

神经形态视觉传感器的研究进展及应用综述

李家宁¹⁾ 田永鸿^{1),2)}

¹⁾(北京大学计算机科学技术系 北京 100871)

²⁾(鹏城实验室 广东 深圳 518055)

摘要 神经形态视觉传感器具有高时域分辨率、高动态范围、低数据冗余和低功耗等优势,近年来在自动驾驶、无人机视觉导航、工业检测及视觉监控等领域(尤其在涉及高速运动和极端光照等条件下)产生了巨大的应用前景.神经形态视觉是一个包含硬件、软件、生物神经模型等的视觉感知系统,终极目标是模拟生物视觉感知结构与机理,以硅视网膜达到、扩展或超越人类智能.作为神经形态工程的重要分支,神经形态视觉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.本文从生物视觉采样模型、神经形态视觉传感器的采样模型及类型、视觉信号处理与特征表达、视觉任务应用等视角进行了系统性地回顾与综述,展望了该领域未来研究的技术挑战与可能发展方向,同时探讨了其对未来机器视觉和人工智能领域的潜在影响.

关键词 类脑智能;人工智能;硅视网膜;仿生视觉;神经形态工程

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2021.01258

Recent Advances in Neuromorphic Vision Sensors: A Survey

LI Jia-Ning¹⁾ TIAN Yong-Hong^{1),2)}

¹⁾(Department of Computer Science and Technology, Peking University, Beijing 100871)

²⁾(Peng Cheng Laboratory, Shenzhen, Guangdong 518055)

Abstract Recently, neuromorphic vision sensors are gaining more and more attention in computer vision owing to the advantages over conventional cameras: high temporal resolution, high dynamic range (HDR), low redundancy and low power consumption. The applications of which include autonomous driving, drone visual navigation, industrial inspection, video surveillance, and the internet of things (IoT), especially in fast motion and challenging illumination conditions. Neuromorphic vision is an important branch of neuromorphic engineering, which uses the silicon retina to mimic the structure and mechanism of the biological visual sensing system and aims to address the shortages of traditional frame-based cameras, and it is also a research hotspot in the field of neuroscience and computer vision. Indeed, neuromorphic vision sensors work in a completely different way and use what it called “spikes” to compute instead of frame-by-frame basis, in which asynchronous spatial-temporal spikes are continuous-time and sparse point-sets in three-dimensional space. This novel technology fundamentally changes the sensing and computing ways in vision information, so that existing computer vision techniques cannot be directly applied to asynchronous spikes from neuromorphic cameras. Nevertheless, neuromorphic vision remains an active area of research in academia and efforts are ongoing to overcome some significant challenges. In this paper, the first part presents a comprehensive overview of the emerging field of neuromorphic vision consisting of the millstones, two bio-inspired vision sampling models (i. e., dynamic vision sampling model and integrating vision sampling model), the types of representative neuro-

morphic sensors (e. g. , DVS, ATIS, DAVIS, CeleX, and Vidar), asynchronous event-based signal processing, event-based feature representation and the typical vision applications. Specially, we introduce the fundamentals of neuromorphic signal processing methods, such as asynchronous spatial-temporal filtering, spike metric, and spike coding. According to the literature, we summarize and report experimental results on the key issue, namely event-based feature representation, from four perspectives consisting of rate-based images, hand-crafted features, spiking neural networks(SNNs), and deep neural networks(DNNs). The second part next highlights the following challenges: (1) Building some large-scale neuromorphic vision datasets for supervised learning approaches; (2) Defining an effective spike metric for neuromorphic signal processing; (3) Learning a better spatial-temporal representation from asynchronous spikes; (4) Exploring high-speed computing paradigm via neuromorphic chips (e. g. , TrueNorth, Loihi, and SpiNNaker); (5) Developing the open-source framework applied to asynchronous spatial-temporal spikes. The last part further discusses the meaningful future research directions on neuromorphic vision sensors. More precisely, we look forward to new materials to enhance the sensing abilities in computational speed, memory, communication overhead, and power consumption. Then, we analyze the potential profound influence that can be extended to multi-band sampling rather than only visible spectrum, multi-task learning just like human vision systems and multi-sensor fusion. Finally, we anticipate a neuromorphic community combing neuroscience and computer vision, which benefits from each other. We believe that the most alluring characteristic of the comprehensive survey paper is that it acts as a bridge between neuromorphic cameras and many computer vision tasks, thus bringing the mainstream of computer vision research to leverage the benefits of these novel sensors and enlarging the footprint of neuromorphic vision towards a broader array of applications.

Keywords brain-inspired intelligence; artificial intelligence; silicon retina; bio-inspired vision; neuromorphic engineering

1 引言

视觉是智慧生物和智能体的基本功能,负责理解与感知外部世界.人类感知系统有超过 80% 的信息量来自视觉系统,远远超过听觉系统、触觉系统以及其他感知系统之和^[1].如何设计强大的视觉感知系统,以应用于计算机视觉技术和人工智能技术,一直是科学界和工业界的研究热点^[2].

视频即静止图像序列,是随着电影电视技术发展起来的一个人造概念,利用了人类视觉系统的视觉暂留现象^[3],将两帧图像刷新的时间间隔小于 50 ms,会产生连续的视觉感受.近年来,传统视频在视觉观赏角度取得了重大进展^[4],但存在数据采样冗余大、感光动态范围小和时域采集低分辨率,在高速运动场景易产生运动模糊等缺点^[5].此外,计算机视觉一直朝着“视频摄像头+计算机+算法=机器视觉”的主流方向^[6-7],却很少人质疑用图像序列(视频)表达视觉信息的合理性,更少人质疑是否凭借该

计算机视觉算法就能实现真正机器视觉.

人类视觉系统具有低冗余、低功耗、高动态及鲁棒性强等优势,可以高效地自适应处理动态与静态信息,且具有极强的小样本泛化能力和全面的复杂场景感知能力^[8].探索人类视觉系统的奥秘,并借鉴人类视觉系统的神经网络结构和视觉信息采样加工处理机理^[9],建立起一套新的视觉信息感知与处理理论、技术标准、芯片和应用工程系统,从而更好模拟、扩展或超越人类视觉感知系统的能力.这是神经科学与信息科学的交叉学科,称之为神经形态视觉^[10-12].

神经形态视觉是一个包含硬件开发,软件支撑,生物神经模型,三者不可缺一的视觉感知系统,其终极目标之一是模拟生物视觉感知结构与机理,以硅视网膜(Silicon Retina)^[13]来以期达到真正的机器视觉.

神经形态视觉传感器的研制是建立在神经科学、生理学等领域对生物视网膜结构与功能机理的研究基础上的.1943年 McCulloch 和 Pitts^[14]提出

一种具有计算能力的神经元模型. 1952 年 Hodgkin 和 Huxley^[15] 对神经元建立动力学微分方程描述神经元动作电位的产生与传递过程, 该动作电位称为脉冲(Spike). 1986 年加州理工学院的博士研究生 Mahowald 所想:“大脑是想象力的发源地, 这让我很兴奋, 我希望可以创造一个想象事物的芯片”, 那时与导师 Carver Mead 教授开始萌发从生物神经科学和工程学角度研究立体视觉的问题. 1990 年 Mead 首次在《Proceedings of the IEEE》上提出神经形态(Neuromorphic)的概念^[16], 利用大规模集成电路来模拟生物神经系统. 1991 年 Mahowald 和 Mead^[13] 在《Scientific American》的封面刊登了一只运动的猫, 标志了第一款硅视网膜的诞生, 其模拟了视网膜上视锥细胞、水平细胞以及双极细胞的生物功能, 正式点燃了神经形态视觉传感器这一新兴领域. 1993 年 Mahowald^[7] 团队为了解决集成电路的稠密三维连线的问题, 提出了一种新型的集成电路通信协议, 即地址事件协议(Address-Event Representation, AER), 实现了事件的异步读出. 2003 年 Culurciello 等人^[18] 设计了一种 AER 方式的积分发放的脉冲模型, 将像素光强编码为频率或脉冲间隔, 称为章鱼视网膜(Octopus Retina). 2005 年 Delbruck 团队^[19] 研制出动态视觉传感器(Dynamic Vision Sensor, DVS), 以时空异步稀疏的事件(Event)表示像素光强变化, 其商业化具有里程碑的意义. 然而, DVS 无法捕捉自然场景的精细纹理图像. 2008 年 Posh 等人^[20] 提出了一种基于异步视觉

的图像传感器(Asynchronous Time-based Image Sensor, ATIS), 引入了基于事件触发的光强测量电路来重构变化处的像素灰度. 2013 年 Delbruck 团队^[21] 开发了动态有源像素视觉传感器(Dynamic and Active pixel Vision Sensor, DAVIS), 即一种双模的技术路线, 增加额外独立的传统图像采样电路弥补 DVS 纹理成像的缺陷, 随后在 2017 年又将其扩展为彩色 DAVIS346^[22]. 2016 年陈守顺教授团队^[23] 采用了增加事件的位宽, 让事件携带像素光强信息输出以恢复场景纹理. 2018 年黄铁军教授团队^[24] 采用了章鱼视网膜^[18] 的光强积分发放采样原理, 将脉冲平面传输替换 AER 方式以节约传输带宽, 验证了积分型采样原理可高速重构场景纹理细节, 即仿视网膜中央凹采样模型(Fovea-like Sampling Model, FSM), 也称 Vidar, 如图 1 所示. 神经形态视觉传感器^[19-23] 模拟生物视觉感知系统, 其具有高的时域分辨率、数据冗余少、低功耗和高动态范围的优势, 在自动驾驶^[25-26]、无人机视觉导航^[27-28]、工业检测^[29] 及视频监控^[30] 等机器视觉领域, 尤其在涉及高速运动和极端光照场景下有着巨大的市场潜力. 此外, 神经形态视觉采样、处理及应用是神经形态工程的重要分支^[31], 为计算神经科学的脑启发视觉模型提供了验证^[32], 也是探索人类智能的有效途径之一. 目前神经形态视觉传感器的研究与应用尚处于初期阶段, 达到或超越人类视觉系统在复杂交互环境下的感知能力还需要大量的探索研究.

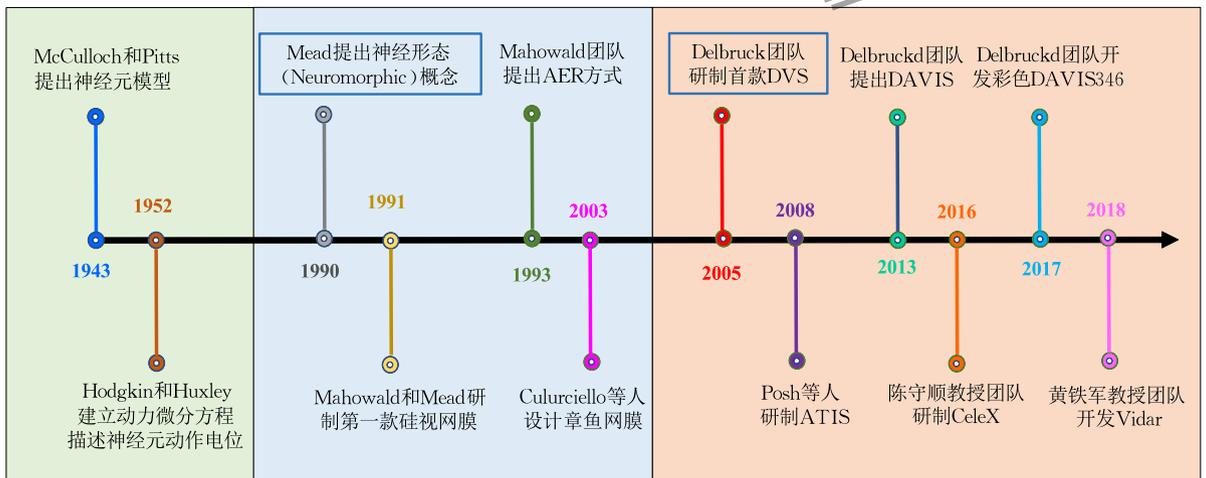


图 1 神经形态视觉传感器的发展历程(蓝色框为里程碑事件)

本文从神经形态视觉的发展历程、生物视觉神经采样模型到神经形态视觉传感器的采样机理及类型、神经视觉信号处理与特征表达及视觉应用的视

角进行系统性回顾与综述, 并展望了该领域未来研究的重大挑战与可能发展方向, 同时探讨了其对未来机器视觉和人工智能领域的潜在影响.

2 神经形态视觉模型与采样机理

神经形态视觉的技术路线总体上分为三个层次:结构层次模仿视网膜,器件功能层次逼近视网膜,智能层次超越视网膜.如果说传统相机是对人类视觉系统的模拟,那么这种仿生物视网膜只是器件功能层次的初级模拟.实际上,传统相机无论是在结构层次、功能层次,甚至智能层次都远不及人类视网膜在各种复杂环境的感知能力.

近年来,各国“脑计划”^[33]相继布局与展开,将从结构层次解析类脑视觉列为重要内容之一支持,主要通过神经科学家采用精细解析与先进探测技术,获取视网膜基本单元的结构、功能及其网络连接,为器件功能层次逼近生物视觉感知系统提供理论支撑.神经形态视觉传感器正是从器件功能层次仿真入手,即采用光电纳米器件模拟生物视觉采样模型与信息处理功能,在有限的物理空间和功耗条件下构造出具有或超越生物视觉能力的感知系统.简而言之,神经形态视觉传感器不用等完全理解视网膜的解析结构与机理再进行模拟,而是借鉴结构层次研究机理并绕过这个更为困难的问题,通过器件功能层次逼近等仿真工程技术手段达到、扩展或超越人类视觉感知系统的能力.

目前,神经形态视觉传感器已经取得了阶段性成果,有模拟视网膜外周感知运动功能的差分型视觉采样模型,如 DVS^[19]、ATIS^[20]、DAVIS^[21-22]、CeleX^[23];也有模拟视网膜中央凹感知精细纹理功能的积分型视觉采样模型,如章鱼视网膜^[18]、Vidar^[24].

2.1 生物视网膜视觉模型

脊椎动物的视网膜是在 6 亿年前由光感受神经细胞演变而成,位于后半眼球的多层复杂神经网络结构,如图 2 所示.灵长类生物视网膜大致可以分为三层结构:光感受器层、内丛状层和外丛状层,其包括光感受器细胞、双极细胞、水平细胞、神经节细胞等主要构成^[34],如图 3 所示.光感受器细胞分为视杆细胞和视锥细胞两类,负责将进入眼球的光信号转化为电信号,并传导至双极细胞和水平细胞.视锥细胞对颜色敏感,主要负责颜色的识别,通常在场景光照较强条件下工作;视杆细胞对光线敏感,能够感受弱光,主要提供夜晚场景下工作,但其没有颜色辨别能力.双极细胞接受光感受器的信号输入,其根据感受野的区域不同分为 ON 型和 OFF 型两类细胞,分别感知光强增加和光强减弱.水平细胞与光感受

器、双极细胞横向互连,对光感受器输出的信号进行亮度调节,同时也负责增强视觉对象边缘凸显轮廓.神经节细胞负责接受双极细胞的视觉信号输入,并以时空脉冲信号(Spatial-Temporal Spike)的形式做出响应,再经视觉纤维传递至视觉皮层^[11].此外,视网膜细胞有多条并行通路传递和处理视觉信号,有极大的带宽传输与速度优势,其中 Magnocellular 和 Parvocellular 通路是最主要的两条信号通路^[35],分别对场景的时域变化敏感和空间结构敏感.

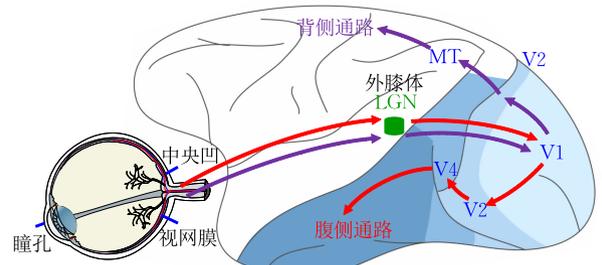


图 2 视觉通路模型

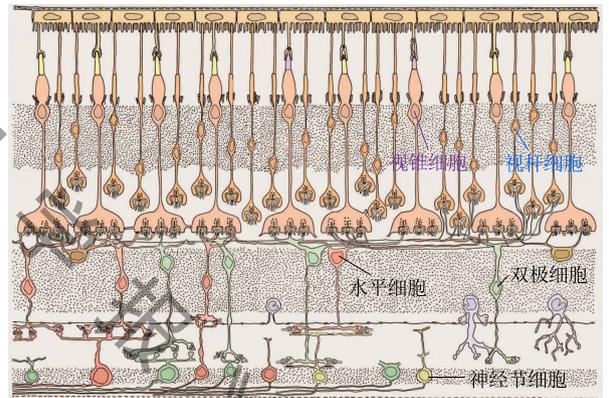


图 3 灵长类生物视网膜横截面示意图^[34]

灵长类生物视网膜具有以下优势^[11]:

- (1) 光感受器的局部自适应增益控制. 以记录光强变化替代绝对光强来消除冗余,对光强感知有高动态范围(High Dynamic Range, HDR);
- (2) 视杆细胞的空间带通滤波器. 滤除低频信息的视觉信息冗余和高频信息的噪声;
- (3) ON 和 OFF 类型. 神经节细胞及视网膜输出均为 ON 和 OFF 脉冲信号编码,降低了单通道的脉冲发放频率;
- (4) 感光功能区. 视网膜中央凹具有高的空间分辨率,可捕捉精细纹理;其外周的区域具有高的时间分辨率,捕获快速运动信息.

此外,依据 Nyquist 采样定理传统相机需要传输超过 20 Gb/s 的数据才能匹配人类视觉的动态范围与空间分辨率,而生物视觉以二进制脉冲信息表

示与编码,视觉神经仅需传输 20 Mb/s 数据至视觉皮层,数据量少将近 1000 倍.因此,视网膜通过神经节细胞将光强信息转换为时空脉冲阵列信号,是一种高效的视觉信息表示与编码的方法^[11, 36],为神经形态视觉传感器提供了理论支撑与功能层次的启发.

生物视觉系统的信息获取、加工和处理主要发生在视网膜、外膝体与视皮层等处^[9],如图 2 所示.视网膜是接收视觉信息的第一站;外膝体是将视网膜视觉信号传输到初级视皮层的信息中转站;视皮层是视觉中央处理器,在学习记忆、思维语言以及知觉意识等高级视觉功能方面发挥至关重要的作用^[37].视皮层信息处理的整个过程由两条并行通路完成:V1、V2 和 V4 等组成腹侧通路主要处理物体形状、颜色等信息识别^[38],也称 what 通路;V1、V2 和 MT 等组成的背侧通路主要处理空间位置、运动等信息^[39],也称 where 通路.因此,借鉴神经计算模型探索人类视觉系统信息处理与分析机制,为计算机视觉与人工智能技术提供借鉴思路与指导方向,进一步启发类脑视觉理论模型与计算方法,从而更好地挖掘视觉特征信息,以逼近生物视觉的高效自适应地处理动静信息、极强的小样本泛化能力与全面视觉分析能力.

2.2 差分型视觉采样及 AER 传输协议

神经元之间的复杂连接,脉冲信号在神经元之间的传递是异步的,那么神经形态工程系统如何模拟这一特性呢?正是 Mahowald 团队^[17]提出新型的通信协议 AER 方式,如图 4 所示,用于脉冲信号多路异步传输,也解决了大规模集成电路的三维稠密连线难题,即“连线问题”(Wiring Problem).

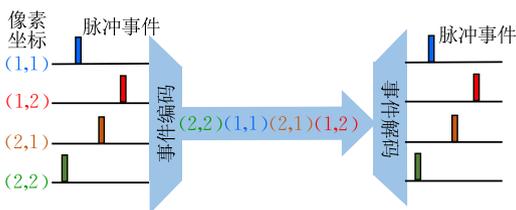


图 4 AER 方式示意图

AER 方式将传感器上每个像素视为独立的,脉冲信号以事件(Event)的形式传输,并按事件产生时间先后顺序异步传出,而不是类似传统相机那样以固定频率的图像传出,解码电路依据地址及时间解析事件属性. AER 方式用于神经形态视觉传感器的主要特点^[40]在于:

- (1) 硅视网膜像素输出事件模拟视网膜的神经元发放脉冲信号的功能;
- (2) 硅视网膜像素之间光强感知、脉冲产生及

传输均异步;

(3) 硅视网膜输出异步事件稀疏时,事件表示及传输更高效.

差分型视觉采样是神经形态视觉传感器感知模型的主流,如 DVS^[19]、ATIS^[20]、DAVIS^[21-22]、CeleX^[23],其对灵长类视网膜外周中光感受器、双极细胞、神经节细胞三层结构的抽象,从器件功能层次上逼近或超越视网膜外周的高时间分辨率感知的能力,如图 5(a).为了提高光强感知动态范围,DVS 系列视觉传感器多采用对数差分模型,即光电流与电压采用对数映射关系,如图 5(b).随着光强的相对变化引起电压变化超过设定阈值 θ ,像素产生一个脉冲信号,如图 5(c)所示,其原理如下:

$$\Delta L \doteq \ln L(u, t) - \ln L(u, t - \Delta t) = p\theta \quad (1)$$

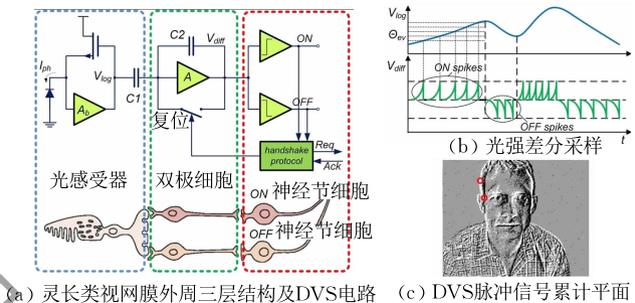


图 5 差分型视觉采样^[11]

差分型视觉传感器采用 AER 方式,每个脉冲信号采用事件形式表示,包括像素位置 $u = (x, y)$ 、发放时间 t 和事件极性 p ,即一个四元组表示 (x, y, t, p) .前三项唯一确定时空域的位置,最后一项极性 $p = \{-1, 1\}$ 分别表示光强的减弱 OFF 和光强增强 ON.差分型神经形态视觉传感器与传统相机相比,其优势在于:

- (1) 输出异步稀疏脉冲也不存在“帧”的概念,不再受限于快门时间和帧率、感知光强的变化,可消除静态不变的视觉冗余;
- (2) 采样具有高时间分辨率,适用于高速运动视觉任务分析;
- (3) 光电流与电压的对数映射关系,增强了高低光照的感知能力进而提升动态范围.

2.3 积分型视觉采样

积分型视觉采样在功能上是对灵长类视网膜中央凹区域的光感受器、双极细胞、神经节细胞三层结构的抽象,如章鱼视网膜^[18]、Vidar^[24].积分型视觉传感器模拟了神经元积分发放模型,将像素光强编码为频率或脉冲间隔^[18, 24, 38-45],具有对视觉场景的精细纹理高速重构的能力^[44],如图 6 所示.光感受

器将光信号转化为电信号,积分器在光照条件 $I(t)$ 条件下进行累计到达累计强度 $A(t)$,当该强度值超过脉冲发放阈值 φ 时,则像素点输出一个脉冲信号,同时积分器复位清空电荷^[45],其原理如下:

$$A(t) = \int_0^t I(t) dt \geq \varphi \quad (2)$$

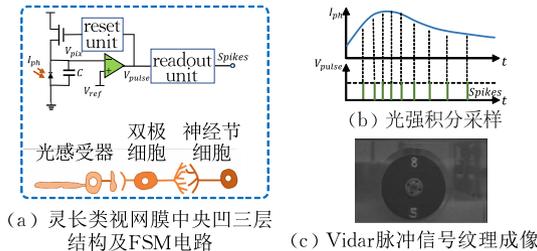


图 6 积分型视觉采样

积分型视觉传感器的像素间彼此独立,章鱼视网膜^[18]采用 AER 方式将脉冲信号进行输出,尤其在光强充足时积分型视觉传感器脉冲发放稠密,事件表示容易出现同一位置及相邻位置多次请求脉冲输出,会出现数据传输的巨大压力,不得不设计总线仲裁机制,为脉冲输出确定优先级,甚至会因带宽限制丢失脉冲信号. Vidar^[24]探索高速轮询的方式以脉冲矩阵的形式传输每个采样时刻的脉冲发放,此方式不需要输出脉冲的坐标与时间戳,只需将像素是否发放标记为“1”和“0”.将脉冲平面轮询的方式代替 AER 方式可以节约传输带宽.

3 神经形态视觉传感器类型

神经形态视觉传感器借鉴生物视觉系统的神经网络结构和视觉信息采样加工处理机理,以器件功能层次模拟、扩展或超越生物视觉感知系统.近年来,一大批代表性的神经形态视觉传感器涌现,是人类在探索仿生视觉技术的雏形,有模拟视网膜外周感知运动功能的差分型视觉采样模型,如 DVS^[19]、ATIS^[20]、DAVIS^[21-22]、CeleX^[23];也有模拟视网膜中央凹感知精细纹理功能的积分型视觉采样模型,如 Vidar^[24].

3.1 DVS

DVS^[19]对灵长类视网膜外周中的光感受器、双极细胞、神经节细胞三层结构的功能抽象,由光电转换电路、动态检测电路和比较器输出电路组成,如图 5 所示.光电转换电路采用对数光强感知模型,提高了光强感知范围,也更加接近生物视网膜的高动态适应的能力.动态检测电路采用差分型采样模型,

即对光强变化做出响应,无光强变化则不响应.比较器依据光强的增加或减弱输出 ON 或 OFF 事件.

传统相机以固定帧率的采样方式,在高速场景易产生运动模糊.而 DVS 采用 AER 异步传输方式的差分型视觉采样模型,以异步时空脉冲信号表示场景光强变化,对场景的变化十分敏感且高时间分辨率,尤其适用于高速运动的视觉任务分析,如图 7 所示. DVS 相比传统相机具有以下优势:高时间分辨率(10^6 Hz)、高动态范围(120 dB)、低功耗、数据冗余少和时延低.

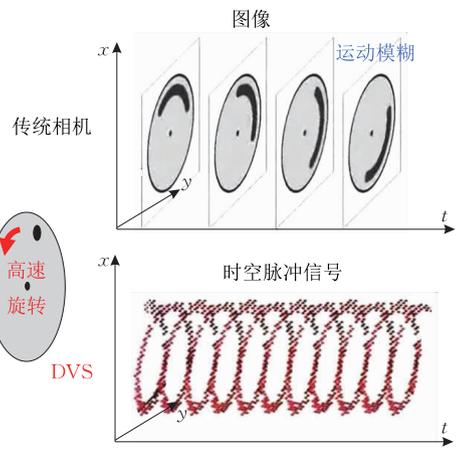


图 7 DVS 时空脉冲信号示意图

Delbruck 团队和 IniVation 公司开发的第一款商用的 DVS128^[19],其空间分辨率为 128×128 ,时域采样频率为 10^6 Hz,动态范围为 120 dB,在高速运动对象识别、检测与跟踪广泛应用.此外,DVS 及其衍生的 ATIS^[20]、DAVIS^[21-22]和 CeleX^[23]等神经形态视觉传感器的研究及产品也备受关注,并逐渐被应用于自动驾驶、无人机视觉导航和工业检测等涉及高速运动视觉任务.例如,Samsung 公司开发了的空间分辨率为 640×480 的 DVS-G2^[46],且像素尺寸为 $9 \mu\text{m} \times 9 \mu\text{m}$. IBM 公司采用了 DVS128 作为类脑芯片 TrueNorth^[47]的视觉感知系统来进行快速手势识别^[48].

DVS 利用差分视觉采样模型可以滤过静止不变或变化较弱的视觉信息以降低数据冗余,同时具有感知高速运动的优势.然而,这种优势带来了视觉重构的劣势,即 ON 或 OFF 事件不携带绝对光强信号,且光强变化较弱时无脉冲信号发放,从而无法重构精细化纹理图像.为了解决 DVS 面向视觉纹理可视化,便衍生出 ATIS^[20]、DAVIS^[21-22]和 CeleX^[23]等神经形态视觉传感器.

3.2 ATIS

ATIS^[20]在 DVS 的基础上巧妙地引入了基于

时间间隔的光强测量电路来实现图像重构,其思路是每次 DVS 电路产生事件时,触发光强测量电路进行工作;光强测量设定了两个不同的参考电压,通过对光强进行积分,并记录达到两个电压发放的事件;由于不同光强的条件下,电压变化相同量所需的时间不同,通过建立光强与时间的映射可以推断出光强大小,从而输出光强变化像素处的光强信息,也称为脉冲宽度调制(Pulse Width Modulation, PWM)^[49]. 此外,为了解决静态区域没有 DVS 脉冲信号发放依然无法获得静态区域的视觉纹理信息,ATIS 引入了一套全局发放机制,即所有像素可被强制发放一次脉冲,这样在 ATIS 初始工作时可获得一整幅图像作为背景,然后运动区域不断产生脉冲进而不断地触发光强测量电路获得运动区域的灰度来更新背景.

Posch 团队和 Prophesee 公司研制的商用 ATIS^[49],其空间分辨率为 304×240 ,时域采样频率为 10^6 Hz,动态范围为 143 dB,在涉及高速视觉任务存在着广泛应用. 另外,Prophesee 公司也受到了 Intel 公司 1500 万美金的项目资助,将 ATIS 应用于自动驾驶汽车的视觉处理系统. 随后,Benosman^[50]团队进一步验证了在 RGB 三个通道的分别用 ATIS 采样脉冲信号再融合彩色的技术方案.

ATIS 在面向高速运动时,依然存在事件与灰度重构更新不匹配的情况,其原因有以下两点:脉冲发放后触发光强测量电路,其测量结果是脉冲发放后一段时间的平均光强,造成运动不匹配;场景稍微变化未引起脉冲发放,从而像素未及时更新,随着时间推移会造成纹理差异明显.

3.3 DAVIS

DAVIS^[21-22]是一种最直观且有效的融合技术思路,将 DVS 和传统相机二者结合,在 DVS 基础上额外引入有源像素传感器(Active Pixel Sensor, APS),用于视觉场景纹理成像.

Delbruck 团队和 IniVation 公司在空间分辨率为 240×180 的 DAVIS240^[21]的基础上进一步推出了彩色 DAVIS346^[22],其空间分辨率达到 346×260 ,时域采样频率为 10^6 Hz,动态范围为 120dB,并将 DVS 所产生的事件坐标的空间位置 (x, y) 携带 RGB 彩色信息,但是 APS 电路采样速度远不如 DVS 电路,APS 模式的帧率为 50FPS,动态范围为 56.7 dB,两套采样电路产生的图像无法做到精准同步,尤其在高速运动场景,且 APS 图像在高速运动场景存在运动模糊.

当前,DAVIS 是神经形态视觉传感器商用产品、工

业应用及学术研究的主流,源于 DVS 系列性传感器(DVS128、DAVIS240、DAVIS346 和彩色 DAVIS346)的学术研究推广,相关视觉任务数据集的公开、代码及软件的开源等营造的良好生态环境. 因此,在本文介绍脉冲信号处理与特征表达、视觉应用等均以差分型视觉采样模型的 DVS 系列传感器为主.

3.4 CeleX

CeleX^[23]在考虑到 ATIS 的光强测量电路存在滞后,在 DVS 电路输出脉冲事件的地址 (x, y) 和发放时间 t 的同时,也及时地将该像素的光强信息 I 输出,即 CeleX 输出事件可用四元组 (x, y, t, I) 表示. CeleX 的设计思路主要包括三个部分^[51]: (1) 引入缓冲和读出开关电路直接将对数光感受器的电路转换为光强信息输出; (2) 采用全局控制信号输出一整帧图像,便于初始工作时可获得整幅图像作为背景和及时全局更新; (3) 专门设计列模拟读出电路输出缓冲器的光强值. CeleX 巧妙地将脉冲事件的位宽设计为 9 bit,既保证了脉冲本身的语义信息,又携带了一定量的光强信息.

陈守顺教授团队和 CelePixel 公司最新发布的第五代 CeleX-V^[52],其空间分辨率达为 1280×800 基本达到了传统相机的水平,同时时域最大输出采样频率为 160 MHz,动态范围为 120 dB,该产品的高空间分辨率、高时域分辨率及高动态范围的“三高”优势引起了当前神经形态工程领域的关注. 此外,CelePixel 公司也受到了 Baidu 公司 4000 万的项目资助. 采用 CeleX-V 用于汽车自动驾驶辅助系统,利用其优势对驾驶异常行为进行实时监测.

CeleX 的脉冲事件采用 9 bit 信息输出,在剧烈运动场景或高速运动场景时,面临数据量大而无法及时传输、甚至丢弃部分脉冲数据以致采样信号无法保真、同时对轻微光照也无法做出响应及时更新等缺点. 但是,CeleX 的“三高”性能及其光流信息的输出等优势,在自动驾驶、无人机视觉导航、工业检测和视频监控等涉及高速运动视觉任务有着巨大的应用潜力.

3.5 Vidar

Vidar^[24]对灵长类视网膜中央凹的光感受器、双极细胞、神经节细胞三层结构的功能抽象,采用了积分型视觉采样模型,将像素光强编码为频率或脉冲间隔^[18, 24, 38-45],具有对视觉场景的精细纹理高速重构的能力^[44]. Vidar 由光电转换电路、积分器电路和比较器输出电路组成,如图 6(a)所示. 光感受器将光信号转换为电信号,积分器将电信号进行积分

累计,比较器将该累计值与脉冲发放阈值比较判断输出脉冲信号,同时积分器被复位,也称为脉冲频率调制(Pulse Frequency Modulation, PFM)^[18]. Vidar 像素间的脉冲信号输出彼此独立,单个像素的脉冲信号按照时间先后顺序排成“脉冲序列”,所有像素按照空间位置相互关系构成“脉冲阵列”,脉冲阵列的每个时刻的截面称为“脉冲平面”,脉冲信号以“1”表示,没有脉冲信号以“0”表示,如图 8 所示.

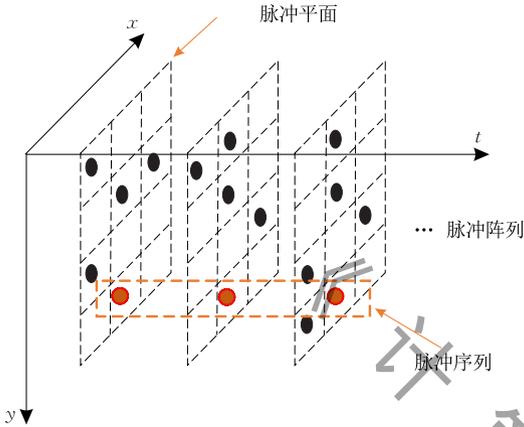


图 8 Vidar 时空脉冲信号示意图

黄铁军教授团队研制的第一款 Vidar^[24],其空间分辨率为 400×250 ,时域采样频率为 4×10^4 Hz,每秒输出 476.3 M 的数据量,并可依据脉冲发放的

时空特性来对静态场景或高速运动场景进行精细化纹理重构,如采用滑动窗口累计法或脉冲间隔映射的方法^[44].此外,Vidar 可自由设定时长的脉冲信号进行影像重构,在成像的动态的范围上具有灵活性.该积分型视觉采样芯片能对高速运动进行精细化纹理重构,可用于高速运动场景的物体检测、跟踪和识别,在自动驾驶、无人机视觉导航、机器视觉等涉及高速视觉任务领域的应用潜力巨大.

Vidar 采用积分型视觉采样模型,将光强信号进行频率或脉冲间隔编码,本质是将光强信息转换为频率编码,相比面向运动感知的 DVS 系列传感器,更加友好地面向视觉精细重构的优势.然而,Vidar 无论在静态场景还是运动区域都会产生脉冲,采样上存在巨大的数据冗余,以及如何控制脉冲发放阈值以便自适应感知不同光照场景和控制数据量是积分视觉采样需要亟待解决的问题.

3.6 仿视网膜视觉传感器性能对比

近期,大量的神经形态视觉传感器涌现并商业化应用,有模拟视网膜外周感知运动功能的差分型视觉采样模型,如 DVS128^[19]、ATIS^[20]、DAVIS346^[21-22]、DVS-G2^[46]、CeleX-V^[23];也有模拟视网膜中央凹功能的积分型视觉采样模型,如 Vidar^[24],具体的各性能参数的对比如表 1 所示.

表 1 神经形态视觉传感器性能参数对比

传感器类型	时间	空间分辨率	最大采样率/Hz	动态范围/dB	功耗/mW	芯片大小/mm ²	像元尺寸/ μm^2	填充因子/%	延迟/ μs	电源电压/V	纹理图像
DVS128 ^[19]	2008	128×128	1×10^6	120	23	6.3×6	40×40	8.1	12	3.3	否
ATIS ^[20]	2011	304×240	1×10^6	143	50~175	9.9×8.2	30×30	20	3	1.8&3.3	灰度
DAVIS346 ^[22]	2017	346×260	1.2×10^7	120	10~170	8×6	18.5×18.5	22	20	1.8&3.3	彩色
DVS-G2 ^[43]	2017	640×480	3×10^9	90	27~50	8×5.8	9×9	100	65~410	1.2&2.8	否
CeleX-V ^[23]	2018	1280×800	1.6×10^8	120	390~470	14.3×11.6	9.8×9.8	9	1	1.2&3.3	灰度
Vidar ^[24]	2018	400×250	4×10^4	—	370	10×6	20×20	13.75	25	1.5&3.3	灰度

神经形态视觉传感器具有两大优势:(1)高速视觉采样的能力,在高速运动视觉任务有着巨大的应用潜力;(2)低功耗,也是 Mead 所提出的神经形态工程^[16]的本质优势和未来可能的终极形态.然而,如何对神经形态视觉传感器输出的时空脉冲信号进行处理、特征表达及高速视觉任务分析是当前神经形态视觉的研究热点;同时如何采样类脑芯片对脉冲信号进行高速处理,应用于涉及高速视觉任务是神经形态工程产业界的关注重点,如 IBM 公司的 TrueNorth^[47]芯片、Intel 公司的 Loihi^[53]芯片和曼切斯特大学的 SpiNNaker^[54]芯片等.

当前,神经形态视觉传感器的空间分辨率从 IniVation 公司开发的第一款商用 DVS128^[19]的

128×128 发展到三星公司 DVS-G2^[46]的 640×480 ,CelePixel 公司 CeleX-V^[52]的 1280×800 ,但相比传统高清与超高清相机在空间分辨率与成像质量上有较大的差距,源于:(1)高空间分辨率与高时域分辨率将给数据传输与存储带来巨大挑战;(2)动态视觉传感器设计的初衷是感知高速运动而非面向高质量视觉观看.总之,神经形态视觉传感器目前尚处于探索的初期阶段,达到人类视觉系统在复杂交互环境下的感知能力还需要大量探索研究.

4 异步时空脉冲信号处理

神经形态视觉传感器模拟了生物视网膜的脉冲

发放机理,如采用差分型视觉采样模型的 DVS 系列传感器,受到视觉场景光强变化刺激发放脉冲信号并记录为地址事件,脉冲信号在空域和时域呈现三维空间的稀疏离散点阵,如图 7 所示。

传统视频信号以“图像帧”范式进行视觉信息表征与信号处理,也是现有机器视觉的主流方向。然而,“异步时空脉冲信号”不同于“图像帧”,现有的图像信号处理机制并不能直接迁移应用。如何建立一套新的信号处理理论与技术体系^[55],是神经形态视觉信号处理领域的研究难点与热点。

4.1 异步时空脉冲信号分析

近年来,异步时空脉冲信号分析^[56]主要集中在滤波、降噪及频域变化分析等方向。

脉冲信号的滤波分析是从信号处理角度的预处理技术,也是神经形态视觉传感器视觉分析任务的应用基础。Benosman^[57]团队提出了一种异步时空脉冲信号的通用滤波方法,即时域或空域构成的层级滤波,可扩展到复杂的非线性滤波器,如边缘检测。Linares-Barranco 等人^[58]将异步时空脉冲信号在 FPGA 上进行滤波降噪和水平特征提取,可显著提升目标识别与跟踪性能。施路平教授团队^[59]针对脉冲信号的发放率进行时空差值滤波,再采用非齐次泊松生成过程实现异步时空脉冲信号的上采样。

神经形态视觉传感器输出的异步时空脉冲信号存在背景噪声和漏电流噪声等干扰^[19-23],Khodamoradi 等人^[60]采用时空相关滤波器来降低 DVS 的背景噪声并在传感器上硬件实现。Orchard 团队^[61]利用 TrueNorth 芯片上实现脉冲神经网络(Spiking Neural Network, SNN)对 ATIS 输出的时空脉冲信号进行降噪,且降噪信号提升了目标对象检测与识别的性能。Wang 等人^[62]将运动对象以外的事件视为噪声,通过光流来判断运动一致性对 DVS 输出的时空脉冲信号进行降噪。

变换域分析是信号处理的基本方法^[63],把时空域变化到频率域,进而研究信号的频谱结构与变化规律。Benosman^[64]团队针对异步时空脉冲信号提出了一种基于事件的快速傅里叶变换,对脉冲信号的频率域有损变换和计算时间做了权衡分析。

异步时空脉冲信号分析与处理存在以下几个探索的方向:

(1) 异步时空脉冲信号在数据分布上可描述为时空点过程^[65],可引入点过程信号处理、学习与推

理理论^[66-67];

(2) 异步时空脉冲信号在时空结构上与点云相似,可利用深度学习在点云网络的结构与方法^[68-70];

(3) 脉冲信号视为图模型的节点,可采用图模型信号处理与学习理论^[71-72];

(4) 异步时空脉冲信号的高时间分辨率的时序优势,挖掘时序记忆模型^[55,73]和借鉴类脑视觉信号处理机理^[74]。

4.2 异步时空脉冲信号度量

异步时空脉冲信号度量是衡量脉冲流之间的相似性,即在度量空间里计算脉冲流之间的距离^[75],是异步时空脉冲信号处理的关键技术之一,在计算神经科学、脉冲编码压缩、机器视觉任务等领域有着广泛且重要的应用。

异步脉冲信号在时空域上呈现为稀疏离散点阵,缺乏在欧式空间的代数运算度量。立体视觉研究^[76]将输出脉冲信号在二维空间投影和提取时空关系的时间面进行度量,应用于三维视觉中深度估计。Gallego 等人^[77]从视觉特征角度系统性地将 DVS 输出的时空脉冲信号进行度量,并在运动补偿、深度估计、光流估计等视觉任务中应用。这些方法是将时空脉冲信号进行时间累积的频率特征,并未充分利用脉冲信号的时域特性。Park 等人^[75]从信号域的角度综述核方法度量,即将离散的时域脉冲信号变换为连续函数,在希尔伯特空间以内积计算脉冲序列之间的距离。Chichinsky 团队^[78]采用卷积神经网络结构将离散脉冲信号映射到特征空间再度量脉冲信号之间距离,并应用到视网膜假体数据测试验证。此类方法在神经生理或模拟生成的脉冲数据上实验,并未考虑真实异步时空脉冲信号的标记属性。

田永鸿教授团队^[79]提出了一种脉冲标记属性独立的脉冲序列度量方法,即 DVS 输出的 ON 和 OFF 的标记属性的脉冲信号分别度量,采用高斯核函数将离散的脉冲序列变换为平滑连续函数,利用希尔伯特空间的内积度量脉冲序列的距离。该方法将脉冲序列作为操作单元,并未考虑脉冲信号的空间结构关系。随后,该团队^[80]进一步将异步时空脉冲信号建模为标记的时空点过程,用条件概率密度函数刻画脉冲信号的空间位置和标记属性,该方法应用于异步脉冲信号有损编码中的运动搜索和失真度量。

异步时空脉冲信号是非结构化数据,不同于可归一化的结构化的“图像帧”,且在主观视觉上差异

无法直接度量.如何将异步脉冲信号度量面向视觉任务和归一化的评价也是亟需解决的难点问题.

4.3 异步时空脉冲信号编码

随着 DVS 系列传感器的空间分辨率不断提高,如三星公司 DVS-G2^[46]的空间分辨率为 640×480 , CeleX-V^[52]的空间分辨率为 1280×800 ,在剧烈运动的场景产生的异步时空脉冲信号面临着传输和存储的巨大挑战,如何对异步时空脉冲信号进行编码压缩,是一个全新的时空数据压缩问题^[81-82].

田永鸿教授团队^[83]首次提出了一种面向时空

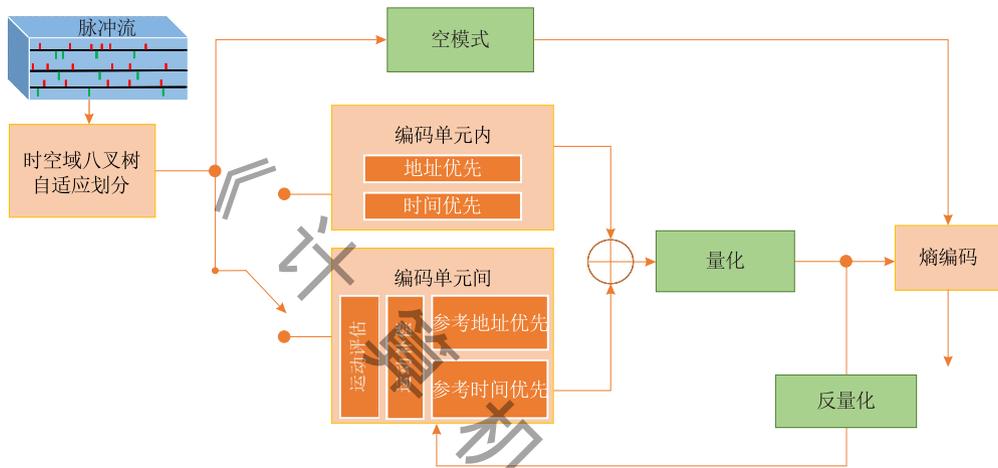


图9 异步时空脉冲信号的编码框架

异步时空脉冲信号的压缩方案是借鉴传统视频编码框架与策略的基础上的初步尝试,但是并未充分分析异步时空脉冲信号的时空特性.端到端的深度学习或受神经计算模型启发的脉冲神经网络的自适应编码器能否应用到异步时空脉冲信号的编码中呢^[85]?当前面临着异步脉冲信号的网络输入、失真度量及足够的数据标注等难题.因此,如何设计鲁棒性的自适应编码器进行异步时空脉冲信号的数据压缩将是一个极具有挑战性且有很高价值的研究课题,可进一步推广到生物脉冲信号的编码与压缩领域.

5 异步时空脉冲信号特征表达

异步时空脉冲信号在时域和空域上呈现为三维空间的稀疏离散点阵,相比传统“图像帧”范式的信号处理与特征表达上更加的灵活,尤其在脉冲信号处理单元的时域长度或脉冲数目上的选择,也增加了异步时空脉冲信号的视觉分析算法输入的难度.因此,如何对异步时空脉冲信号

异步脉冲信号的编码压缩框架,其以脉冲长方体为编码单元,设计了地址优先和时间优先的预测编码策略,可实现对脉冲信号的有效压缩.随后,该团队^[79]进一步探索了更加灵活的时空域八叉树自适应划分、编码单元内的预测及编码单元间的预测等编码策略,进一步提升了时空脉冲信号的压缩效率.此外,该团队^[80]对脉冲空域位置和时域发放时间的失真进行了度量分析,并探讨了面向视觉任务分析^[84]的有损编码的方案,具体的异步时空脉冲信号的编码框架如图9所示.

进行特征表达^[86],如图10所示,挖掘异步时空脉冲信号的时空特性,面向“高精度”与“高速”的视觉分析任务,是神经形态视觉领域最重要和核心的研究问题,也决定着神经形态视觉传感器的推广与应用.

从近些年的研究文献的概况与分布上来看,主要集中在四个方面:频率累计图像(Rate-based Image)、手工设计特征(Hand-crafted Feature)、端到端的深度网络和脉冲神经网络.

5.1 频率累计图像

神经形态视觉传感器的输出为了适用于现有基于“图像帧”的视觉算法,可将异步时空脉冲信号按照时域固定长度或固定数目进行时域投影或频率累计,即频率累计图像.

(1)模型法.将频率累计图像按照图像模式的先验知识进行建模与特征提取. Bischof 团队^[87]将两个 DVS 输出的脉冲信号分别在时域投影为二值图像进行 360° 全景视觉场景重建.陈守顺教授团队^[88]将 CeleX 输出的脉冲信号进行图像插帧和指

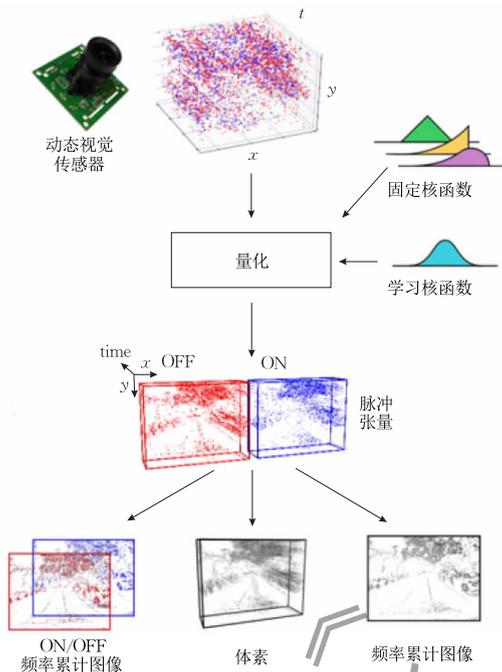


图 10 异步脉冲信号频率累计图像表达^[86]

导运动区域用于高速目标跟踪。Scaramuzza 团队^[89-90]将 DVS 在时域频率累积为灰度图像,再与 APS 图像进行最大似然匹配跟踪。

(2) 深度学习方法. 将频率累计图像输入到基于“图像帧”的深度学习网络。Scaramuzza 团队^[25]将 ON 和 OFF 的脉冲流分别按时域频率累积为灰度图像,再利用 ResNet 对自动驾驶场景的方向盘转角预测。Chen^[91]将 DVS 输出的脉冲流映射为灰度图像,再利用 APS 的伪标签用于自动驾驶场景的车辆检测。田永鸿教授团队^[26]利用积分发放模型将 DVS 输出的脉冲流提取为图像,再联合 APS 图像进行自动驾驶场景的车辆检测。Jiang 等人^[92]按固定时域长度将 DVS 输出的脉冲流累积为灰度图像,并联合 APS 进行行人检测。Zhu 等人^[93]将 DVS 输出的脉冲流按照时域先后顺序映射为灰度图像,利用所提出的 EV-FlowNet 网络进行光流估计。Wang 等人^[94]按照脉冲流按照时域固定长度或固定数目累积为图像,再进行高动态和高帧率的图像重构。Alonso 等人^[95]将 ON 和 OFF 在固定时域长度分别统计直方图信息并融合为灰度图像,采用所提出的 EV-SegNet 网络对自动驾驶场景进行语义分割。Cannici 等人^[96]依据脉冲流的先后顺序映射为灰度图像,采用注意力的机制对目标进行检测。

这些将异步时空脉冲流简洁地时域投影或频率累积转换图像的策略,可直接兼容“图像帧”的视觉

算法,能将神经形态视觉传感器快速且有效应用到涉及高速运动的视觉任务中,也是当前异步时空脉冲信号的特征表达的主流,但该策略并未充分挖掘脉冲流的时空特性。

5.2 手工设计特征

在深度学习算法主导之前,手工设计特征也广泛应用于机器视觉领域,如著名的 SIFT^[97]算子。如何针对异步脉冲信号设计紧凑的手工设计特征^[98],且有尺度和旋转不变形的鲁棒特性,是神经形态视觉传感器应用于视觉任务的重要技术。

一些神经形态视觉领域的学者正在探索如何设计面向视觉任务的手工设计特征。Lin 等人^[99-103]利用异步时空脉冲信号的时域与空域分布特性进行边缘及角点特征提取,用于目标识别、立体视觉的深度估计及目标局部特征跟踪等视觉任务应用。Linares-Barranco 团队与 Delbruck 团队在 CAVIAR 项目^[104]中,采样专用卷积核芯片对脉冲流进行特征提取并用于高速目标跟踪与识别。Bartolozzi 团队^[105]采用时空滤波器进行无监督学习提取视觉感受野特征用在机器人导航的目标识别任务。唐华锦教授团队^[106]提出了一种事件包(Bag Of Events, BOE)的特征,其采用统计学习的方法分析事件的概率,并用于 DVS 采集的手写字体识别。Benosman 团队^[107]将 ATIS 输出的脉冲流,以单个事件为单元并考虑一定空域内的事件发放时间关系来计算时间面(Time Surface),并以层级聚类的方式提取面向识别的 HOTS 特征,随后,该团队^[108]为了消除噪声的对特征提取的干扰,采样邻域内的事件时间面进行平均,再按时间域进行聚合 HATS 特征。Scheerlinck 等人^[109]采用时空滤波器对脉冲流进行边缘算子特征提取,用于目标局部特征检测与跟踪。Zhu 等人^[110]采用时域线性采样核对脉冲流进行权重累积为特征图,采用无监督学习的自编码器网络进行光流估计和深度估计等视觉任务,此外,Scaramuzza 团队^[111]也将该特征输出 LSTM 网络中进行视频重构。Orchard 团队^[98]所提出的 DART 特征算子,其具有尺度和旋转不变形,可应用于目标对象检测、跟踪与识别领域。

手工设计特征在面向特定的视觉任务应用有着较好性能,但手工设计特征需要大量的先验知识,也需要对任务需求及数据特性深入了解并大量调试工作。因此,利用任务驱动的级联方式来监督学习与表

达特征^[86],可更好地挖掘异步时空脉冲信号的时空特性.

5.3 端到端的深度网络

深度学习是当前人工智能的研究热潮,在图像、语音、文本等领域展现出了明显的性能优势.如何让异步时空脉冲信号在端到端的深度网络进行学习,充分挖掘其时空特性,是神经形态视觉研究的热点^[91-103, 112-115]与难点.

(1)卷积神经网络.利用3D卷积对异步时空脉冲信号进行处理. Sekikawa 等人^[112]首次采用端到端的深度网络进行异步时空脉冲信号的视觉任务分析,其采用3D时空分解卷积作为异步时空脉冲信号输入端的计算,即3D卷积核分解为2D空间核和1D运动速度核,再利用递归运算的高效方式处理连续的脉冲流,相比频率累计图像的方法在自动驾驶场景中方向盘转角预测任务中有显著的提升,是神经形态视觉任务领域的里程碑. Tuflyakov 等人^[113]提出了一种事件序列嵌入,即分别采用空间聚合网络和时域聚合网络将离散的脉冲信号提取为连续嵌入特征,其性能相比较于频率累计图像和手工设计特征在深度估计应用存在明显性能优势.

(2)点云神经网络.将异步时空脉冲信号视为三维空间的点云进行处理.神经形态视觉传感器输出的脉冲信号与点云在三维空间数据结构分布上相似,而在时域分布上更加稀疏. Sekikawa 等人^[70]首次采用基于点云网络的 PointNet^[68-69]结构处理异步时空脉冲信号,称为 EventNet,采用时域编码和递归处理的高效处理方式,并应用自动驾驶场景的语义分割和运动评估. Wang 等人^[114]将异步时空脉冲信号视为事件云,并采用了点云网络 PointNet++^[68]的多层级结构提取特征进行手势识别. Yang 等人^[115]将异步脉冲信号视为点云,并提出了一种注意力机制对脉冲信号的领域采样,在手势识别上相比 PointNet^[67-68]等有显著的性能优势.

(3)图神经网络.将脉冲信号视为图模型的节点,采用图模型的处理方式. Welsh 等人^[71]提出将异步时空脉冲信号以概率图模型的相关节点表示,在脉冲信号降噪、光流估计等视觉任务上验证了该方式的时空表示能力. Bi 等人^[72]首次在图神经网络对异步时空脉冲信号进行建模,在手势、字符数字、运动目标等识别任务中较频率累计图像和手工设计特征有大幅度的性能提升.

端到端的深度网络可更好地挖掘异步时空脉冲信号的时空特性,其显著性能优势也备受关注.深度网络在大数据驱动下监督学习获得的性能优势,但异步脉冲信号几乎无法像传统图像那样进行直接主观标注,尤其在目标检测、跟踪及语义分割等高层次视觉任务.此外,异步时空脉冲信号的高速处理能力和低功耗是神经形态视觉传感器广泛应用的前提,而深度学习目前在任务处理速度和功耗上的并无优势.目前,端到端的深度网络对异步脉冲信号的特征表达尚处于萌发阶段,存在大量的研究点及其优化空间.

5.4 脉冲神经网络

脉冲神经网络是第三代神经网络^[116],是模拟生物脉冲信号处理机制的网络结构,其考虑脉冲信号发放的精确时间信息,也是异步时空脉冲信号特征学习的重要研究方向之一.

脉冲神经网络在神经形态视觉应用主要集中在目标分类与识别.唐华锦教授团队^[117]提出了一种多层级联式的前馈脉冲神经网络,应用于DVS记录的数字字符进行分类. Orchard 等人^[74]提出了时域信度分配策略进行脉冲神经网络的反向传播,并采用GPU进行加速运算.施路平教授团队^[118]设计了面向分类任务的深度脉冲神经网络,并在深度学习开源平台上进行监督学习与加速运算.此外,部分研究人员也利用脉冲神经网络在复杂的视觉任务进行尝试. Benosman 团队^[119-120]基于经验信息设计了多层神经元组合的脉冲神经网络,分别应用于双目视差和单目变焦的立体视觉系统. Acharya 等人^[121]提出多层级联式的脉冲神经网络应用固定场景目标的候选区域的检测. Bing 等人^[122]基于STDP学习规则设计了一个端到端的脉冲神经网络应用于机器人视觉导航系统.

目前,脉冲神经网络还停留在理论研究阶段,如异步脉冲时空信号的监督学习梯度优化理论^[123-128]、无监督学习的突触可塑性机理^[129]、深度学习结构启发脉冲神经网络设计^[130]等,其在神经视觉分析任务的性能还远不及端到端的深度学习网络.但是,脉冲神经网络是借鉴神经计算模型更加贴近脑视觉信息处理与分析机制,有着重大的发展潜力与应用前景^[131].因此,如何进一步利用视觉皮层信息加工与处理机理来启发理论模型与计算方法,为脉冲神经网络的设计与优化提供借鉴思路与指导

方向,更好的挖掘视觉特征信息与提升计算效率,如何解决脉冲神经网络的监督学习适用复杂视觉任务,以及如何在硬件电路或神经形态芯片^[47, 53-54]模拟神经元微分方程的高效计算,是脉冲神经网络从理论研究面向实际应用亟需解决的问题。

6 神经形态视觉传感器应用

随着脑认知科学与类脑视觉计算的兴起,计算机视觉成为推动人工智能浪潮的重要方向.神经形态视觉是受生物视觉系统结构与采样机理的启发,达到或超越人类智能的有效途径之一,成为计算神经科学与计算机视觉领域的研究热点.此外,神经形态视觉传感器^[19-23]的高时间分辨率、高动态范围、低功耗、低数据冗余等优势,在自动驾驶^[25-26]、无人机视觉导航^[27-28]、工业检测^[29]及视频监控^[30]等机器视觉领域,尤其在涉及高速摄影、高速运动和复杂光照场景下有着巨大的市场应用潜力。

6.1 数据集及软件平台

6.1.1 仿真数据集

仿真数据是以计算图像学的技术手段来模拟神经形态视觉传感器的采样机理,其采用渲染的形式仿真光学环境、信号传输及电路采样. Mueggler 等人^[132]仿真了在虚三维环境运动的 DAVIS,并生成了 DVS 的脉冲流、APS 图像及场景的深度图,可应用于图像重构、立体视觉的深度估计及视觉导航等视觉任务. Scaramuzza 团队^[133]开源了事件相机仿真器(Event Camera Simulator, ESIM),模拟输出 DVS 脉冲流、APS 图像、场景光流、深度图、相机姿态与路径等数据. Li 等人^[134]提供了大规模多传感器的室内场景的仿真数据集,用于室内导航与定位(Simultaneous Localization and Mapping, SLAM)。

仿真数据集以低成本方式逼近真实数据采集的效果,尤其为端到端的深度学习算法提供数据驱动,能进一步地推动神经形态视觉的研究与发展。

6.1.2 真实数据集

目前,真实数据集中主要是分类与识别任务: N-MNIST^[135]、N-Caltech101^[135]、MNIST-DVS^[136]、CIFAR10-DVS^[137]、DVS-Gesture^[48]、N-CARS^[108]、ALS-DVS^[72]、DVS-PAF^[138]和 DHP19^[139]. Orchard 团队^[125]采用 ATIS 记录 LED 显示器屏幕上运动 MNIST 和 Caltech101 图像. Linares-Barranco 团

队^[136]和施路平教授^[137]团队分别将 DVS128 记录 LED 显示屏上移动 MNIST 数字字符和 CIFAR-10 图片数据. IBM 公司^[48]在不同光照条件下用 DVS128 拍摄 11 种手势动作,构建 DVS-Gesture 手势识别数据集. Benosman 团队^[108]以 ATIS 记录真实车道场景的车辆,组成了 N-CAR 车辆的二分类数据集. Bi 等人^[72]以 DAVIS240 采集了真实的 10 万个美式字母手势的 ALS-DVS 数据集,是识别领域的最大规模的数据集. Miao 等人^[138]采用彩色 DAVIS346 构建了行人检测、行为识别与跌倒检测的 DVS-PAF 数据集. Delbruck 团队^[139]采用 DAVIS346 记录了 17 类 3D 行人姿态的 DHP19 数据集。

神经形态视觉传感器面向场景图像重构任务的数据集: DDD17^[140]、DVS-Intensity^[141]、CED^[142]和 PKU-Spike-High-Speed^[44]. Delbruck 团队^[140]采用了 DAVIS346 记录不同光照场景的脉冲流、图像及方向盘转角等消息, DDD17 数据集也广泛应用于神经形态视觉任务. Scheerlinck 等人^[141-142]分别采用 DAVIS240 和彩色 DAVIS346 构建了用于视觉场景重构的 DVS-Intensity 和 CED 数据集. 田永鸿教授团队^[44]利用 Vidar 记录高速视觉场景及运动对象,构建了 PKU-Spike-High-Speed 数据集。

神经形态视觉传感器面向目标检测、跟踪及语义分割的视觉任务数据集: PKU-DDD17-CAR^[26]、DVS-Pedestrian^[92]、DVS-Benchmark^[143]、EED^[144]、EV-IMO^[145]和 DET^[146]. 田永鸿教授团队^[26]在 DDD17^[140]数据集上对驾驶场景的车辆进行标注,并构建了 PKU-DDD17-CAR 车辆检测数据集. Jiang 等人^[92]利用 DAVIS240 记录校园场景行人,构建了 DVS-Pedestrian 行人检测数据集. Delbruck 团队^[143]采用 DAVIS240 记录 LED 显示屏上目标跟踪数据集,构建可用于目标跟踪、行人识别及目标识别的 DVS-Benchmark 数据集. Aloimonos 团队^[144-145]分别针对目标检测、跟踪与语义分割构建了 EED 和 EV-IMO 数据集. Cheng 等人^[146]采用 CeleX-V 记录了自动驾驶场景中车道线检测的 DET 数据。

神经形态视觉传感器面向深度估计与视觉里程计等任务的数据集: MVSEC^[147]、LMED^[148]、UZH-FPZ^[149]和 ECDS^[132]. Zhu 等人^[147]公开了用于立体视觉的 MVSEC 数据集,采用 DAVIS346 记录大量

自动驾驶及无人机场景,并在神经形态视觉领域广泛应用. Leung 等人^[148]发布了大规模多模态的 LMED 数据集,在移动机器人上用 DAVIS240 和 Astra 深度相机记录室内场景. Scaramuzza 团队^[149]针对 SLAM 任务构建了 UZH-FPV 数据集,包含 DAVIS346 的脉冲流、APS 的图像、光流、相机姿态及路径等内置数据.

总之,神经形态视觉传感器公开的大规模数据集相对较少,尤其在复杂的视觉任务,如在目标检测、跟踪及语义分割等应用中对象及语义级的标注. 开发大规模的神经形态视觉应用数据集是端到端监督学习数据驱动的源泉.

6.2 视觉场景图像重构

ATIS 和 DAVIS 能弥补 DVS 无法直接捕捉场景精细纹理的缺陷,却无法直接对高速运动及极端光照场景图像进行视觉重构,一些视觉场景图像重构方法致力于让 DVS 系列传感器拥有更高的成像质量. Bardow 等人^[150]利用滑动的时空窗口和最小化优化函数来进行场景光流和光强强度评估. Barua 等人^[151]将脉冲信号按固定时域长度进行累计,再采用块稀疏的字典学习方法进行图像重构. Reinbacher 等人^[152]引入光流和构建流形的复杂计算,可实时重构视觉场景图像. Scheerlinck 等人^[141]提出了一种异步滤波策略,结合 APS 图像与 DVS 脉冲流,可重构高帧率与高动态的视频. Pan 等人^[153]提出了一种基于事件的双积分模型,可利用 DVS 输出的脉冲信号对 APS 图像进行去模糊重构. Wang 等人^[62]采用固定时域长度或固定数据的事件累计图像,在利用对抗生成网络生存具有高动态和高帧率的图像视频序列. Scaramuzza 团队^[111]将固定长度的脉冲流进行时域采样为特征图,在利用 LSTM 网络的时序关系进行视频重构.

神经形态视觉传感器是面向机器视觉感知系统,尤其是动态视觉传感器,设计的初衷并非是面向高质量视觉观看. 因而,高质量、高效率和高保真的视觉影像应由视觉传感器直接采样完成,如 CeleX-V. 此外,视觉场景图像重构算法需考虑:如何挖掘神经形态的高时间分辨率及高动态的优势,并利用脉冲信号的时序关系进一步提升面向图像重构的质量.

6.3 光流估计

光流是空间物体在观测成像平面上的像素运动的瞬时速度,不仅涵盖被测物体的运动信息,也有

丰富的三维结构信息,在研究目标检测、跟踪与识别等视觉应用任务有着重要的作用. Benosman 团队^[154-155]提出在像素领域范围内进行光流估计,能对高速及高动态场景中目标进行实时评估. 此外,该团队^[156]利用 4D 时空属性进一步在三维立体视觉中对高速运动目标进行光流估计. Orchard 团队^[157-158]在 TrueNorth 芯片上采样脉冲神经网络进行毫秒级的运动光流评估. Bardow 等人^[150]对 DVS 的脉冲流进行场景光流并用于视觉场景重构. Delbruck 团队^[159-160]采用块匹配的方式在 FPGA 上进行光流评估,进一步^[161]探讨了自适应时间片段的块匹配,可对稀疏或稠密的脉冲流进行实时评估. Zhu 等人^[93]将脉冲流按频率累计为特征图谱,再采用监督学习的 EV-FlowNet 来进行光流评估. Gallego 等人^[162]提出通用的最大对比度框架用于运动补偿、深度估计及光流评估. Croon 团队^[163]采用 DVS128 进行光流估计,能保持无人机平稳着陆. 该团队^[164]进一步以无监督学习的层级脉冲神经网络进行感知全局运动. 陈守顺教授团队^[165]提出了联合像素领域的渲染模式和事件灰度的光流估计方法,并在 CeleX-V 采样输出光流信息.

因异步时空脉冲信号端到端监督学习的特征表达与大规模光流数据集缺失的难题,当前光流评估方法主要是鉴于先验信息的模型,能直接为神经形态视觉传感器的采样芯片提供光流信息输出. 然而,端到端监督学习的方法能充分挖掘异步时空脉冲信号的时空特性,从而进一步提高光流运动评估的性能.

6.4 目标识别

神经形态视觉传感器在字符识别、物体对象识别、手势识别、步态识别及行为识别等领域有着广泛应用,尤其在涉及高速运动、极端光照的场景. 目标识别算法是目前神经形态视觉任务的研究主流,从异步时空脉冲信号的处理角度主要分为:频率累计图像、手工设计特征、端到端的深度网络及脉冲神经网络.

(1) 频率累计图像. 陈守顺教授团队^[166]将脉冲流按照固定时间长度进行投影为图像,并利用前馈网络对人体姿态进行识别. Linares-Barranco 团队^[167]将异步脉冲流按时域固定长度编码为频率累计图像,再利用卷积神经网络来识别人体姿态及高速运动字符. Delbruck 团队^[168]将 DVS 输出的脉冲

流累计为图像与语音信号联合输入到深度置信网络进行字符识别。此外,该团队^[73]采用固定时域长度累计图像序列,用 LSTM 来识别运动字符。Orchard 团队^[169]将 ATIS 输出的脉冲流按照固定时域长度或固定脉冲数据累计为图像,利用卷积网络识别物体对象。IBM 公司^[48]将脉冲信号按时域累积为图像,在神经形态处理芯片 TrueNorth 上用卷积神经网络进行手势识别。Cannici 等人^[96]将脉冲流转为的图像采用注意力的机制对目标检测并识别。Pradhan 等人^[170]将脉冲流累计为图像,再利用卷积神经网络进行行为识别。Wang 等人^[62]先将脉冲信号进行降噪的预处理,再将累计的图像输入到卷积神经网络进行步态识别。

(2)手工设计特征。Lee 等人^[171]采用积分发放模型对运动区域进行分割,利用隐马尔科夫模型对目标区域提取特征进行手势识别。唐华锦教授团队^[106]通过提取 BOE 特征进行手写运动字体识别。Benosman 团队^[172]对脉冲流进行运动特征提取并用于手势识别。此外,该团队^[107-108, 173-174]针对脉冲流分别进行 HOTS、HATS 等时间面特征提取,并采用分类器对手写字体进行识别。Shi 等人^[175]对脉冲流提取二进制特征,并采用统计学习的框架进行字符及姿态等识别。施路平教授团队^[176]以时域编码方法将脉冲流转换为图像,并将卷积神经网络进行

特征在分类器进行数字识别。

(3)端到端的深度网络。Wang 等人^[114]将异步时空脉冲信号视为事件云,并采用了端到端深度点云网络 PointNet++^[68]的层级结构进行手势识别。Yang 等人^[115]采用一种注意力机制对脉冲信号的领域采样,并深度点云网络结构进行手势识别。Bi 等人^[72]首次在图神经网络对异步时空脉冲信号进行建模,用于手势、字符数字、运动目标等识别任务。

(4)脉冲神经网络。唐华锦教授团队^[117]提出了一种多层级联式的前馈脉冲神经网络,应用于手写数字字符识别。Orchard 团队^[177]构建了多层级联式的脉冲神经网络模型对高速运动的字符进行识别。随后,该团队^[74]进一步采用端到端的可监督学习的脉冲神经网络进行字符数字识别。施路平教授团队^[118]设计了面向分类任务的深度脉冲神经网络,并在深度学习开源平台上进行监督学习。

端到端的深度网络在目标的分类与识别任务性能上占据明显优势,如表 2 所示,源于将异步时空脉冲信号以事件点进行建模,能更好地挖掘时空特性。此外,大规模的分类识别数据集为深度网络模型提供了数据驱动,如何进一步利用高时间分辨率的时序优势,挖掘时序记忆模型和借鉴类脑视觉信号处理机理,并在神经形态处理芯片上高速识别,是当前神经形态视觉任务上亟需研究的问题。

表 2 典型的目标识别算法在公开数据集上的测试性能表

表达方式	识别方法	N-MNIST ^[135]	MNIST-DVS ^[136]	N-Caltech ^[135]	CIFAR10-DVS ^[137]	N-CARS ^[205]	DVS-gesture ^[48]	ASL-DVS ^[72]
频率累计图像	VGG_19 ^[72]	0.972	0.983	0.549	0.334	0.728	—	0.806
	RestNet_50 ^[72]	0.984	0.982	0.637	0.558	0.903	—	0.886
	LSTM ^[73]	0.837	0.824	0.196	0.245	0.789	0.769	0.621
手工设计特征	HOTS ^[107]	0.808	0.803	0.210	0.271	0.624	0.785	0.656
	HATS ^[108]	0.991	0.984	0.642	0.524	0.902	0.933	0.871
端到端深度网络	PATs ^[115]	—	—	—	—	—	0.974	—
	RG-CNNs ^[72]	0.990	0.986	0.657	0.540	0.914	0.938	0.901
脉冲神经网络	H-Frist ^[177]	0.712	0.595	0.054	0.077	0.561	0.529	0.479
	Direct-SNN ^[118]	0.995	—	—	0.605	—	—	—
	SLAYER ^[74]	0.992	0.956	0.598	0.532	0.907	0.936	0.896

6.5 目标检测、跟踪与分割

6.5.1 目标检测

近年来,神经形态视觉的目标检测方法从任务角度分为两个方向:基元检测和目标对象检测。

(1)基元检测。Bartolozzi 团队^[178-179]在 iCub 机器人平台上对脉冲流上进行角点检测。Benosman 团队^[180]利用空间几何关系对目标的角点匹配与检测。Aloimonos 团队^[181]采用结构化随机森林对目标

的角点检测并提取目标轮廓。Scaramuzza 团队^[182]对脉冲信号提取时间面并利用边缘算子检测目标的角点。Alzugarav 等人^[102]采用一种异步时空滤波的方法进行角点检测,并应用高速运动的机器人的视觉导航系统。Manderscheid 等人^[103]提出了速度不变形时间平面的特征,并采用随机森林的学习方法进行角点检测。Li 等人^[183]提出了一种快速计算的 FA-Harris 角点检测算子。Seifozzakerini 等人^[184]采

用脉冲神经网络实现 Hough 变换来脉冲流进行直线检测. Benosman 团队^[185]利用迭代优化和最小二拟合方法,对脉冲流进行直线检测.此外, Lee 等人^[186]定义脉冲流的缓存区来进行目标边缘检测.

(2) 目标对象检测. Delbruck 团队^[187]采用 DVS 输出的脉冲流按照频率累计为图像并与 APS 同步,并用 DVS 聚类生成的目标候选区,再采用卷积神经网络将候选区目标进行分类. Chen^[91]利用 APS 的伪标签用于自动驾驶场景的车辆检测. 田永鸿教授团队^[26]将 DVS 脉冲流按照 APS 帧率同步融合,再利用卷积神经网络对自动驾驶场景的车辆检测. Jiang 等人^[92]也采用 DVS 与 APS 同步的策略对行人联合检测. Bartolozzi 团队^[188]采用卷积神经网络将 ATIS 脉冲流的累计图进行目标检测,并应用在 iCub 移动机器人平台. Carnici 等人^[189]采用卷积网络分别将脉冲流转为图像和字符数字的检测.

神经形态视觉的目标检测相关研究与应用尚处于萌发阶段,主要是面向基元检测,即研究视觉的基元特征并检测,为姿态估计及视觉里程计等高级视觉任务提供基础特征.目前,部分目标对象检测方法也是将脉冲流转换为图像特征,并没有充分挖掘脉冲流的时空特性,尤其是高时间分辨率的优势. DVS 时域连续及空间低分辨率的特性,极少规模性的对象标注数据集,较难以实现对目标对象的高精度检测,部分研究方法探索利用时域信息进行检测与跟踪一体化.因此,如何进一步挖掘脉冲流的时空信息,利用监督学习和数据驱动的端到端的深度网络或生物启发性的类脑视觉方法实现高速运动目标检测是亟需解决的问题.

6.5.2 目标跟踪

目标跟踪算法从跟踪任务角度分为两个方向:基元特征跟踪与目标对象跟踪.

(1) 基元特征跟踪. Alzugarav 等人^[101-102]对目标的角点运动路径建立跟踪假设模型^[190],应用于高速运动的机器人的视觉导航系统. Zhu 等人^[191]提出了数据关联的概率模型,并利用光流信息优化关联模型,在高速运动和极端光照条件下对目标特征点持续跟踪. Scaramuzza 团队^[89-90]将 DVS 脉冲流积分图像与 APS 图像梯度特征建立最大似然优化匹配模型,实现对特征点的稳定跟踪.此外, Yuan 等人^[192]提出对直线边缘快速定位与跟踪的方法,用于相机姿态估计. Conradt 团队^[193]在立体视觉系统

进行直线跟踪,并应用移动机器人的视觉导航系统. Li 等人^[194]利用 DVS 的脉冲、APS 的图像及相机 IMU 参数来进行片段特征跟踪.

(2) 目标对象跟踪. Benosman 团队^[195-197]在 ATIS 脉冲流上分别对人脸、交通场景汽车及光学仪器上高速运动粒子等运动目标进行跟踪. Aloimonos 团队^[144]将脉冲流在空间投影为特征面,并采用运动补偿和卡尔曼滤波的方法对高速运动目标稳定跟踪. Delbruck 团队^[187]采用 DVS 和 APS 联合用于运动目标的检测,其将 DVS 脉冲流检测出目标候选区域并用 CNN 分类,再利用粒子滤波对目标定位与跟踪. 陈守顺教授团队^[88]将 CeleX 的脉冲信号来重构插帧来提高图像序列帧率,同时脉冲流可指导运动区域用于高速目标跟踪. Chin 等人^[198]将 DVS 的脉冲流转为图像,并采用旋转评估对天空恒星目标跟踪. Chen 等人^[199]提出了一个自适应时间面将脉冲流转为图像序列,再用检测与跟踪一体化的方法对目标进行跟踪. 此类方法均是将脉冲转为图像或时间面,并未充分挖掘异步脉冲信号的视觉时空特性.部分方法^[200-202]将脉冲信号在时空聚类跟踪, Hinz 等人^[200]对脉冲信号进行在线聚类,应用于交通场景的车辆目标跟踪. Benosman 团队^[201]提出多核并行聚类的目标跟踪方法,适用于方向和尺度变化的多目标高速运动.此外,该团队^[202]进一步采用立体视觉匹配聚类跟踪的算法,可解决高速运动多目标遮挡的问题. 此类方法在简单视觉场景能对多目标有效跟踪,但在聚类区域重叠区域存在鲁棒性差的缺点.

神经形态视觉传感器的高时间分辨率、高动态范围、低冗余及低功耗,尤其适用于高速运动或极端光照条件下的目标跟踪.当前,目标跟踪算法都将脉冲信号转为图像或特征面,未充分挖掘脉冲信号的时空特性,尤其“图像帧”的处理范式很难达到超高速的处理能力,如何借鉴生物视觉信号处理机理及神经形态处理芯片的计算能力,实现“高精度”与“超高速”的目标跟踪是亟需解决的问题.

6.5.3 目标分割

神经形态视觉的目标分割是将脉冲流分成若干个特定且具有特质的区域,并提取感兴趣目标的精确位置的技术. Schraml 等人^[203]将脉冲流进行时空聚类并构建立体视觉系统,可对多行人进行实时分割与行为分析. Chen 等人^[200, 204]提出一种实时聚类跟踪的算法,可实时对交通场景的车辆目标分割与

跟踪. Barranco 等人^[205]提出一种均值漂移的聚类方法,在操控机器人上实时对多目标进行分割与跟踪. Aloimonos 团队^[145]将脉冲流在时域采样为特征序列,用卷积神经网络进行目标分割. Stoffregen 等人^[206]对脉冲流进行时空聚类来分割运动目标及背景,尤其适用于在高速运动及极端光照条件的运动目标分割及背景建模.

当前,脉冲流的目标分割大规模数据集相对较少,部分研究工作尝试经验法的聚类方法. 脉冲流在时空域呈现为三维稀疏点阵,其本质是时空稀疏信号的子空间聚类问题. DVS 在应用到无人机、自动驾驶、机器人等相对运动的 ego-motion 场景,在感知运动目标的同时也触发背景的脉冲流,同时脉冲信号空间分辨率较低且缺乏纹理结构信息. 因此,异步脉冲信号的稀疏子空间聚类是当前神经形态视觉研究的难点,也是神经形态视觉传感器应用于复杂视觉任务的亟需解决的问题.

6.6 三维场景深度估计

当前,利用神经形态视觉传感器的立体视觉应用主要是双目成像与测量系统^[207-218],少数研究者探讨了单目视觉里程计^[219-220]及单目变焦^[127]的方法,从异步时空脉冲信号的处理角度主要分为:频率累计图像、手工设计特征、端到端的深度网络及脉冲神经网络.

(1) 频率累计图像. Lee 等人^[171]将 DVS 输出脉冲流累计为直方图来计算视差,应用于立体视觉系统中手势识别任务. Schraml 等人^[207-208]采用双目的 DVS 模拟激光雷达 360°全景视觉场景重构,将其传感器扫描周期内的脉冲流进行投影为图像计算视差,在自动驾驶场景可实时计算场景深度.

(2) 手工设计特征. Delbruck 团队^[210]首次开发了双目 DVS 的立体视觉系统,其将输出的像素的脉冲序列为特征匹配计算视差,在此基础上进一步^[211]考虑目标的三维几何特征来匹配计算深度信息,并对目标进行三维重建. Benosman 团队^[212]将 ATIS 输出的脉冲流提取时间面等多个特征聚合来计算视差,应用在双目及三目立体视觉系统. Zhu 等人^[213]分别将双目中脉冲流构建视差容器特征,再构建视差匹配优化函数计算视差. Zhou 等人^[214]对脉冲流分别提取角点特征与时空文本特征用于匹配计算深度信息. Xie 等人^[215]利用 DAVIS240 搭建双目的立体视觉系统,对脉冲流提取局部特征用于立体匹配计算深度信息. 此外, Scaramuzza 团队^[219]首

次针对脉冲流构建单目相机多视角三维重建,其利用基于事件的扫描平面方法生成时差图来估计深度信息. Zhou 等人^[220]对脉冲信号提取时间面特征并建立优化函数来计算视差,用于单目相机多视角三维重建.

(3) 端到端的深度网络. Tulyakov 等人^[113]首次针对异步时空脉冲信号提出端到端的深度网络在立体视觉系统进行深度估计,其采用空间聚合网络和时域聚合网络将离散的脉冲信号提取为连续嵌入特征,较频率累计图像和手工设计特征在深度估计的视觉任务上性能与处理速度有明显的性能提升.

(4) 脉冲神经网络. Conradt 团队^[216]提出一种生物启发性的事件驱动的协同网络进行立视差估计. IBM 公司^[217]采用 DAVIS240 搭建双目立体视觉系统,在神经形态处理芯片 TrueNorth 上进行级联式计算视差,对高速运动场景进行实时三维重建. Benosman 团队^[119]采用了多层级联的神经元组合的脉冲神经网络进行双目视觉的深度估计与三维重建,此外该团队^[120]进一步利用脉冲神经网络探索了单目变焦的立体视觉系统.

近年来,深度学习在立体视觉领域的成功应用,显著提升了深度估计与三维重建的计算速度与性能. 如表 3 所示,但在神经形态视觉的立体视觉应用刚处于萌发阶段^[113]. 大规模立体视觉的数据集、激光雷达等测距传感器的融合^[218]、异步脉冲信号的特征表达,以及利用神经形态处理芯片在立体视觉^[217]中高速处理都是亟需解决的难题.

表 3 典型的深度估计算法在 MVESC^[147]数据集测试性能

表达方式	估计方法	平均深度估计误差/cm	
		Sequence1	Sequence3
频率累计图像	TSSES ^[209]	36	36
	SGM ^[113]	93	119
	FCVF ^[113]	99	103
手工设计特征	Semi-Dense ^[219]	13	33
	CopNet ^[209]	61	64
深度网络	DDES ^[113]	13.6	18.4

6.7 姿态估计与视觉里程计

视觉里程计 (Visual Odometry, VO) 是利用单个或多个视觉传感器的输出信号估计智能体的位置与姿态,也称通过视觉信息获取相机姿态的问题. 从相机标定的角度可分为相机姿态估计 (Pose tracking) 和视觉惯性里程计 (Visual-Inertial Odometry, VIO).

(1) 相机姿态估计. Scaramuzza 团队^[221]在无人机上搭载 DVS,将脉冲流按照频率累计为图像并对

直线特征跟踪从而对相机姿态评估。Kim 等人^[222]首次利用单目 DAVIS 相机联合评估三维场景结构、6-DOF 相机姿态及场景光强。Kueng 等人^[223]采用 DAVIS240 中 APS 图像检测角点,并用 DVS 脉冲流进行特征跟踪来对相机姿态估计。Gallego 等人^[224]将脉冲流构建场景深度图,在利用贝叶斯滤波的策略对 6-DOF 相机姿态评估,并能对高速运动及极端光照条件下进行实时估计。此外,Rebecq 等人^[225]进一步集成三维重建方法与并行加速对姿态跟踪。Bryner 等人^[226]提出了一种包含光强度和深度信息的三维光度特征图,并采用最似然优化模型对 6-DOF 姿态评估。Nguyen 等人^[227]将脉冲流累计为图像,并采用空间堆栈式的 LSTM 学习 6-DOF 的相机姿态。

(2) 视觉惯性里程计。Censi 等人^[228]首次将 DVS 与 APS 结合用于视觉里程计。Zhu 等人^[229]将脉冲流提取特征面并采用扩展的卡尔曼滤波进行特征跟踪,再与 IMU 内测参数联合对 6-DOF 相机姿态评估。Mueggler 等人^[230]结合 DVS 脉冲流和 IMU 参数,提出一种连续的视觉惯性里程计。

当前,神经形态视觉传感器在视觉里程计的大部分工作是基于几何运动约束的求解,涉及特征提取、特征匹配、运动估计等过程,较传统相机其处理速度与性能精度均有显著的提升,尤其在高速运动场景及高动态场景。然而,模型法存在设计过程繁琐和相机更换平台重新标定等不足,端到端的深度网络在数据驱动下有潜力大幅度提升处理速度与性能精度,但深度学习将高维观测向量映射到高维姿态向量空间是一个极其困难的问题。此外,利用双目或多目的视觉场景深度信息来进一步提高视觉里程计性能也是亟需解决的问题。

6.8 神经形态工程系统应用

神经形态视觉传感器在工程系统应用主要有以下几个方面:

(1) 消费电子。三星公司^[231]利用 DVS 的快速感知动态变化、低数据冗余及低功耗等优势,应用于手机的触屏唤醒功能。此外,DVS 脉冲流增强传统相机视频采集质量^[153],尤其在高速场景及极端光照条件下,从而进一步提高消费电子类相机的高速摄影及动态感知能力。

(2) 工业检测。Ni 等人^[232]利用 ATIS 的高时间分辨率对工业微型机床的高速运动的抓手进行跟踪与系统反馈。Belbachir 等人^[233]采用 DVS 对高速运

转平台上的工业部件进行实时识别。作为工业检测的视觉感知系统,神经形态视觉的低功耗的优势在万物互联的物联网时代存在巨大的应用潜力^[234]。

(3) 移动机器人。Jorg 团队^[235]将 DVS 作为移动机器人的视觉感知系统,其在高速移动状况下可实时抓取物体。Bartolozzi 团队^[236]将 DVS 用于 iCub 人形机器人的视觉感知系统,其具有低延时、高速运动及极端光照的感知能力。

(4) 自动驾驶。陈守顺教授团队^[23]将 CeleX 用于驾驶员的疲劳检测系统;Scaramuzza 团队^[25]将 DVS 用于自动驾驶场景的方向盘转角预测;田永鸿教授团队^[26]将 DVS 和 APS 结合用于自动驾驶的目标联合检测,其性能在高速运动场景及极端光照条件均有显著性的提升。

(5) 无人机导航。Scaramuzza 团队^[237, 238]将 DVS 作为无人机的视觉导航系统,相比传统相机在高速运动场景与极端光照有更好的跟踪能力。

(6) 高速视觉测量。Hsu 等人^[239]设计一款 64×64 的动态视觉采样芯片,该芯片内集成光流计算模块,能对高速运动目标进行速度测量。

7 研究挑战与可能发展方向

神经形态视觉传感器的研究与应用已取得了阶段性的发展,但达到或超越人类视觉系统在复杂交互环境下的感知能力还存在诸多问题与挑战。此外,本章进一步讨论了主要可能发展的研究方向。

7.1 研究挑战

神经形态视觉传感器具有时域高分辨率的优势,其输出的脉冲信号在时域和空域呈现三维稀疏点阵,视觉信号处理及应用存在以下几个方面的挑战:

(1) 大规模的数据集。监督学习视觉任务提供数据驱动及视觉分析的模型评价。目前,神经形态视觉任务的数据集主要是分类与识别任务,其数据场景简单且规模较小,总体上公开的大规模数据集较少,尤其在复杂的视觉任务,如目标检测、跟踪及对象分割等应用,原因在于脉冲流在时空呈现异步点阵,在主观视觉无法直接手工标注与评价,需借助其他传感器辅助标定。开发大规模的神形态视觉应用数据集是亟需解决的问题,可利用计算图形学以低成本构建仿真数据集^[132-134];从现有大规模视觉数据集映射到脉冲信号的数据集^[240];借助其他传

感器大规模采集与标注的真实数据集。

(2) 异步时空脉冲信号的度量. 度量空间里计算脉冲流之间的距离^[75], 是异步时空脉冲信号处理的基础关键技术之一, 在计算神经科学、脉冲编码压缩、监督学习的视觉任务等众多领域有着广泛应用. 异步时空脉冲信号是非结构化数据, 不同于可归一化的结构化的“图像帧”, 缺乏欧式空间的直接度量, 且在主观视觉上差异无法直接度量. 如何将异步脉冲信号度量面向视觉任务和归一化的评价也是亟需解决的研究难题, 可将异步时空脉冲信号在数据分布上建模为时空点过程^[65], 并引入点过程信号处理、度量与学习理论.

(3) 异步时空脉冲信号的特征表达. 挖掘异步时空脉冲信号的时空特性, 面向高精度的视觉分析任务. “图像帧”范式的机器视觉方法是当前主流方向, 但异步时空脉冲信号不同于“图像帧”, 现有的深度学习算法并不能直接迁移应用. 此外, 脉冲信号相比传统“图像帧”范式的特征表达上更加的灵活, 尤其在脉冲信号处理单元的时域长度或脉冲数目上的选择, 也增加了异步时空脉冲信号的视觉分析算法输入的难度. 如何对异步时空脉冲信号进行高效的特征表达, 是神经形态视觉领域核心的研究问题之一, 可利用数据驱动建立面向具体视觉任务的端到端的神经网络^[68-72]; 挖掘具有时序记忆机制^[55, 73]的脉冲神经网络^[74]进行学习与推理.

(4) 高速处理能力. 挖掘时域高分辨率特性, 面向高速视觉任务. 异步时空脉冲信号的高速处理能力是神经形态视觉传感器广泛应用的前提, 而现有手工设计特征、端到端的深度网络等特征表达方法在任务处理速度上的并无优势. 神经形态处理芯片在脉冲信号上有高速且并行处理的能力, 如 IBM 公司的 TrueNorth^[47] 芯片、Intel 公司的 Loihi^[53] 芯片和曼切斯特大学的 SpiNNaker^[54] 芯片等, 也是当前神经形态工程产业界的关注重点.

(5) 神经形态开源学习框架. 开发适用于异步脉冲信号处理与学习的开源框架. 目前, 神经形态领域类似深度学习的框架较少, 亟需开发适应于非结构化脉冲数据的框架^[241], 如脉冲神经网络或兼容深度学习与脉冲网络的混合框架, 为神经形态视觉研究者提供开源学习工具与生态环境.

7.2 可能发展的方向

神经形态视觉传感器借鉴生物视觉系统的神经网络结构和视觉信息采样加工处理机理, 以器件功

能层次模拟、扩展或超越生物视觉感知系统. 在传感器的器件材料工艺、视觉采样模型、视觉分析任务、感知系统及其应用推广上存在以下可能发展的方向:

(1) 面向新型材料的神经形态视觉传感器. 新材料、新器件等应用构建具有神经元、突触及记忆功能的神经形态视觉传感器. Zhou 等人^[242] 采用光电阻式的随机存取记忆存储器 (Optoelectronic Resistive Random Access Memory, ORRAM) 的突触器件应用于神经形态视觉传感器, ORRAM 阵列具有视觉信号感知、信号记忆与存储进行预处理的功能, 从而提升后端视觉分析任务的性能, 在边缘计算和物联网存在巨大的应用价值. 因此, 研发新材料及器件工艺是神经形态视觉传感器从器件功能层次模拟、扩展或超越生物视觉感知系统的最有效的方式之一.

(2) 面向多频谱波段的神经形态采样模型. 将神经形态视觉传感器的差分型及积分型采样模型应用在可见光以外的频谱波段, 如红外、紫外及微波等. Lenero-Bardallo 等人^[43] 将积分型采样模型用于红外波段, 可在红外波段成像且有高时间分辨率、高动态及低数据冗余的优势. 将生物视觉采样模型及机理借鉴到多频谱的信号采样, 来提升现有单一波段传感器的感知范围与能力.

(3) 面向多视觉任务协同. 同一个模型分析与处理多个视觉分析任务, 其共享底层特征, 任务之间相互利用, 有更强的泛化能力. 目前, 神经形态视觉任务是面向单一视觉任务的特定经验模型及学习方式, 可借鉴传统“图像帧”的多任务协同模型^[243], 在功能上逼近人类多任务视觉信息处理与分析的能力.

(4) 面向多传感器融合的神经形态工程系统. 将视觉、语音及触觉等多神经形态传感器融合, 应用于神经形态工程的智能感知系统. Liu 团队^[244] 首次研制了语音信号的神经形态传感器, 即动态语音传感器 (Dynamic Audio Sensor, DAS). Barolozziy 研制^[245] 了神经形态的触觉传感器, 并在《科学》上发表论文. 如何将视觉、语音及触觉等多个神经形态传感器, 甚至与传统相机、激光雷达等传感器融合^[246], 构建在复杂交互环境下的感知能力系统是神经形态工程应用的目标.

(5) 计算神经科学与神经形态工程的耦合. 计算神经科学对视网膜的精细解析与采样机理的研究, 利用视觉神经反演计算理论研究视网膜编解码

模型,提出新型的神经形态视觉采样与重构模型,为器件功能层次逼近生物视觉感知系统提供理论支撑^[247, 248];借鉴生物视觉系统信息处理与分析机制,进一步启发类脑视觉理论模型与计算方法,为计算机视觉与人工智能技术提供借鉴思路与指导方向^[249]。同时,神经形态视觉在信号采样、处理与分析应用进展为计算神经科学提供模型验证,甚至为生命医学领域提供视网膜假体等仿生医学器件,从而推动计算神经科学的研究与发展。二者彼此耦合,为构建一套新型的视觉信息感知、处理与分析提供了理论与应用支撑。

8 总 结

神经形态视觉是一个包含硬件开发、软件支撑、生物神经模型,三者不可缺一的视觉感知系统,其终极目标之一是模拟生物视觉感知结构与机理,以硅视网膜来以期达到真正的机器视觉。神经形态视觉采样、处理及应用是神经形态工程的重要分支,目前尚处于研究的初期阶段,达到或超越人类视觉系统在复杂交互环境下的感知能力还需要大量的探索研究。距 1986 年 Mahowald 所想“The brain is imagination, and that was exciting to me; I wanted to build a chip that could imagine something”,答案也许如其导师 Carver Mead 教授所说“Listen to the technology, find out what it’s telling you”,我们也相信神经形态工程不久将会是可能的答案。

参 考 文 献

- [1] Ernst M O, Banks M S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 2002, 415(6870): 429-433
- [2] Liu L, Chen J, Fieguth P, et al. From BoW to CNN: Two decades of texture representation for texture classification. *International Journal of Computer Vision*, 2019, 127(1): 74-109
- [3] Amor T A, Reis D S, Campos D, et al. Persistence in eye movement during visual search. *Scientific Reports*, 2016, 6: 20815
- [4] Zare A, Homayouni M, Aminlou A, et al. 6K and 8K effective resolution with 4 K HEVC decoding capability for 360 video streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2019, 15(2s): 68
- [5] Schuler C J, Hirsch M, Harmeling S, et al. Learning to deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(7): 1439-1451
- [6] Hossain M D, Sohel F, Shiratuddin M F, et al. A comprehensive survey of deep learning for image captioning. *ACM Computing Survey*, 2019, 51(6): 118
- [7] Wu Y, Ji Q. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 2019, 127(2): 115-142
- [8] Bays P M, Husain M. Dynamic shifts of limited working memory resources in human vision. *Science*, 2018, 321(5890): 851-854
- [9] Linsley D, Eberhardt S, Sharma T, et al. What are the visual features underlying human versus machine vision?// *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2017: 2706-2714
- [10] Indiveri G, Douglas R. Neuromorphic vision sensor. *Science*, 2000, 288(5469): 1189-1190
- [11] Posh C, Serrano-Gotarredona T, Linares-Barranco B, et al. Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output. *Proceedings of the IEEE*, 2014, 102(10): 1470-1484
- [12] Boahen K. Neuromorphic microchips. *Scientific American*, 2005, 292(5): 56-63
- [13] Mahowald M, Mead C. The silicon retina. *Scientific American*, 1991, 264(5): 76-83
- [14] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 1943, 5(4): 115-133
- [15] Hodgkin A L, Huxley A F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 1952, 117(4): 500-544
- [16] Mead C. Neuromorphic electronic systems. *Proceedings of the IEEE*, 1990, 78(10): 1629-1636
- [17] Lazzaro J, Wawrzynek J, Mahowald M, et al. Silicon auditory processors as computer peripherals//*Proceedings of the Advances in Neural Information Processing Systems*. Denver, USA, 1993: 820-827
- [18] Culurciello E, Etienne-Cummings R, Boahen K A. A biomorphic digital image sensor. *IEEE Journal of Solid-State Circuits*, 2003, 38(2): 281-294
- [19] Lichtsteiner P, Posh C, Delbruck T. A 128 × 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 2008, 43(2): 566-576
- [20] Posh C, Matolin D, Wohlgenannt R. An asynchronous time-based image sensor//*Proceedings of the IEEE International Symposium on Circuits and System*. Seattle, USA, 2008: 2130-2133
- [21] Brandli C, Berner R, Yang M, et al. A 240 × 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 2014, 49(10): 2333-2341

- [22] Moeys D P, Corradi F, Li C, et al. A sensitive dynamic and active pixel vision sensor for color or neural imaging applications. *IEEE Transactions on Biomedical Circuits and System*, 2017, 12(1): 123-126
- [23] Guo M, Ding R, Chen S. Live demonstration: A dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand//*Proceedings of the IEEE International Symposium on Circuits and System*. Montreal, Canada, 2016; 456
- [24] Dong S, Huang T, Tian Y. Spike camera and its coding methods//*Proceedings of the Data Compression Conference*. Salt Lake City, USA, 2017; 437
- [25] Maqueda A I, Loquercio A, Gallego G, et al. Event-based vision meets deep learning on steering prediction for self-driving cars//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018; 5419-5427
- [26] Li J, Dong S, Yu Z, et al. Event-based vision enhanced: A joint detection framework in autonomous driving//*Proceedings of the IEEE International on Multimedia and Expo*. Shanghai, China, 2019; 1396-1401
- [27] Mitrokin A, Sutor P, Fermuller C, et al. Learning sensor/motor control with neuromorphic sensors: Towards hyperdimensional active perception. *Science Robotics*, 2019, 4(30): eaaw6736
- [28] Sanket N J, Parameshwara C M, Singh C D, et al. EVDodgeNet: Deep dynamic obstacle dodging with event cameras//*Proceedings of the IEEE International Conference on Robotics and Automation*. Paris, France, 2020; 1-8
- [29] Afshar S, Hamilton T J, Davis L, et al. Event-based processing of single photon avalanche diode sensors. *IEEE Sensors Journal*, 2020, 20(14): 7677-7691
- [30] Xu L, Xu W, Golyanik V, et al. EventCap: Monocular 3D capture of high-speed human motions using an event camera//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2020; 4968-4978
- [31] Liu S C, Delbruck T. Neuromorphic sensory systems. *Current Opinion in Neurobiology*, 2010, 20(3): 288-295
- [32] Weiland J D, Humayun M S. Retinal prosthesis. *IEEE Transactions on Biomedical Engineering*, 2014, 61(5): 1412-1424
- [33] Poo M, Du J, et al. China brain project: Basic neuroscience, brain diseases, and brain-inspired computing. *Neuron*, 2016, 92(3): 591-596
- [34] Field G D, Chichilnisky E J. Information processing in the primate retina: circuitry and coding. *Annual Review of Neuroscience*, 2007, 30: 1-30
- [35] Rakic P, Riley K P. Regulation of axon number in primate optic nerve by prenatal binocular competition. *Nature*, 1983, 305(5930): 135
- [36] Maass W. To spike or not to spike: That is the question. *Proceedings of IEEE*, 2015, 103(12): 2219-2224
- [37] Nassi J J, Callaway E M. Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience*, 2009, 10(5): 360-372
- [38] Amedi A, Malach R, Hendler T, et al. Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, 2001, 4(3): 324-330
- [39] Freud E, Plaut D C, Behrmann M. What's happening in the dorsal visual pathway. *Trends in Cognitive Sciences*, 2016, 20(10): 773-784
- [40] Costas S J, Serrano G T, et al. A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision system. *IEEE Transactions on Circuits and Systems*, 2007, 54(7): 1444-1458
- [41] Lenero-Bardallo J A, Bryn D H, Hafliger P. Bio-inspired asynchronous pixel event tricolor vision sensor. *IEEE Transactions on Biomedical Circuits and Systems*, 2014, 8(3): 345-357
- [42] Lenero-Bardallo J A, Hafliger P, Carmona-Galan R, et al. A bio-inspired vision sensor with dual operation and readout modes. *IEEE Sensors Journal*, 2016, 16(2): 317-330
- [43] Lenero-Bardallo J A, Guerrero-Rodriguez J M, Carmona-Galan R, et al. On the analysis and detection of flames with an asynchronous spiking image sensor. *IEEE Sensors Journal*, 2018, 18(16): 6588-6595
- [44] Zhu L, Dong S, Huang T et al. A retina-inspired sampling method for visual texture reconstruction//*Proceedings of the IEEE International Conference on Multimedia and Expo*. Shanghai, China, 2019; 1432-1437
- [45] Gao J, Wang Y, Nie K, et al. The analysis and suppressing of non-uniformity in a high-speed spike-based image sensor. *Sensors*, 2018, 18(12): 4232
- [46] Son B, Sun Y, Kim S, et al. 4.1 A 640×480 dynamic vision sensor with a 9 μm pixel and 300 Meps address-event representation//*Proceedings of the IEEE International Solid-State Circuits Conference*. San Francisco, USA, 2017; 66-67
- [47] Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 2014, 345(6197): 668-673
- [48] Amir A, Taba B, Berg D, et al. A low power, fully event-based gesture recognition system//*Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. Hawaii, USA, 2017; 7243-7252
- [49] Posh 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. *IEEE Journal of Solid-State Circuits*, 2010, 46(1): 259-275
- [50] Macireau A, Ieng S H, Simon-Chane C, et al. Event-based color segmentation with a high dynamic range sensor. *Frontiers in Neuroscience*, 2018, 12: 35
- [51] Huang J. Asynchronous High-Speed Feature Extraction Image Sensor Ph. D. dissertation]. Nanyang Technological University, Singapore, 2018

- [52] Chen S, Guo M. Live demonstration: CeleX-V: A 1M pixel multi-mode event-based sensor//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019; 1682-1683
- [53] Davies M, Srinivasa N, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 2018, 38(1): 82-99
- [54] Furber S B, Galluppi F, Temple S, et al. The SpiNNaker project. *Proceedings of the IEEE*, 2014, 102(5): 652-665
- [55] Indiveri G, Liu S C. Memory and information processing in neuromorphic systems. *Proceedings of the IEEE*, 2015, 103(8): 1379-1397
- [56] Lagorce X. Computational Methods for Event-Based Signals and Applications[Ph. D. dissertation]. University of Parris VI, Parris, 2015
- [57] Ieng S H, Posch C, Benosman R. Asynchronous neuromorphic event-driven image filtering. *Proceedings of the IEEE*, 2014, 102(10): 1485-1499
- [58] Linares-Barranco A, Perez-Pena F, Moeys D P, et al. Low latency event-based filtering and feature extraction for dynamic vision sensors in real-time FPGA applications. *IEEE Access*, 2019, 7: 134926-134942
- [59] Li H, Li G, Shi L. Super-resolution of spatiotemporal event-stream image. *Neurocomputing*, 2019, 335: 206-214
- [60] Khodamoradi A, Kastner R. $O(N)$ -space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Transactions on Emerging Topics in Computing*, 2018
- [61] Padala V, Basu A, Orchard G. A noise filtering algorithm for event-based asynchronous change detection image sensors on TrueNorth and its implementation on TrueNorth. *Frontiers in Neuroscience*, 2018, 12: 118
- [62] Wang Y, Du B, Shen Y, et al. EV-Gait: Event-based robust gait recognition using dynamic vision sensors//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019; 6358-6367
- [63] Lee Y, Vo T V, Lim K W, et al. Z-transforms and its inference on partially observable point process//Proceedings of the International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018; 2369-2375
- [64] Sabatier Q, Ieng S H, Benosman R. Asynchronous event-based Fourier analysis. *IEEE Transactions on Image Processing*, 2017, 26(5): 2192-2202
- [65] Gonzalez J A, Rodriguez-Cortes F J, Cronie O, et al. Spatio-temporal point process statistics: A review. *Spatial Statics*, 2016, 18: 505-544
- [66] Xiao S, Yan J, Farajtabar M, et al. Learning time series associated event sequences with recurrent point process networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(10): 3124-3136
- [67] Upadhyay U, De A, Rodriguez M G. Deep reinforcement learning of marked temporal point processes//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018; 3168-3178
- [68] Qi C R, Yi L, Su H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017; 5099-5108
- [69] Li Y, Bu R, Sun M, et al. PointCNN: Convolution on X-transformed points//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018; 820-830
- [70] Sekikawa Y, Hara K, Saito H. EventNet: Asynchronous recursive event processing//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019; 3887-3896
- [71] Welsh T, Zarrella G, Bhattacharya P, et al. Graph based event processing//Proceedings of the Imaging Systems and Applications. Munich, Germany, 2019, IM1B.3
- [72] Bi Y, Chadha A, Abbas A, et al. Graph-based object classification for neuromorphic vision sensing//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea, 2019; 491-501
- [73] Neil D, Pfeiffer M, Liu S C. Phased LSTM: Accelerating recurrent network training for long or event-based sequences//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016; 3882-3890
- [74] Shrestha S B, Orchard G. SLAYER: Spike layer error reassignment in time//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018; 1412-1421
- [75] Park I M, Seth S, Paiva A R, et al. Kernel methods on spike train space for neuroscience: A tutorial. *IEEE Signal Processing Magazine*, 2013, 30(4): 149-160
- [76] Steffen L, Reichard D, Weinland J, et al. Neuromorphic stereo vision: A survey of bio-inspired sensors and algorithms. *Frontiers in Neurobotics*, 2019, 13: 28
- [77] Gallego G, Gehrig M, Scaramuzza D. Focus is all you need: Loss functions for event-based vision//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019; 12280-12289
- [78] Shah N P, Madugula S, Chichinsky E J, et al. Learning a neural response metric for retinal prosthesis//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018; 1-15
- [79] Dong S, Bi Z, Tian Y, et al. Spiking coding for dynamic vision sensor in intelligent driving. *IEEE Internet of Things Journal*, 2018, 6(1): 60-71
- [80] Fu Y, Li J, Dong S, et al. Spike coding: Towards lossy compression for dynamic vision sensor//Proceedings of the Data Compression Conference. Salt Lake City, USA, 2019; 572-572

- [81] Atluri G, Karpatne A, Kumar V. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys*, 2018, 51(4): 83
- [82] Khan N, Martini M G. Bandwidth modeling of silicon retinas for next generation visual sensor networks. *Sensors*, 2019, 19(8): 1751
- [83] Bi Z, Dong S, Tian Y, et al. Spike coding for dynamic vision sensor//Proceedings of the Data Compression Conference. Salt Lake City, USA, 2018: 117-126
- [84] Cohen G, Afshar S, Orchard G, et al. Spatial and temporal downsampling in event-based visual classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 5030-5044
- [85] Keshtkaran M R, Pandarinath C. Enabling hyperparameter optimization in sequential autoencoders for spiking neural data//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 15 911-15 921
- [86] Gehrig D, Loquercio A, Derpanis K G, et al. End-to-end learning of representations for asynchronous event-based data//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea, 2019: 5633-5643
- [87] Schraml S, Belbachir A N, Bischof H. An event-driven stereo system for real-time 3D 360° panoramic vision. *IEEE Transactions on Industrial Electronics*, 2015, 63(1): 418-428
- [88] Huang J, Wang S, Guo M, et al. Event-guided structured output tracking of fast-moving objects using a CeleX sensor. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28(9): 2431-2417
- [89] Gehrig D, Rebecq H, Gallego G, et al. Asynchronous, photometric feature tracking using events and frames//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 750-765
- [90] Gehrig D, Rebecq H, Gallego G, et al. EKLt: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, 2020, 128(3): 601-618
- [91] Chen N F Y. Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA, 2018: 644-653
- [92] Jiang Z, Xia P, et al. Mixed frame-/event-driven fast pedestrian detection//Proceedings of the IEEE International Conference on Robotics and Automation. Montreal, Canada, 2019: 8332-8338
- [93] Zhu A Z, Yuan L. EV-FlowNet: Self-supervised optical flow estimation for event-based cameras//Proceedings of the Robotics, Science and Systems. Pittsburgh, USA, 2018: 1-9
- [94] Wang L, Ho Y S, Yoon K J. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 100081-100090
- [95] Alonso I, Murillo A C. EV-SegNet: Semantic segmentation for event-based cameras//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019: 1624-1633
- [96] Cannici M, Cicoone M, Romanoni A, et al. Attention mechanism for object recognition with event-based cameras//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Hawaii, USA, 2019: 1127-1136
- [97] Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91-110
- [98] Remesh B, Yang H, Orchard G M, et al. DART: Distribution aware retinal transform for event-based cameras. *IEEE Pattern Analysis and Machine Intelligence*, 2019
- [99] Lin S, Xu F, Wang X, et al. Efficient spatial-temporal normalization of SAE representation for event camera. *IEEE Robotics and Automation Letters*, 2020, 5(3): 4265-4272
- [100] Zou D, Shi F, Liu W H, et al. Robust dense depth maps generations from sparse DVS stereos//Proceedings of the British Machine Vision Conference. London, UK, 2017
- [101] Alzugaray I, Chli M. ACE: An efficient asynchronous corner tracker for event cameras//Proceedings of the IEEE International Conference on 3D Vision. Québec, Canada, 2018: 653-661
- [102] Alzugaray I, Chi M. Asynchronous corner detection and tracking for event cameras in real time. *IEEE Robotics and Automation Letters*, 2018, 3(4): 3177-3184
- [103] Manderscheid J, Sironi A, Bourdis N, et al. Speed invariant time surface for learning to detect corner points with event-based cameras//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 10 245-10 254
- [104] Serrano-Gotarredona R, Oster M, Lichtsteiner P, et al. CAVIAR: A 45 k neuron, 5 M synapse, 12 G connects/s AER hardware sensory-processing-learning-actuating system for high-speed visual object recognition and tracking. *IEEE Transactions on Neural Networks*, 2009, 20(9): 1417-1438
- [105] Akolkar H, Panzeri S, Bartolozzi C. Spike time based unsupervised learning of receptive fields for event-driven vision//Proceedings of the IEEE International Conference on Robotics and Automation. Seattle, USA, 2015: 4258-4264
- [106] Peng X, Zhao B, Yan R, et al. Bag of events: An efficient probability-based feature extraction method for AER image sensors. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 28(4): 791-803

- [107] Lagorce X, Orchard G, Galluppi F, et al. HOTS: A hierarchy of event-based time-surfaces for pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(7): 1346-1359
- [108] Sironi A, Brambilla M, Bourdis N, et al. HATS: Histograms of averaged time surfaces for robust event-based object classification//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018; 1731-1740
- [109] Scheerlinck C, Barnes N, Mahony R. Asynchronous spatial image convolutions for event cameras. *IEEE Robotics and Automation Letters*, 2019, 4(2): 816-822
- [110] Zhu A Z, Yuan L, Chaney K, et al. Unsupervised learning event-based learning of optical flow, depth, and egomotion //*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019; 989-997
- [111] Rebecq H, Ranftl R, Koltun V, et al. Events-to-video: Bringing modern computer vision to event cameras//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019; 3857-3866
- [112] Sekikawa Y, Ishikawa K, Saito H. Constant velocity 3D convolution. *IEEE Access*, 2018, 6: 76490-76501
- [113] Tulyakov S, Fleuret F, Kiefel M, et al. Learning an event sequence embedding for dense event-based deep stereo//*Proceedings of the IEEE International Conference on Computer Vision*. Seoul, Korea, 2019; 1527-1537
- [114] Wang Q, Zhang Y, Yuan J, et al. Space-time event clouds for gesture recognition; From RGB cameras to event cameras //*Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. Hawaii, USA, 2019; 1825-1835
- [115] Yang J, Zhang Q, Ni B, et al. Modeling point clouds with self-attention and Gumbel subset sampling//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019; 3323-3332
- [116] Maass W. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 1997, 10(9): 1659-1671
- [117] Zhao B, Ding R, Chen S, et al. Feedforward categorization on AER motion events using cortex-like features in a spiking neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 26(9): 1693-1978
- [118] Wu Y, Deng L, Li G, et al. Direct training for spiking neural networks; Faster, larger, better//*Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019, 33; 1311-1318
- [119] Osswald M, Ieng S H, Benosman R, et al. A spiking neural network model of 3D perception for event-based neuromorphic vision systems. *Scientific Reports*, 2017, 7: 40703
- [120] Haessig G, Berthelon X, Ieng S H, et al. A spiking neural network model of depth from defocus for event-based neuromorphic vision. *Scientific Reports*, 2019, 9(1): 3744
- [121] Acharya J, Padala V, Basu A. Spiking neural network based region proposal networks for neuromorphic vision sensors//*Proceedings of the IEEE International Symposium on Circuits and Systems*. Sapporo, Japan, 2019; 1-5
- [122] Bing Z, Meschede C, Huang K, et al. End to end learning of spiking neural network based on R-STDP for a lane keeping vehicle//*Proceedings of the IEEE International Conference on Robotics and Automation*. Brisbane, Australia, 2018; 1-8
- [123] Wozniak S, Pantazi A, Bohnstingl T, et al. Deep learning incorporating biological inspired neural dynamics and in-memory computing. *Nature Machine Intelligence*, 2020, 2: 325-336
- [124] Gilra A, Gerstner W. Non-linear motor control by local learning in spiking neural network//*Proceedings of the International Conference on Machine Learning*. Vienna, Austria, 2018; 1768-1777
- [125] Huh D, Sejnowski T J. Gradient descent for spiking neural networks//*Proceedings of the Advances in Neural Information Processing Systems*. Montreal, Canada, 2018; 1433-1443
- [126] Jin Y, Zhang W, Li P. Hybrid macro/micro level backpropagation for training deep spiking neural networks//*Proceedings of the Advances in Neural Information Processing Systems*. Montreal, Canada, 2018; 7005-7015
- [127] Zhang W, Li P. Spike-train level backpropagation for training deep recurrent spiking neural networks//*Proceedings of the Advances in Neural Information Processing Systems*. Vancouver, Canada, 2019; 7800-7811
- [128] Neftci E O, Mostafa H, Zenke F. Surrogate gradient learning in spiking neural networks; Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 2019, 36(6): 51-63
- [129] Saunders D J, Patel D, Hazan H, et al. Locally connected spiking neural networks for unsupervised feature learning. *Neural Networks*, 2019, 119: 332-340
- [130] Tavanaei A, Ghodrati M, Kheradpishep S R, et al. Deep learning in spiking neural networks. *Neural Networks*, 2019, 111: 47-63
- [131] Jang H, Simeone O, Gardner B, et al. An introduction to probabilistic spiking neural network; Probabilistic models, learning rules, and applications. *IEEE Signal Processing Magazine*, 2019, 36(6): 64-77
- [132] Mueggler E, Rebecq H, Gallego G, et al. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *The International Journal of Robotics and Research*, 2017, 36(2): 142-149
- [133] Rebecq H, Gehrig D, Scaramuzza D. ESIM: An open event camera simulator//*Proceedings of the Conference on Robot Learning*. Zurich, Switzerland, 2018; 969-982

- [134] Li W, Saeedi S, McCormac J, et al. InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset// Proceedings of the British Machine Vision Conference. Newcastle, UK, 2018
- [135] Orchard G, Jayawant A, Cohen G K, et al. Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in Neuroscience*, 2015, 8: 437
- [136] Serrano-Gotarredona T, Linares-Barranco B. Poker-DVS and MNIST-DVS. Their history, how they were made, and other details. *Frontiers in Neuroscience*, 2015, 9: 481
- [137] Li H, Liu H, Ji X, et al. CIFAR10-DVS: An event-stream dataset for object classification. *Frontiers in Neuroscience*, 2017, 11: 309
- [138] Miao S, Chen G, Ning X, et al. Neuromorphic benchmark datasets for pedestrian detection, action recognition, and fall detection. *Frontiers in Neuroscience*, 2019, 13: 38
- [139] Calabrese E, Taverni G, Awai Easthope C, et al. DHP19: Dynamic vision sensor 3D human pose dataset//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019: 1695-1704
- [140] Binas J, Neil D, Liu S C, et al. DDD17: End-to-end DAVIS driving dataset//Proceedings of the International Conference on Machine Learning. New York City, USA, 2017: 1-5
- [141] Scheerlinck C, Barnes N, Mahony R. Continuous-time intensity estimation using event cameras//Proceedings of the Asian Conference on Computer Vision. Perth, Australia, 2018: 308-324
- [142] Scheerlinck C, Rebecq H, Stoffregen T, et al. CED: Color event camera dataset//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. Long Beach, USA, 2019: 1684-1693
- [143] Hu Y, Liu H, Pfeiffer M, et al. DVS benchmark datasets for object tracking, action recognition, and object recognition. *Frontiers in Neuroscience*, 2016, 10: 45
- [144] Mitrokhin A, Fermüller C, Parameshwara C, et al. Event-based moving object detection and tracking//Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems. Madrid, Spain, 2018: 1-9
- [145] Mitrokhin A, Ye C, Fermüller C, et al. EV-IMO: Motion segmentation dataset and learning pipeline for event cameras//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China, 2019: 6105-6122
- [146] Cheng W, Luo H, Yang W, et al. DET: A high-resolution DVS dataset for lane extraction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019: 1666-1675
- [147] Zhu A Z, Thakur D, Ozaslan T, et al. The multi vehicle stereo event camera dataset: An event camera dataset for 3D Perception. *IEEE Robotics and Automation Letters*, 2018, 3(3): 2032-2039
- [148] Leung S, Shamwell E J, Maxey C, et al. Toward a large-scale multimodal event-based dataset for neuromorphic deep learning applications//Proceedings of the Micro-and Nanotechnology Sensors, Systems, and Applications X. Orlando, USA, 2018: 1-10
- [149] Delmerico J, Cieslewski T, Rebecq H, et al. Are we ready for autonomous drone racing? The UZH-FPV drone racing dataset//Proceedings of the IEEE International Conference on Robotics and Automation. Montreal, Canada, 2019: 6713-6719
- [150] Bardow P, Davision A J, Leutenegger S. Simultaneous optical flow and intensity estimation from an event camera// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 884-892
- [151] Barua S, Miyatani Y, Veeraraghavan A. Direct face detection and video reconstruction from event cameras// Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Lake Placid, USA, 2016: 1-9
- [152] Reinbacher C, Graber G, Pock T. Real-time intensity-image reconstruction for event cameras using manifold regularization. *International Journal of Computer Vision*, 2018, 126(12): 1381-1393
- [153] Pan L, Scheerlinck C, Yu X, et al. Bringing a blurry frame alive at high frame-rate with an event camera//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 6820-6829
- [154] Benosman R, Ieng S H, Clercq C et al. Asynchronous frameless event-based optical flow. *Neural Networks*, 2012, 27: 32-37
- [155] Benosman R, Clercq C, Lagorce X, et al. Event-based visual flow. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 25(2): 407-417
- [156] Ieng S H, Carneiro J, Benosman R B. Event-based 3D motion flow estimation using 4D spatio-temporal subspaces properties. *Frontiers in Neuroscience*, 2017, 10: 596
- [157] Orchard G, Etienne-Cummings R. Bioinspired visual motion estimation. *Proceedings of the IEEE*, 2014, 102(10): 1520-1536
- [158] Haessig G, Cassidy A, Alvarez R, et al. Spiking optical flow for event-based sensors using IBM's TrueNorth neurosynaptic system. *IEEE Transactions on Biomedical Circuits and Systems*, 2018, 12(4): 860-870
- [159] Rueckauer B, Delbruck T. Evaluation of event-based algorithm for optical flow with ground-truth from inertial measurement sensor. *Frontiers in Neuroscience*, 2016, 10: 176
- [160] Liu M, Delbruck T. Block-matching optical flow for dynamic vision sensors: Algorithm and FPGA implementation// Proceedings of the IEEE International Symposium on Circuits and Systems. Baltimore, USA, 2017: 1-4

- [161] Liu M, Delbruck T. Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors//Proceedings of the British Machine Vision Conference. Newcastle, UK, 2018
- [162] Gallego G, Rebecq H, Scaramuzza D. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 3867-3876
- [163] Hordijk B J P, Scheper K Y W, Croon G C. Vertical landing for micro air vehicles using event-based optical flow. *Journal of Field Robotics*, 2018, 35(1): 69-90
- [164] Paredes-Valles F, Scheper K Y W, Croon G C. Unsupervised learning of a hierarchical spiking neural network for optical estimation: from events to global motion perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2051-2064
- [165] Huang J, Guo M, Wang S, et al. A motion sensor with on-chip pixel rendering module for optical flow gradient extraction//Proceedings of the IEEE International Symposium on Circuits and Systems. Florence, Italy, 2018: 1-5
- [166] Chen S, Akselrod P, Zhao B, et al. Efficient feedforward categorization of objects and human postures with address-event image sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 34(2): 302-314
- [167] Perez-Carrasco J A, Zhao B, Serrano C, et al. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing — Application to feedforward ConvNets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2706-2719
- [168] O'Connor P, Neil D, Liu S C, et al. Real-time classification and sensor fusion with a spiking deep belief network. *Frontiers in Neuroscience*, 2013, 7: 178
- [169] Ghosh R, Misha A, Orchard G, et al. Real-time object recognition and orientation estimation using an event-based camera and CNN//Proceedings of the IEEE Conference on Biomedical Circuits and Systems. Lausanne, Switzerland, 2014: 544-547
- [170] Pradhan B R, Bethi Y, Narayanan S, et al. N-HAR: A neuromorphic event-based human activity recognition system using memory surfaces //Proceedings of the IEEE International Symposium on Circuits and Systems. Sapporo, Japan, 2019: 1-5
- [171] Lee J H, Delbruck T, Pfeiffer M, et al. Real-time gesture interface based on event-driven processing from stereo silicon retinas. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(12): 2250-2263
- [172] Clady X, Maro J M, et al. A motion-based feature for event-based pattern recognition. *Frontiers in Neuroscience*, 2017, 10: 594
- [173] Akolkar H, Meyer C, Clady X, et al. What can neuromorphic event-driven precise timing add to spike-based pattern recognition? *Neural Computation*, 2015, 27(3): 561-593
- [174] Lagorce X, Ieng S H, Clady X, et al. Spatiotemporal features for asynchronous event-based data. *Frontiers in Neuroscience*, 2015, 9: 46
- [175] Shi C, Li J, Wang Y, et al. Exploiting lightweight statistical learning for event-based vision processing. *IEEE Access*, 2018, 6: 19396-19406
- [176] Li H, Li G, Ji X, et al. Deep representation via convolutional neural network for classification of spatiotemporal event streams. *Neurocomputing*, 2018, 299: 1-9
- [177] Orchard G, Meyer C, Etienne-Cummings R, et al. HFrist: A temporal approach to object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(10): 2028-2040
- [178] Vasco V, Glover A, Mueggler E, et al. Independent motion detection with event-driven cameras//Proceedings of the IEEE International Conference on Advanced Robotics. Hong Kong, China, 2017: 530-536
- [179] Vasco V, Glover A, Bartolozzi C. Fast event-based Harris corner detection exploiting the advantages of event-driven cameras//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon, Korea, 2016: 4144-4149
- [180] Clady X, Ieng S H, Benosman R. Asynchronous event-based corner detection and matching. *Neural Networks*, 2015, 66: 91-106
- [181] Barranco F, Teo C L, et al. Contour detection and characterization for asynchronous event sensors//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 486-494
- [182] Mueggler E, Bartolozzi C, Scaramuzza D. Fast event-based corner detection//Proceedings of the British Machine Vision Conference. London, UK, 2017
- [183] Li R, Shi D, Zhang Y, et al. FA-Harris: A fast and asynchronous corner detector for event cameras//Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems. Macau, China, 2019: 6223-6229
- [184] Seifozzakerini 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//Proceedings of the British Machine Vision Conference. York, UK, 2016
- [185] Valeiras D R, Clady X, et al. Event-based line fitting and segment detection using a neuromorphic visual sensor. *IEEE Transactions on Neural Network and Learning Systems*, 2018, 30(4): 1218-1230
- [186] Lee S, Kim H, Kim H J. Edge detection for event cameras using intra-pixel-area events//Proceedings of the British Machine Vision Conference. Cardiff, UK, 2019

- [187] Liu H, Moeys D P, et al. Combined frame-and event-based detection and tracking//Proceedings of the IEEE International Symposium on Circuits and Systems. Montreal, Canada, 2016: 2511-2514
- [188] Iacono M, Weber S, Glover A, et al. Towards event-driven object detection with off-the-shelf deep learning//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain, 2018: 1-9
- [189] Cannici M, Ciccone M, Romanoni A, et al. Asynchronous convolutional networks for object detection in neuromorphic cameras//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019
- [190] Alzugaray I, Chli M. Asynchronous multi-hypothesis tracking of features with event cameras//Proceedings of the IEEE International Conference on 3D Vision. Fukuoka, Japan, 2019, 269-278
- [191] Zhu A Z, Atanasov N, Daniilidis K. Event-based feature tracking with probabilistic data association//Proceedings of the IEEE International Conference on Robotics and Automation. Singapore, 2017: 4465-4470
- [192] Yuan W, Ramalingam S. Fast localization and tracking using event sensors//Proceedings of the IEEE International Conference on Robotics and Automation. Stockholm, Sweden, 2016: 4564-4571
- [193] Everding L, Conradt J. Low-latency line tracking using event-based dynamic vision sensors. *Frontiers in Neurorobotics*, 2018, 12: 4
- [194] Li K, Shi D, Zhang Y, et al. Feature tracking based on line segments with the dynamic and active-pixel vision sensor. *IEEE Access*, 2019, 7: 110874-110883
- [195] Valeiras D R, Lagorce X, Clady X, et al. An asynchronous neuromorphic event-driven visual part-based shape tracking. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(12): 3045-3059
- [196] Ni Z, Ieng S H, Posch C, et al. Visual tracking using neuromorphic asynchronous event-based cameras. *Neural Computation*, 2015, 27(4): 925-953
- [197] Berthelon X, Chenegros G, Libert N, et al. Full-field OCT technique for high speed event-based optical flow and particle tracking. *Optics Express*, 2017, 25(11): 12611-12621
- [198] Chin T J, Bagchi S, Eriksson A, et al. Star tracking using an event camera//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019: 1646-1655
- [199] Chen H, Wu Q, et al. Asynchronous tracking-by-detection on advance time surfaces for event-based object tracking//Proceedings of the ACM International Conference on Multimedia. Nice, France, 2019: 473-481
- [200] Hinz G, Chen G, Aafaque M, et al. Online multi-object tracking-by-clustering for intelligent transportation system with neuromorphic vision sensor//Proceedings of the Joint German/Austrian Conference on Artificial Intelligence. Dortmund, Germany, 2017: 142-154
- [201] Lagorce X, Meyer C, et al. Asynchronous event-based multikernel algorithm for high-speed visual features tracking. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 26(8): 1710-1720
- [202] Camunas-Mesa L A, Serrano-Gotarredona T, Ieng S H, et al. Event-driven stereo visual tracking algorithm to solve object occlusion. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 29(9): 4223-4237
- [203] Schraml S, Belbachir A N. A spatio-temporal clustering method using real-time motion analysis on event-based 3D vision//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. San Francisco, USA, 2010: 57-63
- [204] Chen G, Gao H, Aafaque M, et al. Neuromorphic vision based multivehicle detection and tracking for intelligent transportation system. *Journal of Advanced Transportation*, 2018: 1-13
- [205] Barranco F, Fermuller C. Real-time clustering and multi-target tracking using event-based sensors//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain, 2018, 5764-5769
- [206] Stoffregen T, Gallego G, Drummond T, et al. Event-based motion segmentation by motion compensation//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea, 2019: 7244-7253
- [207] Weikersdorfer D, Adrian D B, Cremers D, et al. Event-based 3D SLAM with a depth-augmented dynamic vision sensor//Proceedings of the IEEE International Conference on Robotics and Automation. Hong Kong, China, 2014: 359-364
- [208] Schraml S, Nabil Belbachir A, Bischof H. Event-driven stereo matching for real-time 3D panoramic vision//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 466-474
- [209] Piatkowska E, Kogler J, Belbachir N, et al. Improved cooperative stereo matching for dynamic vision sensor with ground truth evaluation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA, 2017: 53-60
- [210] Rogister P, Benosman R, Ieng S H, et al. Asynchronous event-based binocular stereo matching. *IEEE Transactions on Neural Networks and Learning Systems*, 2011, 23(2): 347-353
- [211] Carneiro J, Ieng S H, Posch C, et al. Event-based 3D reconstruction from neuromorphic retinas. *Neural Networks*, 2013, 45: 27-38
- [212] Ieng S H, Garneiro J, Osswald M, et al. Neuromorphic event-based generalized time-based stereovision. *Frontiers in Neuroscience*, 2018, 12: 442
- [213] Zihao Zhu A, Chen Y, Daniilidis K. Realtime time synchronized event-based stereo//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 433-447
- [214] Zou D, Guo P, Wang Q, et al. Context-aware event-driven stereo matching // Proceedings of the IEEE International

- Conference on Image Processing. Phoenix, USA, 2016; 1076-1080
- [215] Xie Z, Zhang J, Wang P. Event-based stereo matching using semiglobal matching. *International Journal of Advanced Robotic Systems*, 2018, 15(1): 1-15
- [216] Firouzi M, Conrard J. Asynchronous event-based cooperative stereo matching using neuromorphic silicon retinas. *Neural Processing Letters*, 2016, 43(2): 311-326
- [217] Andreopoulos A, Kashyap H J, Nayak T K, et al. A low power, high throughput, fully event-based stereo system// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018, 7532-7542
- [218] Martel J N P, Muller J, Conrard J, et al. Live demonstration an active system for depth reconstruction using event-based sensors// *Proceedings of the IEEE International Symposium on Circuits and Systems*. Florence, Italy, 2018: 1
- [219] Rebecq H, Gallego G, et al. EMVS: Event-based multi-view stereo — 3D reconstruction with an event camera in real-time. *International Journal of Computer Vision*, 2018, 126(12): 1394-1414
- [220] Zhou Y, Gallego G, Rebecq H, et al. Semi-dense 3D reconstruction with a stereo event camera// *Proceedings of the European Conference on Computer Vision*. Munich, Germany, 2018: 235-251
- [221] Mueggler E, Huber B, Scaramuzza D. Event-based, 6-DOF pose tracking for high-speed maneuvers// *Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems*. Chicago, USA, 2014: 2761-2768
- [222] Kim H, Leutenegger S, Davision A J. Real-time 3D reconstruction and 6-DOF tracking with an event camera// *Proceedings of the European Conference on Computer Vision*. Amsterdam, The Netherlands, 2016: 349-364
- [223] Kueng B, Mueggler E, Gallego G, et al. Low-latency visual odometry using event-based feature tracks// *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. Daejeon, Korea, 2016: 16-23
- [224] Gallego G, Lund J E A, Mueggler E, et al. Event-based, 6-DOF camera tracking from photometric depth maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(10): 2402-2412
- [225] Rebecq H, Horstschäfer T, Gallego G, et al. EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2016, 2(2): 593-600
- [226] Bryner S, Gallego G, Rebecq H, et al. Event-based, direct camera tracking from a photometric 3D map using nonlinear optimization// *Proceedings of the IEEE International Conference on Robotics and Automation*. Montreal, Canada, 2019: 325-331
- [227] Nguyen A, Do T T, Caldwell D G, et al. Real-time 6DOF pose relocalization for event cameras with stacked spatial LSTM networks// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach, USA, 2019: 1638-1645
- [228] Censi A, Scaramuzza D. Low-latency event-based visual odometry// *Proceedings of the IEEE International Conference on Robotics and Automation*. Hong Kong, China, 2014: 703-710
- [229] Zhu A Z, Atanasov N, Daniilidis K. Event-based visual inertial odometry// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, USA, 2017: 5816-5824
- [230] Mueggler E, Gallego G, Rebecq H, et al. Continuous-time visual-inertial odometry for event cameras. *IEEE Transactions on Robotics*, 2018, 34(6): 1425-1440
- [231] Won J Y, Ryu H, Delbruck T, et al. Proximity sensing based on a dynamic vision sensor for mobile devices. *IEEE Transactions on Industrial Electronics*, 2014, 62(1): 536-544
- [232] Ni Z, Bolopion A, Agnus J, et al. Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics. *IEEE Transactions on Robotics*, 2012, 28(5): 1081-1089
- [233] Belbachir A N, Hofstatter M, Litzberger M, et al. High-speed embedded-object analysis using a dual-line timed-address-event temporal-contrast vision sensor. *IEEE Transactions on Industrial Electronics*, 2010, 58(3): 770-783
- [234] Vanarse A, Osseiran A, Rassau A. Neuromorphic engineering— A paradigm shift for feature IM technologies. *IEEE Instrumentation & Measurement Magazine*, 2019, 22(2): 4-9
- [235] Chen G, Bing Z, Rohrbein F, et al. Toward brain-inspired learning with the neuromorphic snake-like robot and the neuromorphic platform. *IEEE Transactions on Cognitive and Development Systems*, 2017, 11(1): 1-12
- [236] Rast A D, Adams S V, Davidson S, et al. Behavioral learning in a cognitive neuromorphic robot: an integrative approach. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(12): 6132-6144
- [237] Falanga D, Kim S, Scaramuzza D. How fast is too fast? The role of perception latency in high-speed sense and avoid. *IEEE Robotics and Automation Letters*, 2019, 4(2): 1884-1891
- [238] Falanga D, Kleber K, Scaramuzza D. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 2020, 5(40): 1-14
- [239] Hsu T H, Chen Y K, Wu J S, et al. A 0.8 V multimode vision sensor for motion and saliency detection with Ping-Pong PWM pixel// *Proceedings of the IEEE International Conference on Solid-State Circuits Conference*. San Francisco, USA, 2020: 110-112
- [240] Gehrig D, Gehrig M, Hidalgo-Carrió J, et al. Video to events: Recycling video datasets for event cameras// *Proceedings of the IEEE Conference on Computer Vision and Pattern*. Seattle, USA, 2020
- [241] Hazan H, Saunders D J, Khan H, et al. Bindnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in Neuroinformatics*, 2018, 12(89): 1-18

- [242] Zhou F, Zhou Z, Chen J, et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nature Nanotechnology*, 2019, 14(8): 776-782
- [243] Smith V, Chiang C K, Sanjabi M, et al. Federated multi-task learning//*Proceedings of the Advances in Neural Information Processing Systems*. Long Beach, USA, 2017: 4224-4434
- [244] Yang M, Chien C H, Delbruck T, et al. A 0.5V 55 μm 64 \times 2 channel binaural silicon cochlea for event-driven stereo-audio sensing. *IEEE Journal of Solid-State Circuits*, 2016, 51(11): 2554-2569
- [245] Bartolozzi C. Neuromorphic circuits impart a sense of touch. *Science*, 2018, 360(6392): 966-967
- [246] 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. *IEEE Robotics and Automation Letters*, 2018, 3(2): 994-1001
- [247] Eshraghian J K, Baek S, Levi T, et al. Nonlinear retinal response modeling for future neuromorphic instrumentation. *IEEE Instrumentation & Measurement Magazine*, 2020, 23(1): 21-29
- [248] Yu Z, Liu J K, Jia S, et al. Toward the next generation of retinal neuroprosthesis: Visual computation with spikes. *Engineering*, 2020, 6(4): 449-461
- [249] Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 2019, 575(7784): 607-617



LI Jia-Ning, Ph. D. candidate. His main research interests include spatial-temporal point processes learning, machine learning and neuromorphic vision.

TIAN Yong-Hong, Ph. D., professor. His current research interests include multimedia big data, machine learning and brain-inspired computation.

Background

Neuromorphic vision is an important branch of neuromorphic engineering, which is a concept developed by Caver Mead in the late 1980s, describing the use of the very-large-scale integration (VLSI) systems utilizing electronic analog circuits to mimic the mechanism of the biological visual sensing system. Recently, the emergence of neuromorphic vision sensors (e. g., DVS, ATIS, DAVIS, CeleX, and Vidar) is gaining more and more attention in computer vision due to the advantages over conventional cameras: high temporal resolution, high dynamic range, low redundancy and low power consumption. Furthermore, Neuromorphic vision is also regarded as one of the most important themes towards the next generation of artificial intelligence, and it is also a research hotspot in the field of neuroscience and computer vision. In fact, these novel cameras are bio-inspired vision sensors that, in contrast to frame-based cameras, work in a completely different way: pixels independently respond to intensity changes with asynchronous spatial-temporal spikes, instead of providing frames at a fixed rate. Thus, existing computer vision techniques cannot be directly applied to this new data, and efforts are attempting to address those significant challenges.

In this paper, we provide a comprehensive overview of

the research progress and applications of neuromorphic vision sensors. Then, we next highlight the following challenges consisting of large-scale neuromorphic vision datasets, spike metric, asynchronous feature representation, high-speed computing paradigm, and open-source framework. Finally, we discuss some meaningful future research directions on neuromorphic vision sensors. We believe that this work will open the door to bring the mainstream of computer vision or neuroscience research towards the neuromorphic vision and enlarge the footprint of these new cameras towards a broader array of applications. In the future, we further focus on the two following directions: (1) Investigating the neural inversion computing theory for novel neuromorphic sampling and processing models; (2) Exploring asynchronous event-based learning algorithm towards high-speed neuromorphic vision measurement devices.

This work is partially supported by the National Natural Science Foundation of China under Grant No. 62027804 and No. 61825101, and the National Artificial Intelligence 2.0 Research and Development Strategy from the Chinese Academy of Engineering.