

神经形态视觉传感器及其应用研究

桑永胜¹, 李仁昊¹, 李耀仟¹, 王蔷薇², 毛耀³

(1. 四川大学计算机学院, 四川 成都 610065; 2. 四川旅游学院, 四川 成都 610100;
3. 中国科学院光电技术研究所, 四川 成都 610209)

摘要: 神经形态视觉传感器是一种模拟生物视觉系统工作机理的传感器, 具有高时间分辨率、低时延、低功耗以及高动态范围等特点。首先, 介绍了神经形态工程的研究背景、神经形态芯片以及神经形态视觉传感器的工作机理和主要优点。然后, 详细介绍了神经形态视觉的主要计算方法, 包括概率统计方法、脉冲神经网络方法以及深度神经网络方法等。最后, 综述了神经形态视觉传感器在同步定位与地图构建、图像重建以及特征检测与跟踪等方面的应用研究。对神经形态视觉传感器从硬件、计算方法和应用等方面作了系统概述, 为相关研究者提供了全面的参考。

关键词: 神经形态视觉传感器; 事件相机; 帧相机; 神经网络; 即时定位与地图构建; 图像重构

中图分类号: TP212

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2019.00133

Research on neuromorphic vision sensor and its applications

SANG Yongsheng¹, LI Renhao¹, LI Yaoqian¹, WANG Qiangwei², MAO Yao³

1. College of Computer Science, Sichuan University, Chengdu 610065, China

2. Sichuan Tourism University, Chengdu 610100, China

3. Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China

Abstract: Neuromorphic vision sensor is a biologically inspired artificial neural system that mimics algorithmic behavior of biological vision systems, which has numerous advantages over standard vision sensors, such as high temporal resolution, low latency, low power, high dynamic range, etc. At first, a brief introduction to neuromorphic engineering, neuromorphic chips, and vision sensors was given. Then the main computing methods for neuromorphic vision were reviewed, including probability and statistics, spiking neural network and deep neural network. Finally, several kinds of applications based on neuromorphic vision sensors were given, such as simultaneous localization and mapping (SLAM), image reconstruction (IR), etc. A review of hardware, computing methods and applications for neuromorphic vision sensors was given, which provided a comprehensive reference for researchers.

Key words: neuromorphic vision sensor, event-based camera, frame-based camera, neural network, simultaneous localization and mapping, image reconstruction

1 引言

近年来, 以深度学习为代表的人工智能技术取得突破性进展, 5G 的商用部署逐步加速, 这些新的技术将推动物联网快速发展。作为数据采集的重

要入口, 传感器在未来几年将迎来爆发式的需求增长。随着传感器的大规模应用, 海量数据的处理给算力和能耗都带来了严峻的挑战。在 2016 年著名的人机围棋大战中, Google 公司的 AlphaGo 系统每局围棋博弈的平均耗电费用高达 3 000 美元。在大

收稿日期: 2019-07-22; 修回日期: 2019-10-31

通信作者: 毛耀, yoyo7802@aliyun.com

基金项目: 国家自然科学基金资助项目 (No.U1831121); 四川省重大科技专项 (No.2019ZDZX0006, No.2018GZDZX0028)

Foundation Items: The National Natural Science Foundation of China (No. U1831121), The Key S&T Special Projects of Sichuan (No.2019ZDZX0006, No.2018GZDZX0028)

数据背景下,算力与能耗已成为阻碍智能应用进一步发展的瓶颈。

人类感知系统和大脑经过数十亿年的进化,具有高效、并行以及低功耗等优点,能够可靠地感知各种复杂场景,高效地处理大量数据。类脑科学是受脑神经与人类认知行为机制启发的,以相关模型为桥梁的,结合计算机软/硬件实现的机器智能^[1]。其中,以人脑神经元结构和人脑感知方式为参考进行建模的神经形态学(neuromorphic)无疑是类脑科学中的重要方向之一。“神经形态”这一术语是20世纪80年代由加州理工学院教授 Mead 首次提出的^[2],他提出了以动态变化突触为基本单元的神经形态自适应模拟系统。传统的冯·诺依曼架构系统采用高速时钟频率、同步逻辑与精确匹配硬件,而神经形态系统则模拟生物神经系统的低速信号和非均匀成分,并在其设计中利用噪声作为特征,神经形态系统的容错能力更强、功耗更低^[3]。基于此开创性发现,神经形态芯片与传感器被认为是通向通用人工智能的重要道路^[4-5],各国学者在该领域进行研究并取得了不同程度的进展。

2 神经形态硬件发展现状

业界主流神经形态结构芯片可按实现方式分为3类:基于传统硬件的神经形态电路系统、基于忆阻器实现的系统和基于有机半导体实现的系统。在第一类芯片中,大多数神经和突触回路都是由无机(硅基)电子元件构成,如斯坦福大学研制的可编程神经形态芯片 Neurogrid^[6]和 IBM 公司于2014年发布的低功耗、高性能神经形态芯片 TrueNorth^[7]。IBM 公司的 TrueNorth 在视频行为模式识别等任务上的能效可达传统 CPU 芯片的数百倍。在第二类芯片中,忆阻器因其断电后仍保留值的特性被研究者用于实现芯片网络中的记忆突触。IBM 公司神经形态小组利用一种名为相变存储器的忆阻器在手写数字数据集上测试了反向传播算法,其精度可达83%,与纯软件实现精度相当^[2];亚利桑那州立大学与清华大学的一个研究小组也成功利用数字转换器读出电路实现了大面积、低时延、低功耗的神经形态芯片^[8]。第三类芯片主要是通过诸如金属离子电迁移^[9]、电荷捕获^[10-11]、氧化还原反应^[12-13]等实现突触的加权功能,例如 Nawrocki 等^[14]利用人工有机记忆突触实现了名为聚合物神经形态电路的装置,其已具有基本的模式分类能力。此外,东

芝公司通过开发时域神经网络(TDNN, time domain neural network)进行物联网和大数据分析,采用了时域模拟和数字混合信号处理(TDAMS3)技术,实现了处理单元的小型化和超低功耗神经形态半导体电路,这使得低功耗物联网设备也能使用深度学习。

自从 Mead 提出可以利用晶体管的模拟物理特性来设计诸如硅视网膜和耳蜗等自适应、低功耗传感器以来,大量神经形态模型被学者应用于视觉、听觉、嗅觉等领域的传感器研发过程中^[15]。在视觉领域,研究者们已应用硅视网膜方法克服了传统帧相机高信息冗余与动态模糊等缺点,研制出了高精度便携传感器,相关内容将在第3节进行详细介绍。在听觉领域,由于传统高分辨率数模转换器和听觉帧的数字处理开销过大,Lyon 与 Mead 等^[16]提出了一种使用模拟超大规模集成电路(aVLSI, analogue very large scale integration)模拟人类耳蜗的听觉传感器。通过引入“重叠级联耳蜗”概念,Watts 等^[17]于1992年对上述模型做出了一定改进。基于前人对硅耳蜗的研究,Sarpeshkar 等^[18]为仿生耳蜗开发了一种可利用100 mA 电池连续工作数年的超低功耗听觉处理器,其可为便携式语音识别系统提供算力;另一方面,基于交流输出峰值技术地址事件表示(AER, address event representation),Chan 等^[19]通过级联低通滤波模拟基底膜的生物物理特性开发了一种带有 AER 接口的听觉传感器,Liu 等^[20]改进了现有的 AER 听觉传感器的大部分缺点并研制了精度更高的听觉传感器。

3 神经形态视觉传感器

传统的帧相机(frame-based camera)以预定的帧速率重复扫描整个场景,并按照固定间隔输出静态帧序列,与场景中实际观察到的任何目标活动无关。因此,该类相机的反应速度通常受到帧频限制,并且输出的连续视频帧具有高度冗余性,从而造成存储空间、计算能力和时间上的浪费。受人类视网膜原理的启发,研究者们设计出一种神经形态视觉传感器(NVS, neuromorphic vision sensor)。与传统相机以固定帧速率连续输出整张图像不同,NVS 相机仅在它捕获到场景中的瞬态变化时,才会产生事件(event)输出,这类相机也被称作事件相机(event-based camera)。帧相机与事件相机输出的对比如图1所示。传统的帧相机输出为整个视场的全

部信息,如图 1(a)所示,而事件相机只捕捉场景中运动的手臂,如图 1(b)所示。

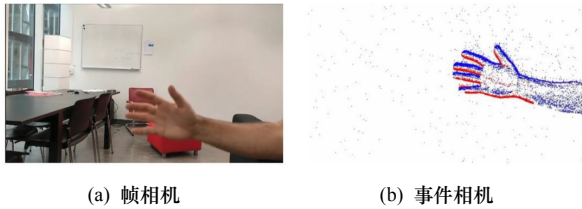


图 1 帧相机与事件相机输出的对比

事件相机中每个像素点只响应像素强度的变化,这种变化主要来自于目标对象或传感器运动使得相同像素位置观察到场景中不同亮度区域而引起的高度变化^[21]。当多个像素点同时请求事件输出时,这些事件将以亚微秒级时延异步输出并构成事件流。帧相机与事件相机信息输出对比如图 2 所示,在相同时间间隔内,传统帧相机(图 2(a))只能输出一幅画面信息,而事件相机(图 2(b))则输出低时延异步事件流。旋转圆盘场景下帧相机与事件相机输出对比如图 3 所示,展示了传统帧相机(图 3(a))与事件相机(图 3(b))对高速转动圆盘上固定点进行一段时间跟踪拍摄得到的结果。图 3(a)显示传统帧相机以固定频率记录整个圆盘信息,信息冗余度高,但帧间丢失了跟踪点的信息。而图 3(b)的事件相机则连续记录跟踪点对应的运动事件,以异步的方式记录有用的运动信息,具有低时延、低带宽需求的特点。

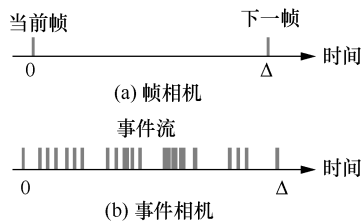


图 2 帧相机与事件相机信息输出对比

基于上述原理性区别,事件相机主要具有如下优点。

1) 高时间分辨率。事件相机的目标是检测亮度变化,其响应速度很快,事件的输出频率可以高达 1 MHz,远远优于传统帧相机 100~1 000 fps 的帧率。因此,被检测到的事件的时间戳分辨率可达微秒级。利用这种特性,事件相机可以捕捉高速运动,而不会受到典型的帧相机运动模糊的影响。

2) 低时延。事件相机每个像素独立工作,不需要等待帧的全局曝光时间。任意像素只要检测到亮

度事件变化,就传事件。因此,事件相机具有最小的时延,实验测试中时延约为 10 μs,在真实应用中可达亚毫秒级。

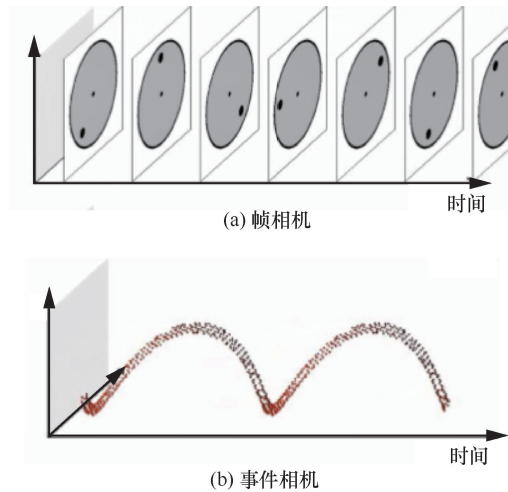


图 3 旋转圆盘场景下帧相机与事件相机输出对比

3) 低功耗。事件相机只传输亮度变化,从而避免了大量冗余数据的传输,因此能耗仅用于处理变化的像素。事件相机的数据传输速率和带宽消耗可以随场景内事件发生的多少而实时变化,具有按需输出的特性和节能优势^[22-23]。传统帧相机数据带宽需求约为 32 Mbit/s,事件相机平均仅为 1 Mbit/s。大多数事件相机的功耗约在 10 mW 级,而有部分相机原型的功耗甚至小于 10 μW。基于事件相机构建的嵌入式系统,其系统整体功耗一般约为 100 mW,甚至更低。

4) 高动态范围。事件相机的动态范围高达 140 dB,远远优于 60 dB 的帧相机。这使得事件相机既能在光照条件良好的白天工作,也能在光线较暗的夜晚采集视场中的动态信息。这是由于事件相机每个像素的光感受器以对数方式独立工作,而非全局快门工作模式。因此,事件相机具有与生物视网膜相似的特性,其像素可以适应非常暗和非常亮的感光刺激。

事件相机的上述优势为其在实时运动分析、粒子追踪、高速移动机器人和无人机等特定应用中提供了新颖有效的解决方案并取得了成功。下面介绍 3 种业界主流的神经形态视觉传感器方案:动态视觉传感器(DVS, dynamic vision sensor)、基于异步时间的图像传感器(ATIS, asynchronous time based image sensor)与 CeleX。

3.1 DVS

Tobi 等^[24]于 1991 年首次完成一种基于帧的硅

视网膜设计,在该设计中,连续事件感光容器被电容耦合到一个每次采集时像素都会重置的输出电路上,此方法启发了 DVS 雏形。每个像素根据其检测到的对数强度信号的变化率调整增量调制器采样率,信号再以亚毫秒的时延传输,这样就获得了超高时间分辨率^[25]。另一方面,为满足真实应用场景对静态输出与绝对亮度的需求,研究者将相同像素的主动式像素传感器(APS, active pixel sensor)与 DVS 结合研制出动态与主动式像素视觉传感器(DAVIS, dynamic and active vision sensor)。该事件相机的优势是像素尺寸较小,其源于基于共享光电二极管的输出电路只占用约 5%的 DVS 像素面积。然而由于 APS 的动态范围有限,DAVIS 仍具有一定局限性。由 iniVation 供应的系列最新款 DAVIS346 视觉传感器拥有 346 dpi×260 dpi 的分辨率,最高达 120 dB 的动态范围与 18.5 μm^2 的像素面积。

3.2 ATIS

Posch 等^[26]提出了一种基于异步时间、全时冗余抑制的视觉传感器设计方法,该方法具有低功耗、高动态范围的特点。这种设计的像素结构将光测量装置与异步相对光照变化检测器结合,利用时域相关的双采样技术和分布式的像素重置方案,弥补了以往基于时间的同步成像的不足。另外,ATIS 利用控制与复位电路可允许选择独立运行的异步或并发运行的同步图像采集模式,其缺点是相较 DVS 像素面积较大,因此同尺寸下分辨率较低。由 Chronocam 供应的系列最新款 ATIS 视觉传感器拥有 304 dpi×240 dpi 的分辨率,最高达 143 dB 的动态范围与 30 μm^2 的像素面积。

3.3 CeleX

CeleX 是上海 CelePixel 公司基于 DVS 针对机器视觉设计的智能图像传感器系列。传感器中的每一像素点能够独立自主地监测相对光强的变化,并在到达阈值时被激发读出信号。行和列仲裁电路实时处理像素激发信号,并确保即使同时接收到多个请求仍能按有序方式逐一处理。传感器依据被激发的事件,输出连续的异步数据流,因此传感器监测的运动物体速度不再受传统的曝光时间和帧速率限制。它可以侦测昂贵的高速相机才能获取到的、帧率高达万帧/s 的高速物体的运动信息,且能够大幅降低后端数据处理量,结合相关配套软件可实现视频高回放帧率的高速运动物体信息捕捉。另一方

面,该传感器能并行输出 3 种模式数据:传统图像帧、动态数据流和全幅光流数据。在结合定向设计的跟踪识别算法后,可在相关领域进行应用,如辅助/自动驾驶、UAV、机器人、监控等。系列最新款 CeleX-V 视觉传感器已拥有可满足自动驾驶领域的 1 280 dpi×800 dpi 的分辨率。

除了上述 3 种神经形态视觉传感芯片,工业界的三星公司也加入了神经形态视觉传感芯片的研制,并于 2017 年推出了 DVS-Gen2,其分辨率达到 640 dpi×480 dpi,并开展了创新性的视觉应用研究。

4 神经形态视觉计算方法研究

随着事件相机在诸多领域的初步成功应用,针对事件流分析与识别的相关问题也在学术界不断获得关注,并亟待解决。由于每个事件在其发生时只产生了非常少的信息量,并且事件型数据不包含图像中常见的灰度特征信息,因此传统机器视觉算法无法直接应用于事件流数据的处理中。

目前,针对事件驱动型相机进行对象识别的方法主要有两类:一类基于统计概率论,另一类基于深度学习理论。基于统计概率论的方法大多在特征袋模型和支持向量机等分类器的基础上,通过改进相关局部特征描述符来处理事件相机的输出^[27-32]。尽管这些方法在一定程度上可以成功地对事件流进行处理,但人工提取的底层特征通常会造空间方面的信息丢失,并且其采用固定事件数量的特征袋方法难以捕捉和刻画事件序列本身隐含的模式特征,所以对复杂和具有相似类别的事件序列模式分类性能并不理想^[33-34]。此外,目前很多视觉描述符模仿人眼编码方式而采用对数极坐标网格,但其在本质上并不符合人类视觉识别能够快速定位重要目标区域并进行集中细致处理的特征^[35],且通常具有计算复杂度高、无仿射不变性和视点变化下错配率高等缺点^[36]。

近年来,深度学习方法的不断发展为事件流序列分析带来新的机遇。由于其在目前任何机器学习基准中都具有先进的性能^[37-38],并且在一些情况下它们已经超过了人类识别的准确率^[39-40],因此当前许多前沿事件流识别算法都结合了深度学习方法。

4.1 基于统计概率论的研究

在统计学习领域,受到传统基于帧的计算机视觉领域最新研究成果的启发,针对事件相机输出的

异步事件流,设计一种高效可用的视觉描述符来进行对象识别和跟踪的研究趋势也正在增加。研究人员通常模仿人类视觉系统的编码方式,在对数极坐标网格的基础上构建描述符^[41],随后通过特征袋模型进行特征提取,最后配合分类器完成整个事件流的对象识别。其中,Lagorce等^[42]提出一种时间曲面描述符的概念,即一种用于提取时空特征的时间导向性方法,并用一个五元组(时空坐标、极性和描述符)表示每一个事件,然而这些时空特征取决于目标对象运动的方向和速度,从而造成特征提取对目标运动的依赖性;而Saeed等^[43]提出的波纹池网络可以对二维图像进行一种变换,将其转换为一维时序模式,但缺陷在于需要将重要的对象进行精确地中心化;针对之前分类方法存在的缺陷,Ramesh等^[44]提出了一种分布式感知视网膜转换描述符。他们使用一种固定的时空对数极坐标网格来获取描述符,并将网格中心定于即将到来的事件上,同时选择事件数量作为特征提取的决定性因素,但这种方法对复杂类别和相似对象的识别准确性较低,在CIFAR10-DVS和N-Caltech101上分别仅获得65.43%和70.33%的准确率。此外,Peng等^[32]从新的视野角度提出一种名为事件袋(BOE, bag of event)的在线特征提取方法,用于对每个事件流记录进行表示,并通过与支持向量机结合来实现对事件流记录的目标分类。该方法在复杂数据集上的分类准确性同样较低,在MNIST-DVS数据集上的平均准确率仅为75.09%,在CIFAR10-DVS数据集上的准确率仅为24.21%。因此到目前为止,以上述概率统计方法为基础构建的事件相机对象识别体系,难以进行后续复杂场景中的运动目标检测、跟踪等任务。

4.2 基于神经网络方法的研究

近年随着深度学习在各种机器学习任务中获得巨大成功^[38-41],越来越多的研究人员开始关注如何利用深度学习网络对事件相机的输出进行对象识别。然而,由于事件相机产生的异步事件流不能直接适应现有的网络模型,所以深度学习在事件相机等神经形态领域的探索应用仍处于早期阶段。针对事件流序列分析识别所设计的神经网络方法,下面分别从脉冲神经网络(SNN, spiking neuron network)和深度神经网络两个类别介绍。

4.2.1 基于SNN的研究

SNN是为解决传统人工神经网络(ANN,

artificial neural network)所需计算量过大与人工神经元的仿生性不足等问题提出的新一代神经网络。其采用诸如IF、LIF等生物神经元模型代替传统ANN的人工神经元,利用特定编码方法将数据转换成脉冲传递信息。另外在SNN中每个神经元独立工作,即没有收到输入信息的神经元不会工作,因此其能耗相较传统ANN更低。SNN被认为是未来传统神经网络的替代方法之一^[45-48],并且通过与事件驱动型传感器结合,可以进一步提高任务处理效率。

神经形态视觉本质上是模拟生物视网膜的工作模式,其事件流的输出与生物神经脉冲十分契合,因此选择脉冲神经网络方法处理事件流具有天然的合理性。早期,Serrano-Gotarredona等^[49]通过实时卷积AER视觉架构,在定制的硬件上使用一种分层脉冲神经网络对不同大小的圆形进行识别和跟踪;随后,Folowosele等^[50]通过在定制的硅硬件上使用一种HMAX启发式算法,进一步区分更多不同的形状。实时手势识别系统被广泛应用于手语识别、虚拟操作、日常辅助、游戏和人机交互等领域,其需要低时延来维持系统平滑性,而传统传感器可能由于帧率不足或动态模糊等问题影响实时手势识别体验,故Amir等^[51]基于IBM公司TrueNorth神经形态芯片开发了利用SNN实现的实时手势识别系统。该系统能在200 mW的功耗下,以105 ms的时延识别一个手势,其在最新数据集DVS Gesture上测试得到的样本外精度高达96.5%。算法流程大致如下:先将时间过滤器级联捕获DVS事件流的一系列快照,之后将串联的快照作为输入特征呈现给使用GPU加速离线训练的卷积层堆栈,再利用名为“赢者通吃”的解码器从输出卷积层找到得分最高的手势,最后通过滑动窗口过滤器清除产生的瞬时手势分类流。

4.2.2 基于深度神经网络的研究

常见的非脉冲类深度神经网络包括卷积神经网络、循环神经网络等,这些网络具有的优势已经在一些研究中得到证明^[52]。在较早的研究中,已经有人考虑到将该类网络与事件相机结合。例如,Farabet等^[53]比较分析基于帧和基于事件的卷积神经网络进行对象识别之间的差异,并且提出一种使用特定硬件实现基于事件的卷积处理方法。其中,最具有意义的工作是Chen等^[27]提出的一种将传统基于帧的数据所训练的卷积神经网络映射为事件

驱动型卷积神经网络的方法，他们通过对旋转的人体轮廓和高速扑克牌符号进行对象识别进一步验证了所提出的方法。此外，Nguyen 等^[54]提出了一种通过事件计数来可视化事件流的方法，并用其完成对长短期记忆网络的训练；Park 等^[55]通过一种基于事件的“时间戳图像”方法，利用像素阵列之间的时空相关性，对基于深度卷积神经网络的远距离手势识别实现了显著的性能改进；Zhu 等^[56]提出了一种新的适应神经网络架构的事件流信息表示形式，可以充分保留事件的时空信息，并为今后深度学习在事件相机输出处理上提供了新的思路。

4.3 事件相机的应用研究

随着神经形态传感器硬件与相应算法的不断发展，研究者已在机器视觉领域的诸多传统应用场景中取得进展。接下来回顾事件相机在图像重建 (IR, image reconstruction)、同步定位与地图构建 (SLAM, simultaneous localization and mapping)、特征检测与跟踪以及面向隐私保护的监控等方向的应用研究成果。

4.3.1 SLAM

SLAM 是实现全自主移动机器人的关键技术，一些研究表明，事件相机技术在 SLAM 应用研究中具有良好前景。Vidal 等^[57]提出了一种融合事件、标准帧和惯性测量状态的估计流水线，通过紧密耦合的方式发挥了传统帧与 DVS 的互补优势。其在公开的事件相机数据集上进行状态估计测试，结果相比仅应用事件管道的精度提高了 130%，相比仅应用标准帧惯性系统的精度提高了 85%。为了解神经形态传感器在极端场景下的相对表现，分别使用搭载传统帧方案和事件帧混合方案的自主四旋翼飞行器在光线情况与飞行器速度不同的 3 种场景下进行特征跟踪实验。结果如图 4 所示，暗光环境下的帧式方案已经无法跟踪特征，而加入了神经形态传感器的方案能随飞行器运动捕捉到一定数量的特征，表现出神经形态传感器高动态范围的特点；明亮高速旋转环境下的帧式方案已经严重受到动态模糊影响，无法捕捉足量特征，而加入了神经形态传感器的方案在此环境下的表现几乎未受速度增大的影响，捕捉到的特征足以完成跟踪任务，表现出神经形态传感器高帧率、低时延的特点。以上实验明显的对比结果让结合了神经形态传感器的飞行器可在传统视觉惯性测距仪无法达到的飞行场景（如低光环境和高动态范围场

景）进行跟踪测距。可以预见，该方案展现的高动态范围、低时延特点，在高速移动机器人方向具有极大的开发应用潜力。

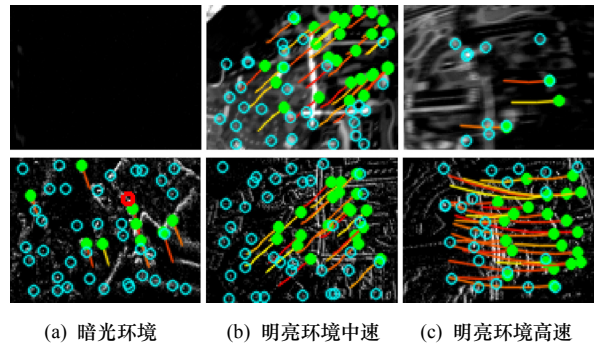


图 4 各种环境下的特征跟踪示例（第一行是基于标准帧的方案，第二行是基于混合事件帧的方案。每列对应相同时间戳，绿色实心点代表持久性特征，蓝色点对应候选特征，彩色线条为跟踪轨迹）

4.3.2 IR

Rebecq 等^[58]开发了一种基于事件的视觉测距算法 (EVO, event-based visual odometry)，该算法利用事件摄像机的独特性质，每秒输出高达数百个姿势估计值，以高精度（小于 0.2% 的位置误差和小于 3% 的方向误差）重建半密集环境 3D 地图。得益于事件相机的性质，该算法不受运动模糊影响，且在具有强烈照明变化的高动态范围环境下运行良好。不同方式的 IR 如图 5 所示，借助 EVO 得到的位置估计可用于室外逆光环境下的 IR，可以看到，利用事件相机所摄图像进行 IR 得到的样张远比传统帧相机在相同位置摄制的图像清晰，利用事件相机特性开发的算法在 IR 方向展现出惊人的潜力。

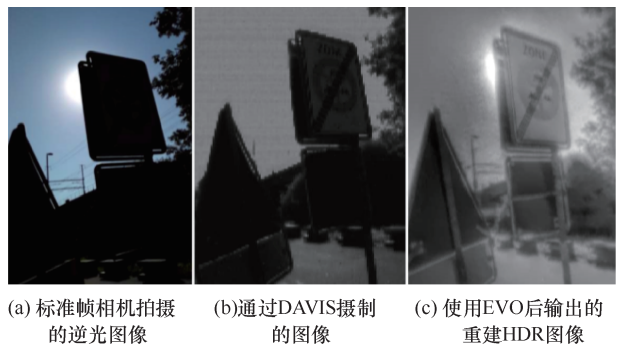


图 5 不同方式的 IR

4.3.3 特征检测与跟踪

特征检测与跟踪一直是传统视觉领域中解决方案众多的高热度应用，但在暗光或超高速移动等严苛条件下，传统视觉跟踪方案具有明显局限性。

事件相机独特机制产生的优势恰恰弥补了这些缺点，因此许多研究者正致力于开发基于事件流的特征检测与跟踪方法。

Ni 等^[59]开发了一种使用异步 AER 相机的新型高速视觉系统。利用此框架，研究者可根据霍夫圆变换对移动微粒产生的事件流进行跟踪，该技术能在几千赫兹的频率下以低计算成本进行基于实时事件的稳健多目标位置检测。经过试验，该方法也可用于高精度高速布朗运动检测。该框架展示出的良好性能满足微机器人系统远程操作对极快视觉反馈的需求，因此是微观测量的一种非常有前景的解决方案。连续霍夫变换跟踪轨迹、真实移动轨迹、基于帧的霍夫跟踪轨迹对比如图 6 所示。由图 6 可以看到，基于事件流的连续霍夫变换跟踪相较传统帧霍夫跟踪更贴近真实跟踪轨迹，也具有更小的抖动率，即更高的稳定性，展现了动态视觉传感器的独特运作机制在粒子追踪方向的优势。

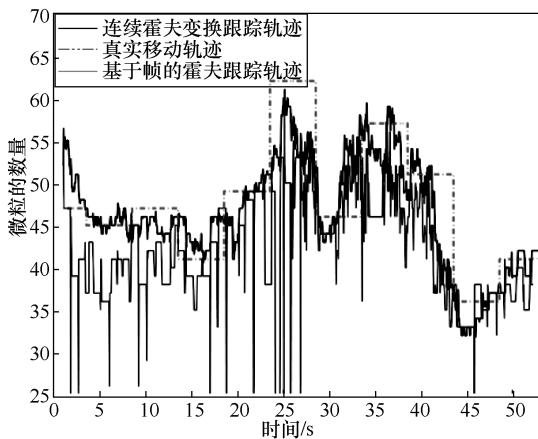


图6 连续霍夫变换跟踪轨迹、真实移动轨迹、基于帧的霍夫跟踪轨迹对比

4.3.4 隐私保护的监控应用

与传统帧相机不同，事件相机只捕捉动态变化，而不会采集颜色等细节信息，对隐私保护要求高

的应用非常适合采用这种相机。对隐私敏感的家庭场景进行了研究测试，对于家庭环境中老人的常见行为进行监测，对站立、蹲下、坐下、跌倒等行为进行识别，识别精度可达 90%。图 7 给出了家庭场景下的行为检测与识别例子，从左至右分别为蹲下、弯腰、坐下以及行走行为。从该仿真测试可以看出，相机只采集运动像素的位置，而不记录隐私细节，对于用户来讲是一种易于接受的家庭安全监测方案。

5 结束语

神经形态作为类脑科学的重要研究方向已被研究者广泛应用于各个领域，其在机器视觉领域的革命性研究成果——神经形态传感器（事件相机），相较传统相机具有低时延、低功耗、高帧率、高动态范围等优势，因此拥有巨大研究价值与广阔的应用前景。本文在介绍神经形态研究背景与各领域硬件发展现状的基础上，对神经形态视觉的计算方法及其在 SLAM、IR、特征检测与跟踪以及面向隐私保护的监控等应用领域的研究进行了概述。

在传感器与算法发展之外，针对事件相机所衍生的科研生态环境也正在逐步完善。许多基于流行语言开发的事件相机环境接口不断涌现，如 Java 平台的 jAER（如图 8 所示）、C/C++ 平台的 cAER 或是基于 Python 的 pyAER；另一方面，受制于当前事件相机的高昂价格，标准数据集的采集与模拟器开发也对科研工作有所裨益。继 IEEE 机器人与自动化国际会议（ICRA2017）特别增设了第一届基于事件的视觉研讨会后，IEEE 国际计算机视觉与模式识别会议（CVPR2019）举行了第二届基于事件的视觉研讨会。目前，越来越多的科研工作者开始认可并重视神经形态传感器的优势与应用前景。希望本文能让读者对神经形态视觉方向有全面的了解，促进更多的研究者进入神经形态视觉研究领域。

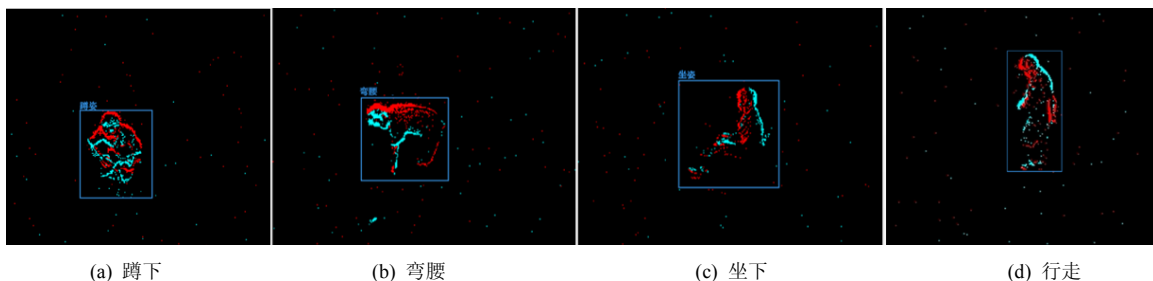


图7 隐私保护的家庭环境行为检测与识别



图8 事件相机环境接口jAER运行样例

参考文献:

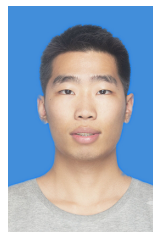
- [1] 陶建华, 陈云霁. 类脑计算芯片与类脑智能机器人发展现状与思考[J]. 中国科学院院刊, 2016, 31(7): 803-811.
TAO J H, CHEN Y J. Current status and consideration on brain-like computing chip and brain-like intelligent robot[J]. Bulletin of Chinese Academy of Sciences, 2016, 31(7): 803-811.
- [2] MEAD C. Neuromorphic electronic systems[J]. Proceedings of the IEEE, 1990, 78(10): 1629-1636.
- [3] NAWROCKI R A, VOYLES R M, SHAHEEN S E. A mini review of neuromorphic architectures and implementations[J]. IEEE Transactions on Electron Devices, 2016, 63(10): 3819-3829.
- [4] BOAHEN K. Neuromorphic microchips[J]. Scientific American, 2005, 292(5): 56-63.
- [5] HASLER J, MARR B. Finding a roadmap to achieve large neuromorphic hardware systems[J]. Frontiers in Neuroscience, 2013, 7: 118.
- [6] GAO P, BENJAMIN B V, BOAHEN K. Dynamical system guided mapping of quantitative neuronal models onto neuromorphic hardware[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2012, 59(10): 2383-2394.
- [7] MEROLLA P A, ARTHUR J V, ALVAREZ-ICAZA R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. Science, 2014, 345(6197): 668-673.
- [8] YU S, CHEN P Y, CAO Y, et al. Scaling-up resistive synaptic arrays for neuro-inspired architecture: challenges and prospect[C]//Electron Devices Meeting. IEEE, 2016.
- [9] MA L, XU Q, YANG Y. Organic nonvolatile memory by controlling the dynamic copper-ion concentration within organic layer[J]. Applied Physics Letters, 2004, 84(24): 4908-4910.
- [10] BLOM P W, SMIT C W, HAVERKORT J E, et al. Carrier capture into a semiconductor quantum well[J]. Physical Review B, Condense Matter, 1993, 47(4): 2072-2081.
- [11] SON D I, YOU C H, KIM W T, et al. Electrical bistabilities and memory mechanisms of organic bistable devices based on colloidal ZnO quantum dot-polymethylmethacrylate polymer nanocomposites[J]. Applied Physics Letters, 2009, 94(13): 132103.
- [12] HA H, KIM O. Electrode-material-dependent switching characteristics of organic nonvolatile memory devices based on Poly (3, 4-ethylenedioxythiophene): poly (styrenesulfonate) film[J]. IEEE Electron Device Letters, 2010, 31(4): 368-370.
- [13] MÜLLER S, PERLOV C, JACKSON W, et al. A polymer/semiconductor write-once read-many-times memory[J]. Nature, 2003, 426(6963): 166-169.
- [14] NAWROCKI R A, VOYLES R M, SHAHEEN S E. Neurons in polymer: hardware neural units based on polymer memristive devices and polymer transistors[J]. IEEE Transactions on Electron Devices, 2014, 61(10): 3513-3519.
- [15] VANARSE A, OSSEIRAN A, RASSAU A. A review of current neuromorphic approaches for vision, auditory, and olfactory sensors[J]. Frontiers in Neuroscience, 2016, 10(275): 115.
- [16] SARPESHKAR R, LYON R F, MEAD C. A low-power wide-dynamic-range analog VLSI cochlea[J]. Analog Integrated Circuits and Signal Processing, 1998, 16(3): 245-274.
- [17] WATTS L, KERNS D A, LYON R F, et al. Improved implementation of the silicon cochlea[J]. IEEE Journal of Solid-State Circuits, 1992, 27(5): 692-700.
- [18] SARPESHKAR R, SALTHOUSE C, SIT J J, et al. An ultra-low-power programmable analog bionic ear processor[J]. IEEE Transactions on Biomedical Engineering, 2005, 52(4): 711-727.
- [19] CHAN V, LIU S C, VAN SCHAIK A. AER EAR: a matched silicon cochlea pair with address event representation interface[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2007, 54(1): 48-59.
- [20] LIU S C, VAN SCHAIK A, MINCH B A, et al. Asynchronous binaural spatial audition sensor with 2×64×4 channel output[J]. IEEE transactions on biomedical circuits and systems, 2013, 8(4): 453-464.
- [21] POSCH C, SERRANO-GOTARREDONA T, LINARES-BARRANCO B, et al. Retinomorph event-based vision sensors: bioinspired cameras with spiking output[J]. Proceedings of the IEEE, 2014, 102(10): 1470-1484.
- [22] LICHTSTEINER P, POSCH C, DELBRUCK T. A 128×128 120dB 15μs latency asynchronous temporal contrast vision sensor[J]. IEEE Journal of Solid-State Circuits, 2008, 43(2): 566-576.
- [23] BERNER R, BRANDLI C, YANG M, et al. A 240×180 10mW 12us latency sparse-output vision sensor for mobile applications[C]//VLSI Circuits (VLSIC), 2013 Symposium on IEEE. IEEE, 2013.
- [24] TOBI D, MEAD C A. Time-derivative adaptive silicon photoreceptor array[J]. Proceedings of SPIE—the International Society for Optical Engineering, 1993, 1541: 92-99.
- [25] GALLEGRO G, DELBRUCK T, ORCHARD G, et al. Event-based vision: a survey[J]. arXiv: Computer Vision and Pattern Recognition, 2019.
- [26] POSCH C, MATOLIN D, WOHLGENANNT R. An asynchronous time-based image sensor[C]//IEEE International Symposium on Circuits & Systems. IEEE, 2008.
- [27] CHEN S, AKSELROD P, ZHAO B, et al. Efficient feedforward categorization of objects and human postures with address-event image sensors[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(2): 302.
- [28] ZHAO B, DING R, CHEN S, et al. Feedforward categorization on AER motion events using cortex-like features in a spiking neural network[J]. IEEE Transactions on Neural Networks, 2015, 26(9): 1963-1978.
- [29] BICHLER O, QUERLIOZ D, THORPE S J, et al. Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity[J]. Neural Networks, 2012, 32(2): 339-348.
- [30] LI H, LI G, SHI L. Classification of spatiotemporal events based on random forest[C]//International Conference on Brain Inspired Cognitive Systems. 2016: 138-148.
- [31] SHAISTA H, ARINDAM B. Multiclass classification by adaptive network of dendritic neurons with binary synapses using structural plasticity[J]. Frontiers in Neuroscience, 2016, 10: 113.
- [32] PENG X, ZHAO B, YAN R, et al. Bag of events: an efficient probability-based feature extraction method for AER image sensors[J]. IEEE

- Transactions on Neural Networks and Learning Systems, 2016: 1-13.
- [33] ZHANG D, LU G. Review of shape representation and description techniques[J]. Pattern recognition, 2004, 37(1): 1-19.
- [34] LI H, LI G, JI X, et al. Deep representation via convolutional neural network for classification of spatiotemporal event streams[J]. Neurocomputing, 2018: S0925231218301486.
- [35] 吴金建. 基于人类视觉系统的图像信息感知和图像质量评价[D]. 西安: 西安电子科技大学, 2014.
- WU J J. Image information perception and quality assessment based on the human visual system[D]. Xi'an: Xidian University, 2014.
- [36] QU X, ZHAO F, ZHOU M, et al. A novel fast and robust binary affine invariant descriptor for image matching[J]. Mathematical Problems in Engineering, 2014, 2014(1): 1-7.
- [37] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436.
- [38] SCHMIDHUBE R, JÜRGE N. Deep learning in neural networks: an overview[J]. Neural Networks, 2015, 61: 85-117.
- [39] CIREŞAN D, MEIER U, SCHMIDHUBER J. Multi-column deep neural networks for image classification[J]. arXiv: Computer Vision and Pattern Recognition, 2012: 1202.2745.
- [40] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on image net classification[C]//IEEE International Conference on Computer Vision. IEEE, 2015: 1026-1034.
- [41] SCHWARTZ E L. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception[J]. Biological Cybernetics, 1977, 25(4): 181-194.
- [42] LAGORCE X, ORCHARD G, GALLUPPI F, et al. Hots: a hierarchy of event-based time-surfaces for pattern recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1346-1359.
- [43] SAEED A, COHEN G K, WANG R M, et al. The ripple pond: enabling spiking networks to see[J]. Frontiers in Neuroscience, 2013, 7: 212.
- [44] RAMESH B, YANG H, ORCHARD G M, et al. DART: distribution aware retinal transform for event-based cameras[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
- [45] GERSTNER W, KISTLER W. Spiking neuron models: an introduction[M]. Oxford: Cambridge University Press, 2002.
- [46] IZHIKEVICH E M, EDELMAN G M. Large-scale model of mammalian thalamocortical systems[J]. Proceedings of the National Academy of Sciences, 2008, 105(9): 3593-3598.
- [47] ANANTHANARAYANAN R, ESSER S K, SIMON H D, et al. The cat is out of the bag: cortical simulations with 10^9 neurons, 10^{13} synapses[C]//Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. IEEE, 2009.
- [48] ELIASMITH C, STEWART T C, CHOO X, et al. A large-scale model of the functioning brain[J]. Science, 2012, 338(6111): 1202-1205.
- [49] SERRANO-GOTARREDONA T, LINARES-BARRANCO B, GALLUPPI F, et al. ConvNets experiments on SpiNNaker[C]//IEEE International Symposium on Circuits and Systems. IEEE, 2015.
- [50] FOLOWOSELE F, VOGELSTEIN R J, ETIENNE-CUMMINGS R. Towards a cortical prosthesis: implementing a spike-based HMAX model of visual object recognition in silico[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2011, 1(4): 516-525.
- [51] AMIR A, TABA B, BERG D, et al. A low power, fully event-based gesture recognition system[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
- [52] EVANGELOS S, DANIEL N, MICHAEL P, et al. Robustness of spiking deep belief networks to noise and reduced bit precision of neuro-inspired hardware platforms[J]. Frontiers in Neuroscience, 2015, 9(00222): 222.
- [53] FARABET C, RAFAEL P, PÉREZ-CARRASCO J, et al. Comparison between frame-constrained fix-pixel-value and frame-free spiking-dynamic-pixel ConvNets for visual processing[J]. Frontiers in Neuroscience, 2012, 6: 32-32.
- [54] NGUYEN A, DO T T, CALDWELL D G, et al. Real-time 6DOF pose relocalization for event cameras with stacked spatial LSTM networks[J]. arXiv: Computer Vision and Pattern Recognition, 2017.
- [55] PARK P K J, CHO B H, PARK J M, et al. Performance improvement of deep learning based gesture recognition using spatiotemporal demosaicing technique[C]//IEEE International Conference on Image Processing. IEEE, 2016.
- [56] ZHU A Z, YUAN L, CHANEY K, et al. EV-FlowNet: self-supervised optical flow estimation for event-based cameras[J]. arXiv: Computer Vision and Pattern Recognition, 2018.
- [57] VIDAL A R, REBECQ H, HORSTSCHAEFER T, et al. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 994-1001.
- [58] REBECQ H, HORSTSCHÄFER T, GALLEGO G, et al. EVO: a geometric approach to event-based 6-DOF parallel tracking and mapping in real time[J]. IEEE Robotics and Automation Letters, 2016, 2(2): 593-600.
- [59] NI Z, PACORET C, BENOSMAN R, et al. Asynchronous event-based high speed vision for microparticle tracking[J]. Journal of Microscopy, 2012, 245(3): 236-244.

[作者简介]



桑永胜(1974-), 男, 四川绵阳人, 四川大学计算机学院副教授, 主要研究方向为类脑智能、机器视觉以及深度学习等。



李仁昊(1997-), 男, 北京人, 四川大学计算金融专业学生, 主要研究方向为计算机视觉、数据挖掘等。

李耀仟(1997-), 男, 广西崇左人, 四川大学计算金融专业学生, 主要研究方向计算机视觉、大数据分析。

王蔷薇(1981-), 男, 四川绵阳人, 四川旅游学院讲师, 主要研究方向为智能数据分析。

毛耀(1978-), 男, 四川眉山人, 中国科学院光电技术研究所研究员, 主要研究方向为光束控制及计算机管理技术, 包括预测跟踪、惯性稳定、自主控制、机器视觉、强化学习等。