

基于事件相机的目标检测与跟踪算法综述

邱嘉钰, 张雅声, 方宇强*, 李鹏举, 郑开源

航天工程大学, 北京 101416

摘要 事件相机是一种受生物启发的新型视觉传感器,是计算神经科学与计算机视觉领域的交叉学科与研究热点。不同于传统相机,事件相机能够异步输出有关亮度变化的事件流,具有高时间分辨率、高动态范围、低延迟、低带宽和低功耗等优势,适合高速运动目标的动态实时感知,是计算机视觉检测与跟踪领域新的研究方向。首先,简要介绍了事件相机的类型、工作原理与优缺点;其次,着重分析了现有的基于事件相机的目标检测与跟踪算法;然后,介绍了视觉检测跟踪相关的事件数据集;最后,对该方向未来的发展趋势进行了展望。

关键词 事件相机; 动态视觉传感器; 目标检测; 目标跟踪; 数据集

中图分类号 TP212 文献标志码 A

DOI: 10.3788/LOP241073

Review of Event Camera-Based Target Detection and Tracking Algorithms

Qiu Jiayu, Zhang Yasheng, Fang Yuqiang*, Li Pengju, Zheng Kaiyuan

Space Engineering University, Beijing 101416, China

Abstract Event cameras are novel visual sensors inspired by biology, representing an interdisciplinary research hotspot in computational neuroscience and computer vision. Unlike traditional cameras, event cameras can asynchronously output event streams related to brightness changes, offering advantages such as high time resolution, wide dynamic range, low latency, low bandwidth, and low power consumption. They are suitable for real-time dynamic perception of high-speed moving targets and represent a new research direction in computer vision detection and tracking. This review first introduces the types, working principles, advantages, and disadvantages of event cameras, followed by an in-depth analysis of existing object detection and tracking algorithms based on event cameras. Subsequently, event datasets related to visual detection and tracking are introduced. Finally, future development trends in this field are discussed.

Key words event camera; dynamic visual sensor; target detection; target tracking; dataset

1 引言

目标检测与跟踪是计算机视觉领域的重要技术,在实时监控、自动驾驶、增强现实等领域都有着广泛的应用^[1]。目标检测通过识别图像或视频中的目标物体定位其位置,而目标跟踪则在视频序列中持续追踪目标对象的位置并实现目标位置信息的更新。当前的目标检测跟踪算法是基于传统帧相机的,并且这些算法已经相当成熟。自 20 世纪 50 年代起,虽然已经研究出大量检测跟踪算法,但传统相机以固定帧率生成图像,在具有挑战性的场景(例如快速运动、极端光照、背景杂乱)下存在一些局限性,难以保证检测跟踪的实时性和精度^[2]。

为了应对上述挑战,一些研究人员采用受生物启发的事件相机来进行目标检测与跟踪^[3]。事件相机是一种新型传感器,与其生物对应物一样由场景中发生的事件驱动。与以帧方式记录强度的常规可见光相机不同,事件相机能够输出连续的异步事件流,仅当视觉强度发生变化时,事件相机的像素才会独立发送信息(也称为产生事件)。事件相机具有高动态范围、高时间分辨率、低功耗和低带宽的出色性能^[4],并且不会受到运动模糊的影响,这些特性很好地弥补了现有光学设备的局限性。因此,事件相机广泛应用于高速运动目标感知、实时监控、视觉检测与跟踪和机器人视觉中^[5]。

本文在充分参考相关研究文献的基础上:首先介

收稿日期: 2024-04-10; 修回日期: 2024-05-28; 录用日期: 2024-06-17; 网络首发日期: 2024-06-24

基金项目: 国家自然科学基金(61906213)

通信作者: *fangyuqiang@nudt.edu.cn

绍了事件相机基础概述;其次对基于事件相机的目标检测方法进行归纳综述;再次对基于事件相机的目标跟踪方法进行归纳综述;随后介绍事件相机数据集;最后对未来事件相机检测跟踪方法值得关注的热点问题进行了展望。

2 事件相机基础概述

2.1 事件相机类型及工作原理

2.1.1 DVS

目前被广泛应用的事件相机可分为 3 类:动态视觉传感器(DVS)、异步时基图像传感器(ATIS)、动态主动像素视觉传感器(DAVIS)。其中,ATIS 和 DAVIS 均是在 DVS 的基础上发展演化而来的^[3]。

2006 年,Lichtsteiner 等^[5-6]提出第一个商业用途的 DVS,不同于传统帧相机,事件相机仅以局部像素级亮度变化的相关信息作为输出。DVS 工作原理如图 1 所示,当某一像素的对数光强度变化在某段时间超过敏感度阈值 C 时,相机输出一个事件 $e = \{x, y, p, t\}$ 。其中, (x, y) 为事件在二维像素空间的坐标, t 为时间戳, p 为极性。

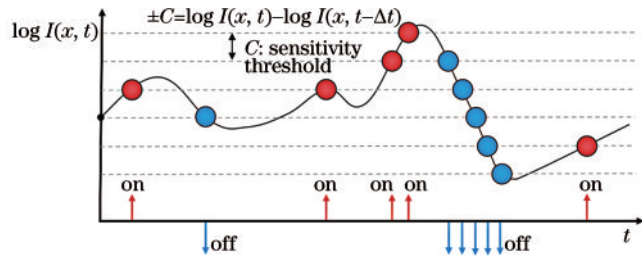


图 1 动态视觉传感器采样原理

Fig. 1 Sampling principle of dynamic vision sensor

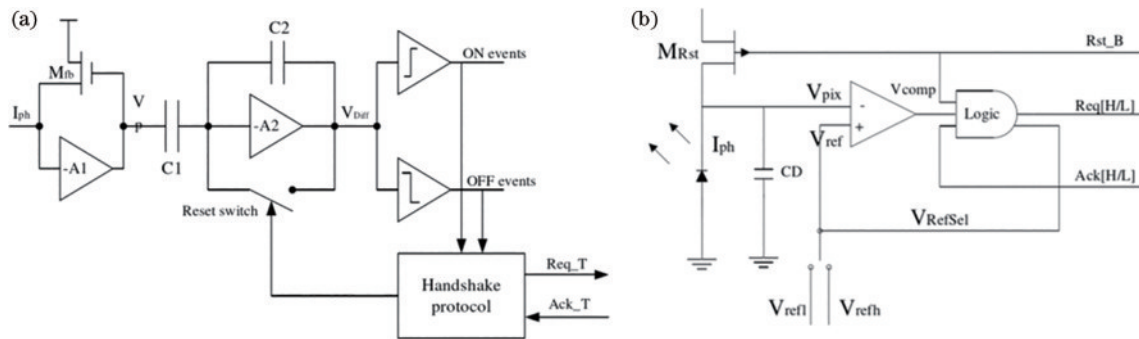


图 2 ATIS 像素电路结构^[9]。(a)A 部分;(b)B 部分

Fig. 2 Circuit structure of ATIS pixel^[9]. (a) Part A; (b) part B

2.1.3 DAVIS

为了解决 DVS 难以直接适配现有算法的问题,2014 年 Brandli 等^[10-11]开发了 DAVIS。DAVIS 传感器融合了 DVS 和主动像素传感器(APS)^[12]的优势,实现了对场景动态事件和灰度信息的同步输出。

DAVIS 的像素电路结构如图 3 所示^[13],分成两个部分:APS 和 DVS。在其电路设计上,DVS 部分负责

DVS 响应 t_k 时刻的对数光强度变化,产生有效事件的过程可以描述为

$$L(x_k, y_k, t_k) = \log_{10} I_k \quad (1)$$

$$\Delta L(x_k, y_k, t_k) = L(x_k, y_k, t_k) - L(x_k, y_k, t_k - \Delta t_k) \quad (2)$$

$$\Delta L(x_k, y_k, t_k) = p_k C \quad (3)$$

式中: t_k 为某一时刻; $L(x_k, y_k, t_k)$ 为 t_k 时刻的对数光强度; I_k 为 t_k 时刻某一像素点的光照强度或亮度; Δt_k 表示在同一像素处上次事件发生经过的时间; $\Delta L(x_k, y_k, t_k)$ 表示对数光强度的变化量; C 为阈值 ($C > 0$); p_k 为极性, $p_k = 1$, 表示相应像素的光照强度在该事件发生时增加, $p_k = -1$, 表示光照强度相应像素的光照强度在该事件发生时减少。

2.1.2 ATIS

2008 年,Posch 等^[7-8]提出了 ATIS,其在提供事件信息的同时还提供一定的灰度信息,满足了对事件类型数据可视化的需求,减轻了后期数据处理的难度。

ATIS 像素电路结构如图 2 所示^[9],主要分为两个主要区域(A 和 B),每个区域都装有感应器。区域 A 内嵌了完备的 DVS 系统,能够感知对数级别光强变化并触发相应的事件输出。而区域 B 内的感应器则负责测量光强变化并进行曝光作业。不同于传统相机的连续帧曝光方式,ATIS 像素采取一种基于时间的曝光策略。这种独特的曝光机制允许 ATIS 即使在高动态范围和复杂光照条件下也能够输出带有纹理的灰阶图像信息,同时仅对那些发生事件的局部像素进行输出,有效地节省了数据资源。但是,若环境光线太暗,可能导致曝光时间过长。如果曝光时间过长,而且超过了下一个事件的激发时刻,电容器会被迫重置,这可能中断曝光进程,引发图像信息的丢失^[3]。

监测光线强度的动态变化并产生事件流,而 APS 部分则负责进行曝光以获取场景的灰度图像。与 ATIS 的设计不同之处在于,DAVIS 中的 DVS 和 APS 共享同一个光敏二极管,且 APS 电路采用更小的尺寸设计,这导致 DAVIS 像素的面积比 ATIS 像素的面积小了 60%。这种紧密集成的设计使得 DAVIS 具有较高的空间效率^[14]。通过 DAVIS 获得的灰度图像与通过传

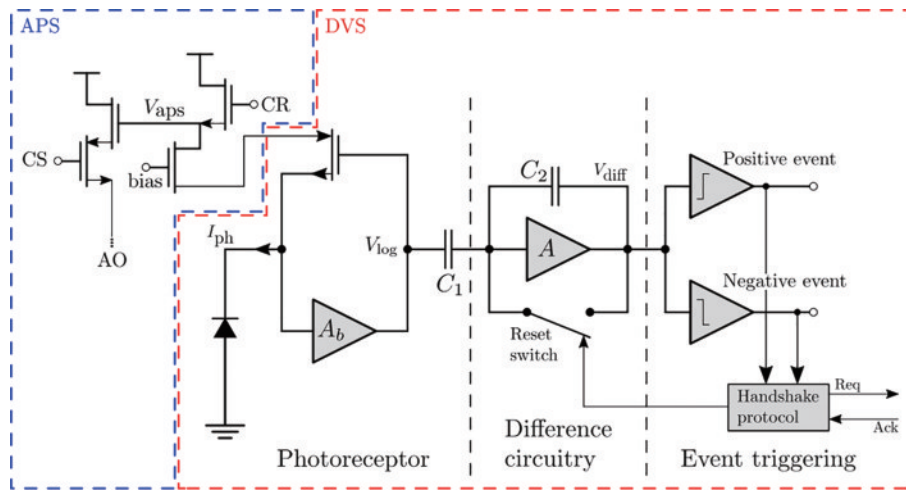


图3 DAVIS 像素电路结构^[13]
Fig. 3 Circuit structure of DAVIS pixel^[13]

统帧相机获得的灰度图像一样,读取时有动态范围限制,存在时间冗余和空间冗余的问题^[3]。

事件相机由于其特殊的属性,可适用于多种场合。
表1概括了3大类传感器的独有特性及其应用领域。

表1 3种传感器的特点和用途对比
Table 1 Comparison of the characteristics and uses of the three sensors

Type	Advantage	Disadvantage	Application scenario
DVS	It only outputs event information and is able to maintain the advantages of low latency, low power consumption and high dynamic range	Unable to output any grayscale information, poor visibility	It is suitable for situations with low requirements for visualization and high requirements for low latency, such as tracking, counting or motion monitoring of high-speed moving objects, real-time positioning
ATIS	It is capable of outputting event information and gray scale information of pixels where light intensity changes occur, with some visibility	It is prone to exposure anomalies and information loss when the ambient brightness changes slowly	It is suitable for frequent changes in environmental brightness and the existence of high-speed movement occasions, mainly used for real-time monitoring of products in industrial manufacturing
DAVIS	It can combine event information and grayscale information with good visibility, high temporal resolution, and has the ability to obtain detailed information about movements	It can be affected by defects such as redundancy of information in APS cameras, low temporal resolution, and low dynamic range	It is suitable for occasions with high visualization requirements and small dynamic range, such as target recognition, target detection, tracking and localization of targets, especially in the field of robotics and unmanned vehicles

2.2 事件相机的优势与挑战

相比于传统相机,事件相机主要有以下优点:

1) 高时间分辨率:由于事件相机检测光照强度的变化,因此事件相机响应速度快,输出频率可高达1 MHz,远高于传统帧相机100~1000 frame/s的帧率^[15],这使得事件相机适合高速运动目标的捕捉,并且不会受到运动模糊的影响。

2) 高动态范围:由于事件相机的光感受器是对数响应的,因此事件相机的动态范围可高达140 dB,远高于传统帧相机60 dB的动态范围^[16]。这使得事件相机可以在极端光照场景下采集到视场中目标的动态信息。

3) 低时延:由于事件相机的异步特性,使其对输入变化的延迟仅在 μs 数量级^[3]。

4) 低带宽、低功耗:由于事件相机仅传输亮度变化,避免了大量冗余数据的传输,事件相机平均数据带宽需求仅为1 Mbit/s,远远低于传统帧相机的32 Mbit/s^[15]。大多数事件相机的功耗约在10 mW级^[16]。

事件相机因具备以上优点,非常适用于存在高动态范围和高速运动目标的场景。然而,事件相机作为一种新型相机,在硬件造成的挑战和数据处理范式等方面还存在很大的挑战。

1) 噪声大、空间分辨率低:由于事件相机差分的工作原理,噪声对事件相机的输出有较大影响,这是限

制事件相机发展的一个重要因素^[3]。并且由于 DVS 的输出缺乏目标纹理信息,空间分辨率远低于传统帧相机。

2) 数据的处理:与传统帧相机的输出不同,事件相机输出的事件流通常包含像素坐标、时间戳和极性,难以适应现有成熟的计算机视觉算法,因此需要进行事件流数据处理以及开发适用于事件相机的新算法^[17]。

3) 事件流的稀疏特性:由于事件相机的稀疏特性,事件相机输出的事件仅包含存在相对运动的目标,信息量较少,目前仅利用这些信息去还原场景或提取场景中感兴趣的对象仍很困难^[3]。

3 目标检测方法现状研究

根据所使用的数据类型,基于事件相机的目标检测算法可以分为两类:一类是基于纯事件流的算法,另一类是基于事件流与传统图像帧相结合的算法。

3.1 基于纯事件流的目标检测算法

基于事件流的目标检测算法直接利用事件的输出进行目标检测,可以实现更快的响应速度和更高的动态范围。为了突出事件流高时间分辨率的特性,许多学者使用事件流直接进行目标检测。

角点检测是计算机视觉领域一种常用的特征检测方法,用于检测图像中的角点位置,具有简单、计算成本低等优点。Benosman 团队^[18]利用空间几何关系对行人进行角点匹配和检测,基于事件流的时间特征,将传统角点检测的问题转化为局部速度估计问题,但是该方法对远距离目标检测效果较差。

针对边界目标的检测问题,Aloimonos 团队^[19]采用结构化随机森林对目标的角点进行检测并提取目标轮廓。该方法考虑了数据之间的空间相关性,提取多个事件数据特征,编码后对结构化随机森林进行训练,最后基于高斯核滤波器进行匹配,从而实现目标检测。

Bartolozzi 团队^[20]在 iCub 机器人平台上利用脉冲流进行角点检测。该方法在局部尺度上实现了相同的固定事件表面角点检测,标记为角点的事件被聚集到角点轨迹中,从而可以估计每个角点的速度。

Han 等^[21]提出一种基于多层感知器的角点检测器,首先将传入事件 e 和时空邻域 s 作为输入来预测 e 是否为角点事件,随后将角点输入至多层感知机中进行目标分类检测。同时,该团队提出一种增强时间表面事件的表示方法,提高了分类器的准确性和鲁棒性。实验结果表明,该方法比其他方法具有更高的稳定性和准确性。

Li 等^[22]提出自适应时间阈值的异步事件角点检测算法,图 4 为检测框架。首先,基于快速傅里叶变换和带通滤波器,计算适应不同事件流的速度和纹理级别的时间阈值,用于去噪和角点检测;然后,构建自适

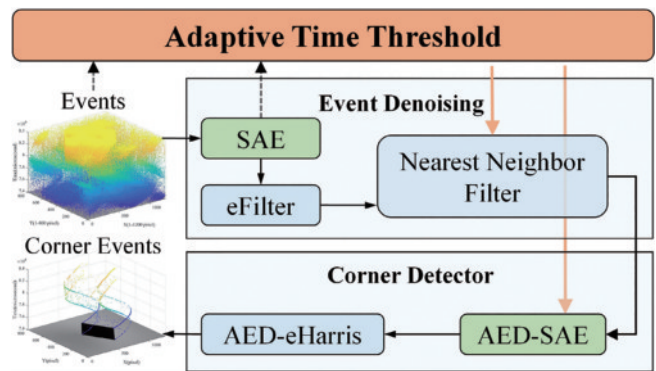


图 4 基于事件的自适应事件阈值的角点检测器^[22]

Fig. 4 Corner detectors of event-based adaptive event threshold^[22]

应指数衰减的活动事件表面(AED-SAE),提高角点检测的质量和速度;最后,提出一种结合 AED-SAE 和基于事件的 Harris 检测器的异步角点检测算法,使检测综合性能得到增强。

基于角点检测的方法对噪声比较敏感,且对尺度变换和旋转变换的目标不敏感,可能导致检测结果不准确。因此,许多研究人员提出基于直线检测的方法。Seilozakerini 等^[23]用脉冲神经网络实现 Hough 变换对脉冲流进行直线检测。该方法首先将事件数据传入脉冲神经网络中,通过网络的神经元动态响应生成边缘响应图。然后,利用这些边缘响应图进行直线检测,在 Hough 空间中找到直线参数,并利用迭代方法优化直线参数,以实现多条直线的检测和跟踪。

Benosman 团队^[24]利用迭代优化和最小二乘拟合方法对脉冲流进行直线检测,同时引入了指数衰减函数来隐式忘记较旧事件的贡献,以减少计算负载。

上述算法适用场景较为简单,对复杂场景下的目标检测性能低。随后许多研究者基于深度学习的方法检测更为复杂场景下的目标。图神经网络处理步骤如图 5 所示。浙江大学的张教授团队^[25]提出一种基于图的事件相机框架——SlideGCN,基于半径搜索法进行快速图的构建,利用卷积的局部空间连通性,避免了所有节点的处理,降低了目标检测复杂度。

Schaefer 等^[26]将事件处理为稀疏的静态时空图,并设计了异步的基于事件的图神经网络(AEGNNs)。通过为每个新事件制定有效的更新规则,将网络激活的重新计算限制在少数几个节点,在目标检测任务中,计算复杂度(FLOP)相较于人工神经网络大大降低,与最先进的异步方法具有相似甚至更好的性能,这为基于事件的低延迟处理打开了大门。

基于图的检测方法会受到计算复杂度和对图模型的依赖性的限制。Ryan 等^[27]采用三角线性函数将脉冲流编码为体素,在 GR-YOLO 模型上检测人脸并跟踪眼睛的运动。

Wang 等^[28]提出了一种双内存聚合网络(DMANet),

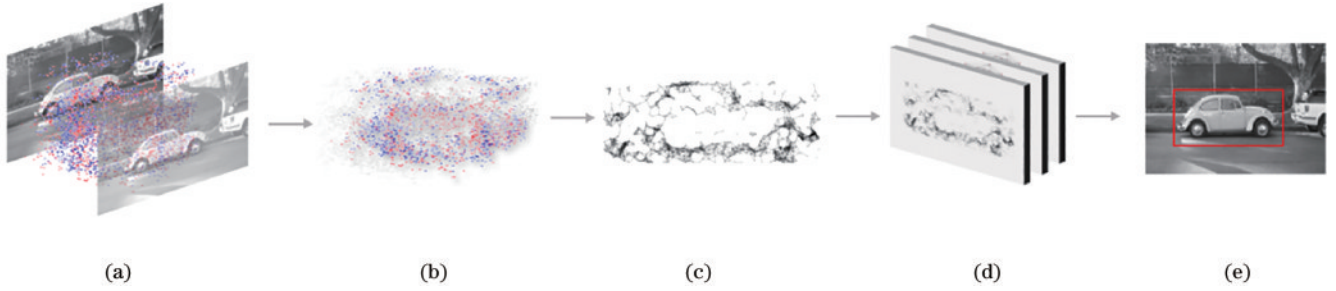


图 5 图神经网络处理步骤。(a)事件流;(b)二次采样;(c)图生成;(d)图神经网络;(e)预测

Fig. 5 Processing steps of graph neural network. (a) Event stream; (b) subsampling; (c) graph generation; (d) GNN; (e) prediction

利用事件流的长期记忆和短期记忆机制,能够提取有用的特征信息,进一步提升目标检测性能。图 6 展示了 DMANet 的框架,其中:远程内存被编码在自适应 ConvLSTM 的隐藏状态中,可用于捕获较长时间范围内聚合的信息特征以进行对象定位;而短程内存则通

过计算相邻时间间隔的事件支柱之间的时空相关性来建模,捕获在较短时间范围内聚合的信息特征以进行对象定位。由于 DMANet 需要处理大量的事件数据,并进行复杂的时空信息聚合,因此其计算复杂度较高,需要较大的计算资源。

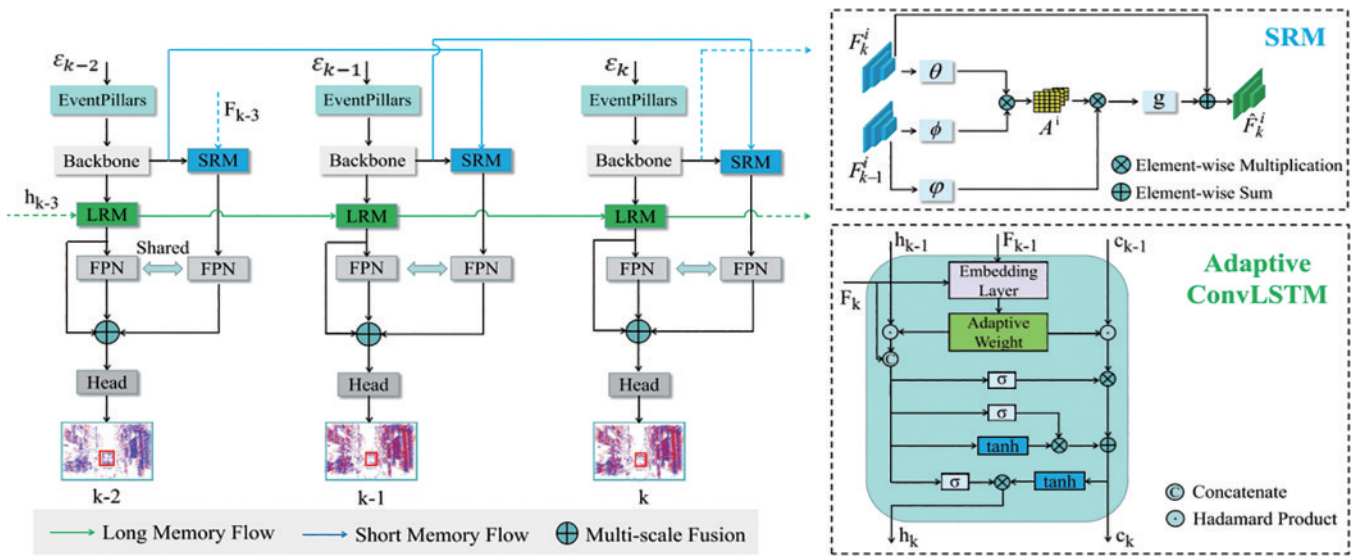


图 6 DMANet 目标检测框架^[28]

Fig. 6 DMANet target detection framework^[28]

近期,一些研究工作通过探索循环结构挖掘事件流的时序信息。北京大学的 Li^[29]提出了一种基于异步时空记忆网络 (ASTMNet) 的脉冲流式目标检测方法。ASTMNet 通过自适应时间采样策略和时域注意力卷积模块,从连续事件流中学习异步注意力嵌入,保留来自连续事件流的时空信息。此外,轻量级循环卷积模块通过设计循环交互层来实现目标检测骨干网络之间在时域上的记忆状态传递,提高了目标检测精度和推理速度,但是对于复杂背景下的缓慢运动或静止目标存在漏检的情况。

Nagaraj 等^[30]提出一种名为 DOTIE 的脉冲神经网络架构,利用事件相机的时间信息可有效检测移动物体。该方法通过物体速度对事件进行分离,然后在空间上进行进一步聚类以确定物体边界。DOTIE 具有异步性和对相机噪声的鲁棒性,在存在静态物体的背景下表现良好。该方法在需要快速准确物体检测的应

用如自主导航系统中具有潜在的应用价值。然而,由于该方法根据速度范围进行事件分离,可能无法准确检测速度范围较大的物体。

综上所述,基于纯事件流的算法在处理效率、能耗等方面具有优势,特别适合简单无杂乱场景下的目标检测,具有较高的准确性和实时性。然而,由于事件数据稀疏性和 DVS 低空间分辨率的特性,可能会在目标密集区域或复杂背景纹理下出现性能下降,也较难实现对远处目标的高精度检测。

3.2 基于事件流和传统图像帧的目标检测算法

基于深度学习的目标检测算法是目前的主流算法,为了增加复杂背景下目标的检测能力,Delbruck 团队^[31]将 DVS 输出的脉冲流按照频率累积为图像并与 APS 同步进行目标检测。该方法首先利用 DVS 的输出对潜在的目标区域进行聚类分析,形成预选区域。随后,在 APS 产生的图像上运用卷积神经网络来提取

潜在的目标位置,进而将感兴趣的区域(ROI)识别为前景或背景。最后,通过粒子滤波技术,从ROI中定位到目标的具体位置。由此,该团队有效解决了复杂背景中使用相机自主移动进行目标跟踪的挑战,显著提升了目标检测和追踪的效能。

Chen^[32]提出一种基于伪标签和监督学习方法的目標检测算法。该算法以 10 ms 的时间间隔将 DVS 事件流累积为伪帧图,同时利用卷积神经网络处理 APS 传感器的图像,生成 DVS 数据的目标检测结果。这些生成的结果被称为伪标签,并用来有监督地训练 YOLO 检测网络。该算法可用于自动驾驶场景中的车辆目标检测,提高了在相机运动和背景杂乱条件下的目标检测精度。

Iacono 等^[33]使用现有的深度学习技术来实现视觉事件的目标检测。首先,他们将 ATIS 脉冲流在时域方向投影为二值图像平面,然后利用 SSD 检测网络进行目标检测,并将其部署在 iCub 移动机器人平台上,成功解决了背景杂乱情况下的目标检测问题。此外,他们还提出了一种创新的处理流程,可借助成熟的基于帧的目标检测算法来引导基于事件的数据集进行数

据标注,以更快地生成所需的数据集。

Chen 等^[34-35]采用 3 种不同的事件流编码方法,包括基于频率、活动事件表面(SAE)和泄漏积分与触发(LIF),然后通过通道级融合和决策级融合,并最终通过动态置信融合函数进行融合,以实现更好的检测性能。其中:基于频率的编码方式统计每个像素点在给定时间间隔内发生的事件次数,并根据次数计算像素值;活动事件表面编码方法根据事件发生的时间来确定每个像素的像素值,反映时间信息而忽略频率信息;泄漏积分与触发编码方法基于泄漏积分与触发神经元模型,将每个像素视为一个神经元,根据膜电位和发放计数器来编码事件。

Jiang 等^[36]混合事件和传统帧通道,提出一种基于 CNN 的行人检测算法,算法框架如图 7 所示。该算法训练了两种 YOLO 网络模型,分别用于检测 DVS 和 APS 通道中的行人,以收集具有各自置信图的行人边界框,随后基于置信图融合方法将检测结果合成。相较于仅使用 APS 的相机,该方法帧速率提高了 2.28 倍、平均精度提高了 3%~18%。

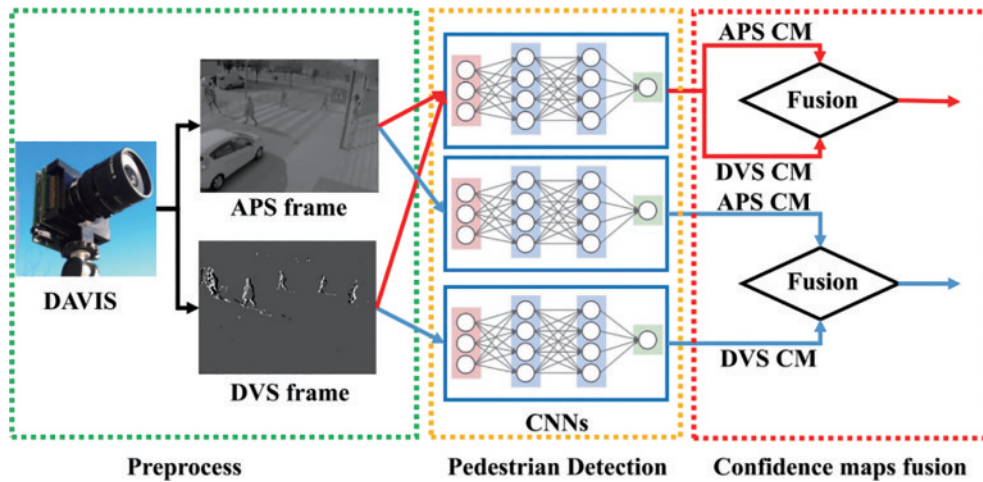


图 7 混合帧/事件驱动的行人检测框架^[36]

Fig. 7 Framework of mixed frame-/event-driven pedestrian detection^[36]

Sharghi 等^[37]使用事件直方图的方法将事件流转换为事件帧,在稀疏卷积神经网络对象检测模型的基础上,结合传统基于图像(左右翻转和裁剪)和基于事件(均匀噪声、去噪、空间抖动和极性翻转)的数据增强策略进行事件增强,以解决数据集大小有限的问题,然而,该算法检测缓慢移动的小物体时存在困难。

苏黎世大学 Gehrig 等^[38]提出了用于事件相机目标检测的循环视觉注意力模型,如图 8 所示。通过空间特征局部和全局自注意力机制以及保留时态信息的递归时态特征聚合,提升了参数效率。其在 GEN1 Automotive Detection 汽车数据集上实现 47.2% 的平均精度均值(mAP),同时推理速度比现有的技术大大提升。

中国科学院大学 Fang^[39]提出一种针对伪帧图像的目标检测算法,首先基于自适应时间分辨率的事件信息可视化方法将事件转化为伪帧图像,以 YOLO 网络为基础,优化网络的各部分结构,在较小规模的深度学习模型下就可以获得很好的检测效果。

综上所述,基于事件流与传统图像帧的方法主要将事件流表征为事件帧,或与传统帧进行联合检测,可以快速适应现有基于图像范式的目标检测算法,适用于背景复杂、高速运动、尺度变化目标的检测。然而该类方法没有充分挖掘事件流的时空特性,尤其是高时间分辨率的优势。

表 2 全面总结了基于事件相机的目标检测算法。进一步分析表明,深度学习被认为是未来的主流,在这些算法中起着重要作用。

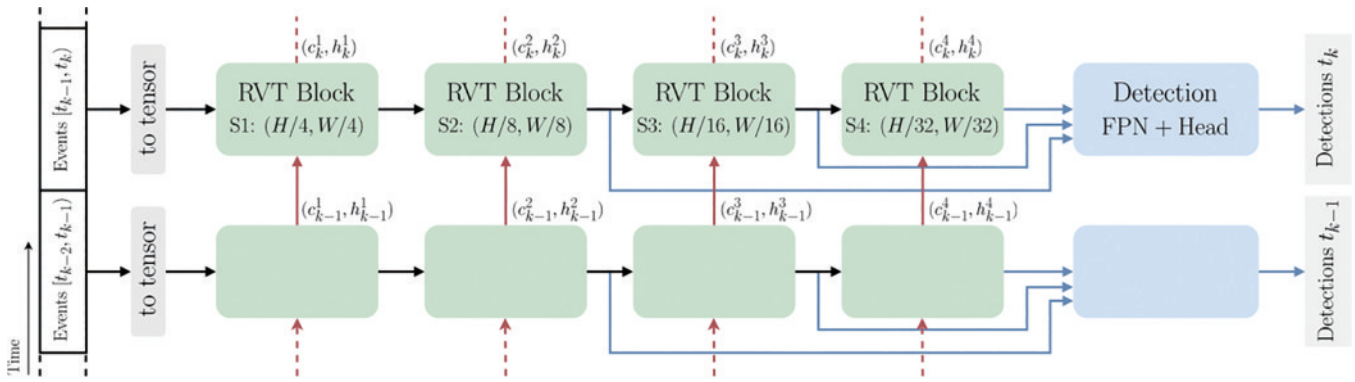


图 8 多阶段循环注意力模型^[38]

Fig. 8 Multistage recurrent attention model^[38]

表 2 基于事件相机的目标检测算法

Table 2 Object detection algorithms based on event camera

Type	Category	Reference	Year/source	Contribution
Pure stream of events	Angle detection	[18]	2015/Neural Networks	Transforming the traditional problem of corner point detection into a local velocity estimation problem
		[19]	2015/IEEE	Consideration of spatial correlation between data
		[20]	2017/IEEE	Realization of the estimation of the velocity of a corner point
		[21]	2023/ICVR	Design of an Enhanced Representation of Time Surface Events
	Linear Detection	[22]	2023/IEEE	Proposing a novel asynchronous corner point detector Arc*
		[23]	2016/BMVC	Development of an event-based Hough transform method
Pure stream of events	Deep learning	[24]	2018/IEEE	Proposing weighted least squares fitting method based on iterative events
		[25]	2021/IEEE	Proposing fast graph construction method based on radius search method
		[26]	2022/ IEEE	Design of an asynchronous event-based graph neural networks (AEGNNs)
		[27]	2021/Neural Networks	Proposing fully Convolutional Recurrent Neural Network Architecture
		[28]	2023/arXiv	Proposing a dual-memory aggregation network (DMANet)
		[29]	2023/Peking University	Proposing pulsed streaming target detection method based on asynchronous spatio-temporal memory network (ASTMNet)
Combination of event streams and traditional frames	Deep learning	[30]	2023/ICRA	Proposing a lightweight spiking neural network that can separate events based on the speed of the corresponding objects
		[31]	2016/ IEEE	Combining CNN and Particle Filtering
		[32]	2018/ IEEE	Training CNNs on pseudo-labels
		[33]	2018/ IEEE	Applying existing deep learning methods to the iCub robotics platform
		[34-35]	2019/Frontiers in Neurorobotics	Using three different event stream encoding methods and two fusion methods
		[36]	2019/IEEE	Proposing a confidence fusion method based on CNN detection results
Combination of event streams and traditional frames	Deep learning	[37]	2022/IEEE	Applying Asynet to maritime incident datasets for target recognition
		[38]	2023/IEEE	Exploring a multistage design approach and proposing Recurrent Vision Transformers for object detection
		[39]	2023/Chinese Academy of Sciences	Proposing an adaptive temporal resolution method for visualizing event information

4 目标跟踪方法现状研究

基于事件相机的目标跟踪方法正处于起步研究阶段,根据使用的数据类型,分为基于纯事件流的目标跟踪算法和基于事件流与传统图像帧相结合的目标跟踪算法。

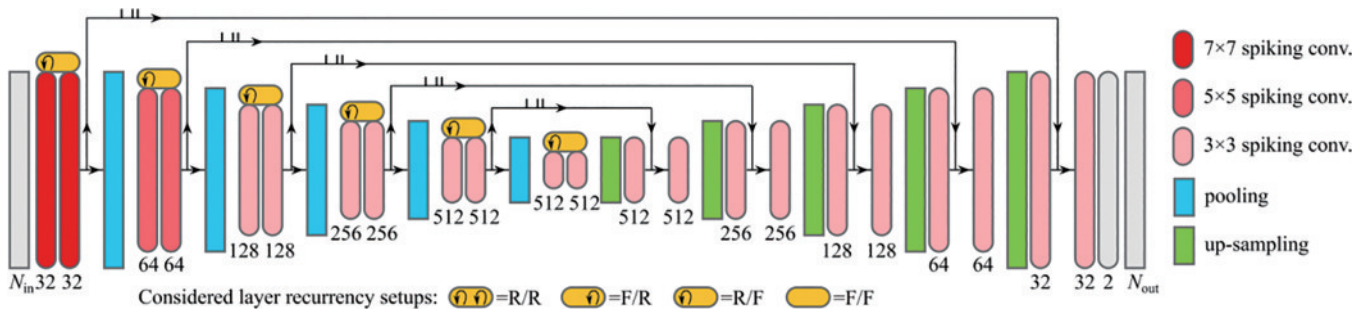
4.1 基于纯事件流的目标跟踪算法

基于角点的检测方法的目标跟踪算法不仅在传统图像帧上得到广泛应用,而且在事件流上也有着良好的跟踪效果。Alzugaray 等^[40]在 e-Fast 角点检测算法基础上提出了全新的 Arc 角点检测算法。在事件流中实时建立异步检测之间的对应关系,由此来实现跟踪。同年,Alzugaray 等^[41]还提出了基于事件流的 ACE* 角点检测跟踪算法,采用一种新颖的基于事件的视觉前端管道,通过预过滤事件流,检测其中的角点并以完全异步的方式实时跟踪它们,可以充分利用事件相机的优势。在此基础上,Liu^[42]提出一种基于异步事件流的角点检测与跟踪方法,在 ACE 跟踪器的空间约束和时间约束基础上添加速度约束,提高了目标跟踪的精度。

基于聚类目标跟踪算法得到广泛发展与研究。Pikatowska 等^[43]采用高斯混合模型来实现严重遮挡情况下的多行人跟踪。该算法通过高斯混合建模形成的事件簇来表征对象,基于期望最大化(EM)算法实现 GNN 参数估计,使用最大后验概率方法来进行簇更新,从而满足时间平滑度约束。然而,EM 估计对新数据非常敏感,可能导致跟踪错误。为了解决由物体遮挡引起的跟踪混淆问题,Benosman 团队^[44]采用立体视觉匹配聚类跟踪的算法,同时集成了 3D 重建和聚类跟踪,在两个任务中引入反馈信息以提高各自的性能。

针对大多数聚类技术需要先了解簇的数量或形状的问题,Barranco 等^[45]使用均值偏移和实时聚类的方法,通过最小化梯度下降实现了实时的跟踪算法。Ramesh^[46]提出一种基于事件的动态物体跟踪框架 e-TLD,该框架结合了事件相机和支持向量机(SVM)分类器,通过局部滑动窗口搜索和全局滑动窗口搜索相结合的方式实现物体的跟踪和检测。e-TLD 在动态场景(平移、旋转和六自由度)中表现出较好的跟踪性能。

光流估计通过计算图像序列中的相邻帧之间的像素运动的方向和速度,从而实现目标跟踪。Benosman 等^[47]提出一种基于事件的异步无帧光流估计算法,该算法基于 LK 光流估计的局部不变性假设,采用事件相邻活动像素的瞬时活动值代替基于帧的灰度值梯度,为光流估计提供了空间梯度。该算法在高数据稀疏性和高时间分辨率条件下以微秒精度和非常低的计算量进行光流计算,可对高速事件进行快速、实时分析,例如在成像移动射弹中或在基于视觉的高速自主飞机的引导中。Ye 等^[48]提出一种无监督学习的稀疏事件数据的光流估计算法,建立了一种全新的轻量级高性能网络架构(ECN),使用双线性插值调整特征大小,可应用于车辆的自动驾驶。Sekikawa 等^[49]提出基于事件的增量光流估计的切向拉长高斯置信传播,从基于稀疏法向流量的测量中实现增量全流量估计。同年,Schnider 等^[50]在 Timelens 的启发下提出基于脉冲神经网络的神经形态光流实时估计算法,该算法的框架如图 9 所示。该策略实现了高速目标的光流估计,其复杂性降低了近两个数量级,同时保持了准确性,为实时部署开辟了道路。



尽管这些跟踪算法在简单场景中表现出色,但它们在大规模跟踪基准测试中的性能仍待进一步验证。此外,在事件流过于稀疏或不足的情况下,这些基于纯事件流的跟踪算法通常无法准确定位目标并进行有效跟踪。

4.2 基于事件流与传统图像帧的目标跟踪算法

Kueng 等^[53]提出了事件流和灰度图相结合的特征轨迹跟踪算法。该算法首先在强度图像中使用 Canny 边缘提取和 Harris 角点提取得到特征点集,然后通过 ICP 算法的变体实现姿态估计。这项研究利用特征点跟踪技术,成功地在自然场景中对基于事件的 DAVIS 传感器进行六自由度运动的跟踪,是事件相机领域里程碑式的进展。然而,该算法在对传感器的非线性多重偏差建模方面存在一定的不足,这可能会导致对帧和事件数据的噪声敏感性增加。

Zhu 等^[54]对 Kueng 等的方法进行了改进,提出一种基于概率数据关联的异步事件传感器特征跟踪方法。传统的特征跟踪方法通常采用硬数据关联方式,即将事件分组作为固定的模型,并将未来事件分配给

这些模型后计算光流。然而,这种硬性数据关联容易降低光流质量并导致较短的光流轨迹。该方法引入了一种新颖的软数据关联模型,通过与光流计算相结合的交织 EM 算法计算关联概率,从而最大化对所有关联的期望(边缘化)。此外,为了实现更长的轨迹,该方法计算了相对于初始点的仿射变形,并将结果残差用作持久性的度量。通过计算得到的光流,每个特征的时间积分不同,并且与光流长度成反比。该方法在自我运动和非常快速的车辆序列中展示了相对于传统基于帧的相机的优越性。

由于目前现有的特征跟踪方法存在需要大量参数调整和对噪声敏感的问题,为了克服这些缺点,苏黎世大学的 Messikommer 等^[55]提出一个用于事件摄像机的数据驱动特征跟踪器,该跟踪器框架如图 10 所示。将灰度图像的模板补丁和事件流联合来跟踪特征,并引入了一种新颖的帧注意模块,该模块可以在同一图像中的特征之间共享信息,实现特征轨迹的更新,利用事件摄像机的高时间分辨率与标准帧相结合来最大限度地提高跟踪性能。

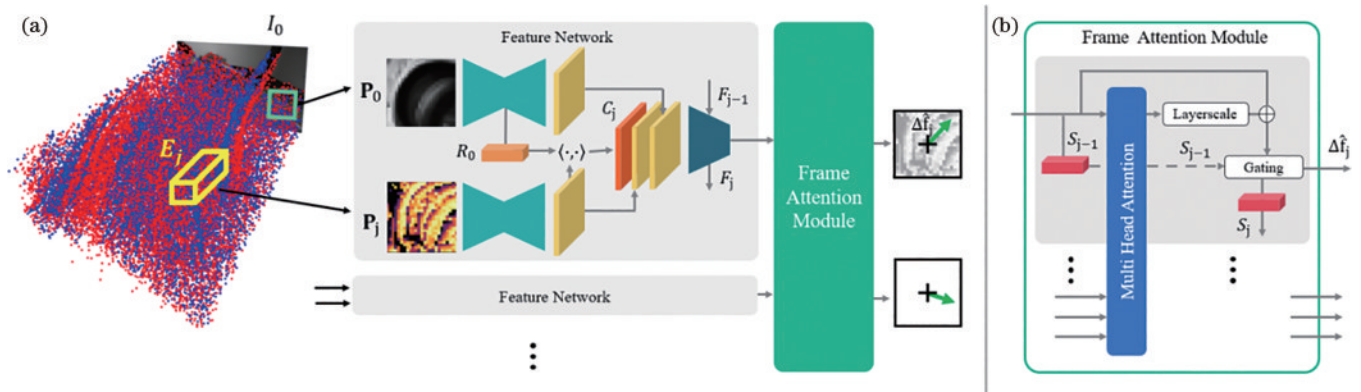


图 10 事件摄像机的数据驱动特征跟踪器^[55]。(a)跟踪器框架;(b)帧注意力模块

Fig. 10 Framework of mixed frame-/event-driven pedestrian detection^[55]. (a) Tracker framework; (b) frame attention module

相关滤波的目标检测跟踪就是设计一个滤波模板,利用该模板与目标候选区域做相关运算,最大输出响应的位置即为当前帧的目标位置。2021年,浙江工业大学 Liu^[42]使用滑动窗口将 DVS 累积为伪帧图像,采用核相关滤波的方法(KCF)实现对背景干扰、快速运动和遮挡目标的稳定跟踪,但该方法不适用于连续变形目标的检测跟踪。

许多研究者利用粒子滤波算法实现了对复杂、高对比度预设形状的逐事件自适应追踪。Arren 团队^[56]开发了一项提升目标追踪技术的方法,基于改良的粒子滤波器和多重假设的滤波手段,在时空维度能有效采样。这种改良能自动优化目标监测的时间窗口,确保在目标运动速度较慢的情况下依然能持续进行跟踪。此方法在追踪捕捉事件流中的目标时能够有效排除由相机运动产生的背景干扰,对相机和目标的相对速度变化显示出强大的鲁棒性,适宜在相机自由活动

的环境中使用。然而,该方法在对目标形状进行估计时存在一定误差,特别是在目标形状与背景相似或光照条件发生变化时,跟踪精度较低。

基于聚类的目标跟踪方法通过对目标的特征或运动轨迹进行聚类分析,将目标从背景和噪声中分离并进行跟踪。西安电子科技大学的 Wu 等^[57]提出了基于空间关系的聚类方法用于跟踪运动目标。该方法对于给定的事件在其附近建立待搜索区域,并设置时间阈值,通过待搜索区域事件数量与阈值的判断,实现感兴趣区域的确定,从而不断更新目标位置,实现高速运动网球目标的跟踪。Bolten 等^[58]提出层次聚类的目标跟踪算法,该算法首先用一个基于相邻事件的滤波器进行预处理,然后处理层次聚类事件,最后跟踪聚类质心。该算法通过将事件流分割成固定时间段的独立片段,并逐一对这些片段进行处理来完成目标的跟踪工作。该算法简单快速,几乎未考虑到每个事件的时间

信息。此类方法对尺度变化和遮挡目标都能够实现稳定跟踪,但对高速运动目标检测跟踪精度较低。

上述算法中事件相机均是固定的,背景对前景的影响较小。针对事件相机对高机动的小目标检测精度低的问题,Shu 等^[59]提出基于配准和前景增强模型进行目标检测与跟踪的方法,实现了运动场景下的目标检测与跟踪。采用自适应高斯混合模型(AGMM)去除背景噪声,并通过连通域分析和目标形态学分析,精度达到 90.2%。

早期的研究使用了各种方法,随着循环神经网络、卷积神经网络等深度学习模型的发展,大量研究人员

使用神经网络进行目标跟踪,提高跟踪精度。

Chen 等^[60]提出一种基于事件的对象运动估计的目标跟踪算法,算法框架如图 11 所示。该算法设计了具有线性时间衰减的同步时间表面 TSLTD 表示,有效地将异步视网膜事件的时空信息编码到具有清晰运动模式的 TSLTD 帧中。随后,将 TSLTD 帧序列提供给新颖的视网膜运动回归网络(RMRNet),以执行端到端 5 自由度对象运动回归。这种方法在处理快速运动和低光照条件等具有挑战性的环境方面具有优越性。

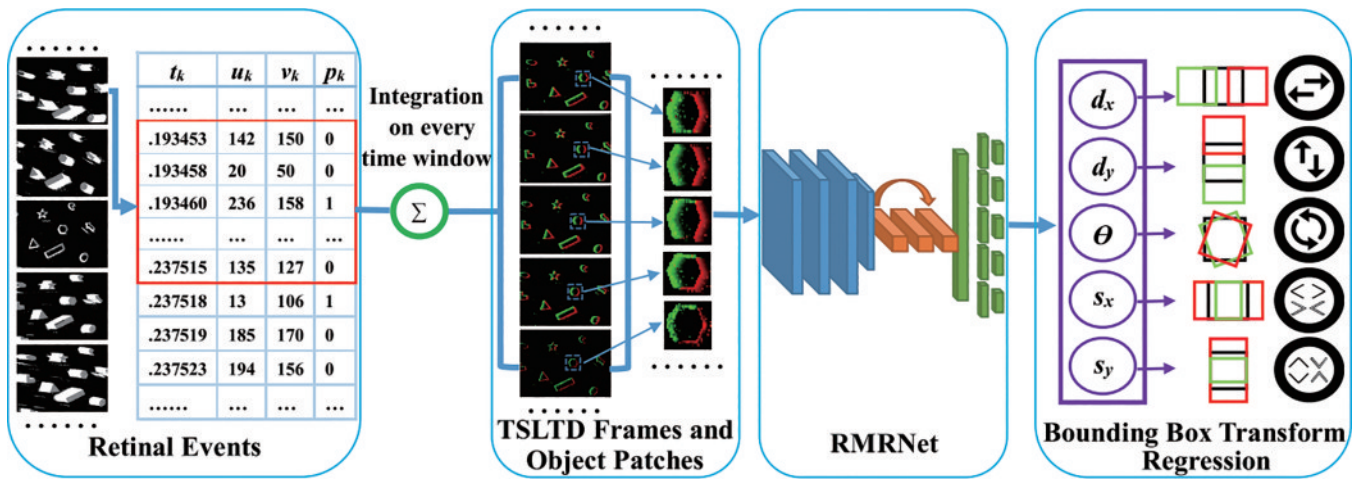


图 11 基于事件的对象运动估计的目标跟踪方法框架^[60]

Fig. 11 Framework for event-based object motion estimation for target tracking methods^[60]

Zhu^[61]设计了一种基于 DVS 与 APS 的双模态融合跟踪网络,APS 域与事件域的特征提取网络采用预训练好的 VGG16 网络,将 VGG16 的第十个卷积层(低维特征)与第十三个卷积层(高维特征)输出的中间结果提取出来,分别得到参考分支的 APS 域和事件域的低维特征与高维特征。在特征融合模块中加入通道注意力机制,将对应维度的 APS 域特征与事件域特征融合,实现了在正常光照、运动模糊、过曝光和低光照条件下对无人机的稳定跟踪,但仍存在检测框轻微漂移的问题。

大连理工大学 Zhang 等^[62]提出了一个脉冲注意力机制网络(STNet),从时间域和空间域动态提取和融合信息,不需要事件预处理,即可实现单目标跟踪。但当事件过于稀疏时,该方法的识别性能明显下降。

此外,该团队^[63]利用传统相机和事件相机之间的互补性,提出了一种多模态融合的方法,以增强单目标跟踪性能。多模态融合的目标跟踪算法框架如图 12 所示,为了将 DVS 和 APS 同步结合起来,该方法将异步事件离散化为时间切片,并将每个离散化的时间切片累积到传统帧中,从而可以轻松地通过基于 CNN 的模型处理。同时,设计了一种新颖的跨域特征集成器,通过自域和跨域注意方案自适应地进行事件域和帧域

视觉线索的有效融合,从而提高跟踪性能。然而,该方法没有利用基于事件的相机的高测量速率来实现低延迟跟踪。

在此基础上,该团队进一步提出一种用于高帧率跟踪的帧-事件对齐和融合网络(AFNet)^[64],该网络通过事件引导的跨模态对齐和交叉相关融合,实现了传统帧和事件的有效融合,提高了跟踪的准确性和鲁棒性。AFNet 包含两个关键组件:对齐模块负责在事件提供的移动线索的指导下,帧和事件的模态之间的跨样式和跨帧率对齐;而融合模块通过学习一个动态滤波器来将两种模态的互补信息融合,从而稳健地融合帧和事件的相关性,实现目标的高速率跟踪。但是该方法的计算复杂度较高,并且对于复杂背景下的目标跟踪性能有待进一步提升。

Wang 等^[65]通过设计一种新的跨模态分层知识蒸馏方案,提出一种新颖的基于事件的视觉跟踪框架。该方法同时输入 RGB 帧和事件流来训练基于 teacher Transformer 的多模式跟踪框架,随后通过新的分层知识蒸馏策略,包括成对相似性、特征表示和基于响应图的知识蒸馏,指导 student Transformer 网络的学习,从而实现从多模态到单态跟踪的知识迁移。该团队通过开发跨模态转换器(CMT)模块进一步构建了一种新

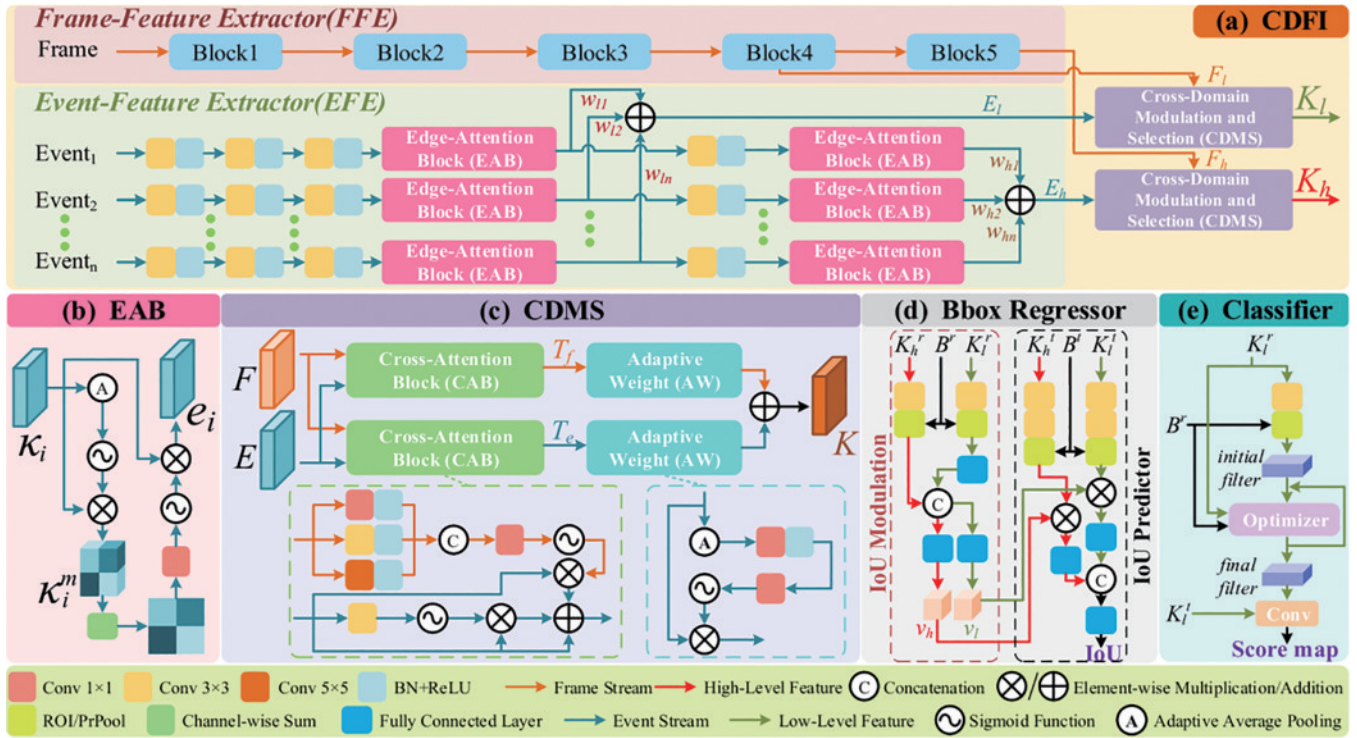


图 12 多模态融合的目标跟踪算法框架^[63]。(a)跨域特征集成器；(b)边缘注意模块；(c)跨域调制和选择模块；(d)Bbox 回归器；(e)分类器

Fig. 12 A framework for target tracking algorithms with multimodal fusion^[63]. (a) Overview of cross-domain feature integrator; (b) edge-attention block; (c) cross-domain modulation and selection block; (d) Bbox regressor; (e) classifier

颖有效的基线方法——VisEvent^[66]。这个转换器通过一个跨注意力层来连接双模态信息之间的交互,然后采用自注意力层从全局视角学习和增强特征表示。CMT 促进了内部和跨模态特征的交互,并显著改进了最终的跟踪结果。

Zhang 等^[67]提出一种用于动态对象跟踪的事件驱动时空方法。采用自适应策略来调整事件数据的时空域实现事件帧重建,应用 Siamese 网络从事件流中提取特征。在摄像机和物体同时剧烈旋转和晃动情况下,比其他最新模型的准确性提高了 30%。

综上所述,二进制事件数据是由动态视觉传感器产生的离散型事件流信息,记录光强的变化位置、时间戳和极性。与传统相机输出连续像素强度值不同,二进制事件以二元方式表示光强变化的极性,即光照强度增加或减少。在 DVS 中,二进制事件数据的异步采集方式意味着传感器根据光强的实际变化情况立即记录事件的发生,而非按照固定的时间间隔采集数据。

这种异步采集方式使得传感器能够实时响应场景变化,适应各种复杂视觉环境。然而,由于缺乏规模性的对象标注数据集,这也为高精度目标检测跟踪带来一定挑战。因此,大多数现有的目标跟踪方法都需要通过将事件流转化为类图像平面或将 DVS 与 APS 进行关联融合,进而快速适应现有目标跟踪算法,对复杂背景下高速运动目标跟踪性能好。然而,这种策略通常需要更多的计算资源,尤其在实时应用中,可能需要更大的计算能力来处理这两种数据源,从而增加硬件成本和功耗。另一方面,在固定时域窗口长度或固定事件数目的累积图像上,会呈现缓慢运动、弱纹理和高速运动模糊效应。

如表 3 所示,本文对基于事件相机的目标跟踪算法进行了汇总。其中,包括方法的分类、发布者、年份、来源和贡献。基于深度学习的目标跟踪算法正逐渐成为未来的主流算法,这一趋势将推动目标跟踪领域的进一步发展和应用。

表 3 基于事件相机的目标跟踪算法

Table 3 Event camera-based target tracking algorithm

Type	Category	Reference	Year/source	Contribution
Pure stream of events	Angle detection	[40]	2018/IEEE	Establishing correspondence between asynchronous detections in the event stream in real time
		[41]	2018/IEEE	Establishing local correspondences between corner events

表 3 (续)

Type	Category	Reference	Year/source	Contribution
Pure stream of events	Clustering	[42]	2021/Zhejiang University of Technology	Adding speed constraints to the ACE tracker
		[43]	2012/IEEE	Gaussian mixture modeling to form event clusters for object characterization
		[44]	2017/IEEE	Combination of stereo matching and object tracking
		[45]	2018/IEEE	mean-drift-clustering-based approach for event labeling
		[46]	2021/IEEE	Applying SVM to event cameras, combining local and global sliding window search
	Optical flow estimation	[47]	2012/Neural Networks	Replacing the gray value gradient by comparing the instantaneous activity values of neighboring active pixels
		[48]	2020/IEEE	Proposing an unsupervised learning algorithm for optical flow estimation of sparse event data
		[49]	2023/IEEE	Proposing incremental full flow estimation from sparse normal flow based measurements
		[50]	2023/IEEE	Proposing a real-time estimation algorithm for neuromorphic optical flow based on impulse neural networks
		Deep learning	[51]	2019/Frontiers in Neurorobotics
[52]	2022/Advances in Neural Information Processing Systems		Downsampling strategy to mine key events and embed the irregular spatio-temporal information of key events into a high-dimensional feature space	
Combination of event streams and traditional frames	Feature tracking	[53]	2016/IEEE	Proposing a feature trajectory tracking algorithm that combines event streams and grayscale maps
		[54]	2017/IEEE	An expectation maximization (EM) algorithm is used to quantify the probabilistic correlation between feature point sets and event data
		[55]	2023/IEEE	A novel frame-attention module is introduced
	Correlation filtering	[42]	2021/Zhejiang University of Technology	Kernel-correlation filtering is introduced
	Particle filtering	[56]	2017/IEEE	Multi-hypothesis filtering technology is introduced
	Clustering	[57]	2018/CHREOC	Clustering based on spatial relationships
		[58]	2019/ICCS	Hierarchical clustering
		[59]	2021/ICCC	Alignment and foreground enhancement models are introduced
		[60]	2020/AAAI	Designing a temporal surface representation of TSLTD
		[61]	2021/Dalian University of Technology	Channel Attention Mechanism added to Feature Fusion Module is introduced
[62]		2022/IEEE	STNET is proposed to dynamically extract and fuse information from the temporal and spatial domains	
Deep learning		[63]	2021/IEEE	Discretizing asynchronous events into time slices accumulated in conventional frames and designing a cross-domain feature integrator
	[64]	2023/IEEE	Design of an event-guided cross-modal alignment and crossover based module	
	[65]	2023/arXiv	Cross-modal hierarchical knowledge distillation scheme is introduced	
	[66]	2024/IEEE	Development of a cross-modal converter for bimodal information interaction	
	[67]	2022/Sensors	Adaptive strategy to adjust the spatial and temporal domain of event data for event frame reconstruction	

5 基于事件的目标检测与跟踪数据集

基于事件流的视觉跟踪近年来引起了越来越多的关注,凸显了基于事件的基准数据集日益增长的重要性。事件相机的发展尚未成熟,大规模的公开数据集仍较为缺乏,尤其是涉及小目标和高速运动场景的。由于事件相机能够连续输出异步事件流,因此对于事件数据集构建的难点在于数据的标注。本小节主要介绍应用于目标检测跟踪领域的相关事件数据集。

应用于目标检测领域的数据集包括:N-MNIST、N-Caltech101、MNIST-DVS、CIFAR10-DVS、N-CARS、Gen1 Automotive Detection Dataset 和 1Mpx Automotive Detection Detection 等。

Orchard 等^[68]采用 ATIS 相机记录液晶显示器上现有的 MNIST 和 Caltech101 数据集的图像,生成 N-MNIST 和 N-Caltech101 数据集。Serrano-Gotarredona 等^[69]通过传感器记录显示器中的移动 MNIST 示例,转换了 10000 个样本数据,生成 MNIST-DVS 数据集。Li 等^[70]使用 DVS128 相机记录显示器上 10 个不同类别的 10000 张基于帧的图像,构建了 CIFAR10-DVS 事件流数据集。Sironi 等^[71]使用安装在汽车挡风玻璃后面的 ATIS 摄像头记录了基于真实世界事件的大型汽车分类 N-CARS 数据集,数据集包括 12336 个汽车样本和 11693 个非汽车样本(背景)。Prophesee 公司采用 ATIS Gen1 相机采集自动驾驶场景车辆与行人构建了大规模的 Gen1

Automotive Detection Dataset^[72]。随后,该公司采用 Gen4 相机采集自动驾驶场景构建了 1Mpx Automotive Detection Detection^[73]。

应用于目标跟踪领域的数据集包括:VOT-DVS、TD-DVS、Ulster、EED、FE108、VisEvent 等。

VOT-DVS^[74]由 60 个单对象跟踪视频组成,每个视频的平均长度为 358 帧,其中有许多具有挑战性的移动背景。TD-DVS^[74]由 77 个单目标跟踪视频组成,每个视频的平均长度为 605 帧,该数据集旨在用于视觉任务中基于事件的学习,如跟踪和动作识别等。Ulster^[31]是由安装在 Pioneer 四轮机器人平台(捕食者机器人)顶部的 DAVIS 传感器捕获的数据集,共包含 20 min 的数据、9000 APS 帧和 1.6 亿个 DVS 事件。EED^[75]是使用 DAVIS240 对不同移动速度的多个大小不同的物体在不同的光照条件下采集的数据集,共包含 7 个视频,234 APS 帧。

FE108^[62]包含在一致光照条件下捕获的 108 个视频,总长度为 1.5 h,包含 21 个不同类别的目标对象,并引入了 4 个具有挑战性的因素。整个数据集总计 208672 帧,平均每个视频 1932 帧。FE108 在帧域和事件域上提供地面实况注释,注释频率分别高达 40 Hz 和 240 Hz。Wang 等^[66]构建了一个大规模可见事件基准 VisEvent,其由 820 个视频对组成,总计 371127 帧,每个视频的平均长度为 453 帧。该数据集包含 17 个挑战因素,包括 RGB 模式挑战(例如快速运动、低照度和曝光)以及事件模式挑战(例如缺乏目标移动、背景混乱等)。

表 4 基于事件相机的视觉跟踪数据集对比

Table 4 Comparison of event camera-based visual tracking datasets

Dataset	Year	Videos	Frames	Class	Attributes	Resolution	Aim	Public
VOT-DVS	2016	60				240×180	Eval	✓
TD-DVS	2016	77				240×180	Eval	✓
Ulster	2016	1	9000			240×180	Eval	×
EED	2018	7	234			240×180	Eval	✓
FE108	2021	108	208672	21	4	346×260	Train&Eval	✓
VisEvent	2021	820	371127		17	346×260	Train&Eval	✓

6 结论与展望

本文简要概括了事件相机的基本原理,重点介绍了基于事件相机的检测跟踪算法及相关数据集。事件相机具有高动态范围、高时间分辨率、低功耗和低带宽的优点,能够为极端光照场景下高速目标的运动感知提供潜在的解决方案。然而,基于事件相机的检测跟踪算法正处于研究起步阶段,与基于传统帧的检测跟踪算法相比还有很大的进步空间。因此,本文在对现有的基于事件相机检测跟踪方法进行总结的基础上,针对该领域做出以下展望。

1) 由于二进制事件和强度信息的异步获取,仅使用事件流进行检测跟踪会在复杂的场景中受到限制,

需要克服事件传递场景的外观变化以及传感器噪声和数据关联问题。因此,有效处理事件流信息并减少噪声是下一步研究的重点。

2) 当前的检测与跟踪算法主要针对事件相机处于静态或近静态状态的场景,然而当事件相机和目标同时处于运动状态时,将面临背景复杂、目标遮挡、混淆等挑战。因此,解决动态复杂场景下的目标检测与跟踪问题是必然的发展方向之一。

3) 基于可见事件的目标检测跟踪基准数据集还比较少,与基于可见相机的跟踪相比,该领域的发展相当缓慢。目前的评价指标主要基于传统帧的检测和跟踪评价方法,但基于事件相机的检测和跟踪方法缺乏有效的评价指标。因此,建立统一的基于事件的检测

和跟踪评价指标是未来研究的重点之一。

4) 即使事件相机在高动态范围、低能耗和高时间分辨率方面比传统 RGB 相机表现更好,但仅依靠事件相机在某些场景中仍然表现不佳,例如静止或静止的物体,因为只能感知运动的信号,而无法捕获静止的物体。另外,事件相机也无法捕获颜色信息或非常细粒度的纹理信息,这是 RGB 相机的优势所在。因此,同时使用 RGB 和事件相机进行跟踪具有很大的落地价值和潜在应用。

5) 事件相机具有高速、低带宽和低功率的显著优点,能够充分满足航空航天领域对传感器的严苛要求^[3]。未来,在追踪如导弹和飞机这类高速移动目标时,使用事件相机有望超越传统遥感相机,提供更优秀的性能表现。目前,事件相机在航天领域的研究和应用较少,具有很大的发展潜力。

参 考 文 献

- [1] 周小龙,刘倩倩,产思贤,等.基于事件相机的视觉跟踪算法综述[J].小型微型计算机系统,2020,41(11):2325-2332.
Zhou X L, Liu Q Q, Chan S X, et al. Event camera-based visual tracking algorithms: a survey[J]. Journal of Chinese Computer Systems, 2020, 41(11): 2325-2332.
- [2] Zhang J Q, Dong B, Fu Y K, et al. A universal event-based plug-In module for visual object tracking in degraded conditions[J]. International Journal of Computer Vision, 2024, 132(5): 1857-1879.
- [3] Fang Y H, Xu W, Piao Y J, et al. Development status and trend of event-based vision sensor[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(12): 1664-1673.
- [4] 邹浩东,昌敏,郭金库,等.基于事件的视觉SLAM综述[J].无人系统技术,2023,6(3):59-80.
Zou H D, Chang M, Guo J K, et al. Review of event-based visual SLAM[J]. Unmanned Systems Technology, 2023, 6(3): 59-80.
- [5] Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 30mW asynchronous vision sensor that responds to relative intensity change[C]//2006 IEEE International Solid State Circuits Conference, February 6-9, 2006, San Francisco, CA, USA. New York: IEEE Press, 2006: 2060-2069.
- [6] Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor[J]. IEEE Journal of Solid-State Circuits, 2008, 43(2): 566-576.
- [7] Posch C, Matolin D, Wohlgenannt R. An asynchronous time-based image sensor[C]//2008 IEEE International Symposium on Circuits and Systems (ISCAS), May 18-21, 2008, Seattle, WA, USA. New York: IEEE Press, 2008: 2130-2133.
- [8] Posch C, Matolin D, Wohlgenannt R. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS[J]. IEEE Journal of Solid-State Circuits, 2011, 46(1): 259-275.
- [9] Posch C, Matolin D, Wohlgenannt R. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless Pixel-level video compression and time-domain CDS[J]. IEEE Journal of Solid-State Circuits, 2011, 46(1): 259-275.
- [10] Brandli C, Berner R, Yang M H, et al. A 240×180 130 dB 3 μs latency global shutter spatiotemporal vision sensor[J]. IEEE Journal of Solid-State Circuits, 2014, 49(10): 2333-2341.
- [11] Brandli C, Muller L, Delbruck T. Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor[C]//2014 IEEE International Symposium on Circuits and Systems (ISCAS), June 1-5, 2014, Melbourne, VIC, Australia. New York: IEEE Press, 2014: 686-689.
- [12] Fossum E R. CMOS image sensors: electronic camera-on-a-chip[J]. IEEE Transactions on Electron Devices, 1997, 44(10): 1689-1698.
- [13] Cho D D, Lee T J. A review of bioinspired vision sensors and their applications[J]. Sensors and Materials, 2015, 27(6): 447-463.
- [14] 孔德磊,方正.基于事件的视觉传感器及其应用综述[J].信息与控制,2021,50(1):1-19.
Kong D L, Fang Z. A review of event-based vision sensors and their applications[J]. Information and Control, 2021, 50(1): 1-19.
- [15] 桑永胜,李仁昊,李耀仟,等.神经形态视觉传感器及其应用研究[J].物联网学报,2019,3(4):63-71.
Sang Y S, Li R H, Li Y Q, et al. Research on neuromorphic vision sensor and its applications[J]. Chinese Journal on Internet of Things, 2019, 3(4): 63-71.
- [16] Gallego G, Delbrück T, Orchard G, et al. Event-based vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 154-180.
- [17] Ramesh B, Zhang S H, Lee Z W, et al. Long-term object tracking with a moving event camera[EB/OL]. [2024-02-05]. <http://bmvc2018.org/contents/papers/0814.pdf>.
- [18] Clady X, Ieng S H, Benosman R. Asynchronous event-based corner detection and matching[J]. Neural Networks, 2015, 66: 91-106.
- [19] Barranco F, Teo C L, Fermüller C, et al. Contour detection and characterization for asynchronous event sensors[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 486-494.
- [20] Vasco V, Glover A, Mueggler E, et al. Independent motion detection with event-driven cameras[C]//2017 18th International Conference on Advanced Robotics (ICAR), July 10-12, 2017, Hong Kong, China. New York: IEEE Press, 2017: 530-536.
- [21] Han S P, Zhao P, Chen X M, et al. A stable learning-based corner detector for event-based cameras[C]//2023 9th International Conference on Virtual Reality (ICVR), May 12-14, 2023, Xianyang, China. New York: IEEE

- Press, 2023: 315-320.
- [22] Li J J, Su L, Guo C D, et al. *Asynchronous event-based corner detection using adaptive time threshold*[J]. *IEEE Sensors Journal*, 2023, 23(9): 9512-9522.
- [23] Seilozzakerini S, Yau W Y, Zhao B, et al. *Event-based Hough transform in a spiking neural network for multiple line detection and tracking using a dynamic vision sensor*[C]//*Proceedings of the British Machine Vision Conference*, September 19-22, 2016, York, UK. London: British Machine Vision Association, 2016.
- [24] Valeiras D R, Clady X, Ieng S H, et al. *Event-based line fitting and segment detection using a neuromorphic visual sensor*[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 30(4):1218-1230.
- [25] Li Y J, Zhou H, Yang B B, et al. *Graph-based asynchronous event processing for rapid object recognition*[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 914-923.
- [26] Schaefer S, Gehrig D, Scaramuzza D. *AEGNN: asynchronous event-based graph neural networks*[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12361-12371.
- [27] Ryan C, O'Sullivan B, Elrasad A, et al. *Real-time face & eye tracking and blink detection using event cameras*[J]. *Neural Networks*, 2021, 141: 87-97.
- [28] Wang D S, Jia X, Zhang Y, et al. *Dual memory aggregation network for event-based object detection with learnable representation*[EB/OL]. (2023-03-17)[2024-02-05]. <https://arxiv.org/abs/2303.09919>.
- [29] Li J N. *Neuromorphic object detection in asynchronous visual streams*[D]. Beijing: Peking University, 2023: 39-59.
- [30] Nagaraj M, Liyanagedera C M, Roy K. *DOTIE: detecting objects through temporal isolation of events using a spiking architecture*[C]//2023 IEEE International Conference on Robotics and Automation (ICRA), May 29-June 2, 2023, London, United Kingdom. New York: IEEE Press, 2023: 4858-4864.
- [31] Liu H J, Moeys D P, Das G, et al. *Combined frame- and event-based detection and tracking*[C]//2016 IEEE International Symposium on Circuits and Systems (ISCAS), May 22-25, 2016, Montreal, QC, Canada. New York: IEEE Press, 2016: 2511-2514.
- [32] Chen N F Y. *Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion*[C]-2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 757-75709.
- [33] Iacono M, Weber S, Glover A, et al. *Towards event-driven object detection with off-the-shelf deep learning*[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018.
- [34] Chen G, Cao H, Ye C B, et al. *Multi-cue event information fusion for pedestrian detection with neuromorphic vision sensors*[J]. *Frontiers in Neurobotics*, 2019, 13: 10.
- [35] Miao S, Chen G, Ning X Y, et al. *Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection*[J]. *Frontiers in Neurobotics*, 2019, 13: 38.
- [36] Jiang Z Y, Xia P F, Huang K, et al. *Mixed frame-/event-driven fast pedestrian detection*[C]//2019 International Conference on Robotics and Automation (ICRA), May 20-24, 2019, Montreal, QC, Canada. New York: IEEE Press, 2019: 8332-8338.
- [37] Sharghi E, Rodriguez J, Mauger J, et al. *Maritime object detection with event-based cameras*[C]//2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), October 11-13, 2022, DC, USA. New York: IEEE Press, 2022.
- [38] Gehrig M, Scaramuzza D. *Recurrent vision transformers for object detection with event cameras*[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 13884-13893.
- [39] Fang Y H. *Research on object detection technology based on dynamic vision sensor*[D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2023: 35-40.
- [40] Alzugaray I, Chli M. *Asynchronous corner detection and tracking for event cameras in real time*[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4): 3177-3184.
- [41] Alzugaray I, Chli M. *ACE: an efficient asynchronous corner tracker for event cameras*[C]//2018 International Conference on 3D Vision (3DV), September 5-8, 2018, Verona, Italy. New York: IEEE Press, 2018: 653-661.
- [42] Liu Q Q. *Research on object tracking based on event camera*[D]. Hangzhou: Zhejiang University of Technology, 2021: 34-48.
- [43] Piątkowska E, Belbachir A N, Schraml S, et al. *Spatiotemporal multiple persons tracking using Dynamic Vision Sensor*[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 35-40.
- [44] Camunas-Mesa L A, Serrano-Gotarredona T, Ieng S H, et al. *Event-driven stereo visual tracking algorithm to solve object occlusion*[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(9): 4223-4237.
- [45] Barranco F, Fermuller C, Ros E. *Real-time clustering and multi-target tracking using event-based sensors*[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018: 5764-5769.
- [46] Ramesh B, Zhang S H, Yang H, et al. *e-TLD: event-*

- based framework for dynamic object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3996-4006.
- [47] Benosman R, Ieng S H, Clercq C, et al. Asynchronous frameless event-based optical flow[J]. Neural Networks, 2012, 27: 32-37.
- [48] Ye C X, Mitrokhin A, Fermüller C, et al. Unsupervised learning of dense optical flow, depth and egomotion with event-based sensors[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 24-January 24, 2021, Las Vegas, NV, USA. New York: IEEE Press, 2020: 5831-5838.
- [49] Sekikawa Y, Nagata J. Live demonstration: tangentially elongated Gaussian belief propagation for event-based incremental optical flow estimation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 3931-3932.
- [50] Schnider Y, Woźniak S, Gehrig M, et al. Neuromorphic optical flow and real-time implementation with event cameras[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 4129-4138.
- [51] Li H M, Shi L P. Robust event-based object tracking combining correlation filter and CNN representation[J]. Frontiers in Neurobotics, 2019, 13: 82.
- [52] Zhu Z Y, Hou J H, Lü X Q. Learning graph-embedded key-event back-tracing for object tracking in event clouds [J]. Advances in Neural Information Processing Systems, 2022, 35: 7462-7476.
- [53] Kueng B, Mueggler E, Gallego G, et al. Low-latency visual odometry using event-based feature tracks[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 9-14, 2016, Daejeon, Republic of Korea. New York: IEEE Press, 2016: 16-23.
- [54] Zhu A Z, Atanasov N, Daniilidis K. Event-based feature tracking with probabilistic data association[C]//2017 IEEE International Conference on Robotics and Automation (ICRA), May 29-June 3, 2017, Singapore. New York: IEEE Press, 2017: 4465-4470.
- [55] Messikommer N, Fang C, Gehrig M, et al. Data-driven feature tracking for event cameras[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 5642-5651.
- [56] Glover A, Bartolozzi C. Robust visual tracking with a freely-moving event camera[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 24-28, 2017, Vancouver, BC, Canada. New York: IEEE Press, 2017: 3769-3776.
- [57] Wu J J, Zhang K, Zhang Y X, et al. High-speed object tracking with dynamic vision sensor[M]//Wang L H, Wu Y R, Gong J Y. Proceedings of the 5th China high resolution earth observation conference. Lecture notes in electrical engineering. Singapore: Springer, 2019, 552: 164-174.
- [58] Bolten T, Pohle-Fröhlich R, Tönnies K D. Application of hierarchical clustering for object tracking with a dynamic vision sensor[M]//Rodrigues J M F, Cardoso P J S, Monteiro J, et al. Computational science-ICCS 2019. Lecture notes in computer science. Cham: Springer, 2019, 11540: 164-176.
- [59] Shu Y J, Sui Y F, Zhao S X, et al. Small moving object detection and tracking based on event signals[C]//2021 7th International Conference on Computer and Communications (ICCC), December 10-13, 2021, Chengdu, China. New York: IEEE Press, 2021: 792-796.
- [60] Chen H S, Suter D, Wu Q Q, et al. End-to-end learning of object motion estimation from retinal events for event-based object tracking[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10534-10541.
- [61] Zhu Q. Design and implementation of UAV target tracking algorithm based on event camera[D]. Dalian: Dalian University of Technology, 2021: 33-51.
- [62] Zhang J Q, Dong B, Zhang H W, et al. Spiking transformers for event-based single object tracking[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 8791-8800.
- [63] Zhang J Q, Yang X, Fu Y K, et al. Object tracking by jointly exploiting frame and event domain[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 13023-13032.
- [64] Zhang J Q, Wang Y C, Liu W X, et al. Frame-event alignment and fusion network for high frame rate tracking [EB/OL]. (2023-05-25)[2024-05-06]. <https://arxiv.org/abs/2305.15688>.
- [65] Wang X, Wang S A, Tang C M, et al. Event stream-based visual object tracking: a high-resolution benchmark dataset and a novel baseline[EB/OL]. (2023-09-26)[2024-05-06]. <https://arxiv.org/abs/2309.14611>.
- [66] Wang X, Li J N, Zhu L, et al. VisEvent: reliable object tracking via collaboration of frame and event flows[J]. IEEE Transactions on Cybernetics, 2024, 54(3): 1997-2010.
- [67] Zhang S X, Wang W M, Li H L, et al. EVtracker: an event-driven spatiotemporal method for dynamic object tracking[J]. Sensors, 2022, 22(16): 6090-6110.
- [68] Orchard G, Jayawant A, Cohen G K, et al. Converting static image datasets to spiking neuromorphic datasets using saccades[J]. Frontiers in Neuroscience, 2015, 9: 437.
- [69] Serrano-Gotarredona T, Linares-Barranco B. A 128×128 1.5% contrast sensitivity 0.9% FPN 3 μs latency 4 mW asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers[J]. IEEE Journal of Solid-State Circuits, 2013, 48(3): 827-838.
- [70] Li H M, Liu H C, Ji X Y, et al. CIFAR10-DVS: an event-stream dataset for object classification[J]. Frontiers

- in Neuroscience, 2017, 11: 309.
- [71] Sironi A, Brambilla M, Bourdis N, et al. [HATS: histograms of averaged time surfaces for robust event-based object classification](#)[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1731-1740.
- [72] de Tournemire P, Nitti D, Perot E, et al. [A large scale event-based detection dataset for automotive](#)[EB/OL]. (2020-01-23)[2024-05-04]. <http://arxiv.org/abs/2001.08499> v3.
- [73] Perot E, de Tournemire P, Nitti D, et al. [Learning to detect objects with a 1 megapixel event camera](#)[EB/OL]. (2020-09-28)[2024-05-04]. <http://arxiv.org/abs/2009.13436> v2.
- [74] Hu Y H, Liu H J, Pfeiffer M, et al. [DVS benchmark datasets for object tracking, action recognition, and object recognition](#)[J]. *Frontiers in Neuroscience*, 2016, 10: 405.
- [75] Mitrokhin A, Fermüller C, Parameshwara C, et al. [Event-based moving object detection and tracking](#)[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018.