合操作。将相同尺度的事件特征和图像特征进行融合,并将每种尺度下融合后的特征输入到FPN处理,增强对多尺度目标的感知能力。图6给出了关键模块替换后算法的主要网络结构。

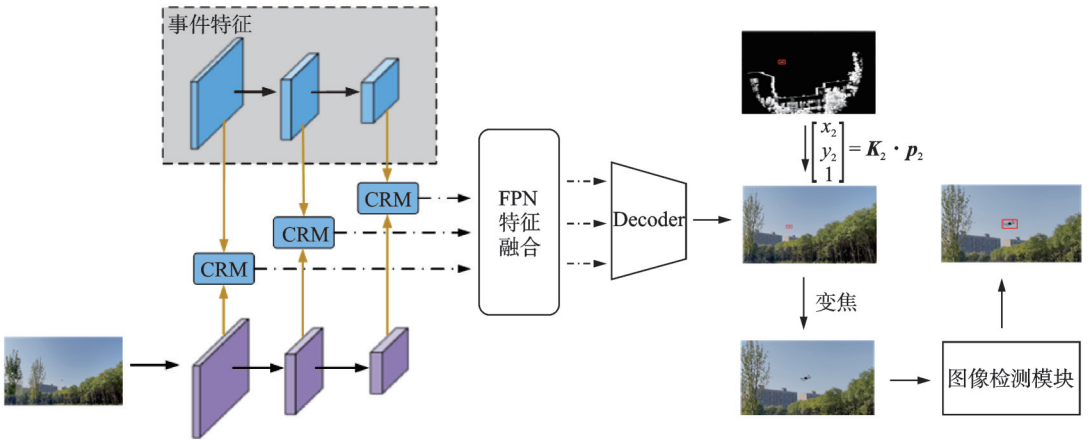

图6 基于图像的低慢小目标凝视识别算法框架

Fig.6 Image based gaze recognition algorithm framework for LSS targets

3.3.1节已进行画面配准,并利用空间信息融合。当事件检测模块检测到目标时,将空间位置信息传递给RGB相机;同时,RGB相机也会进行检测,随后将两路分别获取的位置信息进行数据融合处理,基于空间配准技术,利用加权平均等算法消除单一传感器可能存在的噪声与误差,进行高精度的目标位置定位。上述方法初步获取到小目标位置后,利用RGB相机的高分辨率特性,进行“凝视”来精细化的识别。根据目标在RGB相机画面中的占比实时评估目标的视场占用情况,并通过智能算法动态控制云台的变焦功能,对目标进行放大或缩小操作,确保目标始终位于画面中央并占据适宜的观测比例,避免目标过大或过小导致的观测偏差。对于变焦后的图片,仅采用可见光单个模态来进行图像检测识别,得到目标信息。通过这一过程,不仅提高了目标细节的捕获能力,还为目标的精细化识别提供了更加清晰的影像数据支撑。

4 实验结果分析

在第3节构建的低慢小目标数据集上,对本文所提出的基于事件检测的算法和跨模态融合模块进

行了一系列的实验验证。首先将本文所提方法与当前先进的检测方法进行比较，指标采用IoU阈值为0.5时模型的平均精度mAP@0.5。然后，在消融实验中评估所提方法主要组成部分的重要性。

4.1 实验设置

本文所提出的模型均采用PyTorch1.10框架、CUDA11.3以及Ubuntu20.04系统进行实验。基于事件检测网络采用Adam优化器，初始学习率设为0.002，并在每个epoch指数衰减0.98，批次大小设置为4，总共训练150轮。实验均在一张RTX3090显卡上进行。

4.2 对比实验

为了进一步验证本文方法的先进性，表3列出了所提方法以及其他检测方法在IoU阈值为0.5和0.75时模型的检测精度AP@0.5和AP@0.75。本文选择了基于两种模态的方法进行对比，一种是基于可见光图像的目标检测算法，另一种是基于事件流的目标检测算法。对于基于可见光图像的检测算法，仍采用按照时间窗口 Δt 划分得到的事件图像作为模型输入，而对于基于事件流的目标检测算法，直接用事件流作为模型的输入。

表3 所提方法与其他检测方法的实验结果对比

Table 3 Comparison of experimental results between the proposed method and other detection methods

对比方法	AP@0.5	AP@0.75
Yolov5	0.762	0.703
Yolov7	0.771	0.711
RED	0.685	0.633
ASTMNet	0.723	0.671
Our method	**0.786**	**0.722**

基于可见光图像的检测算法，选取了Yolov5和Yolov7[30]算法。Yolo系列是目标检测领域最流行的模型之一，被广泛应用在图像检测任务中[31-32]，具有优秀的检测效果。在本文构建的低慢小目标数据集上，由于只有一类无人机，且相对而言背景并不复杂，Yolov5和Yolov7的AP@0.5分别为0.762和0.771，其AP@0.75分别为0.703和0.711，而所提方法的AP@0.5和AP@0.75分别为0.786和0.722，相比于Yolov5分别提升了2.4%和1.9%，相比于Yolov7分别提升了1.5%和1.1%。

基于事件流的检测方法，选取RED[33]和ASTMNet[21]进行对比。由表3可以看出，在低慢小目标数据集上，所提方法相比于RED，AP@0.5和AP@0.75分别提升了10.1%和8.9%；相比于ASTMNet，AP@0.5和AP@0.75分别提升了6.3%和5.1%。

图7给出了本文方法和基于可见光图像的检测算法中的Yolov7、基于事件流检测算法中的ASTMNet之间的可视化对比。相比于Yolov7和ASTMNet方法，本文方法通过同时保留并更好地协同深层与浅层特征信息，使语义信息更加丰富，并利用事件图像序列的时序信息，有效改善了目标的误检漏检问题，准确识别出目标信息，可以看出本文方法对于低慢小目标检测的良好性能。

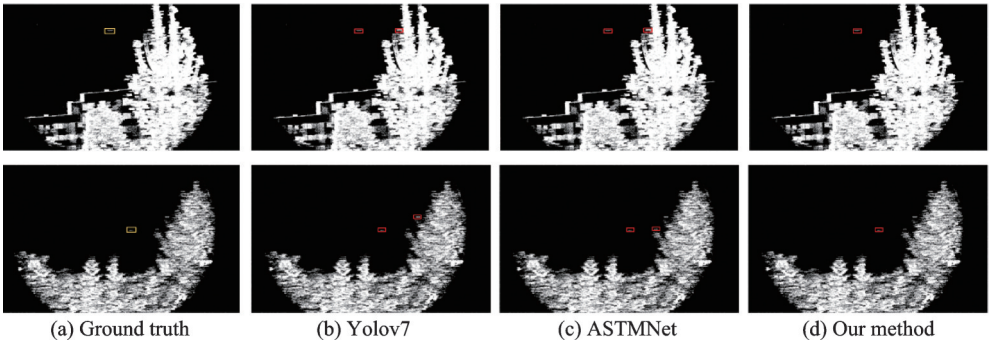

图7 所提方法与其他检测方法的可视化结果对比

Fig.7 Comparison of visual results of the proposed method and other detection methods

4.3 消融实验

本节分别在事件检测算法和跨模态融合识别算法上进行消融实验,分析各部分的贡献,以验证所提方法优化措施的有效性。

4.3.1 事件检测算法

首先验证提出的基于事件检测算法在低慢小目标检测任务上的有效性。表4列出了不同组成部分的消融实验结果。本文的基线模型采用3层普通卷积层堆叠进行特征提取,并使用Smooth L_1-Loss[10]进行回归框预测和Softmax Focal Loss[28]进行分类预测。之后,将普通卷积替换为SCConv,SCConv的mAP达到0.668,相比于基线模型,提升了5.5%,说明了SCConv在处理空间结构复杂或具有多通道特征输入时的有效性。进一步,在SCConv的基础上,引入了基于层次尺度的HS-FPN,mAP达到0.722,提升了5.4%。结果表明了HS-FPN能够使用高级特征作为权重来过滤低尺度特征中包含的必要语义信息,将高层和低层信息协同集成,有效提升了检测能力。最后,在上一步的基础上加入时序特征提取模块,mAP达到0.786,提升了6.4%。结果表明利用事件流特有的时序信息,建立事件图像之间的时间关联性,能有效地提升小目标检测效果。

表4 不同组成部分的消融实验结果

Table 4 Ablation experiment results of different components

组合方案	Baseline	SCConv	HS-FPN	ConvLSTM	mAP
1	√				0.613
2	√	√			0.668
3	√	√	√		0.722
4	√	√	√	√	0.786

4.3.2 跨模态融合算法

本节验证了提出的跨模态融合方法在低慢小目标检测任务的有效性。表5列出了基于不同融合策略下的消融实验结果。在单独使用RGB图像作为模型的输入时,mAP为0.623;在单独使用Event图像作为模型输入时,mAP为0.609。随后,两种模态的数据均被输入到3.4节的图像识别网络中,在主干网络特征提取时,将两个模态图像在各个特征层提取到的特征进行简单的串联拼接(Concatenation),将得到的融合特征送入FPN网络,此时的mAP达到了0.689,相比于单模态的检测结果有了较为明显的提升。进一步,将跨模态特征之间的简单串联拼接替换为3.3.2节中提到的CRM模块,mAP的值为0.735。相比于特征的简单拼接,取得了4.6%的性能提升。这表明CRM模块的通道注意力机制在处理跨模态特征时能够有效地进行语义信息融合,增强模型的表达能力。

表5 不同融合策略的消融实验结果

Table 5 Ablation experiment results of different fusion strategies

融合策略	RGB	Event	Concatenation	CRM	mAP
1	√				0.623
2		√			0.609
3	√	√	√		0.689
4	√	√		√	0.735

图8展示了本文跨模态融合方法与单一可见光模态、事件与可见光模态特征简单拼接方法的对比结果。图8(a)为原图像。图8(b)为单一可见光模态的检测结果,可以看出发生漏检的情况较多。图8(c)为跨模态简单拼接的方法,虽然使用事件特征作为可见光特征的补充,但这种拼接的方法只是简单将特征合并,并未获得各模态重要信息的特征,因此误检的情况比较常见。图8(d)为本文的CRM融合方法,通过引入注意力机制,聚焦对于跨模态融合有价值的信息,实现精准的信息融合,进而提升对低慢小目标的检测效果,成功检测到目标。

图8 不同融合策略的可视化结果对比

Fig.8 Comparison of visual results of different fusion strategies

4.3.3 结果可视化

图9展示了本文方法在实际场景中的部分可视化结果。图9(a)为基于事件相机的无人机目标检测效果,采用本文提出的基于事件的检测方法,利用CeleX相机在室外环境实现了实时目标检测。可以看出,该方法在无人机目标检测任务中表现出良好的效果,能够有效减少误检的发生,显著提升了检测的可靠性和精确性。图9(b)为本文设计的跨模态融合模块在RGB相机数据上的识别效果。通过对事件相机与RGB相机数据的有效融合,该方法在不同场景中实时准确地识别并跟踪无人机目标。实验表明,本文提出的融合方法能够适应复杂背景和不同光照条件下的目标检测需求,具有较强的鲁棒性。图9(c)为系统在连续多帧检测到目标后,通过智能算法动态控制云台的变焦功能,实现对目标的精准放大以及在可见光单模态下的进一步检测与跟踪结果。结果表明变焦后仍然能够较为准确地识别并跟踪上目标。

图9前两行分别展示了在无背景天空、以高楼为背景的环境下,目标距离约为300 m时本文方法的检测识别结果;第3行展示了在同时存在树木和建筑物背景下,目标距离约为80 m时的检测识别结果,

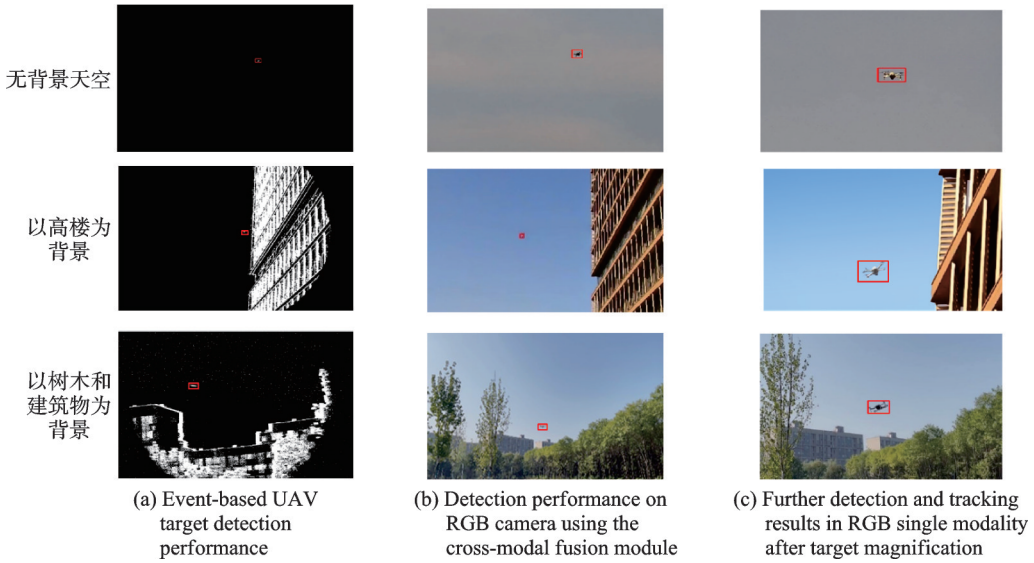

图9 所提方法在实际场景中的检测识别结果

Fig.9 Detection and recognition results of the proposed method in actual scenes

可以看出在不同环境背景、不同探测距离时,本文方法均能取得较好的结果。同时,本文方法的检测速度也满足要求,实现了精度和实时性的双重需求,满足了低慢小目标检测与跟踪的实际应用需求,进一步证明了本文方法在实际应用中的可行性与高效性。

5 结束语

本文提出了一种基于事件相机与可见光相机协同的低慢小目标检测系统。该系统首先利用事件相机的大视场范围与高速成像特点,对场景全局扫视,并采用基于事件的目标检测算法实现初始检测与粗定位。随后,通过空间与语义信息的双重协同融合技术,进一步精确小目标的定位与检测。在此基础上,可见光相机借助高分辨率成像与智能变焦功能进行"凝视",从而获得更细致的目标形态与特征。该"事件驱动-图像精细"的双路协同模式兼具快速响应与高精度,能在多目标、复杂背景和突发运动等多种场景下保持较强的检测与跟踪能力,既提升了检测效率,又满足了低慢小目标检测与跟踪的实际应用需求,展现出良好的鲁棒性与实用性。

参考文献:

[1] 金永光,叶方伟,卢晓珍,等.区块链赋能的低空智联网[J].数据采集与处理,2024,39(1):2-14.
JIN Yongguang, YE Fangwei, LU Xiaozhen, et al. Low-altitude intelligent network empowered by blockchain[J]. Journal of Data Acquisition and Processing, 2024, 39(1): 2-14.

[2] 汤新民,顾俊伟,刘冰,等.低空监视技术及其发展趋势综述[J].南京航空航天大学学报,2024,56(6):973-993.
TANG Xinmin, GU Junwei, LIU Bing, et al. Review on low-altitude surveillance technology and its development trend[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2024, 56(6): 973-993.

[3] 徐辰宇,曹杰,杨峰,等.远距离"低慢小"目标探测技术研究进展(特邀)[J].激光与光电子学进展,2024,61(20):27-40.
XU Chenyu, CAO Jie, YANG Feng, et al. Advances in long-range low-slow-small target detection technology (Invited) [J]. Laser & Optoelectronics Progress, 2024, 61(20): 27-40.

[4] XU D, ZHANG H. Study of low-altitude slow and small target detection on radar[C]//Proceedings of 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017). Beijing, China: Atlantis Press, 2017: 529-532.

[5] 李赟玺.面向"低慢小"目标探测与识别的激光雷达关键技术研究[D].哈尔滨:哈尔滨工业大学,2020.
LI Yunxi. Research on key technologies of lidar for detection and recognition of "low, slow, and small" targets [D]. Harbin: Harbin Institute of Technology, 2020.

[6] FAN S, WU Z, XU W, et al. Micro-doppler signature detection and recognition of UAVs based on OMP algorithm[J]. Sensors, 2023, 23(18): 7922.

[7] HU Y, WU X, ZHENG G, et al. Object detection of UAV for anti-UAV based on improved YOLO v3[C]//Proceedings of 2019 Chinese Control Conference (CCC). Guangzhou, China: IEEE, 2019: 8386-8390.

[8] NALAMATI M, KAPOOR A, SAQIB M, et al. Drone detection in long-range surveillance videos[C]//Proceedings of 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Taipei, China: IEEE, 2019: 1-6.

[9] REN S, HE K, GIRSHICK R B, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39: 1137-1149.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. [S.l.]: Springer Cham, 2016: 21-37.

[11] WANG W, WANG P, NIU Z. A real-time detection algorithm for unmanned aerial vehicle target in infrared search system [C]//Proceedings of 2018 IEEE International Conference on Signal Processing, Communications and Computing. Qingdao,

China: IEEE, 2018: 1-5.

[12] ZHANG P, WANG X, WANG X, et al. Infrared small target detection based on spatial-temporal enhancement using quaternion discrete cosine transform[J]. IEEE Access, 2019, 7: 54712-54723.

[13] GALLEGO G, DELBRÜCK T, ORCHARD G, et al. Event-based vision: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 154-180.

[14] LICHTSTEINER P, POSCH C, DELBRÜCK T. A 128 × 128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change[C]//Proceedings of 2006 IEEE International Solid State Circuits Conference—Digest of Technical Papers. CA, USA: IEEE, 2006: 2060-2069.

[15] KIM J, BAE J, PARK G, et al. N-imageNet: Towards robust, fine-grained object recognition with event cameras[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 2146-2156.

[16] NGUYEN A, DO T, CALDWELL D G, et al. Real-time 6DOF pose relocalization for event cameras with stacked spatial LSTM networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. CA, USA: IEEE, 2019: 1638-1645.

[17] LI Y, WANG Z, WANG L, et al. Actions as moving points[C]//Proceedings of Computer Vision—ECCV. Glasgow, UK: Springer International Publishing, 2020: 68-84.

[18] LI J, WEN Y, HE L. SCConv: Spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Vancouver, Canada: IEEE, 2023: 6153-6162.

[19] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[20] CHEN Y, ZHANG C, CHEN B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in Biology and Medicine, 2024, 170: 107917.

[21] LI J, LI J, ZHU L, et al. Asynchronous spatio-temporal memory network for continuous event-based object detection[J]. IEEE Transactions on Image Processing, 2022, 31: 2975-2987.

[22] LIU B, XU C, YANG W, et al. Motion robust high-speed light-weighted object detection with event camera[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-13.

[23] FINN C, GOODFELLOW I, LEVINE S. Unsupervised learning for physical interaction through video prediction[J]. Advances in Neural Information Processing Systems, 2016, 29: 64-72.

[24] LIU M, ZHU M. Mobile video object detection with temporally-aware feature maps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 5686-5695.

[25] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [J]. Advances in Neural Information Processing Systems, 2015, 28: 802-810.

[26] TOMY A, PAIGWAR A, MANN K S, et al. Fusing event-based and RGB camera for robust object detection in adverse conditions[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Philadelphia, USA: IEEE, 2022: 933-939.

[27] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of European Conference on Computer Vision. [S.l.]: Springer Cham, 2018: 3-19.

[28] ROSS T Y, DOLLÁR G. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV).Venice, Italy: IEEE, 2017: 2999-3007.

[29] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

[30] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 7464-7475.

[31] ZHU X, WANG S, SU J, et al. High-speed and accurate cascade detection method for chip surface defects[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-12.

[32] ZHANG H, DENG L, BI L, et al. Small object detection algorithm based on improved Yolov5[C]//Proceedings of 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT). Jilin, China: IEEE, 2023: 280-283.

[33] PEROT E, DE TOURNEMIRE P, NITTI D, et al. Learning to detect objects with a1 megapixel event camera[J]. Advances in Neural Information Processing Systems, 2020, 33: 16639-16652.

作者简介:

常宇轩(2001-),男,硕士研究生,研究方向:仿生动态视觉处理、基于事件相机的目标检测,E-mail: xd_cyx@163.com。

杨文(1993-),男,助理研究员,研究方向:基于事件相机的视觉处理、图像质量增强/评估,E-mail: yangwen@xidian.edu.cn。

吴金建(1986-),**通信作者**,男,教授,研究方向:图像智能处理、高质量成像、目标检测,E-mail: jinjian.wu@mail.xidian.edu.cn。

(编辑:张黄群)