

基于事件相机的图像重构综述

徐 齐^① 邓 洁^① 申江荣^{*②} 唐华锦^② 潘 纲^②

^①(大连理工大学人工智能学院 大连 116024)

^②(浙江大学计算机科学与技术学院 杭州 310027)

摘 要: 事件相机是一种新型仿生视觉传感器, 当像素点的亮度变化超过阈值后, 会输出一系列事件信息。该类视觉传感器异步输出像素的坐标、时间戳以及事件的极性, 因此具有低延迟、低功耗、高时间分辨率和高动态范围等特点。它能够捕捉到高速运动和高动态场景中的信息, 可以用来重构高动态范围和高速运动场景。图像重构后可以应用在物体识别、分割、跟踪以及光流估计等任务中, 是视觉领域重要的研究方向之一。该文从事件相机出发, 首先简要叙述事件相机的现状、发展过程、优势与挑战, 然后介绍了各种类型事件相机的工作原理和一些基于事件相机的图像重构算法, 最后阐述了事件相机面临的挑战和未来趋势, 并对文章进行了总结。

关键词: 事件相机; 动态视觉传感器; 图像重构

中图分类号: TN911.73; TP751

文献标识码: A

文章编号: 1009-5896(2023)08-2699-11

DOI: [10.11999/JEIT221456](https://doi.org/10.11999/JEIT221456)

A Review of Image Reconstruction Based on Event Cameras

XU Qi^① DENG Jie^① SHEN Jiangrong^{*②} TANG Huajin^② PAN Gang^②

^①(School of Artificial Intelligence, Dalian University of Technology, Dalian 116024, China)

^②(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

Abstract: Event cameras are bio-inspired sensors that outputs a stream of events when the brightness change of pixels exceeds the threshold. This type of visual sensor asynchronously outputs events that encode the time, location and sign of the brightness changes. Hence, event cameras offer attractive properties, such as high temporal resolution, very high dynamic range, low latency, low power consumption, and high pixel bandwidth. It can capture information in high-speed motion and high-dynamic scenes, which can be used to reconstruct high-dynamic range and high-speed motion scenes. Brightness images obtained by image reconstruction can be interpreted as a representation, and be used for recognition, segmentation, tracking and optical flow estimation, which is one of the important research directions in the field of vision. This survey first briefly introduces event cameras from their working principle, developmental history, advantages, and challenges of event cameras. Then, the working principles of various types of event cameras and some event camera-based image reconstruction algorithms are introduced. Finally, the challenges and future trends faced by event cameras are described, and the article is concluded.

Key words: Event camera; Dynamic vision sensors; Image reconstruction

收稿日期: 2022-11-21; 改回日期: 2023-06-06; 网络出版: 2023-06-19

*通信作者: 申江荣 jrshen@zju.edu.cn

基金项目: 国家自然科学基金(62206037), 科技创新2030-新一代人工智能重大项目(2021ZD0109803), 人工智能与数字经济广东省实验室(深圳)开放课题(GML-KF-22-11), 大连理工大学中央高校基本科研业务费(DUT21RC(3)091)

Foundation Items: The National Natural Science Foundation of China (62206037), The National Key R&D Program of China (2021ZD0109803), The Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) (GML-KF-22-11), The Fundamental Research Funds for the Central Universities (DUT21RC(3)091)

1 引言

如今传统相机已经广泛应用在人类生活的各个领域,包含军事国防、航天航空、船舶运输、道路交通、体育、自动驾驶等领域。传统的基于帧的视觉传感器可以获取丰富的图像信息,在深度学习的助力下,计算机视觉领域相关算法蓬勃发展。然而,传统相机由于其自身成像原理,在延时、功耗、动态范围等方面有一些缺陷。传统相机由人工创建的定时和控制信号驱动,以固定速率和固定曝光时间输出强度图像,即便像素上的值未发生变化,也会重复曝光,降低了数据传输效率、膨胀了数据体积,导致数据处理延迟和数据冗余等问题。此外,在记录快速运动物体时易产生运动模糊,在高动态范围场景下易出现过曝或者欠曝的情况,这些都会导致细节信息的丢失。

与传统相机相比,事件相机能够发送具有高动态范围、高时间分辨率和低延迟的异步事件流,可以在高速运动和具有挑战性的光照条件下发挥强大优势。事件相机与传统相机有着不同的工作方式,由视野中发生的事件驱动,输出各个像素上的亮度变化。在像素阵列中,当像素级亮度变化超出一定阈值时,事件相机每个像素以微秒级分辨率独立地输出,输出的异步稀疏数据被称为“事件”(event),包括事件的像素位置、时间戳以及极性。图1为理想情况下传统相机和事件相机的输出比较,当圆盘匀速转动时,传统相机输出的帧之间有一定时间间隔,会有一定的延迟,而事件相机的输出是连续的;当圆盘静止时,事件相机没有事件信息输出,而传统相机仍会输出图像,某些情况下会导致数据冗余;当圆盘转速加快时,传统相机出现了运动模糊,而事件相机的响应仍是正常的。

由于事件相机异步输出事件流,这是一种与传

统图像完全不同的表示方式,现有的、成熟的计算机视觉技术不能直接应用于这些数据。因此,需要研究新的算法来处理事件数据。从低级视觉任务,如视觉里程计^[2]、特征跟踪^[3]和光流^[4],到高级任务,如物体分类^[5,6]、姿态识别^[7]以及目标跟踪^[8]等。除了逐事件地处理数据^[1],还可以将一组事件映射为2D的表示形式,如TS(Time Surface)^[6]、事件图像^[2]、3D体素网格^[9]或亮度图像^[10-12]。然而,这些都不是自然图像,现有的大部分计算机视觉方法性能受到限制。更重要的是,用真实图像数据训练的深度网络不能直接应用于此类表示形式。

基于事件相机的图像重构是指通过将事件流集成转换为强度图像。动态视觉传感器逐像素编码场景中的视觉信息,这种事件流信息是一种非冗余的、“压缩的”可视数据,事件的集成可以理解为“解压”事件流中编码的可视数据,在无噪声和传感器最佳响应状态的理想条件下,事件的集成可以产生“绝对”亮度。通过图像重构获得亮度图像可以解决现有算法受限问题,能够将重构图像直接用于现有的计算机视觉任务中,因此重构具有更高动态范围和更少模糊伪影的高分辨率图像具有十分重要的意义。同时,事件表示连续图像帧之间的强度变化,与传统图像提供的信息相关联,能够生成具有超分辨率的高帧率视频,在工业界有非常强的应用前景。

本文将对事件相机及其在图像重构领域的算法作详细的介绍,文章结构如下:第1节、第2节简要介绍事件相机的发展概况、优势以及面临的挑战,第3节简要介绍5类事件相机的工作原理并对几种相机进行比较,第4节介绍一些图像重构算法,第5节、第6节概括算法面临的挑战和发展方向,并进行全文总结。

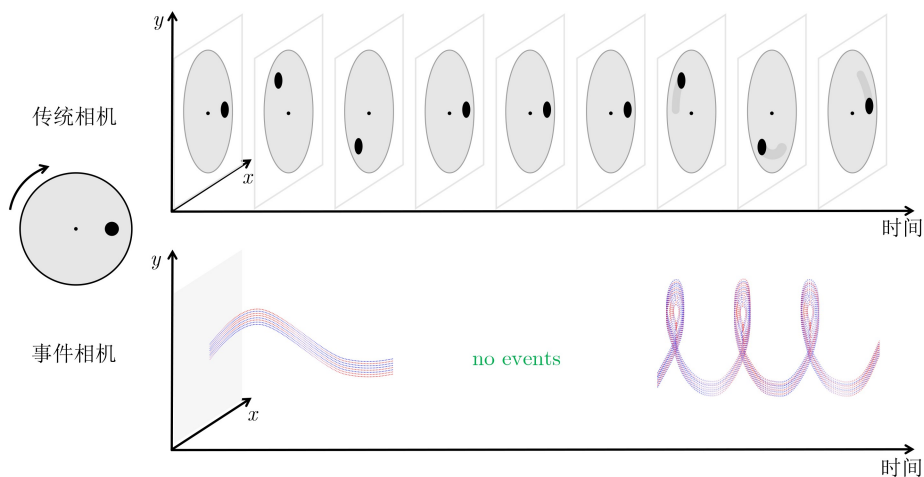


图1 传统相机与事件相机的输出比较^[1]

2 事件相机简介

2.1 事件相机概况

最早的事件相机可以追溯到1992年,被称作硅视网膜,是Mahowald^[13]基于脊椎动物视网膜模型设计的,在此研究中提出了新型集成电路通信协议——地址事件协议,异步输出带地址的数字脉冲来响应光照强度的变化。然而,硅视网膜不同像素之间的响应存在相当大的失配问题,且像素面积太大,因此,只能作为演示设备,无法用于实践。随后十几年,基于神经形态的视觉传感器不断发展,研究者相继研发出了时间对比度传感器、基于梯度的传感器、边缘方向敏感传感器和光流传感器。但是,这些传感器由于噪声、像素面积、对比度灵敏性以及电路复杂性等原因,在实际任务的应用中仍存在很大问题。

直到2006年,Delbruck团队^[14]研发出动态视觉传感器(Dynamic Vision Sensor, DVS),使事件相机的发展进入新阶段。后续,为了满足可视化的要求,Posch等人^[15]在2008年研发出了基于异步时间的图像传感器(Asynchronous Time-based Image Sensor, ATIS),可以输出事件信息和灰度信息。2013年Delbruck团队^[16]又提出了动态主动像素视觉传感器(Dynamic and Active Pixel Vision Sensor, DAVIS),在2017年被扩展为彩色版本DAVIS346^[17],2018年5月在机器人和自动化国际会议上DAVIS346首次展示给顶尖的研究人员,标志着适用于真实环境的高分辨率、高性能传感器的诞生。不同于Delbruck团队的工作,南洋理工大学团队^[18]开发了一种新型动态视觉传感器CeleX,不仅能够显示物体的运动变化,还能够重构纹理。2018年北京大学的黄铁军教授团队^[19]研发了一款具有独特视频捕捉方案的脉冲相机——Vidar,该传感器可以重构任意时刻的图像,从而实现高速运动回放。

2.2 优势与挑战

与传统的基于帧的视觉传感器相比,神经形态视觉传感器异步测量像素的亮度变化,因此,事件相机具有以下优势:

(1)低时延。事件相机的每个像素独立工作,不需要等待全局曝光时间:一旦检测到亮度变化超过阈值,它就会生成事件信息。因此,事件相机具有微妙级别的低时延特性。

(2)低能耗。事件相机仅捕捉亮度的变化,与背景信息无关,能够减少冗余数据的处理能耗。

(3)高时间分辨率。事件相机以微秒级别的时间分辨率检测事件并加盖时间戳。因此,事件相机可以捕获高速运动场景下的目标,可有效避免传统

视觉传感器所产生的运动模糊等情况。

(4)高动态范围。事件相机具有非常高的动态范围(> 120 dB),而传统相机一般只有70 dB。与生物视网膜类似,DVS像素可以适应极暗和极亮的刺激。在高曝光和低曝光的情况下,传统相机均会产生信息丢失,但是对于事件相机来说,若像素亮度变化存在,仍会记录信息。

尽管如此,目前事件相机的应用仍然在硬件和算法处理方面面临很多挑战:

(1)需要研究新的视觉信息处理算法。事件相机与标准相机有着完全不同的输出方式,一个是异步的稀疏的事件流,另一个是同步密集的程度信息。因此,基于帧的视觉算法不能直接用于事件数据。

(2)需要克服运动依赖性引起的场景外观变化。对于同一个静止场景,事件相机相对于场景的移动方向不同,输出的事件信息会有很大差异。每个事件都包含二进制的亮度变化信息,亮度变化不仅取决于场景亮度,还取决于场景和摄像机之间的相对运动方式。在不同时间,建立事件信息之间的数据关联具有挑战性。

(3)需要对噪声进行处理。视觉传感器都会有一定的噪声,可以分为外部噪声和内部噪声,外部噪声是由系统外部的干扰造成的,内部噪声是由光与电的基本性质,电器的机械运动,器材本身以及内部电路等原因引起的。目前事件相机的噪声等级还比较高,如何克服噪声的影响是研究者目前要面对的问题之一。

3 事件相机的类型及工作原理

3.1 DVS

DVS是Lichtsteiner等人^[14]提出的 128 像素 \times 128 像素的CMOS视觉传感器,每个像素独立地量化光强变化,对光强变化的部分生成脉冲事件,而对无变化的部分不做出响应。脉冲事件以异步数字像素地址流的形式被输出,其输出速率与场景的光强变化有关。图2显示了该相机的像素结构,分为3个部分:快速对数光感受器电路(photoreceptor circuit)、差分电路(differencing circuit)和双晶体管比较器(comparators)。光感受器通过对数转换自动控制单个像素的增益,并且快速响应照明的变化。这种光感受器电路的缺点是晶体管阈值变化导致像素之间的大量直流失配,当直接使用此输出时需要进行校准。在光强变化到达阈值输出事件后,通过将差分电路的输出平衡到复位电平来消除直流失配,电容比 $C1/C2$ 决定变化放大的增益,通过差分电路的精确增益,减小了不可避免的比较器失配的影响。比较器将反向放大器的输出与全局阈值进

要额外曝光时间，无需在空间和时间领域内匹配事件和强度信息。脉冲只表示强度变化导致没有“OFF”事件，但是附加的模数转换器值可以弥补“OFF”事件的缺失。

3.5 Spike Camera

2017年北京大学的黄铁军教授团队^[21]研发了一款脉冲相机(Spike Camera)，它有一组传感器阵列，每个像素都能独立地捕获亮度信息，如果累积的强度达到阈值，则将触发脉冲，并将脉冲发送到总线。产生的脉冲流表示亮度的变化，分析脉冲流的模式，可以重构任意时刻，因此可以连续播放高速运动视频。2022年提出了一种新的视觉表示Vidar^[19]，采用积分型视觉采样模型，电路包括3个部分：脉冲触发电路，复位电路和读出电路，如图6所示。像素中的光电二极管连续捕获光子并将入射光照明转换为连续的光电流 I_{ph} ，光电二极管电压 V_{pix} 在光电收集过程中降低，当光电二极管电压达到某个阈值 V_{ref} 时，比较器的输出切换，并产生脉冲信号。锁存器检测到脉冲信号之后， V_{pix} 被重置为预先设定的复位电压。Vidar利用脉冲序列来表示光在时空域的变化，可以准确地保留物理光流的时间，因此可以在任何给定时刻重构场景。

3.6 事件相机的比较

事件相机是一种神经形态视觉传感器，可以分为差分型和积分型两类视觉采样模型。

基于差分型视觉采样模型的传感器种类繁多，如：DVS128, ATIS, DAVIS346, CeleX-IV, CeleX-V等，表1列出了近几年推出的几种视觉传感器的

性能比较。虽然ATIS和DAVIS都是事件流和图像帧的结合，但ATIS与DAVIS相比，能耗较高且像素面积较大。此外，ATIS只提供事件检测像素的强度测量，而DAVIS提供所有像素的强度测量，这些缺点使ATIS难以商业化。CeleX系列传感器不断发展，推出新的像素结构，实现更高分辨率、更低延迟，现在已经推出CeleX-V，具有1280像素×800像素。

基于积分型视觉采样模型的传感器是模拟视网膜中央凹功能，典型的是北京大学的Vidar相机。DVS的光敏单元只在亮度变化超过阈值时产生事件，Vidar的每个光敏单元独立捕获光子，当累积强度超过阈值时产生脉冲。因此，Vidar有效地记录了每个采样位置的场景信息。Vidar相机输出的是脉冲流，研究者将其与脉冲神经网络(Spiking Neural Network, SNN)相结合，可以进行实时处理，实现了比人类视觉快1000倍的高速移动目标检测和跟踪。该相机的另一个优势是提供了一种更加灵活的图像采集方式，可以重构任意时刻的图像。但是，Vidar在静态和动态区域都会产生脉冲，与DVS系列传感器相比，存在数据冗余问题。

4 图像重构算法与应用

事件相机和传统相机是两种不同的范式，它们之间存在固有的差异，现有的针对传统相机的计算机视觉算法不能直接应用于事件相机。虽然ATIS与DAVIS能输出强度图像，但是仍然存在限制：一是低帧率强度图像，相对于事件数据的高时间分辨率，目前的事件相机只能输出低帧率的强度图像。二是运动模糊，当记录高速运动场景时，由于相机和场景之间的相对运动，输出的强度图像往往是模糊的。

为了应对这些挑战，研究者已经提出了重构高质量图像和高帧率视频的各种方法。本节根据论文中使用的数据类型作为分类依据，分为纯事件流以及事件流与图像结合两类，对基于事件相机的图像

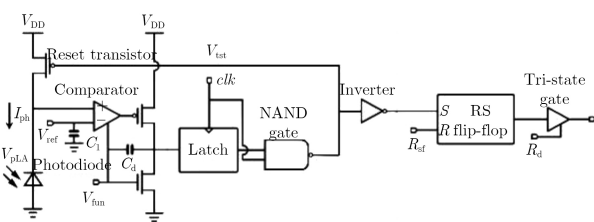


图6 Vidar电路^[19]

表1 几种事件相机的性能比较

类型	分辨率	时间分辨率(μ s)	最小延时(μ s)	宽度(MEPS)	功耗(mA)	重量(无镜头)(g)	动态范围(dB)	最小对比敏感度(%)	是否能输出强度图
DVXplorer Lite	320×240	65-200	<200	100	<140	75	Events: > 90	0.13	否
DAVIS346	346×260	1	<100	12	<180	100 g	DVS: 120 APS: 56.7	14.3 (on) 22.5 (off)	是
DVXplorer	640×480	65-200	<200	165	<140	100 g	Events: > 90	0.13	否
Gen 4 CD	1280×720	—	20-150	1066	32-84	—	>124	11	否
CeleX-IV	768×640	—	10	200	—	—	90	30	是
CeleX-V	1280×800	—	8	140	400	—	120	10	是

重构算法做详细介绍,在表2和表3中进行了不同重构算法的比较。

4.1 基于纯事件流

在基于事件流的图像重构算法研究初期,多数研究通过在静态场景中移动事件相机实现图像重构,此类研究受限于静态场景。最早的从事件中重构图像的研究是开始于Cook等人^[22]的交互式地图研究,受生物系统分布式处理事件的启发,设计了相互连接的网络,从旋转的事件相机中恢复强度图像,光流和角速度。Kim等人^[23]开发了一种扩展卡尔曼滤波器,用于从旋转事件相机重构2D全景梯度图像,并通过泊松积分升级为全强度帧,文献^[24]

利用相机绕单轴旋转的约束运动来重构图像,然后用于立体深度估计,在自然光照下以5000像素×1024像素的图像分辨率和较高的动态范围实时生成3D全景视图。但这些研究都是在静态场景中实现重构,并且有严格的运动限制。

随后,为了解决运动限制问题,在更多运动场景中实现图像重构,许多研究工作者引入了正则化。2016年Bardow等人^[4]提出变分能量最小化框架,从事件滑动窗口同时恢复强度图像和密集光流,不需要估计相机运动和场景特征来重构强度图像。论文表明,光流对于恢复图像边缘是必要的,特别是在快速运动场景,这是第1个适用于动态场景的视频

表 2 基于纯事件流的不同重构的比较

文献	应用场景	适用范围	适配相机	特点
[23]	精确跟踪的相机旋转	静态场景 相机轻微旋转运动	DVS128	能够进行高时间分辨率和高动态范围场景;对环境 and 相机的运动有严格的限制。
[24]	360°深度全景成像	静态场景, 高速旋转	360°HDR深度相机 TUCO-3D	实现高分辨率和高动态范围的自然场景 实时3维360°HDR全景成像;
[5]	极端、快速运动和高动态范围场景的跟踪和重建	动态场景, 普通运动	DVS128	对环境或相机的运动有严格的限制。 不会对相机运动和场景内容施加任何限制;
[11]	实时重建高质量的图像	动态场景, 普通运动	DVS128	重构中具有严重的细节损失。
[25]	基于事件的视频重建、 基于事件的直接人脸检测	-	DVS128	重构的图像出现了一些噪声像素。 从事件流中以超过2000 Frames/s的速度重构图像和视频, 第1个从事件流中直接检测人脸的方法。
[26]	基于事件的物体分类 和视觉惯性测程	任意运动 (包括高速运动)	DAVIS240C	在模拟事件数据上训练的循环卷积网络的事件到视频重构, 优于以往最先进的方法。 重构具有更精细的细节和更少的噪声。
[27]	基于事件的物体分类 和视觉惯性测程	高动态范围场景 和弱光场景	DAVIS240C, Samsung DVS Gen3, Color-DAVIS346	合成快速物理现象的高帧率视频、 高动态范围视频和彩色视频。
[28]	基于事件的物体分类 和视觉惯性测程	高动态范围场景	DAVIS240, DAVIS346, Prophesee, CeleX sensors	性能几乎与最先进的(E2VID ^[29,30])一样好, 但是计算成本只有一小部分, 速度提高了3倍, FLOPs降低了10倍, 参数减少了280倍。
[30]	生成极高帧率视频	快速运动, 高动态范围, 极端照明条件	DAVIS	构建了具有更多细节的HDR图像和高帧率视频, 在极端照明条件和快速运动时也不模糊。
[31]	重构超分辨率图像	快速运动 高动态范围	DAVIS	从模拟和真实数据的事件中重构高质量的超分辨率图像, 可以恢复非常复杂的物体, 如人脸。
[32]	语义分割、对象识别和检测	高动态范围	Samsung DVS Gen3, DAVIS240C	进一步扩展到锐化图像重建 和彩色事件图像重建。
[33]	高速视觉任务	常速和高速场景	spike camera	提出一种脉冲神经模型, 在高速运动和静态场景下都能重建出高质量的视觉图像。

表 3 基于事件流和传统图像的不同重构的比较

文献	应用场景	适用范围	适配相机	特点
[37]	生成高速视频	前景是高速运动 背景是静态的	DVS(事件流)	只能为形状简单的物体恢复大而平滑的运动, 不能处理有变形或经历3维刚性运动的物体。
[12]	重构高动态范围, 高时间分辨率图像和视频	高速、高动态范围	DAVIS240C	在夜间行驶的高速、弱光条件下, 能够恢复运动模糊的物体; 相机直接对准太阳过度曝光的场景, 能恢复树叶和树枝等特征。
[38]	从单个模糊帧及其事件数据 重构高帧率、清晰的视频	低光照和复杂 动态场景	DAVIS	在低光照和复杂动态场景等不同条件下 高效地生成高质量的高帧率视频。
[39]	消除模糊重构高时间 分辨率视频	低光照和复杂的 动态场景	DAVIS346	极端的灯光变化会降低该方法在高动态场景下的性能, 事件误差累积的噪声会降低重构图像的质量。

重构框架。然而, 此框架需要多个手工制作的正则化器, 这会导致严重的细节损失。Munda等人^[11]的工作与文献^[4]类似, 在不明确估计光流的情况下将重构问题转换为能量最小化问题, 使用变分方法优化这种能量, 在一定程度上降低了复杂度。2016年, Barua等人^[25]提出了一种基于学习的方法, 设计了基于补丁的模型从事件中重构图像, 在模拟数据上使用K-SVD来学习一个字典, 该字典将集成事件的小块映射到图像梯度, 并使用泊松积分来恢复强度图像。

早期的研究使用了各种方法, 如流形正则化、光流和滤波器等, 随着深度学习的发展, 基于卷积神经网络的工作取得了非常优秀的重构效果^[26-28]。2019年, Rebecq等人^[26]提出了一种新颖的重构模型E2VID, 采用循环网络架构UNet。该模型的重构函数由递归卷积神经网络实现, 训练网络的事件数据是由事件模拟器合成的。这种带循环连接的网络不需要在每个时间步从头开始重构新图像, 而只需要使用新的事件序列增量更新以前的重构。之后Rebecq对E2VID进行了改进, 文献^[27]提出的新网络具有记忆功能, 会从任意长的过去事件序列中更新, 而不是只使用最后几张图像。其次, 新网络使用了堆叠的ConvLSTM, 这可以防止长序列在时间反向传播期间梯度消失, 增加了网络处理可变数量事件的能力, 也适用于更长事件序列的训练。但上述工作在实现视频重构时使用的神经网络模型结构较为复杂且计算成本高昂。

2020年, Scheerlinck等人^[28]提出了一种新的神经网络架构FireNet, 从事件中执行快速图像重构。FireNet比E2VID^[27]小得多, 需要更少的参数(38 k vs 10 M), 更少的内存(0.16 MB vs 43 MB)和更少的浮点运算数(12.6 G vs 147.2 G), 运行速度比E2VID快3倍。FireNet是一个完全卷积的网络, 它依赖于循环连接来随着时间的推移建立一个状态, 在较小的网络上重用以前的结果, 能在非常小的循环网络上快速运行。

以上基于卷积神经网络的研究证明了用合成事件数据训练神经网络是可实现的。ESIM^[29]是一个事件模拟器, 可以将图像的强度信息转换为事件信息。大多数商业化的事件相机虽然能够生成全帧图像(类似传统相机的图像帧), 但是这些图像存在过曝光和模糊的问题, 不适合训练, 事件模拟器可以提供事件信息对应的真实的强度图像, 用合成数据训练的模型在真实事件相机数据上具有较好的泛化性。

2019年, Wang等人^[30]提出了基于事件相机的

条件生成对抗网络, 学习观测图像与随机噪声到输出图像之间的映射关系, 将纯事件信息转换为高动态范围图像和高帧率视频。2020年Wang等人^[31]又提出了一种端到端的模型EventSR, 从低质量事件流中重构低质量强度图像, 然后再从低质量强度图像中重构超分辨强度图像。整个模型分3个阶段进行训练, 首先从事件流中重构低质量图像, 然后增强图像质量, 最后对增强的图像进行上采样。该方法在没有真实高分辨图像的情况下进行无监督训练, 并且部署对抗学习。此外, 为了训练EventSR, 论文提出了包含真实场景和模拟场景的开放数据集。2021年作者又将论文中的方法扩展到了高动态范围和彩色的图像重构^[32], 并且证明了该论文的成果可以作为高级任务事件的中间表示, 如语义分割、对象识别和检测。

基于传统神经网络的事件图像重构模型的一个潜在缺点是计算成本, 这与事件相机低功耗和低延迟的特性不符, 无法充分发挥事件相机的优势。而SNN作为仿生神经网络, 异步输出二进制脉冲, 相较于传统神经网络在功耗和计算成本上有很大优势, 因此, 一些学者将其应用在事件相机的图像重构领域。

2020年Zhu等人^[33]提出了一种新的3层脉冲神经模型对脉冲相机的输出重构。该模型由局部运动激发层、脉冲细化层和视觉重构层组成, 这些层由Leaky integrate-and-fire神经元组成, 神经元之间的突触连接采用的是脉冲时序依赖可塑性规则。局部运动激发层接收脉冲并标记神经元的运动状态: 静止或运动, 然后脉冲细化层根据每个神经元的运动状态调整不应期, 最后的视觉重构层采用自适应机制, 将每个神经元的动态阈值映射到灰度图像中作为输出。2021年Zhu等人^[34]提出了一种基于深度SNN的视频重构框架EVSNN。为了更好地利用时间信息, 作者提出了自适应膜电位神经元PA-EVSNN, 可以根据输入的脉冲自适应地更新膜电位。作者提出的基于SNN的视频重构方法达到了与基于ANN的模型相当的性能, 并且在能耗和计算效率上有着更好的表现。

2021年Duwek等人^[35]提出了一个CNN与SNN相结合的神经网络。第1阶段的CNN被训练进行拉普拉斯预测, 第2阶段将经过训练的CNN转换为SNN进行推理, 在推理过程中, 第1阶段的拉普拉斯被驱动到SNN中进行泊松积分。每个事件数据都转换为事件帧张量, 每个事件帧都应用空间中值滤波器进行预处理, 使用卷积核降低噪声。该方法从CNN开始, CNN预测给定帧张量的拉普拉斯, 然

后利用泊松积分优化的SNN来处理,将事件帧划分为多个块,分别预测和重构每个块,最后将它们拼接在一起合成完整的图像。

早期的研究使用了各种方法,如流形正则化、光流和异步滤波器等,近几年卷积神经网络、循环神经网络与生成对抗网络等深度学习模型不断发展,此类研究使用真实与合成事件数据训练神经网络,基于传统神经网络图像重构方法取得了更加优秀的重构效果。但是,与传统神经网络相比较,SNN更具有生物可解释性,神经元之间通过离散的脉冲传输信号。传统神经网络方法不能充分利用时间信息,而时间信息是事件流的重要信息,SNN可以将时间信息进行编码,将稀疏的事件数据与SNN结合,可以有效地解决分类与识别任务并且具有较高的能源效率。

4.2 基于事件流和传统图像

上述方法仅使用事件数据重构图像,场景中的静态和无纹理区域不会触发较多事件,缺乏强度图像中包含的补充信息,这导致无法利用场景中静态区域和无纹理区域的信息,使得模型重构能力有限,将事件流与传统图像相结合进行图像重构可以克服这种局限性。

2014年,Brandli等人^[36]将DAVIS相机输出的图像帧与事件流结合,提出了第1个针对DAVIS的实时事件数据解压缩算法,动态估计每个事件的时间对比度阈值来计算两帧之间的强度信息。所提出的算法任务分为两部分:基于事件的图像解压缩和强度估计。每一帧的新图像都会重置强度信息,防止整合误差的过度增长,但也丢弃了重要的累积信息。2017年,Liu等人^[37]使用仿射运动模型重构高速静态背景视频,能够恢复快速运动的场景,例如下落或旋转的物体以及弦的振动。

2018年Scheerlinck等人^[12]提出了基于事件强度估计的连续时间公式,使用互补滤波将图像帧与事件结合,滤波器采取异步的、事件驱动的更新方案,以合并最新的事件信息,消除窗口延迟,不依赖于运动模型,可以在高动态、复杂的环境中运行。连续时间公式不是随着新帧的到来重置强度估计值,而是保留来自事件的高动态范围信息,保持比图像帧具有更大时间分辨率和动态范围的图像状态。

2019年Pan等人^[38]提出了一种基于事件的双积分(Event-based Double Integral, EDI)模型,从单个模糊图像及其事件数据中重构高帧率的、清晰的视频。为了减少事件累积的噪声,作者将积分限制在曝光时间中心周围的时间间隔内,能够重构较短

的视频片段,将片段拼接获得最终视频。但是,当相机和物体具有较大的相对运动时,会导致闪烁。因此,2020年作者将EDI模型扩展到多重双积分(mEDI)^[39],以处理视频片段边界处的不连续性,作者进行了大量实验,此方法可以在不同条件下(例如低光照和复杂的动态场景)生成高质量、高帧率的视频,但极端的照明变化会降低该方法的性能,此外,事件误差积累也会降低重构图像的质量。

使用事件流和强度图像,二者之间相互作用,在重构图像上可以获得较好的效果,但无法解决图像帧模糊的问题,重构高帧率视频时,会有一定的模糊。文献^[40]对相机和观察到的场景之间的相对运动进行编码,在一定程度上改善了模糊帧导致的图像退化问题,充分利用了编码的运动信息进行高帧率视频的重构。

目前,图像帧和事件流相结合的图像重构算法多数是基于DAVIS相机,目标是在高速、弱光、高动态范围条件下重构高帧率、清晰的视频,后续可以应用在车辆夜间行驶等场景,但是还没有适配与所有相机的统一的重构框架。

5 挑战与展望

事件相机是一种神经形态视觉传感器,随着事件相机的不断发展,在硬件和算法上的不断进步,有关事件相机的研究已经取得阶段性进展,但想要比肩人类视觉系统在复杂交互环境中的感知能力,还面临着巨大的挑战,本节讨论事件相机所面临的挑战以及未来发展方向。

在硬件方面,事件相机是一类受人脑启发创建的神经形态视觉传感器。人脑具有速度快、能耗低、可扩展等优点,但是,让硅电路模仿细胞复杂且丰富的机制,使其能像大脑一样不断调整以适应复杂多变的环境是十分困难的。受外界环境因素的影响,晶体管电流的可变性会限制相机的精度,同时电路内部的元器件之间也会产生不必要地响应。神经形态芯片研发人员需要进行巧妙的电路和系统设计来降低这些干扰,以提高事件相机的输出精度,减少输出的噪声。

数据集对模型的训练至关重要,而且在大规模数据集上实验得到的结果更具有说服力。当前,基于事件相机的数据集主要是针对分类和识别任务,这些数据集场景简单且数据规模较小。尽管各个研究机构正在尝试收集更多的数据,但是由于事件流是时间和空间上的异步点阵,无法直接手工标注,只能借助其他传感器进行辅助标注,这导致上述数据集规模不够,无法直接利用标签学习方法实现大模型的训练。因此,亟需开发一个更大规模的事件

相机数据集, 以提升模型在实现相关任务时的学习能力。

在事件流的图像重构方面, 对事件数据的处理方式主要有两类, 一类是逐事件处理, 另一类是将事件聚合成帧的形式。第1种逐事件的数据处理方式虽然可以实现较小的延迟, 但是每个事件携带的信息量少, 而且受噪声影响, 计算量较大, 算法的实用性低。第2种将事件流累积成帧的数据处理方式, 牺牲了事件相机低延时的特性。因此, 现阶段迫切需要研究能有效提高运算效率且保留事件相机低延迟低功耗特性的事件流处理算法。

6 总结

本文介绍了事件相机的几种类型和工作原理, 然后介绍了一些基于事件相机的图像重构算法。事件相机是一种神经形态视觉传感器, 具有延迟低、功耗低、动态范围高、时间分辨率高等优点, 按照其信息采集方式, 可以分为差分型视觉采样和积分型视觉采样两类。事件相机在高动态范围和捕获快速移动物体等场景中具有巨大的潜力, 为了从事件流中提取有用的信息, 事件流通常被转换为中间表示, 图像和视频是一种非常有用的表达形式。以重构算法中使用的数据类型作为分类依据, 介绍了一些图像重构算法, 当前主流的重构算法是对纯事件流处理, 图像重构有助于建立事件的对应关系, 对事件进行可视化和解释, 让研究者更直观地了解事件编码信息。此外, 图像重构算法可以作为目标检测识别和跟踪等计算机视觉任务的中间过程, 因此, 是类脑视觉领域重要的研究方向之一。

参 考 文 献

- [1] KIM H, LEUTENEGGER S, and DAVISON A J. Real-time 3D reconstruction and 6-DoF tracking with an event camera[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 349–364. doi: [10.1007/978-3-319-46466-4_21](https://doi.org/10.1007/978-3-319-46466-4_21).
- [2] VIDAL A R, REBECQ H, HORSTSCHAEFER T, *et al*. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios[J]. *IEEE Robotics and Automation Letters*, 2018, 3(2): 994–1001. doi: [10.1109/LRA.2018.2793357](https://doi.org/10.1109/LRA.2018.2793357).
- [3] LAGORCE X, MEYER C, IENG S H, *et al*. Asynchronous event-based multikernel algorithm for high-speed visual features tracking[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(8): 1710–1720. doi: [10.1109/TNNLS.2014.2352401](https://doi.org/10.1109/TNNLS.2014.2352401).
- [4] BARDOW P, DAVISON A J, and LEUTENEGGER S. Simultaneous optical flow and intensity estimation from an event camera[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 884–892. doi: [10.1109/CVPR.2016.102](https://doi.org/10.1109/CVPR.2016.102).
- [5] LAGORCE X, ORCHARD G, GALLUPPI F, *et al*. HOTS: A hierarchy of event-based time-surfaces for pattern recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(7): 1346–1359. doi: [10.1109/TPAMI.2016.2574707](https://doi.org/10.1109/TPAMI.2016.2574707).
- [6] 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, Salt Lake City, USA, 2018: 1731–1740. doi: [10.1109/CVPR.2018.00186](https://doi.org/10.1109/CVPR.2018.00186).
- [7] 张远辉, 许璐钧, 徐柏锐, 等. 基于事件相机的人体动作识别方法[J]. *计量学报*, 2022, 43(5): 583–589. doi: [10.3969/j.issn.1000-1158.2022.05.04](https://doi.org/10.3969/j.issn.1000-1158.2022.05.04).
ZHANG Yuanhui, XU Lujun, XU Borui, *et al*. Human action recognition method based on event camera[J]. *Acta Metrologica Sinica*, 2022, 43(5): 583–589. doi: [10.3969/j.issn.1000-1158.2022.05.04](https://doi.org/10.3969/j.issn.1000-1158.2022.05.04).
- [8] GALLEGO G, LUND J E A, MUEGGLER E, *et al*. Event-based, 6-DOF camera tracking from photometric depth maps[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(10): 2402–2412. doi: [10.1109/TPAMI.2017.2769655](https://doi.org/10.1109/TPAMI.2017.2769655).
- [9] GEHRIG D, LOQUERCIO A, DERPANIS K, *et al*. End-to-end learning of representations for asynchronous event-based data[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 5633–5643. doi: [10.1109/ICCV.2019.00573](https://doi.org/10.1109/ICCV.2019.00573).
- [10] CORDONE L, MIRAMOND B, and THIERION P. Object detection with spiking neural networks on automotive event data[C]. International Joint Conference on Neural Networks, Padua, Italy, 2022. doi: [10.1109/IJCNN55064.2022.9892618](https://doi.org/10.1109/IJCNN55064.2022.9892618).
- [11] MUNDA G, REINBACHER C, and POCK T. Real-time intensity-image reconstruction for event cameras using manifold regularisation[J]. *International Journal of Computer Vision*, 2018, 126(12): 1381–1393. doi: [10.1007/s11263-018-1106-2](https://doi.org/10.1007/s11263-018-1106-2).
- [12] SCHEERLINCK C, BARNES N, and MAHONY R. Continuous-time intensity estimation using event cameras[C]. The 14th Asian Conference on Computer Vision, Perth, Australia, 2018: 308–324. doi: [10.1007/978-3-030-20873-8_20](https://doi.org/10.1007/978-3-030-20873-8_20).
- [13] MAHOWALD M. VLSI analogs of neuronal visual processing: A synthesis of form and function[D]. [Ph. D. dissertation], California Institute of Technology, 1992.
- [14] LICHTSTEINER P, POSCH C, and DELBRUCK T. A 128×128 120 dB 15 μ s latency asynchronous temporal

- contrast vision sensor[J]. *IEEE Journal of Solid-State Circuits*, 2008, 43(2): 566–576. doi: [10.1109/JSSC.2007.914337](https://doi.org/10.1109/JSSC.2007.914337).
- [15] POSCH C, MATOLIN D, and WOHLGENANT 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. doi: [10.1109/JSSC.2010.2085952](https://doi.org/10.1109/JSSC.2010.2085952).
- [16] BRANDLI C, BERNER R, YANG Minhao, *et al.* A 240×180 130 db 3 μ s latency global shutter spatiotemporal vision sensor[J]. *IEEE Journal of Solid-State Circuits*, 2014, 49(10): 2333–2341. doi: [10.1109/JSSC.2014.2342715](https://doi.org/10.1109/JSSC.2014.2342715).
- [17] MOEYS D P, CORRADI F, LI Chenghan, *et al.* A sensitive dynamic and active pixel vision sensor for color or neural imaging applications[J]. *IEEE Transactions on Biomedical Circuits and Systems*, 2018, 12(1): 123–136. doi: [10.1109/TBCAS.2017.2759783](https://doi.org/10.1109/TBCAS.2017.2759783).
- [18] HUANG Jing, GUO Menghan, and CHEN Shoushun. A dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand[C]. 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, USA, 2017: 1–4. doi: [10.1109/ISCAS.2017.8050546](https://doi.org/10.1109/ISCAS.2017.8050546).
- [19] HUANG Tiejun, ZHENG Yajing, YU Zhaoifei, *et al.* 1000×faster Camera and machine vision with ordinary devices[EB/OL]. <https://www.sciencedirect.com/science/article/pii/S2095809922002077>. doi: [10.1016/j.eng.2022.01.012](https://doi.org/10.1016/j.eng.2022.01.012).
- [20] FOSSUM E R. CMOS image sensors: Electronic camera-on-a-chip[J]. *IEEE Transactions on Electron Devices*, 1997, 44(10): 1689–1698. doi: [10.1109/16.628824](https://doi.org/10.1109/16.628824).
- [21] DONG Siwei, HUANG Tiejun, and TIAN Yonghong. Spike camera and its coding methods[C]. 2017 Data Compression Conference, Snowbird, USA, 2021: 437. doi: [10.1109/DCC.2017.69](https://doi.org/10.1109/DCC.2017.69).
- [22] COOK M, GUGELMANN L, JUG F, *et al.* Interacting maps for fast visual interpretation[C]. 2011 International Joint Conference on Neural Networks, San Jose, USA, 2011: 770–776. doi: [10.1109/IJCNN.2011.6033299](https://doi.org/10.1109/IJCNN.2011.6033299).
- [23] KIM H, HANDA A, BENOSMAN R, *et al.* Simultaneous mosaicing and tracking with an event camera[J]. *J Solid State Circ*, 2008, 43: 566–576.
- [24] BELBACHIR A N, SCHRAML S, MAYERHOFER M, *et al.* A novel HDR depth camera for real-time 3D 360° panoramic vision[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, USA, 2014: 425–432. doi: [10.1109/CVPRW.2014.69](https://doi.org/10.1109/CVPRW.2014.69).
- [25] BARUA S, MIYATANI Y, and VEERARAGHAVAN A. Direct face detection and video reconstruction from event cameras[C]. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, USA, 2016: 1–9. doi: [10.1109/WACV.2016.7477561](https://doi.org/10.1109/WACV.2016.7477561).
- [26] REBECQ H, RANFTL R, KOLTUN V, *et al.* Events-to-video: Bringing modern computer vision to event cameras[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 3852–3861. doi: [10.1109/CVPR.2019.00398](https://doi.org/10.1109/CVPR.2019.00398).
- [27] REBECQ H, RANFTL R, KOLTUN V, *et al.* High speed and high dynamic range video with an event camera[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(6): 1964–1980. doi: [10.1109/TPAMI.2019.2963386](https://doi.org/10.1109/TPAMI.2019.2963386).
- [28] SCHEERLINCK C, REBECQ H, GEHRIG D, *et al.* Fast image reconstruction with an event camera[C]. 2020 IEEE Winter Conference on Applications of Computer Vision, Snowmass, USA, 2020: 156–163. doi: [10.1109/WACV45572.2020.9093366](https://doi.org/10.1109/WACV45572.2020.9093366).
- [29] REBECQ H, GEHRIG D, and SCARAMUZZA D. ESIM: An open event camera simulator[C]. The 2nd Conference on Robot Learning, Zürich, Switzerland, 2018: 969–982.
- [30] WANG Lin, MOSTAFAVI I S M, HO Y S, *et al.* Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 10073–10082. doi: [10.1109/CVPR.2019.01032](https://doi.org/10.1109/CVPR.2019.01032).
- [31] WANG Lin, KIM T K, and YOON K J. EventSR: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 8312–8322. doi: [10.1109/CVPR42600.2020.00834](https://doi.org/10.1109/CVPR42600.2020.00834).
- [32] WANG Lin, KIM T K, and YOON K J. Joint framework for single image reconstruction and super-resolution with an event camera[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 7657–7673. doi: [10.1109/TPAMI.2021.3113352](https://doi.org/10.1109/TPAMI.2021.3113352).
- [33] ZHU Lin, DONG Siwei, LI Jianing, *et al.* Retina-like visual image reconstruction via spiking neural model[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 1435–1443. doi: [10.1109/CVPR42600.2020.00151](https://doi.org/10.1109/CVPR42600.2020.00151).
- [34] ZHU Lin, WANG Xiao, CHANG Yi, *et al.* Event-based video reconstruction via potential-assisted spiking neural network[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 3584–3594. doi: [10.1109/CVPR52688.2022.00358](https://doi.org/10.1109/CVPR52688.2022.00358).
- [35] DUWEK H C, SHALUMOV A, and TSUR E E. Image reconstruction from neuromorphic event cameras using Laplacian-prediction and Poisson integration with spiking

- and artificial neural networks[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Nashville, USA, 2021: 1333–1341. doi: [10.1109/CVPRW53098.2021.00147](https://doi.org/10.1109/CVPRW53098.2021.00147).
- [36] BRANDLI C, MULLER L, and 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), Melbourne, Australia, 2014: 686–689. doi: [10.1109/ISCAS.2014.6865228](https://doi.org/10.1109/ISCAS.2014.6865228).
- [37] LIU Hanchao, ZHANG Fanglue, MARSHALL D, *et al.* High-speed video generation with an event camera[J]. *The Visual Computer*, 2017, 33(6): 749–759.
- [38] PAN Liyuan, SCHEERLINCK C, YU Xin, *et al.* Bringing a blurry frame alive at high frame-rate with an event camera[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 6813–6822. doi: [10.1109/CVPR.2019.00698](https://doi.org/10.1109/CVPR.2019.00698).
- [39] PAN Liyuan, HARTLEY R, SCHEERLINCK C, *et al.* High frame rate video reconstruction based on an event camera[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(5): 2519–2533. doi: [10.1109/TPAMI.2020.3036667](https://doi.org/10.1109/TPAMI.2020.3036667).
- [40] SHEDLIGERI P and MITRA K. Photorealistic image reconstruction from hybrid intensity and event-based sensor[J]. *Journal of Electronic Imaging*, 2019, 28(6): 063012. doi: [10.1117/1.JEI.28.6.063012](https://doi.org/10.1117/1.JEI.28.6.063012).
- 徐 齐：男，副教授，研究方向为类脑计算、神经形态计算、神经信号编解码。
- 邓 洁：女，硕士生，研究方向为类脑视觉。
- 申江荣：女，助理研究员，研究方向为类脑计算、神经形态计算、神经信号编解码。
- 唐华锦：男，教授，研究方向为类脑计算、神经形态芯片与传感器、智能感知。
- 潘 纲：男，教授，研究方向为脑机接口、类脑计算、普适计算。
- 责任编辑：马秀强