论文全译 / REVIEW ARTICLE

事件视觉：综述

Event-based Vision: A Survey

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, Davide Scaramuzza

本文已获《IEEE Transactions on Pattern Analysis and Machine Intelligence》接收，2020 年。© IEEE
arXiv:1904.08405v3［cs.CV］，2020 年 8 月 8 日

摘要

事件相机是一种仿生传感器。与传统帧相机不同，它不以固定速率拍摄图像，而是异步测量各像素的亮度变化，输出的事件流编码了亮度变化的时间、位置和符号。相较于传统相机，事件相机具备高时间分辨率（μs 量级）、极高动态范围（140 dB 对 60 dB）、低功耗和高像素带宽（kHz 量级）；高带宽有助于减少运动模糊。在传统相机难以应对的低延迟、高速度和高动态范围场景中，事件相机因此在机器人与计算机视觉领域很有潜力。然而，要发挥这些优势，必须用新方法处理这类传感器不同寻常的输出。本文全面综述新兴的事件视觉领域，重点介绍事件相机的应用以及为发挥其卓越特性而开发的算法。本文从工作原理、现有器件及所处理的任务介绍事件相机，涉及从低层视觉（特征检测与跟踪、光流等）到高层视觉（重建、分割、识别）的任务。我们还讨论事件处理技术，包括基于学习的技术，并介绍面向这类新型传感器的专用处理器，例如采用脉冲神经网络的处理器。最后，我们指出，要让机器以更高效、更接近生物的方式感知世界并与之交互，还有哪些挑战需要解决，以及有哪些机会值得探索。

关键词：事件相机、仿生视觉、异步传感器、低延迟、高动态范围、低功耗。

目录

1 引言与应用	2	6 资源	34
2 事件相机的工作原理	3	7 讨论	36
3 事件处理	10	8 结论	37
4 算法与应用	15	致谢	37
5 事件系统与应用	32	参考文献	38

作者单位与资助

G. Gallego 就职于德国柏林工业大学。Tobi Delbruck 任职于苏黎世联邦理工学院信息技术与电气工程系，工作地点为苏黎世大学与苏黎世联邦理工学院联合神经信息学研究所（瑞士苏黎世）。Garrick Orchard 就职于美国加利福尼亚州英特尔研究院。Chiara Bartolozzi 就职于意大利热那亚的意大利理工研究院。Brian Taba 就职于美国加利福尼亚州 IBM 研究院。Andrea Censi 任职于瑞士苏黎世联邦理工学院机械与过程工程系。Stefan Leutenegger 与 Andrew Davison 就职于英国伦敦的帝国理工学院。Jörg Conradt 就职于瑞典斯德哥尔摩的 KTH 皇家理工学院。Kostas Daniilidis 就职于美国宾夕法尼亚州宾夕法尼亚大学。D. Scaramuzza 任职于苏黎世大学信息学系及苏黎世大学和苏黎世联邦理工学院的神经信息学系（瑞士）。

G. Gallego 和 D. Scaramuzza 获得 SNSF-ERC 启动基金，以及瑞士国家科学基金会通过国家机器人研究能力中心（NCCR Robotics）提供的资助。

联系邮箱：guillermo.gallego@tu-berlin.de

1 引言与应用

“大脑意味着想象力，这让我兴奋；我想造出一块能想象出某种东西的芯片[1]。”

原注 1 · https://youtu.be/FKemf6Idkd0?t=67

1986 年，加州理工学院研究生 Misha Mahowald 正是带着这个想法，与 Carver Mead 教授开始从生物学和工程学相结合的角度研究立体视觉。几年后，即 1991 年，《科学美国人》封面刊登了一幅猫的图像 [1]；它由一种模仿眼睛神经结构的新型“硅视网膜”采集，展示了一种崭新而强有力的计算方式，推动了神经形态工程这一新兴领域的兴起。今天，我们仍在追求同一个富有远见的目标：弄清大脑如何工作，并在计算机芯片上造出一个大脑。当前的努力包括耗资十亿美元量级的旗舰项目，如欧洲的人脑计划（Human Brain Project）和蓝脑计划（Blue Brain Project），以及美国 BRAIN（通过推进创新神经技术开展脑研究）计划。

本文概述“硅视网膜”或“事件相机”这一仿生技术，例如 [2]、[3]、[4]、[5]，重点介绍它们在传统和新兴计算机视觉及机器人任务中的应用。对人类来说，视觉是感知世界最重要的感觉，并与大脑协作学习新事物。近年来，这项技术受到学界和工业界的广泛关注。这既因为已有可用的事件相机原型，也因为事件相机在处理标准帧式图像传感器难以解决的问题上有优势，例如高速运动估计 [6]、[7] 和高动态范围（HDR）成像 [8]；标准帧式传感器输出的是按固定节拍同步采集的图像序列。

事件相机是异步传感器，改变了视觉信息获取的方式：它根据场景动态对光进行采样，而不是根据与所见场景无关的时钟采样。它的优点包括极高的时间分辨率和低延迟（二者均为微秒量级）、极高的动态范围（140 dB，而标准相机为 60 dB），以及低功耗。因此，在高速、高动态范围等传统相机难以应对的场景中，事件相机对机器人及可穿戴应用很有潜力。虽然事件相机直到 2008 年才有商用产品 [2]，近年来有关这类新传感器的文献[2]，以及三星 [5] 和 Prophesee[3] 等公司的量产计划，都表明商业界有很大兴趣将这类新型视觉传感器用于移动机器人、增强现实与虚拟现实（AR/VR）以及电子游戏。然而，事件相机的工作方式根本不同于标准相机：它异步测量逐像素亮度变化（称为“事件”），而不以恒定速率测量“绝对”亮度。因此需要新方法来处理其输出，发挥它的潜力。

原注 2 · https://github.com/uzh-rpg/event-based_vision_resources [9]

原注 3 · http://rpg.ifi.uzh.ch/ICRA17_event_vision_workshop.html

事件相机的应用：事件相机比其他传感方式更有优势的典型场景包括实时交互系统，例如机器人或可穿戴电子设备 [10]。能否在不可控光照下运行、延迟和功耗对这些系统很重要 [11]。事件相机可用

于目标跟踪 [12]、[13]，监视和监测 [14]，以及目标或手势识别 [15]、[16]、[17]。它也适用于深度估计 [18]、[19]、结构光三维扫描 [20]、光流估计 [21]、[22]、HDR 图像重建 [8]、[23]、[24] 和同步定位与建图（SLAM）[25]、[26]、[27]。事件视觉是一个不断发展的研究领域。随着事件相机的普及，还会出现更多应用 [9]，如图像去模糊 [28] 和恒星跟踪 [29]、[30]。

文章结构： 第 2 节介绍事件相机的工作原理、优点，以及这种新型视觉传感器带来的挑战。第 3 节讨论从事件相机输出中提取信息的几种常用方法，以及一些方法背后的生物启发。第 4 节回顾事件相机在从低层到高层视觉任务中的应用，以及为发挥其潜力而设计的部分算法，并指出各主题未来的研究机会和开放挑战。第 5 节介绍神经形态处理器和嵌入式系统。第 6 节回顾用于事件相机的软件、数据集和模拟器，以及其他信息来源。最后是讨论（第 7 节）和结论（第 8 节）。

2 事件相机的工作原理

标准相机以外部时钟规定的速率（如 30 fps）采集完整图像。事件相机不同，例如动态视觉传感器（DVS）[2]、[31]、[32]、[33]、[34]，每个像素对场景亮度变化独立而异步地响应（图 1b）。因此，事件相机输出的是一串数据率可变的数字“事件”或“脉冲”。每个事件表示一次亮度（对数光强）[4]变化：某个像素在特定时刻[5]发生了预定幅度的变化（图 1b；第 2.4 节）。这种编码借鉴了生物视觉通路的脉冲特性（第 3.3 节）。

原注 4 · 亮度是知觉量；为行文简洁，本文用它指代对数光强，因为在均匀照明的场景中，两者密切对应。

原注 5 · 命名：“事件相机”输出数据驱动的事件，用于指示位置和时间。这个名称在过去十年不断演变：最初称为采用地址事件表示（AER）的硅视网膜，后来称为基于事件的相机。一般而言，事件可以指示任何种类的信息（光强、局部空间对比度等）。但遗憾的是，在过去约五年里，“事件相机”一词实际上已成为与 DVS 输出的特定亮度变化表示同义的术语。

每个像素每次发送事件时都会记住当时的对数光强，并持续监测相对于该记忆值的变化是否达到足够大的幅度（图 1a）。变化超过阈值时，相机便发送一个事件；芯片传出的信息包括变化的位置 x, y、时间 t 和 1 位极性 p（即亮度增加“ON”或下降“OFF”）。事件输出见图 1b、1e 和 1f。

图 01 DAVIS 相机与 DVS 像素 / 原文第 2 页

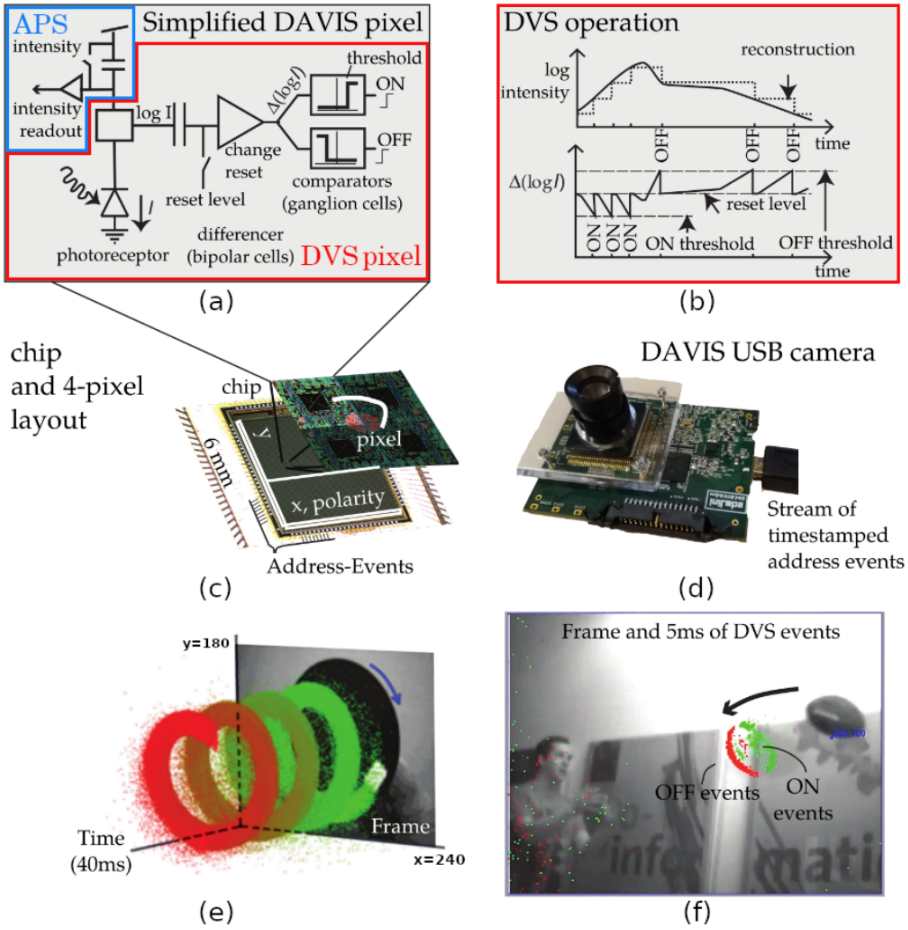


图 1. DAVIS 相机 [4] 概览。基于事件的动态视觉传感器（DVS [2]）与基于帧的主动像素传感器（APS）置于同一个像素阵列，每个像素中共享同一光电二极管。（a）DAVIS 像素的简化电路图（DVS 像素为红色，APS 像素为蓝色）。（b）DVS 像素把光信号转换为事件的工作示意图。（c-d）DAVIS 芯片及 USB 相机照片。（e）用 DAVIS 观察旋转黑色圆盘上的白色方块，会产生灰度帧以及时空中的螺旋形事件轨迹。时空事件用从绿色（过去）到红色（现在）的颜色编码。（f）自然场景的帧与叠加的事件；帧滞后于低延迟事件，事件按极性着色。图改编自 [4]、[35]。有关 DVS、DAVIS 和 ATIS 像素设计的更深入对比，见 [36]。

图中术语：intensity readout = 光强读出；photoreceptor = 光感受器；change reset = 变化复位；reset level = 复位电平；differencer = 差分器（双极细胞）；comparators = 比较器（神经节细胞）；threshold = 阈值；reconstruction = 重建；address-events = 地址事件；Frame and 5ms of DVS events = 帧与 5 ms 内的 DVS 事件。

事件通过共享的数字输出总线，从像素阵列传到芯片外围，再传出相机，通常使用地址事件表示（AER）读出 [37]、[38]。总线可能达到饱和，从而扰动事件的发送时刻。视芯片和硬件接口类型而定，事件相机的读出速率介于 2 MHz [2] 和 1200 MHz [39] 之间。

事件相机是一种数据驱动的传感器，输出取决于场景中的运动或亮度变化程度。运动越快，每秒产生的事件越多，因为每个像素都会使其增量调制器的采样率适应其所监测的对数光强信号的变化速率。事件时间戳的分辨率达微秒量级，传输延迟不到 1 毫秒，这使传感器能够迅速响应视觉刺激。

像素接收的入射光取决于场景照明与表面反射率的乘积。若照明近似恒定，对数光强的变化就表示反射率发生变化。这种反射率变化主要是视野中物体运动引起的。因此，DVS 的亮度变化事件本身对场景照明具有不变性 [2]。

DVS 像素与帧相机的带宽比较：DVS 像素虽然很快，但与物理换能器一样，带宽有限。如果入射光强变化过快，前端光感受器电路会滤除这些变化 [40]。相当于标准图像传感器曝光时间的上升和下降时间，是该带宽的倒数。图 2 给出了 DVS 像素的实测频率响应示例（[2] 中的 DVS128）。测量装置（图 2a）用产生的正弦变化信号测量响应。图 2b 显示，在低频下，DVS 像素每个周期产生一定数量的事件。超过某个截止频率后，光感受器的动态响应会滤除变化，周期内的事件数随之下降。截止频率随光强单调增加。在较亮光照下，DVS 像素的带宽约为 3 kHz，相当于约 300 μs 的曝光时间。光强降至原来的 1/1000 时，DVS 带宽降至约 300 Hz。即使 LED 亮度降低 1000 倍，DVS 像素的频率响应仍为 60 fps 图像传感器的奈奎斯特频率 30 Hz 的十倍。此外，帧相机会将高于奈奎斯特频率的分量混叠回基带，DVS 像素则因其连续时间响应而不发生这种混叠。

图 02 DVS 像素与帧相机的频率响应 / 原文第 3 页

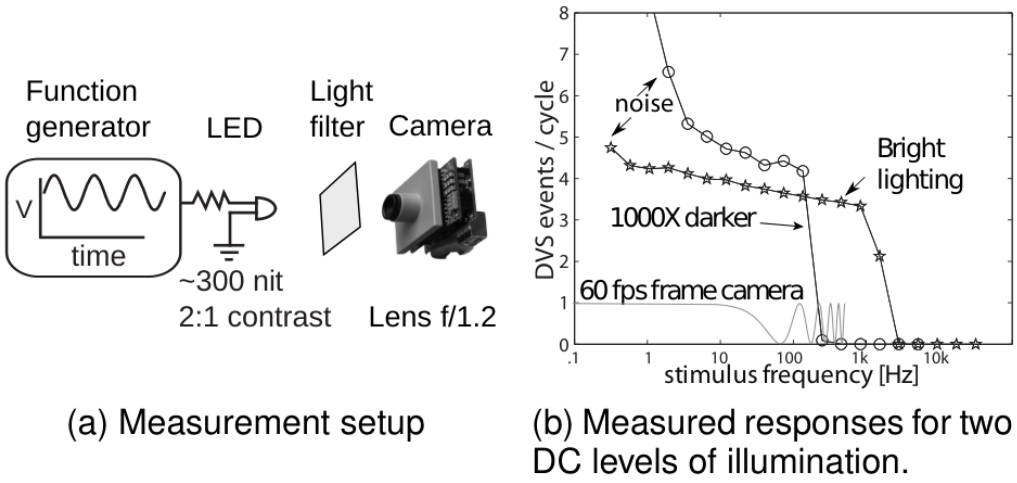


图 2. 正弦 LED 刺激下单个 DVS 像素的“事件传递函数”。频率极低时，背景事件引起额外的 ON 事件。60 fps 相机曲线表明包含奈奎斯特频率以上频率的混叠效应时的传递函数。图改编自 [2]。

图中术语：（a）测量装置：function generator = 函数发生器，light filter = 滤光片，lens = 镜头，contrast = 对比度。（b）两种直流照明水平下的实测响应：DVS events / cycle = 每周期 DVS 事件数，stimulus frequency [Hz] = 刺激频率 [Hz]，noise = 噪声，bright lighting = 明亮照明，1000× darker = 光强降至 1/1000，60 fps frame camera = 60 fps 帧相机。

2.1 事件相机的设计

本节介绍最常见的事件相机设计。实际器件（商用或原型相机，例如 DAVIS240）汇总于第 2.5 节。

1986–1992 年间，Mahowald 和 Mead 在加州理工学院研制了首个硅视网膜；这项博士论文研究 [41] 获得了颇具声望的 Clauser 奖[6]。他们的传感器采用对数响应像素，仿照 Kufler 三层视网膜，并以 AER 协议输出脉冲事件。但它存在几个缺点：每块绕线连接的视网膜电路板都要精密调节偏置电位器；不同像素的响应有明显失配；像素太大，难以做成实用器件。之后十年，神经形态工程界研制了一系列硅视网膜，这些进展见 [36]、[38]、[42]、[43]。

原注 6 · http://www.gradoffice.caltech.edu/current/clauser

DVS 事件相机 [2] 源自一种帧式硅视网膜设计，其中连续时间光感受器通过电容耦合到读出电路，每次采样像素时，读出电路都会复位 [44]。电子学和神经科学文献 [10]、[36]、[38]、[45]、[46]、[47] 回顾了较新的事件相机技术。虽然仅靠处理 DVS 事件（即亮度变化）就能解决出人意料的多种应用问题，但有些应用显然还需要某种形式的静态输出（即“绝对”亮度）。为弥补这项不足，研究者开发了多种能同时输出动态和静态信息的相机。

异步时间型图像传感器（ATIS）[3]、[48] 的像素内含一个 DVS 子像素（称为变化检测，CD），它触发另一个子像素读出绝对光强（曝光测量，EM）。触发时，电容被复位到高电压，随后由另一只光电二极管将电荷逐步泄放。光越亮，电容放电越快。与 [49] 一样，ATIS 的光强读出另外传送两个事件，用跨过两个电压阈值之间的时间间隔编码光强。这样，只有发生变化的像素才提供新的光强值。照明越亮，这两个事件的间隔越短。ATIS 实现了很大的静态动态范围（>120 dB）。不足之处在于其像素面积至少为 DVS 像素的两倍；而且在暗场景中，两个光强事件间隔可能较长，新的事件可能打断光强读出（[50] 提出一种变通方法）。

图 1 所示的动态与主动像素视觉传感器（DAVIS）[4]、[51] 得到广泛使用，它将传统主动像素传感器（APS）[52] 与 DVS 集成在同一个像素内。与 ATIS 相比，它的优点是像素尺寸小得多：二者共享光电二极管，读出电路仅使 DVS 像素面积增加约 5 %。光强帧（APS）可以按固定帧率触发，也可由 DVS 事件分析按需触发，但后一种方式很少使用[7]。不过，APS 读出的动态范围有限（55 dB）；与标准相机一样，当像素未发生变化时，读出是冗余的。

原注 7 · https://github.com/SensorsINI/jaer/blob/master/src/eu/seebetter/ini/chips/davis/DavisAutoShooter.java

因为 ATIS 和 DAVIS 的设计均包含 DVS 像素（变化检测器）[36]，文中通常用“DVS”统称输出二值极性事件的输出端或电路，不区分其源自 DVS、ATIS 还是 DAVIS 设计。

2.2 事件相机的优势

与标准相机相比，事件相机有许多潜在优势：

高时间分辨率：模拟电路可以快速监测亮度变化，事件则用 1 MHz 时钟数字读出。这意味着事件检测和时间戳的分辨率可达到微秒量级。因此，事件相机能捕捉极快的运动，而不受帧相机常见的运动模糊影响。

低延迟：各像素独立工作，无须等待整个帧的全局曝光；一旦检测到变化，便立即传输。因此事件相机的延迟极小：在实验台上约 10 μs，在实际环境中不到 1 毫秒。

低功耗：事件相机只传输亮度变化，从而去除冗余数据，只需为处理发生变化的像素耗电。在裸片层面，大多数相机的功耗约为 10 mW，有的原型低于 10 μW。传感器直接连接处理器的嵌入式事件相机系统已实现不超过 100 mW 的系统级功耗（即传感加处理）[17]、[53]、[54]、[55]。

高动态范围（HDR）：事件相机极高的动态范围（>120 dB）显著超过高质量帧相机的 60 dB，能够采集从月光到日光条件下的信息。这是因为像素的光感受器以对数尺度工作，且每个像素独立运行，不用等待全局快门。与生物视网膜类似，DVS 像素既能适应很暗的刺激，也能适应很亮的刺激。

2.3 新传感范式带来的挑战

事件相机改变了视觉信息获取的范式。要发挥其优势，就必须设计新方法（算法与硬件）来处理采集的数据并提取信息。具体来说：

1）应对不同的时空输出：事件相机与标准相机的输出根本不同。事件是异步且空间稀疏的，图像则是同步且稠密的。因此，为图像序列设计的帧式视觉算法不能直接用于事件数据。

2）应对不同的光度传感方式：标准相机给出灰度信息，每个事件则仅包含二值的亮度变化信息（增加或下降）。亮度变化不仅取决于场景亮度，还取决于场景与相机当前和过去的相对运动。

3）应对噪声与动态效应：所有视觉传感器都有噪声，包括光子固有的散粒噪声和晶体管电路噪声，也都存在非理想效应。对事件相机尤其如此：时间对比度的量化过程复杂，尚未得到完整表征。

因此，新方法必须重新考虑事件数据的时空、光度和随机性质。由此产生两个问题：怎样才能最好地从事件中提取与给定任务有关的信息？又应怎样对噪声和非理想效应建模，以便更好地提取有意义的信息？

2.4 事件生成模型

事件相机 [2] 的像素各自独立，对光电流的对数 $L \doteq \log(I)$（“亮度”）的变化作出响应。具体而言，在无噪声情形下，事件 $e_k \doteq (\mathbf{x}_k, t_k, p_k)$ 在像素 $\mathbf{x}_k \doteq (x_k, y_k)^\top$ 和时刻 t_k 触发的条件是：该像素自上一次事件以来的亮度增量，即

$$\Delta L(\mathbf{x}_k, t_k) \doteq L(\mathbf{x}_k, t_k) - L(\mathbf{x}_k, t_k - \Delta t_k), \qquad (1)$$

达到时间对比度阈值 $\pm C$（图 1b），即

$$\Delta L(\mathbf{x}_k, t_k) = p_k\, C, \qquad (2)$$

其中 $C > 0$，Δt_k 为同一像素自上次事件以来经过的时间，极性 $p_k \in \{+1, -1\}$ 是亮度变化的符号 [2]。

对比度灵敏度 C 由像素偏置电流决定 [56]、[57]。偏置电流由片上的数字可编程偏置发生器生成，用于设定图 1 所示变化检测器的速度和阈值电压。已知这些电流，即可估计灵敏度 C [56]。实际中，正（“ON”）事件和负（“OFF”）事件可能分别按不同的阈值 C_+、C_- 触发。典型 DVS [2]、[5] 可把阈值设置在 10 %-50 % 的照明变化范围。C 的下限由噪声及像素间失配（变化性）决定；C 设得过低会引发大量噪声事件，最先出现于 C 值较低的像素。光感受器增益较高的实验型 DVS 能达到更低的阈值，如 1 % [58]、[59]、[60]；但只有在光照很强、条件理想时才能得到这些数值。根本上，像素必须在光电流存在散粒噪声的情况下，对光电流中的微小变化作出响应。在给定照明水平和目标检测可靠性时，这一散粒噪声限制决定了 DVS 阈值与速度之间的关系 [60]、[61]。

事件与亮度的时间导数：式（2）说明，事件相机像素对自上次事件以来的亮度变化幅度设定阈值。当 Δt_k 很小时，可通过泰勒展开将式（2）中的增量近似写为 $\Delta L(\mathbf{x}_k, t_k) \approx \frac{\partial L}{\partial t}(\mathbf{x}_k, t_k)\Delta t_k$。因此，可将事件理解为提供关于时间导数的信息：

$$\frac{\partial L}{\partial t}(\mathbf{x}_k, t_k) \approx \frac{p_k\, C}{\Delta t_k}. \qquad (3)$$

这是对亮度的间接测量，不同于我们通常用标准相机测量绝对亮度的做法。注意，触发 DVS 事件的是亮度幅值的变化（式 2），而不是亮度导数超过阈值（式 3）。上述解释可用于设计有物理依据的事件视觉算法，如 [7]、[23]、[24]、[28]、[62]、[63]、[64]、[65]；这与仅将事件视作光度意义模糊的点集合的算法不同。

移动边缘产生事件：假定照明恒定，对式（2）线性化并应用亮度恒常假设，可证明事件由移动边缘产生。Δt 很小时，式（2）的光强增量可近似为[8]：

$$\Delta L \approx -\nabla L \cdot \mathbf{v} \Delta t, \qquad (4)$$

原注 8 · 将亮度恒常假设（即光流约束）$\frac{\partial L}{\partial t}(\mathbf{x}(t),t)+\nabla L(\mathbf{x}(t),t)\cdot\dot{\mathbf{x}}(t)=0$（图像点的速度 $\mathbf{v}\equiv\dot{\mathbf{x}}$）代入泰勒近似式 $\Delta L(\mathbf{x},t)\doteq L(\mathbf{x},t)-L(\mathbf{x},t-\Delta t)\approx\frac{\partial L}{\partial t}(\mathbf{x},t)\Delta t$，即可得到式（4）[66]。

也就是说，亮度梯度 $\nabla L(\mathbf{x}_k,t_k)=(\partial_x L,\partial_y L)^\top$ 以速度 $\mathbf{v}(\mathbf{x}_k,t_k)$ 在图像平面上移动了位移 $\Delta\mathbf{x}\doteq\mathbf{v}\Delta t$，从而产生这一增量。式（4）的点积表明：（i）运动与边缘平行时，由于 $\mathbf{v}\cdot\nabla L=0$，不产生事件；（ii）运动垂直于边缘时（$\mathbf{v}\parallel\nabla L$），事件生成率最高（即达到幅度为 $|C|$ 的亮度变化所需时间最短）。

概率性事件生成模型：式（2）是理想化的事件生成模型。更符合实际的模型考虑传感器噪声和晶体管失配，因此包含静态不变与随时间变化的随机触发条件，并且用概率函数表示。该函数本身是局部照明水平和传感器工作参数的复杂函数。[2] 给出了这一概率密度的测量结果（使用 DVS128），表明可能为以对比度阈值 C 为中心的正态分布。分布的 1σ 宽度通常对应 2-4 % 的时间对比度。这种事件生成模型可以用于事件相机仿真器（emulator）[73]、模拟器（simulator）[74] 和事件处理算法 [24]、[66]。还有其他概率性事件生成模型，例如让事件生成的似然与图像梯度幅值成正比 [75]（用于大梯度产生大部分事件数据的场景），或用混合分布对似然建模以抵抗传感器噪声 [7]。未来更贴近实际的模型可能还纳入不应期（即像素产生一次事件后忽略对数亮度变化的时长；不应期越长，快速移动物体产生的事件越少）和总线拥塞 [76]。

上述事件生成模型较简单，只在一定程度上依据传感器噪声表征建立。与标准图像传感器一样，DVS 也有固定模式噪声（FPN[9]），但在 DVS 中，表现为像素之间事件阈值不同。标准 DVS 可实现最低阈值 $C\approx\pm15\,\%$，像素间对比度阈值的标准差约为 2.5 %-4 % [2]、[77]。也有研究比较 DVS 事件所指示的亮度变化与 DAVIS 相邻 APS 帧差分所给出的亮度变化，尝试测出逐像素阈值 [78]。然而，对 DVS 像素时变噪声和读出噪声的认识尚处于初步阶段 [2]、[59]、[76]、[79]；现有噪声滤除方法主要着眼于计算效率，假定真实物体产生的事件在时空上的相关性强于噪声事件 [38]、[80]、[81]、[82]、[83]。目前我们离建立一个可在任意照明与偏置条件下预测事件相机噪声统计量的模型还有很大距离。解决这一难题将带来更好的估计方法。

原注 9 · https://en.wikipedia.org/wiki/Fixed-pattern_noise

2.5 现有事件相机

表 1 汇总了最常用或较新的相机。数值为近似值，因为它们并非在同一个测试平台上测得。事件相机的特性与其他 CMOS 图像传感器（CIS）技术相差很大，需要形成统一的规格标准，方便研究者使用。表 1 表明，从首款实用事件相机 [2] 问世以来，主要趋势是提高空间分辨率和读出速度，并加入新功能，如灰度级输出（ATIS 和 DAVIS）、集成惯性测量单元（IMU）[84] 和多相机时间戳同步 [85]。IMU 的作用类似前庭感觉，可能改善相机位姿估计，例如在视觉惯性里程计中。直到最近，重心才更多转向缩小像素尺寸这一困难任务，以便经济地量产大像素阵列传感器。在这方面，三维晶圆堆叠工艺在缩小像素和提高填充因子上作用最大。

表 01 商用与原型事件相机 / 原文第 6 页

	Supplier	iniVation			Prophesee				Samsung			CelePixel		Insightness
	Camera model	DVS128	DAVIS240	DAVIS346	ATIS	Gen3 CD	Gen3 ATIS	Gen 4 CD	DVS-Gen2	DVS-Gen3	DVS-Gen4	CeleX-IV	CeleX-V	Rino 3
Sensor specifications	Year, Reference	2008 [2]	2014 [4]	2017	2011 [3]	2017 [67]	2017 [67]	2020 [68]	2017 [5]	2018 [69]	2020 [39]	2017 [70]	2019 [71]	2018 [72]
	Resolution (pixels)	128 × 128	240 × 180	346 × 260	304 × 240	640 × 480	480 × 360	1280 × 720	640 × 480	640 × 480	1280 × 960	768 × 640	1280 × 800	320 × 262
	Latency (µs)	12μs @ 1klux	12μs @ 1klux	20	3	40 - 200	40 - 200	20 - 150	65 - 410	50	150	10	8	125μs @ 10lux
	Dynamic range (dB)	120	120	120	143	> 120	> 120	> 124	90	90	100	90	120	> 100
	Min. contrast sensitivity (%)	17	11	14.3 - 22.5	13	12	12	11	9	15	20	30	10	15
	Power consumption (mW)	23	5 - 14	10 - 170	50 - 175	36 - 95	25 - 87	32 - 84	27 - 50	40	130	-	400	20-70
	Chip size (mm^2)	6.3 × 6	5 × 5	8 × 6	9.9 × 8.2	9.6 × 7.2	9.6 × 7.2	6.22 × 3.5	8 × 5.8	8 × 5.8	8.4 × 7.6	15.5 × 15.8	14.3 × 11.6	5.3 × 5.3
	Pixel size (μm^2)	40 × 40	18.5 × 18.5	18.5 × 18.5	30 × 30	15 × 15	20 × 20	4.86 × 4.86	9 × 9	9 × 9	4.95 × 4.95	18 × 18	9.8 × 9.8	13 × 13
	Fill factor (%)	8.1	22	22	20	25	20	> 77	11	12	22	8.5	8	22
	Supply voltage (V)	3.3	1.8 & 3.3	1.8 & 3.3	1.8 & 3.3	1.8	1.8	1.1 & 2.5	1.2 & 2.8	1.2 & 2.8		1.8 & 3.3	1.2 & 2.5	1.8 & 3.3
	Stationary noise (ev/pix/s) at 25C	0.05	0.1	0.1	-	0.1	0.1	0.1	0.03	0.03		0.15	0.2	0.1
	CMOS technology (nm)	350	180	180	180	180	180	90	90	90	65/28	180	65	180
		2P4M	1P6M MIM	1P6M MIM	1P6M	1P6M CIS	1P6M CIS	BI CIS	1P5M BSI			1P6M CIS	CIS	1P6M CIS
	Grayscale output	no	yes	yes	yes	no	yes	no	no	no	no	yes	yes	yes
	Grayscale dynamic range (dB)	NA	55	56.7	130	NA	> 100	NA	NA	NA	NA	90	120	50
	Max. frame rate (fps)	NA	35	40	NA	NA	NA	NA	NA	NA	NA	50	100	30
Camera	Max. Bandwidth (Meps)	1	12	12	-	66	66	1066	300	600	1200	200	140	20
	Interface	USB 2	USB 2	USB 3		USB 3	USB 3	USB 3	USB 2	USB 3	USB 3			USB 2
	IMU output	no	1 kHz	1 kHz	no	1 kHz	1 kHz	no	no	1 kHz	no	no	no	1 kHz

表 1. 商用或原型事件相机对比。由于没有统一的测量测试平台，表中数值为近似值。

表中术语：Supplier = 供应商；camera model = 相机型号；sensor specifications = 传感器规格；year, reference = 年份、文献；resolution = 分辨率；latency = 延迟；dynamic range = 动态范围；min. contrast sensitivity = 最低对比度灵敏度；power consumption = 功耗；chip size = 芯片尺寸；pixel size = 像素尺寸；fill factor = 填充因子；supply voltage = 供电电压；stationary noise = 静态噪声（表中条件 25 °C、单位为事件数 / 像素 / 秒）；CMOS technology = CMOS 工艺；grayscale output / dynamic range = 灰度输出 / 灰度动态范围；max. frame rate = 最高帧率；max. bandwidth = 最大带宽（Meps：百万事件 / 秒）；interface = 接口；IMU output = 惯性测量单元输出；NA = 不适用。

像素尺寸：应用最广的事件相机像素相当大：DVS128 为 40 µm，ATIS 为 30 µm，DAVIS240、DAVIS346 为 18.5 µm（表 1）。已发表的最小 DVS 像素为 4.86 µm [68]；相比之下，传统工业全局快门 APS 通常在 2 µm–4 µm 范围。低空间分辨率无疑限制应用，不过许多开创性研究都使用 128 × 128 像素的 DVS128 [2]。已发表阵列规模最大的 DVS，空间分辨率也只有约 100 万像素（1280 × 960 像素 [39]）。事件相机像素几乎与工艺特征尺寸同步缩小。考虑到 DVS 像素是混合信号电路，而这种电路通常不会随制程等比例缩小，这一进展相当显著。然而，要实现更小的像素并不容易，可能必须放弃相机最初采用的严格异步电路设计理念 [86]。相机成本受裸片尺寸（量产时硅的成本约为 5–10 美元/cm^2）和光学系统的约束（为另一种传感器画幅设计可量产的小型光学系统，可能耗费数千万美元）。

填充因子：像素填充因子有限，是早期事件相机量产面临的一大障碍。填充因子是像素感光面积占总面积的比例。像素电路复杂，因而可留给收集光线的光电二极管的面积较小。例如，填充因子为 20 % 的像素会丢掉五个光子中的四个。这对追求最佳性能显然不可接受，但即使最早的事件相机也可在月光下感知高对比度特征 [2]。早期 CIS 传感器通过微透镜将光聚焦到像素的光电二极管来缓解这一问题。更好的选择可能是使用背照式技术（BSI）。BSI 把芯片翻转，从背面照明，因此原则上整个像素面积都可以收集光子。如今几乎所有智能手机相机都采用背照方式；但 BSI 工艺成本较高，直到最近才有 BSI 事件相机的演示 [39]、[68]、[69]、[87]。BSI 也有问题：光可产生额外的“寄生”光电流，导致虚假的“漏电”事件 [56]。

成本：目前，采用事件相机技术的现实障碍是每台数千美元的高成本，这与早期飞行时间、结构光和热成像相机的处境类似。成本高，一是因为硅芯片设计和制造有一次性工程成本（即便其中很大一部分由科研经费负担），二是原型试产可提供的样品有限。预计一旦量产，价格就会大幅下降。消费级家用监控设备 “Samsung SmartThings Vision” 是一例：它内置事件相机 [5]，售价为 100 美元。

先进事件相机

更先进的事件相机正处于积极研发中，只能通过与开发者开展科研合作获得。下面讨论先进相机开发中涉及的问题，以及正在研制的新型相机。

彩色：多数昼行性动物具有某种色觉，多数传统相机也能感知颜色。早期彩色事件相机的尝试 [88]、[89]、[90]，试图采用由 Foveon 开创的“vertacolor”原理，根据不同波长的光在硅中穿透深度的差异分色 [91]、[92]。但其分色性能不佳。截至成文，关于实用彩色事件相机的发表成果很少；已有方案或集成彩色滤光阵列（CFA）[93]、[94]、[95]，或使用分色棱镜 [96]，后者成本远高于 CFA。

更高的对比度灵敏度：已有研究尝试提高事件相机的时间对比度灵敏度，研制出了灵敏度更高的实验型传感器 [58]、[59]、[60]（在实验室条件下可低至约 1 %）。这些传感器采用热辐射计（thermal bolometer）[97] 思路的变体，即在变化检测器（图 1）之前提高增益，以降低输入折算固定模式噪声。不过，这级中间前置放大器需要主动增益控制，防止信号削波。提高对比度灵敏度是可行的，但代价是减小动态范围（如 [5]）。

3 事件处理

事件相机引起的范式转变带来一个关键问题：如何从事件数据中提取有意义的信息，以完成给定任务？这个问题很宽泛，答案取决于应用，同时又决定了解决任务时所用算法的设计。

事件相机以异步、稀疏的方式获取信息，同时具备高时间分辨率和低延迟。因此，时间因素，尤其是延迟，在事件处理方式中起关键作用。按同时处理的事件数量，可分为两类算法：（i）逐事件方法，每到一个事件，系统状态（所估计的未知量）就可能改变，从而达到最低延迟；（ii）按事件组或事件包工作的方法，会引入一定延迟。不考虑延迟时，成组事件（即时间窗）方法若每次将窗滑动一个事件，也可在每个新事件到达时更新状态。因此，两者的区别更微妙：单个事件无法提供足够的估计信息，还需要过去的事件或额外知识。下文回顾这一分类。

另一种独立的划分方式按事件处理方法，区分基于模型的方法与无模型方法（即数据驱动、机器学习方法）。若在优化框架中处理事件，还可以按目标函数或损失函数的类型分类：基于几何、时间或光度的方法（例如以事件极性或事件活动为变量的函数）。每类方法都有利弊，当前的研究正探索各类方法所能提供的能力。

3.1 事件表示

为便于提取解决任务所需的有意义信息（“特征”），事件经过处理，常被转换为其他表示形式（图 3）。这里回顾几种常用的事件数据表示。没有额外知识时，单个事件所传递的信息有限，其中若干表示正是为了汇聚这些信息而产生的。有些表示只是简单的人工设计数据变换，有些则更复杂。

图 03 同一事件序列的六种表示 / 原文第 7 页

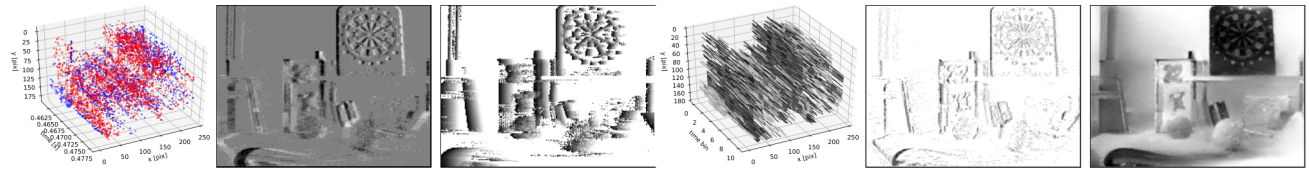


图 3. slider depth 序列 [98] 的几种事件表示（第 3.1 节）。（a）时空事件，按极性着色（正极性为蓝色，负极性为红色）。（b）事件帧（亮度增量图像 $\Delta L(\mathbf{x})$）。（c）仅由负事件构成的时间表面，每个像素存最近一次事件的时间戳（色深的像素表示时间较近）。（d）插值体素网格（240 × 180 × 10 个体素），按极性着色，从深色（负）到浅色（正）。（e）运动补偿事件图像 [99]（事件累加得到的锐利边缘比没有事件的白色像素深）。（f）按 [8] 得到的重建光强图像。网格状表示兼容传统计算机视觉方法 [100]。

图中术语：x / y [pix] = 像素横 / 纵坐标；time [s] = 时间 [s] ；time bin = 时间分箱。

单个事件 $e_k \doteq (\mathbf{x}_k, t_k, p_k)$ 用于逐事件方法，例如概率滤波器和脉冲神经网络（SNN，第 3.3 节）。滤波器或 SNN 具有从过去事件积累而来、或由额外知识给出的信息，并与新到事件异步融合，产生输出。例见 [7]、[24]、[62]、[101]、[102]。

事件包：将时空邻域内的事件 $\mathrm{E} \doteq \{e_k\}_{k=1}^{N_e}$ 一同处理以获得输出。该表示保留精确的时间戳和极性。选取合适的包大小 N_e，对满足算法假设至关重要（如包所跨时段内的运动速度恒定）；选择因任务而异。例见 [18]、[19]、[103]、[104]。

事件帧 / 事件图像或二维直方图：用简单的方法（如计数或逐像素累加极性）把时空邻域内的事件变成一幅图像（二维网格），供基于图像的计算机视觉算法使用。虽然事件帧与自然图像的统计特性不同，有些算法仍能工作。这类直方图可自然地按事件活动决定采样率；关于如何累积这类帧来计算光流，见 [105]。这种做法并不完全符合事件范式，因为它量化了事件的时间戳，可能丢失稀疏性（但见 [106]），而且所得图像对所用事件数极其敏感。尽管如此，事件帧在文献 [23]、[26]、[64]、[105]、[107]、[108] 中影响显著，因为：（i）它以简单方式把陌生的事件流变成熟悉的二维表示，其中包含场景边缘的空间信息，而边缘是自然图像中最富含信息的区域；（ii）它不仅表明何处有事件，也表明何处没有事件（同样有信息）；（iii）它可直观解释为边缘图或亮度增量图像等；（iv）其数据结构兼容传统计算机视觉。

时间表面（TS）：TS 是二维图，每个像素存一个时间值，例如该像素最近一次事件的时间戳 [80]、[109]。事件因此变成一幅图像，其“强度”是所在位置运动历史的函数，值越大对应越近的运动。经典计算机视觉将 TS 称为运动历史图像 [110]。它显式呈现丰富的事件时间信息，且可异步更新。使用指数核时，TS 相对过去的事件更强调近期事件。为使表示对运动速度不变，文献 [111]、[112] 提出了归一化。与其他网格状事件表示相比，TS 每个像素只保留一个时间戳，因此高度压缩信息；在纹理丰富、像素频繁发出脉冲的场景中，其有效性下降。为降低对噪声的敏感性，可对时空窗口内的事件滤波，计算各像素值 [113]。其他例子见 [21]、[114]、[115]、[116]。

体素网格：事件的时空（三维）直方图，每个体素代表指定的像素和时间区间。由于不将事件压缩到二维网格，该表示能较好地保留时间信息（图 3）。若使用极性，体素网格可以直观地看作图像平面上标量场（极性 $p(x,y,t)$ 或亮度变化 $\partial L(x,y,t)/\partial t$）的离散化，无事件处用零极性标出。每个事件的极性可累加到一个体素 [117]、[118]，也可用核分配到离它最近的若干体素 [8]、[119]、[120]。两种方案都对事件时间戳作量化，但后者（插值体素网格）提供亚体素精度。

三维点集：时空邻域内的事件被视为三维空间内的点 $(x_k, y_k, t_k) \in \mathbb{R}^3$，时间维因而成为一个几何维度。这种稀疏表示用于基于点的几何处理方法，如平面拟合 [21] 或 PointNet [121]。

图像平面上的点集：将事件看作图像平面上随时间演化的二维点集。这是早期基于均值漂移或迭代最近点（ICP）的形状跟踪方法常用的表示 [122]、[123]、[124]、[125]、[126]；这些方法只需事件即可跟踪边缘模式。

运动补偿事件图像 [127]、[128]：该表示不仅取决于事件，还取决于运动假设。运动补偿的原理是：边缘在图像平面上移动时，会在掠过的像素触发事件；将事件变换到同一参考时刻并使其尽可能对齐，可估计边缘运动并得到锐利的变换后事件图像（即直方图，IWE）[128]。IWE 据此给出判断事件与候选运动符合程度的准则：变换事件所得边缘越锐利，拟合越好 [99]。所得运动补偿图像也容易理解（即产生事件的边缘模式），而且以比事件更熟悉的方式呈现视觉信息。某种意义上，运动补偿揭示了事件流中潜藏的、“对运动不变”的边缘图。这些图像可供特征跟踪等后续处理使用 [64]、[129]。点集 [130]、[131] 和时间表面 [132]、[133] 也有运动补偿版本。

重建图像：图像重建（第 4.6 节）得到的亮度图像可以理解为比事件帧或 TS 更不受运动影响的表示，并可用于推断 [8]，取得一流结果。

处理事件极性有两种方式：分别处理正、负事件后合并结果（如用 TS [109]），或在共同的表示中一起处理（如亮度增量图像 [64]），后者常把邻近事件的极性聚合起来。极性取决于运动方向，因此会干扰目标识别等本应与运动无关的任务（可用多个运动方向的训练数据缓解）。对于运动估计，极性或能提供帮助，尤其是检测方向突变时。

[100] 给出了将事件数据转换为上述若干网格表示的通用框架。它还考察了将不同表示送入人工神经网络（ANN）如何影响任务性能，并据此提出自动学习使该性能最大化的表示。

3.2 事件处理方法

事件处理系统由若干阶段组成：预处理（适配输入）、核心处理（特征提取和分析）与后处理（生成输出）。第 3.1 节的事件表示可以出现在不同阶段。例如 [127] 将事件包用于预处理，核心处理则以运动补偿事件图像为内部表示。其他情况下，上述表示可能仅用于预处理。例如 [22] 将事件转换成事件图像和时间表面，再由 ANN 处理。

事件处理方法受所选表示与可用硬件平台的影响，三者相互影响。举例说，很自然可以选用稠密表示，设计相应算法并在标准处理器（例如 CPU 或 GPU）上执行；同样，也可在神经形态硬件（第 5.1 节）上实现 SNN（第 3.3 节），逐个处理事件，寻求更高效且延迟更低的解法。逐事件方法主要有滤波器（确定性或概率性）和 SNN。处理事件包也有许多方法，如人工设计的特征提取器、深度神经网络（DNN）等。下面回顾部分最常见的方法。

逐事件方法：确定性滤波器，如（时空）卷积和活动滤波器，已用于降噪、特征提取 [134]、图像重建 [62]、[135] 和亮度滤波 [63] 等。卡尔曼滤波和粒子滤波等概率滤波器（贝叶斯方法）已用于

SLAM 系统的位姿跟踪 [7]、[24]、[25]、[75]、[101]。这些方法需要额外信息，通常是“外观”信息，例如灰度图像或场景地图，可由过去的事件或附加传感器提供。随后将每个新到事件与这些信息 比较，以所得差异为新息，更新滤波器状态。滤波器是逐事件处理的主要方法之一，因为它自然地做 到：（i）处理异步数据，从而尽可能减小处理延迟，保留传感器特性；（ii）汇聚多个小信息源（ 如事件）所提供的信息。

另一类主要方法采用多层 ANN（不论是否为脉冲式），其中许多参数必须从事件数据中计算。使用无监督学习训练的网络通常为分类器（如 SVM）提取特征，分类器训练仍需一些有标注的数据 [15]、[109]、[136]。若有足够的标注数据，则可采用反向传播等监督学习方法训练网络，不需要单独的分类器。许多方法在训练时使用事件包（在帧上进行深度学习），此后把训练好的网络转换为逐事件处理数据的 SNN [137]、[138]、[139]、[140]、[141]。逐事件无模型方法主要用于物体分类 [15]、[109]、[137]、[138] 或动作分类 [16]、[17]、[142]，目标通常是嵌入式应用 [137]，常采用定制的 SNN 硬件 [15]、[17]（第 5.1 节）。通过深度学习训练的 SNN 通常比依靠无监督学习提取特征的 SNN 更准确，但研究者越来越关注如何直接在 SNN [142]、[143] 和嵌入式设备 [144] 上高效实现监督学习。

事件组处理方法：因为每个事件所带信息很少而且会受噪声影响，常要同时处理多个事件，才能为所考虑的问题提供足够的信噪比。事件组方法用上述表示（事件包、事件帧等）汇集事件所含的信息，以估计问题的未知量，通常不需额外的数据。因此，处理方式随事件表示而变。

许多表示只对数据作预处理，以便复用基于图像的计算机视觉工具。在这方面，事件帧是一种实用表示，多种方法已将其用于不同任务。在 [107]、[145] 中，事件帧使传统立体视觉方法得以复用，取得了有限的效果。事件帧还给出自适应帧率的信号，有利于相机位姿估计 [26]（图像对齐）或计算光流 [105]（块匹配）。事件帧也是基于图像的学习方法（DNN、SVM、随机森林）简单且有效的输入 [22]、[108]、[146]、[147]。很少有研究在设计算法时考虑事件帧的光度含义（式 4）。[23] 考虑了这一点，表明用这样简单的表示就能联合计算多个感兴趣的视觉量（光流、亮度等）。光强增量图像（式 4）也用于特征跟踪 [64]、图像去模糊 [28] 或相机跟踪 [65]。

时间表面（TS）对场景边缘和运动方向敏感，已应用于许多涉及运动分析和形状识别的任务。例如，对 TS 作局部平面拟合可获得光流信息 [21]、[148]。TS 可作为类似神经网络的层级特征提取器的基本构件，逐级聚合越来越大的时空邻域中的信息，再送给分类器识别 [109]、[113]。在立体视觉方法中，TS 提供用于匹配的代理光强图像 [116]、[149]，把光度匹配准则变成时间匹配准则：根据不同图像平面上事件的同时发生情况以及时间戳的相似程度匹配像素。近期也有人研究把 TS 作为卷积 ANN（CNN）的输入来计算光流 [22]，即同时让网络提取特征、回归速度。TS 常用于角点检测，既可使用改造后的图像方法（Harris、FAST）[111]、[114]、[115]，也可用新的基于学习的方法 [112]。不过，由于“运动覆盖”问题 [110]，它们在纹理丰富的场景中性能下降 [115]。

基于体素网格的方法包括变分优化和 ANN（如 DNN）。相比使用低维表示的方法，它们需要更多内存，往往也需要更多计算；但时间信息保留得更好，因此能得到更好的结果。在这些方法中，体素网格用作内部表示 [117]（如计算光流），或作为 DNN 的多通道输入、输出 [119]、[120]。体素网格因而通过卷积 [119]、[120]，或按目标函数的最优条件导出的运算 [117] 处理。

一旦将事件转换为网格状表示，传统视觉中大量工具都可用于提取信息，既包括特征提取器（如 CNN），也包括相似度度量（如互相关），用于衡量数据与任务模型假设的贴合或一致程度（如事件对齐程

度）。这些度量用作分类（SVM、CNN）、聚类、数据关联、运动估计等任务的目标函数。神经科学文献也在尝试设计直接作用于脉冲（如事件流）的度量，以免引入数据转换所带来的问题。

处理事件组的深度学习方法采用深度神经网络（DNN）。应用例子包括分类 [150]、[151]，图像重建 [8]、[118]，转向角预测 [108]、[152]，以及光流 [22]、[119]、[153]、深度 [153] 或自运动 [119] 的估计。这些方法主要区别在于输入表示以及训练时优化的损失函数。已有输入表示包括事件图像 [108]、[147]、TS [22]、[133]、[153]、体素网格 [119]、[120] 和点集 [121]（第 3.1 节）。分类任务的损失函数使用人工标注的标签，而基于事件做回归的网络可由第三方真值（如位姿）监督 [108]、[147]，也可用对应灰度图测量光度一致性来监督 [22]，或完全无监督（只依赖训练输入事件）[119]、[153]。[99] 研究了从事件无监督学习的损失函数。架构方面，多数网络采用编码器-解码器结构，如图 4。这种结构可只用卷积，以尽量减少网络权重数目。此外，可在解码器的各个空间尺度上施加损失函数。

图 04 从体素网格到光流与自运动 / 原文第 9 页

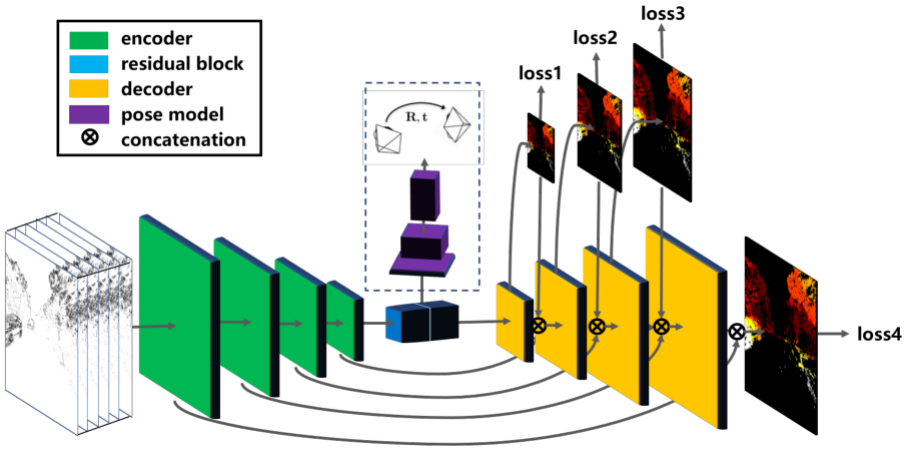


图 4. 时空体中的事件被转换为插值体素网格（左），再送入 DNN，以无监督方式计算光流和自运动 [119]。这样就能用适应事件数据的新损失函数（如运动补偿损失）复用现代基于张量的 DNN 架构。

图中术语：encoder = 编码器，residual block = 残差块，decoder = 解码器，pose model = 位姿模型，concatenation = 拼接，loss 1-4 = 各尺度损失；R、t 表示旋转与平移。

最后，运动补偿用于估计与一组事件最匹配的运动参数。它采用连续时间变换模型，可利用事件的精细时间分辨率（第 3.1 节），有别于传统的图像算法。运动补偿可用于估计自运动 [127]、[128]、光流 [119]、[128]、[130]、[154]、深度 [19]、[99]、[128]、运动分割 [132]、[154]、[155]，或估计视觉惯性里程计（VIO）中的特征运动 [129]、[131]。[104] 也采用连续时间运动模型，不过是用来融合事件数据和 IMU 数据，并不用于运动补偿。求连续时间运动模型参数 [99]、[104] 时，可以使用共轭梯度、高斯-牛顿等标准优化方法。

每组事件数（即时空邻域的大小）是许多方法的重要超参数。它高度依赖事件处理算法和可用资源，可采用不同的选取策略 [11]、[105]、[118]、[127]，例如固定事件数、固定观察时长（即固定帧率），或更自适应的方案（在图像平面各区域对事件数设阈值）[105]。采用固定事件数自然适应相机输出率，却没有考虑输出率随空间位置而变化。固定帧率选取的事件数会变化，随场景不同，可能过少也可能过多。通常更希望采用更能适应场景动态（时间与空间）的准则，但这类准则不易设计。

3.3 仿生视觉处理

生物学原理和计算基元指导着事件相机像素及某些事件处理算法和硬件的设计，例如脉冲神经网络（SNN）。

视觉通路： DVS [2] 的灵感来自生物视觉通路的功能。生物视觉的“在哪里”通路[10] 中，有专门处理动态视觉信息的“瞬态”通路。从昆虫到人类，动物都有这些瞬态通路。在人类，瞬态通路约占视觉系统的 30 %。它从瞬态神经节细胞开始，这些细胞大多位于视网膜中央凹之外；通路经丘脑的大细胞层和 V1 区的特定亚层，延伸至 MT 和 MST 区，后二者属于背侧通路，含有许多运动选择性细胞 [45]。DVS 对应瞬态通路中截至视网膜神经节细胞的部分。类似地，ATIS 的灰度（EM）事件对应经大脑小细胞层传导的“持续”或“是什么”通路 [36]、[43]。

原注 10 · https://en.wikipedia.org/wiki/Two-streams_hypothesis

使用 SNN 处理事件： 泄漏积分发放、自适应指数等人工神经元，是受哺乳动物视皮层神经元启发的计算基元，也是人工 SNN 的基本构件。一个神经元接收视觉空间一个小区域（感受野）的输入脉冲（“事件”），随之改变其内部状态（膜电位）；当状态超过阈值时，神经元产生输出脉冲（动作电位）。神经元按层级相连，形成 SNN。脉冲既可由事件相机像素产生，也可由 SNN 神经元产生。信息沿层级传递，从事件相机像素进入 SNN 的最初几层，再传到更高（更深）的层。第一层的感受野大多基于高斯差（对中心-周边对比敏感）、Gabor 滤波器（对特定方向的边缘敏感）及其组合。随信息深入网络，感受野越来越复杂。ANN 中的内层计算近似为卷积。人工 SNN 的一种常见做法是假定：若某神经元未收到前一层的输入脉冲，就不会产生输出脉冲。这样就可跳过这些神经元的计算。视觉处理结果几乎与刺激呈现同时出现 [156]；传统 CNN 则按固定时间间隔在所有位置同时计算卷积，二者很不一样。

任务： 仿生模型已用于多种低层视觉任务。例如，利用按时空方向排列的滤波器 [80]、[134]、[157] 模仿初级视皮层感受野的工作原理 [158]、[159]，可以估计事件光流。同类方向性滤波器也用于按 [161] 的生物学模型实现脉冲式选择性注意模型 [160]。源自双目视觉的仿生模型，如循环式横向连接与兴奋-抑制神经连接 [162]，被用于解决事件立体视觉的对应问题 [41]、[163]、[164]、[165]、[166]，或控制人形机器人的双眼辐辏 [167]。视皮层也为 [168] 的层级特征提取模型提供灵感，该模型用 SNN 实现并用于目标识别。这样的网络越善于从精确脉冲时序提取信息，性能越好 [169]。早期网络是人工设计的（如使用 Gabor 滤波器）[53]；近期研究则让网络通过脉冲时序依赖可塑性（STDP）等脑启发的学习方式形成感受野，从而提高识别率 [136]。作为补充，另一些方法在深层网络中采用更偏计算启发的监督学习方式（如反向传播），以高效实现脉冲深度卷积网络 [143]、[170]、[171]、[172]、[173]。相较传统的视觉方法，上述方法延迟更低、效率更高。

为构建小型、高效、反应灵敏的计算系统，昆虫视觉也是事件处理的灵感来源。为此，研究者为小型机器人开发了快速高效的避障和目标捕获系统 [174]、[175]、[176]。这些系统采用由 DVS 输出驱动的神经元模型，对迅速迫近的物体作出响应并触发逃避反射。

4 算法与应用

本节按所解决的任务回顾多项事件视觉研究。首先是图像平面上的低层视觉，例如特征检测、跟踪和光流估计；然后讨论与场景三维结构有关的任务，如深度估计、视觉里程计（VO），以及在研究发展中与之相关的课题，例如光强图像重建。最后，讨论运动分割、识别以及感知与控制的耦合。

4.1 特征检测与跟踪

图像平面上的特征检测和跟踪是许多视觉任务的基本构件，例如视觉里程计、目标分割和场景理解。事件相机可随场景动态异步跟踪，并具备低延迟、高动态范围和低功耗（第 2.2 节）。因此它能在标准相机两帧之间看不见的时间段内跟踪。为此，所开发的方法必须处理这一视觉信号独特的时空和光度特性：事件仅异步报告亮度变化（第 2.3 节）。

挑战：事件表示依赖运动方向的亮度变化，因此，使用事件相机进行特征检测与跟踪的一个主要难点，是克服运动依赖造成的场景外观变化（图 5）。跟踪要求在不同时刻的事件（或由事件构建的特征）间建立对应，即数据关联；外观变化使这一过程困难。另一主要挑战是应对传感器噪声和相机运动可能引起的杂乱事件。

图 05 运动方向变化使事件难以对应 / 原文第 10 页

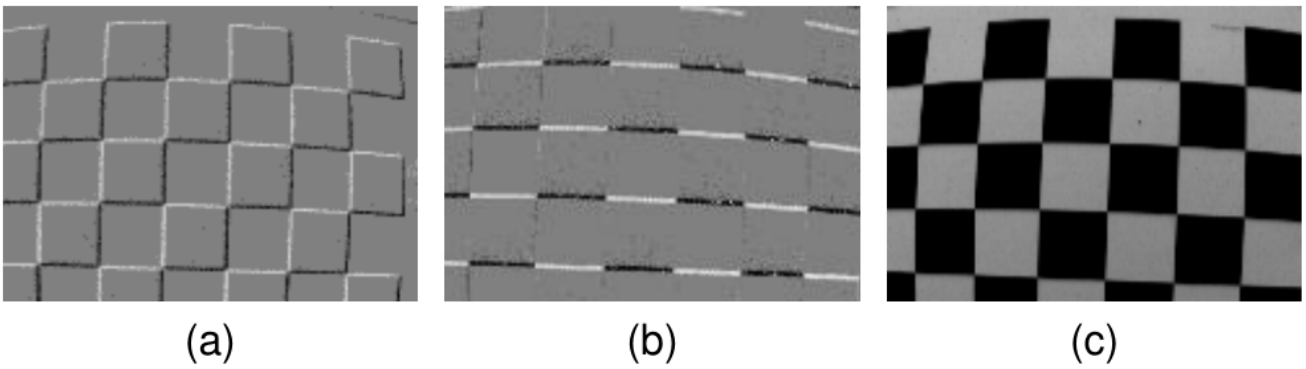


图 5. 数据关联的挑战。（a）和（b）显示场景（c）沿两个不同运动方向产生的事件：（a）斜向；（b）上下方向。光强增量图像（a）和（b）由短时间区间内事件极性累积而成：光强未变的像素为灰色，光强增加与下降的像素分别为亮色和暗色。由于（c）中边缘模式的外观随运动变化，（a）与（b）之间的事件对应关系显然不易建立。图改编自 [64]。

图中术语：（a）斜向运动；（b）上下运动；（c）原场景。本图的灰、亮、暗分别表示无变化、光强增加与光强下降。

文献回顾：早期的事件特征方法很简单，重在证明事件驱动视觉系统具有低延迟和低处理需求。因此它们假定相机固定，把移动物体作为呈团块聚集的事件源 [6]、[12]、[14]、[122]、[177]、圆 [178] 或直线 [54] 跟踪。只需处理产生事件的像素。简单的高斯相关滤波器足以检测事件斑团，斑团可用高斯混合模型表示 [179]。跟踪时，将每个新事件关联到最近的现有斑团或特征，异步更新其参数（位置、尺寸等）。圆 [178] 与直线 [54] 在霍夫变换空间中被视为斑团。这些方法用于交通监测与监视 [14]、[122]、[179]、机器人高速跟踪 [6]、[12]、流体中的粒子跟踪 [177] 以及微型机器人 [178]，但适用的目标形状有限。

对更复杂、高对比度、用户定义形状的跟踪，已采用迭代最近点（ICP）算法 [123]、梯度下降 [124]、均值漂移和蒙特卡洛方法 [180] 的逐事件实现，或粒子滤波 [181] 实现。[123]、[124] 的迭代方法用最近邻策略把新到事件关联到目标形状，并更新其变换参数，实现很高的跟踪速度（等效帧率 200 kHz）。其他研究 [180] 对目标形状（亦称“核”）建立一组旋转、缩放版本，与事件匹配以处理几何变换；预定义的核借助内置排斥机制，在跟踪时彼此不重叠。人脸、人体等复杂物体已用基于部件的形状模型跟踪 [182]；物体表示为由弹簧连接的一组基本元素 [183]。各部件跟踪器只需追随近似椭圆形状产生的事件斑团，由虚拟机械系统的弹性能给出跟踪质量判据。多数跟踪方法将

事件视为单独的点（不使用极性），以最低延迟异步更新系统状态。方法性能高度依赖多个模型参数的调节，通常须按所跟踪物体通过实验确定 [180]、[182]。

上述方法需要先验知识或用户输入来确定跟踪目标。在高速公路上跟踪汽车、跟踪奔向球门的球等场景中，这项限制可以接受，因为已知物体可大幅简化计算。然而当可能的目标种类增多时，就需要能确定更贴近实际场景的特征的方法。[125]、[130] 提出的特征是以点集表示的局部边缘模式，新到事件通过某种 ICP 方法与之配准。另一些方法 [27]、[129] 在运动补偿事件图像（第 3.1 节）的图块上复用成熟的特征检测器 [184] 和跟踪器 [185]，得到良好结果。这些方法都能跟踪在自然场景中运动的相机所观测到的特征，从而可在真实场景中估计自运动 [126]、[129]、[131]。由运动补偿事件构建的特征（图像 [129] 或点集 [130]）是边缘模式的有效表示。但它们依赖运动方向，所以随事件外观变化，跟踪器会发生漂移 [64]。要无漂移跟踪，必须使用对运动不变的特征。

结合事件与帧：如果可以获得待跟踪模式的绝对光强（图 5c，即对运动不变的特征表示或“地图”），数据关联（图 5）就更容易。利用一体式帧-事件传感器（如 DAVIS [4]）长处的工作采用这一做法。[64]、[125]、[126] 中的算法在帧上自动检测任意边缘模式（特征），利用事件异步跟踪。Harris 角点检测器 [184] 给出特征位置，角点周围的边缘模式给出特征描述符：[125]、[126] 将 Canny 边缘转成点集，作为 ICP 跟踪的模板，因此假定事件大多由强边缘触发。相比之下，[64] 通过帧光强给出边缘模式，并用生成模型（式 4）根据帧预测事件，寻找使实测与预测事件间光度误差最小的运动参数。[64] 比较了五种特征跟踪器，发现生成模型最准确，达到亚像素精度，尽管计算开销较大。最后，[64] 还表明仅用事件传感器就足够了：以事件重建的图像（第 4.6 节）代替帧，仍能达到相似的检测与跟踪效果。

角点检测与跟踪：由于事件相机天然响应场景边缘，它能加快较低层基元（关键点或“角点”）的检测。这些基元标出感兴趣的像素，可在其周围提取局部特征而不受孔径问题影响，因而提供可靠的跟踪信息。[186] 在时空事件流中拟合平面，得到两条移动边缘，并以两者的交点为角点。为应对事件噪声，在最小二乘之外辅以类似 RANSAC 的采样技术。这种在时间表面局部拟合平面的方法也有利于估计光流 [21] 和“事件寿命”[148]，二者均由平面系数获得。近期研究 [111]、[114]、[115] 将 Harris [184]、FAST [187] 等常用帧式关键点检测器扩展到事件相机，把时间表面（TS）当作自然光强图像处理。[114] 先将 TS 二值化，再应用 Harris 检测器的导数滤波。[115] 则为加快检测，用当前事件周围 TS 上两条同心圆上的逐像素比较替代导数滤波。运动角点所产生的局部 TS 有两个分界明显的区域，分别对应新近与较早的事件。因此，通过搜索 TS 值比其余像素高的连续像素弧段得到角点。[111] 改进了 [115] 的检测器，并提出角点跟踪策略。假定角点在图像平面上沿连续轨迹运动、检测到的事件角点准确，则利用树形假设图，按距离将它们沿轨迹串联。以上人工设计的 TS 角点检测器会受运动方向改变引起的 TS 变化影响。为克服这种影响，[112] 提出数据驱动方法，学习光强图像角点在 TS 上的外观；为此使用灰度输入（来自 DAVIS 或 ATIS 相机）提供角点标注的监督信号。为权衡精度与速度，使用随机森林分类器。事件角点可用于视觉里程计、自运动分割 [188] 等多种任务，不过已有演示很少。

机会：检测和跟踪方法虽多，却很少用共同数据集对比性能。建立基准数据集 [189] 和评估流程将推动本领域及其他相关领域的发展。此外，多数算法按跟踪目标通过实验设定参数。若能自适应调参，将有望扩大跟踪器的适用范围。基于学习的特征检测与跟踪也有较大的研究空间。

4.2 光流估计

光流估计是指在不知道场景几何或运动的情况下，计算物体在图像平面上的速度。这个问题是不适定的，因此需要正则化才能求解。

事件光流估计的难点在于事件以不熟悉的方式编码视觉信息（第 2 节）。传统相机通过分析相邻两幅图像得到光流。两幅图像提供空间和时间导数，代入亮度恒常假设（第 4 页），再结合平滑性假设，就有足够的方程求解每个图像像素的光流。事件既不提供绝对亮度，也不提供空间连续的数据。单个事件所含信息不足以确定光流，需要汇聚事件才能估计，于是出现了一个不同寻常的问题：在事件张成的 x–y–t 图像时空中，应该在哪些位置计算光流？理想情况下，人们想知道整个空间的流场，但计算代价高。实践中只在指定点计算光流：在事件位置，或按人为选定的时刻生成的图像上。尽管如此，从事件计算光流仍有吸引力：测量对象是边缘，而场景边缘处的光流歧义较小；精细的时间信息也使高速光流测量成为可能 [11]。另一挑战是设计一种光流算法，它既符合已知的灵长类早期视皮层处理机理（具备生物合理性），又可在神经形态处理器上高效实现。

文献回顾： 表 2 按不同准则列举了一些事件光流方法。早期研究 [191] 尝试将经典计算机视觉方法改用于事件数据（图 6b）。这些方法基于亮度恒常假设 [185]，讨论集中在事件是否带有足以用此类方法估计光流的信息 [134]。事件允许估计亮度时间导数（式 3）；要应用这类经典方法 [185]，还需附加假设来近似空间导数 ∇L。但边缘掠过像素时，每个像素可能只产生极少的事件，可靠估计导数 $(\nabla L, \partial L/\partial t)$ 很难 [134]，使 [191] 之类基于梯度的方法难以给出确定的光流估计。像 [21] 那样考虑事件在 x–y–t 空间中的局部分布更稳健，因而更可取。

表 02 光流方法分类 / 原文第 12 页

Reference	N/F?	S/D?	Model?	Bio?
Delbruck [80], [190]	Normal	Sparse	Model	Yes
Benosman et al. [190], [191]	Full	Sparse	Model	No
Orchard et al. [157]	Full	Sparse	ANN	Yes
Benosman et al. [21], [190]	Normal	Sparse	Model	No
Barranco et al. [192]	Normal	Sparse	Model	No
Barranco et al. [193]	Normal	Sparse	Model	No
Conradt et al. [194]	Normal	Sparse	Model	No
Brosch et al. [134], [195]	Normal	Sparse	Model	Yes
Bardow et al. [117]	Full	Dense	Model	No
Liu et al. [105]	Full	Sparse	Model	No
Gallego [128], Stoffregen [154]	Full	Sparse	Model	No
Haessig et al. [196]	Normal	Sparse	ANN	Yes
Zhu et al. [22], [119]	Full	Dense	ANN	No
Ye et al. [153]	Full	Dense	ANN	No
Paredes-Vallés [102]	Full	Sparse	ANN	Yes

表 2. 按输出与设计对若干光流方法进行分类。某些方法给出完整光流（F），另一些只给出沿局部亮度边缘法线的光流分量（N）。输出可以是稠密（D）流场，即某一时刻每个像素的光流；也可以是稀疏（S）流场，即只在选定的像素处计算光流。按设计区分，方法可基于模型或无模型（人工神经网络，ANN），也可按是否受神经生物学启发区分。

表中术语：Reference = 文献；N/F = 法向 / 完整光流；S/D = 稀疏 / 稠密光流；Model? = 算法设计（Model：基于模型；ANN：人工神经网络）；Bio? = 是否受神经生物学启发。

图 06 稀疏与稠密光流示例 / 原文第 12 页

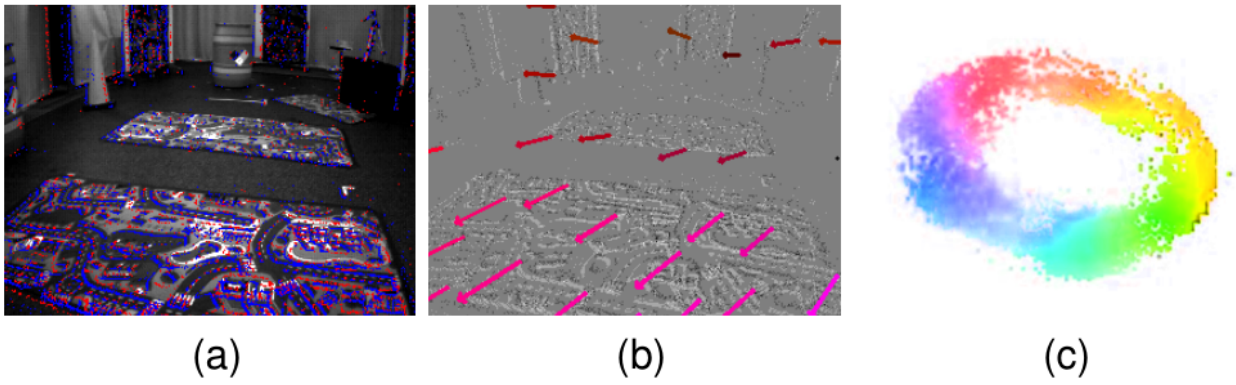


图 6. 两个光流估计示例。（a）和（b）为室内飞行场景 [197]。（a）将事件（极性以红色 / 蓝色标示）叠加在来自 DAVIS 的灰度帧上；（b）显示用 [185] 从亮度增量图像算出的稀疏光流（按大小和方向着色）。（c）另一个场景：黑暗环境下以 750 °/s 旋转的指尖陀螺的稠密光流 [119]。事件使复杂条件下的光流估计成为可能。

图中术语：（a）红 / 蓝区分事件极性；（b）颜色表示光流大小与方向；（c）旋转陀螺的稠密光流。

[21] 用时间表面和平面近似，从几何角度分析事件的局部分布。边缘移动时产生的事件类似于时空中一个表面上的点（时间表面，第 3 节）。表面在 x–t 和 y–t 截面上的斜率编码了边缘运动，因此可对表面拟合平面，从平面系数读出斜率，从而估计光流。虽然这种方法仅给出法向光流（即垂直于边缘的分量），但即使只有少量事件也适用。当然，拟合质量取决于时空邻域大小，这也适用于其他方法。邻域太小，拟合平面可能不确定；邻域太大，局部平面可能无法很好地近似事件流。

[134] 根据灵长类视觉系统的实验发现，提出一种层级光流估计架构。它在事件流上应用一组时空滤波器，对不同运动速度和方向产生选择性（类似 Gabor 滤波器），同时保留事件的稀疏表示。这类滤波器在形式上等价于时空相关检测器。其他仿生方法 [102]、[157] 也可解释为滤波器组，沿不同的时空方向对事件流采样；[157] 与 [134] 使用人工设计的滤波器，而 [102] 用新型 STDP 规则从事件数据学习滤波器。[157] 的 SNN 经突触连接延迟事件，使神经元充当符合检测器，检测运动模式。神经元在 5 × 5 像素的感受野内对 8 种速率和 8 个方向（即 64 种速度）敏感。这些方法可在神经形态硬件中实现，以低功耗高效计算。

[193] 针对纹理区域的光流估计。它先将事件转换为事件帧，再应用 Gabor 滤波器组。假设事件帧的（时空）边缘保持不变，则由滤波器组输出的相位梯度 [198] 得到光流。受孔径问题限制，只能计算沿空间相位梯度方向的速度。

[23]、[117] 等方法将光流与其他量（尤其是图像光强）联合估计，使各个量能引入成熟的方程，并在求解过程中相互促进收敛。已知图像光强或等价地已知 $(\nabla L, \partial L/\partial t)$ 是有利的，因为可将其代入亮度恒常定律，约束光流。[117] 把多个关系（式 2、亮度恒常、平滑性先验等）作为惩罚项纳入同一个目标函数，通过变分法优化。该方法在图像平面上求使目标函数最小的光流与图像光强，即找到最能解释事件在 x–y–t 空间中分布的解（使用体素网格）。它因此输出稠密光流（即每个像素都有光流）。未产生事件的像素（对应亮度均匀区域）上的光流向量来自平滑性先验，因此不如触发了事件的像素（即边缘）处的向量可靠。

[105] 以自适应速率计算事件帧（第 3 节），采用视频编码技术（块匹配）估计光流。它可解释为寻找最能匹配两个长方体内事件分布的光流向量，两个长方体各被压缩成事件帧。因此，光流问题被表述为寻找事件之间的对应，即同一个场景点在不同时间触发的事件。该方法定义两个事件集合（“

块”）以及比较它们的相似度度量。假定事件帧的外观在短时间内变化不大，所以用绝对距离和等简单度量即可比较。它可以在 FPGA 上实现，在效率和精度之间进行权衡。

[99]、[128]、[154] 的框架将长方体内的事件变换为运动补偿图像，使所得图像块的锐度最大，以求得光流（第 3 节）。这可理解为对事件施加一个自适应滤波器，滤波器系数定义使其响应最大的时空方向。[130] 也用运动补偿计算光流，但使用点集。

近期出现了深度学习方法 [22]、[119]、[153]。这些方法将大量可用的事件数据与 ANN 结合。[22] 采用自监督方案训练一个编码器-解码器 CNN 来估计稠密光流。损失函数度量用网络输出的光流对齐之后，DAVIS 灰度图像间的误差。训练后的网络只凭事件即可准确预测光流，输入为时间表面和事件帧。[153] 给出首个仅凭事件估计稠密光流、深度及自运动（即学习运动恢复结构）的单目 ANN 架构。ANN 输入为跨多个时间片的事件，形式是事件帧以及采用平均时间戳的时间表面。这样比 [22] 更能降低事件噪声、保留事件流结构。网络以无监督方式训练，度量用所估光流对齐后的相邻时间片间的事件光度误差。随后，[119] 使用基于时间表面的运动补偿损失函数，把 [22] 扩展为对光流和自运动的无监督学习。

评估：估计光流的计算开销很高。某些方法 [22]、[117]、[119]、[153] 需要 GPU；另一些方法 [105] 较轻量，但精度不及前者。只有少数算法 [21]、[105]、[134]、[157] 已移植到硬件逻辑电路，以减轻 CPU 负载并尽量降低延迟。综述 [190] 比较过早期若干事件光流方法 [21]、[80]、[191]，但仅比较了由旋转相机生成的流场，即缺乏运动视差与遮挡。对于较新的方法，精度、效率与延迟之间的多重权衡尚未得到适当量化。

机会：要评估这个关键低层视觉任务的进展及可复现性，有必要建立综合数据集，在多种场景（不同纹理、速度、视差、遮挡、照明等）下提供准确的光流真值，并建立共同的评估方法。为真实场景提供事件光流真值不容易，尤其是对不符合相机自运动所致运动场的移动物体。全面量化比较已有的事件光流方法，可帮助识别改进方法的关键思路。

4.3 三维重建：单目与立体视觉

用事件相机估计深度是一个宽广的领域。可按场景、相机配置或运动分类，因为这些条件决定着问题的假设。

瞬时立体视觉：多数事件相机深度估计研究针对“瞬时”立体视觉，即用两台或更多刚性相连且同步的相机，在极短时段内采集的事件（理想情况下逐事件）进行三维重建。因相机同步，不同图像平面的事件共享同一时钟。此类研究遵循立体视觉的经典两步法：先解决跨图像平面的事件对应问题（极线匹配），再三角测量三维点的位置 [199]。主要难点是确定事件对应关系，这一步计算量最大。事件匹配可以：（i）在事件帧 [145]、[200] 或时间表面 [149]（第 3 节）上使用传统立体视觉度量（如归一化互相关）；（ii）利用不同传感器事件的同时性和时间相关性 [149]、[201]、[202]；也可以并用两者。这些是局部方法，通过比较事件邻域进行匹配，因为仅凭单个时间戳无法匹配事件 [170]、[203]。可利用极线约束 [204]、顺序、唯一性、边缘方向及极性等附加约束，降低匹配歧义和错误对应，从而改善深度估计 [18]、[170]、[205]。也可比较两个立体图像平面上事件空间分布的局部上下文描述符 [206]、[207] 进行匹配。

全局方法考虑附加的规则性约束，深度估计比局部方法更好，即对歧义更不敏感。这类方法包括将 Marr 和 Poggio 的协同立体视觉算法 [162] 扩展到事件相机 [41]、[164]、[165]、[166]、[208]。此类方法构建对视差敏感的神经元网络，接收两台相机的事件，通过放大、抑制等操作实现匹配约束

（唯一性、连续性），提取视差。它们既用时间相似性匹配事件，也用时空邻域，并通过迭代的非线性运算得到整体的全局最优解。协同立体视觉的讨论见 [43]。[209]、[210]、[211] 也属此类，它们用马尔可夫随机场上的置信传播或半全局匹配 [212] 改善立体匹配。这些方法主要基于优化，试图定义性质良好的能量函数，使其最小化解对应正确的对应图。能量函数包含规则性约束，耦合邻近 点的对应关系，使解比局部方法更不易受歧义影响，代价是更多计算。[213] 给出了一张比较不同立 体视觉方法的表，但须谨慎解读，因为这些方法并未用同一数据集作基准测试。

近期 [214] 提出用专用硬件（GPU）进行穷举式空间扫描。方法采用与 [19]、[128] 相似的思路：在视差空间图像 [19]、[215] 中，正确深度表现为经位移的事件落在“聚焦”体素。相比之下，另一些方法结合事件相机与神经形态处理器（第 5.1 节），构成完全基于事件、低功耗（100 mW）的高速立体视觉系统 [165]、[213]。效率与准确率之间的权衡尚未量化。

上述多数方法在相机固定且只有少数物体运动的场景中进行演示，因事件数据不杂乱，较容易建立对应。事件匹配延迟低、速率高（约 1 kHz）且功耗低，说明事件相机有望对运动目标或无杂乱干扰的场景进行高速三维重建。

多视点全景图：也有研究 [216]、[217] 针对瞬时立体视觉（用极短时间区间内的事件生成深度图），但采用两台非同时采集的事件相机。这些方法利用受约束的硬件配置（两台运动已知的旋转事件相机），或者（i）恢复光强图像，再应用传统立体视觉 [216]，或者（ii）以时间度量匹配事件 [217]。

单目深度估计：[19]、[25]、[128] 已演示用一台事件相机估计深度。这与前述问题显著不同，因为不能利用多个图像平面间事件的时间相关性。这些方法累积运动相机不同时刻的事件信息，恢复场景的半稠密三维重建（即三维边缘图），因而需要知道相机运动。因此，它们并不追求瞬时深度，而是为 SLAM 估计深度 [218]。

[25] 的方法属于一个流程，采用三个并行工作的滤波器，联合估计事件相机运动、场景三维地图和光强图像。其深度估计需要用附加量--光强图像--完成数据关联。相较之下，[19]（图 7）提出一种空间扫描方法，利用事件流的稀疏性进行三维重建，不需要建立事件匹配或恢复光强图像。它把事件反投影到空间中，得到射线密度体 [219]，再以射线密度的局部极大值确定场景结构。该方法计算高效，[26] 将其用于视觉里程计。

图 07 手持单目事件相机的深度图 / 原文第 14 页



(a) (b)

图 7. 手持事件相机单目深度估计的示例。（a）场景；（b）半稠密深度图，以红（近）到蓝（远）的伪彩色标记。图片由 [19] 提供。

图中术语：半稠密深度图只在有估计的位置着色；红色近，蓝色远。

SLAM 中的立体深度：近期，受小基线多视图立体视觉［220］的启发，［116］提出一个面向 SLAM 的立体深度估计方法。它借助时间表面，优化跨图像平面事件的局部时空一致性，获得半稠密场景三维重建。与经典的事件匹配加三角测量范式［149］不同，它采用前向投影，可在不显式建立事件对应关系的情况下估计深度。该方法有望将事件相机的优势带到自动驾驶汽车等事件立体 SLAM 应用中。

用结构光估计深度：上述三维重建方法均是被动式的，即不干预场景。相比之下，也有基于事件的主动三维重建研究，向场景投光，并以事件相机测量反射［20］、［221］、［222］。例如，［221］把 DVS 与脉冲线激光器结合，像三维线扫描仪一样快速重建地形。运动对比度三维扫描［20］是一种结构光技术，在复杂的三维扫描环境（如强光、高反射表面或运动表面）下可同时实现高分辨率、高速度和稳健性能。使用脉冲激光的主动系统利用了事件相机的高时间分辨率和冗余抑制特性，但其适用场景有限，而且未必安全（是否安全取决于扫描远处物体所需的激光功率）。

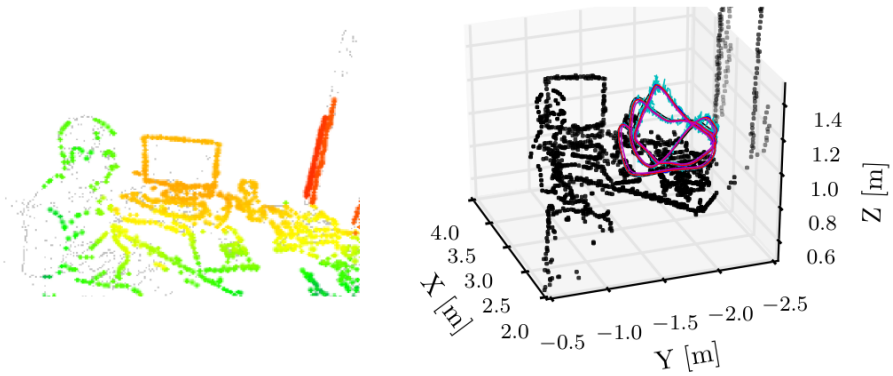
机会：事件深度估计方法虽多，却未在同一数据集上评估，很难比较性能。因此希望：（i）提供全面的事件深度评估数据集与测试平台；（ii）在该数据集上对多种已有方法进行基准测试，以便比较性能。

4.4 位姿估计与 SLAM

采用事件相机解决同步定位与建图（SLAM）问题一直很困难，因为大多为传统相机开发的方法与概念（特征检测、匹配、迭代图像对齐等）不适用，或当时尚不可用；事件与图像有根本区别。挑战是设计新的 SLAM 技术，充分发挥相机的优势（第 2.3 和 2.2 节），显示它有助于应对当前帧相机难以处理的场景。历史上，这类技术的设计着重保留数据的低延迟特性，即能为每个到达的事件产生状态估计（第 3 节）。然而每个事件所含信息不足以从头估计状态（如已标定相机的六自由度〔DOF〕位姿），所以目标变成让每个事件能异步更新系统状态。概率（贝叶斯）滤波器［223］自然符合这一描述，因此在事件 SLAM 中很常用［7］、［24］、［75］、［224］。为适配事件相机，主要需要根据事件生成过程（第 2.4 节）设计合理的似然函数。

事件是由光强边缘的表观运动引起的，因此多数 SLAM 系统生成的地图自然只包含场景边缘，即半稠密地图（图 8 与［19］）。但应注意，事件相机不直接测量光强梯度，只测量其时间变化（式 2）；因此边缘的存在与否、方向和强度（在图像平面和三维空间中）必须与相机运动一起估计。场景上一点的光强梯度幅值与该点对应事件的触发率相关，可帮助实现可靠跟踪［103］。用于跟踪的边缘信息也可从生成模型（第 2.4 节）采用的亮度图［7］、［24］、［25］的梯度中获得。

图 08 半稠密地图与相机轨迹 / 原文第 16 页



(a) Events & projected map. (b) Camera trajectory & 3D map.

图 8. 事件 SLAM。（a）[98] 中的重建场景：将按深度着色的重投影半稠密地图叠加在事件（灰色）上，可以看出地图与事件对齐良好。（b）估计的相机轨迹（多种方法）与半稠密三维地图（即点云）。图片由 [104] 提供。

图中术语：Events & projected map = 事件与投影地图；Camera trajectory & 3D map = 相机轨迹与三维地图；X、Y、Z 的单位为 m。

最一般情形的事件 SLAM（6 自由度运动和自然三维场景）颇具挑战，研究按难度递增的场景逐步推进。可沿三个维度描述复杂性：问题维数、运动类型和场景类型。文献中多数方法先处理定位子问题（估计运动），因为需估计的自由度较少。运动类型方面，在解决最复杂的自由相机运动（6 自由度）之前，研究先考虑了受约束的运动，例如旋转或平面运动（二者均为 3 自由度）。场景方面，先有针对光度（高对比度）或结构（线条地图或二维地图）具有人工设定的场景的解法，之后才转向最难的自然场景（三维且有任意光度变化）。有些方案需要额外传感（如 RGB-D）以降低问题复杂度，但这会引入一些帧系统中的瓶颈（如延迟和运动模糊）。表 3 按上述复杂性维度对相关工作分类。

表 03 位姿跟踪与建图方法 / 原文第 15 页

Reference	Dim	Track	Depth	Scene	Event	Additional requirements
Cook [23]	2D	✓	✗	natural	✓	rotational motion only
Weikersdorfer [224]	2D	✓	✗	B&W	✓	scene parallel to motion
Kim [24]	2D	✓	✗	natural	✓	rotational motion only
Gallego [127]	2D	✓	✗	natural	✓	rotational motion only
Reinbacher [103]	2D	✓	✗	natural	✓	rotational motion only
Censi [75]	3D	✓	✗	B&W	✗	attached depth sensor
Weikersdorfer [225]	3D	✓	✓	natural	✗	attached RGB-D sensor
Mueggler [226]	3D	✓	✗	B&W	✓	3D map of lines
Gallego [7]	3D	✓	✗	natural	✗	3D map of the scene
Rebecq [19]	3D	✗	✓	natural	✓	pose information
Kueng [126]	3D	✓	✓	natural	✗	intensity images
Kim [25]	3D	✓	✓	natural	✓	image reconstruction
Rebecq [26]	3D	✓	✓	natural	✓	–

表 3. 使用事件相机进行位姿跟踪和 / 或建图的事件方法。运动类型以“2D”（3 自由度运动，如平面运动或旋转）和“3D”（三维空间中自由的 6 自由度运动）标记。各列指出方法是否做跟踪（Track）和深度估计（Depth），是否只用事件（Event），考虑的场景类型（Scene），以及任何附加要求。只有 [25]、[26] 在最一般的场景中仅用事件解决问题。

表中术语：Reference = 文献；Dim = 运动维度类别；Track = 跟踪；Depth = 估计深度；Scene = 场景（natural：自然；B&W：黑白）；Event = 仅使用事件；Additional requirements = 附加要求（仅旋转运动、场景平行于运动平面、附加深度 / RGB-D 传感器、三维线图、场景三维地图、位姿信息、光强图像、图像重建、无）。

相机跟踪方法：[101] 首次提出用事件相机跟踪位姿。它用粒子滤波器跟踪地面机器人运动；机器人观察的是与运动平面平行、由人工黑白线条图案构成的平面场景。主要创新是设计似然函数，在给定机器人位姿与场景地图时量化事件生成概率。该函数基于事件位置和地图最近边缘之间的重投影误差。[75] 把标准灰度相机装在 DVS 上，用贝叶斯滤波器估计当前事件和标准相机上一帧之间的微小位移。系统针对平面运动和黑白场景；似然函数与事件所在位置的地图梯度幅值成正比（第 2.4 节）。[227] 提出非完整约束下平面运动的位姿跟踪，支持闭环检测及拓扑正确的轨迹。该方法先将事件转换成帧，再应用传统方法 SeqSLAM。

[24]、[103]、[127] 研究了事件相机的三维朝向估计。[24] 用粒子滤波器跟踪相机的旋转运动，似然函数为以对比度灵敏度 $C \approx 0.2$ 为中心的近似高斯分布（第 2.4 节）。[103] 通过最小化事件位置的光度误差来估计相机旋转，使用表示场景边缘强度的事件概率图 [224]。最后，[127] 引入运动补偿优化框架（第 3 节），用于估计相机角速度，而非绝对旋转。这些系统限于旋转运动，不考虑平移或深度，但为解决更复杂的问题提供了思路，例如 [25]、[128]、[129]。

6 自由度的相机跟踪最难。[226] 提出事件算法，在极高速运动中跟踪 DVS 位姿。这是针对人工黑白线条地图的人工设计方法，而不是概率滤波器。它假定事件仅由一条地图直线（离事件最近的直线）生成，对这些直线进行跟踪并求交，为 PnP 方法（相机后方交会）提供点。[228] 给出这种方法的连续时间表述和扩展，通过非线性优化事件到直线的重投影误差来求相机轨迹，而不是独立的位姿。相比之下，[7]、[65] 证明了自然场景中的 6 自由度高速跟踪能力。它们使用生成模型：[7] 提出采用混合密度构造稳健似然函数的概率滤波器（第 2.4 节）；[65] 则对亮度增量图像（第 3.1 节）与由场景地图预测的图像之间的光度误差做非线性优化。后者结果略好。

跟踪与建图：下面来看同时解决跟踪与建图的方法。Cook 等人 [23] 提出交互网络中的通用消息传递算法，从事件中联合估计自运动、图像光强和光流，但系统仅适用于旋转运动。联合估计有吸引力，因为可以尽可能利用联系各变量的方程（如式 4 与旋转先验），以期得到更好的解。

[224] 扩展 [101] 得到二维事件 SLAM 系统，因而也限于平面运动和高对比度场景。该方法使用粒子滤波器跟踪，事件似然与事件相对地图的重投影误差负相关。算法同时构建场景边缘地图，形式为占据图 [223]，每个像素代表该像素触发事件的概率。[225] 将方法扩展到三维，但依赖与事件相机相连的外部 RGB-D 传感器估计深度。深度传感器引入瓶颈，使系统失去事件相机低延迟和高速的优势。

[24] 的滤波方法展示了如何在跟踪旋转事件相机三维朝向的同时，构建自然场景的高分辨率全景图。它并行运行概率滤波器以解决两个子任务。逐像素的卡尔曼滤波器估计各位置场景边缘的方向和强度，构成全景梯度图。再通过泊松积分将梯度图转化为绝对光强图，具有超分辨率和 HDR 特性。[103] 同样展示了旋转运动下的 SLAM：给定概率边缘图，最小化事件位置上的光度误差以跟踪相机。地图同步构建，每个地图点表示该位置产生事件的概率 [224]。因此它是度量场景边缘强度的全景占据图。

近期，[25]、[26]（表 3）给出不依赖附加传感、面向自然场景与 6 自由度运动的完整三维事件 SLAM 解法。[25] 扩展 [24]，以三个交错运行的概率滤波器跟踪位姿、估计深度和光强。但它假设深度、光强梯度与相机运动互不相关，稳健性有限（尤其在初始化时）。此外，其计算量大，实时运行需要 GPU。相比之下，[26] 的半稠密方法表明，深度估计或位姿跟踪并不需要重建光强。其方法以几何为基础：进行空间扫描来重建三维结构 [19]，通过边缘图对齐跟踪位姿（用每帧少量事件进行非线性优化）。由此得到的 SLAM 系统可在 CPU 上实时运行。

以一定延迟换效率，概率滤波器 [24]、[25]、[224] 可以处理小组事件。其他方法一开始就是为事件组设计的，例如基于非线性优化的方法 [26]、[127]、[128]，可在 CPU 上实时运行。同时处理多个事件也有利于降噪。

机会：上述 SLAM 方法缺少用于减少漂移的闭环能力。目前演示事件 SLAM 的场景尺度明显小于帧式 SLAM。不过，追求两者相同的尺度或许不合理，因为事件相机未必与标准相机解决同样的问题；两类传感器是互补的，如 [7]、[27]、[64]、[75] 所述。立体事件 SLAM [229] 尚未得到充分探索，研制比现有单目方案更准确、高效、稳健的方法也是如此。与 IMU 进行传感器融合可提高 SLAM 的稳健性 [27]、[218]。

4.5 视觉惯性里程计（VIO）

将事件相机与刚性连接的惯性测量单元（IMU）结合，可以提高事件视觉里程计与 SLAM 的稳健性。VIO 与 VO 有本质区别：VO 系统可能丢失跟踪（即无法输出位姿）；VIO 系统则不会“失效”（总有位姿输出），而是会漂移。出于这个原因及其他原因，一些事件相机集成了 IMU（见表 1）。

VIO 的一个关键问题，是怎样在时间上融合来自同步高采样率 IMU（例如 1 kHz）与异步事件相机的数据。文献有三种做法：（i）采用异步概率滤波器 [131]；（ii）用预积分理论 [230] 在所需的事件采集时刻把 IMU 数据转成较低频率的信息片段 [27]、[129]、[131]；（iii）采用连续时间框架，将两个传感器的测量放到同一条时间轴上，而该时间轴也用于对解（即相机位姿）建模 [104]。

基于特征的 VIO：已有事件 VIO 系统大多“基于特征”，可分两步：首先从事件提取特征轨迹（第 4.1 节），随后使用现代 VIO 算法（如 [230]、[231]、[232]）将图像平面上的点轨迹与 IMU 数据融合。即前端将事件传递的光度信息转为几何信息，然后由高度优化的几何 VIO 流程处理。例如，[131] 用 [130] 跟踪特征，并用卡尔曼滤波器 [231] 将其与 IMU 数据结合。近期，[129] 提出由时空窗口内事件合成运动补偿事件图像（第 3 节），然后用经典方法 [185]、[187] 检测和跟踪特征。通过基于关键帧的非线性优化 [232]，将特征轨迹与惯性数据融合，恢复相机轨迹和稀疏三维路标地图。随后，[27] 扩展 [129] 的非线性优化，使之也能融合光强帧，并在资源受限的平台（四旋翼）上演示，借助事件相机的优势使其能在低光照和 HDR 场景中飞行。上述方法在 6 自由度运动数据集 [98] 上进行了基准评估，每种方法均优于所承接的前一方法。

基于重投影误差的 VIO：[104] 提出不同的方案，通过连续时间框架 [233] 融合事件和惯性数据。与上述基于特征的方法不同，它沿一段相机轨迹优化一个同时包含惯性误差项和事件重投影误差项的目标函数，做法类似视觉惯性光束法平差。

机会：上述研究先把事件转成几何信息，再适配面向传统相机的先进 VIO 方法。但如 [128] 所提示，或许可以避免转换，直接从事件恢复相机运动和场景结构。例如，类似面向标准相机的 VI-DSO [234]，优化一个包含光度（即事件触发率 [26]）和惯性误差项的函数。

立体事件 VIO 仍待探索，将事件深度估计的思路与 SLAM 和 VIO 结合，值得研究。

上述所有问题也可尝试采用基于学习的方法。目前文献主要采用基于模型的方法，但如同帧视觉，我们预计基于学习的方法也将在事件处理中发挥重要作用。相关尝试见 [108]、[109]、[113]、[119]。

4.6 图像重建

事件表示亮度变化，因此在理想条件下（无噪声、传感器响应理想等），对事件积分可以得到“绝对”亮度。这一点直观，因为事件只是对场景视觉内容进行逐像素的非冗余（即“压缩”）编码。事件积分，或更一般地说图像重建（图 9），可理解为对事件流所编码的视觉数据进行“解压”。由于事件的时间分辨率极高，可以用极高帧率（如 2 kHz–5 kHz [8]、[235]）重建亮度图像，甚至连续地重建 [62]。

图 09 隧道出入口与高速爆裂场景 / 原文第 17 页



图 9. 图像重建。在汽车驶出隧道的场景中，消费级相机（Huawei P20 Pro）的帧（左）存在欠曝或过曝。事件能捕捉更宽的场景动态范围，通过图像重建方法可将其恢复（中）。事件还能用于重建高速场景，例如爆裂的马克杯（右）。图片由 [8]、[237] 提供。

图中术语：Tunnel scene = 隧道场景；Exploding mug = 爆裂的马克杯。

文献显示，事件图像重建的关键是需要正则化。事件相机各像素独立报告亮度变化，在一个时间区间内逐像素积分这些变化，只能得到亮度增量图像。要恢复区间末端的绝对亮度，需要在增量上加一幅偏置图像（即区间起始时的亮度图像）[78]、[98]。值得注意的是，有些工作利用空间或时间平滑，或同时利用两者 [62]、[135]、[235]、[236]，从零初值重建亮度，即不必知道偏置图像。利用从自然场景学习得到的特征做正则化 [8]、[118]、[120]、[235] 也有效。

文献回顾：[23] 首次在假定相机旋转、场景静止的条件下完成从事件重建图像。该方法在视觉地图网络中基于这些假定及亮度恒常关系（式 4），以像素间消息传递算法联合估计场景亮度等多个量。同样在上述运动和场景假定下，[24] 展示了如何从事件重建高分辨率全景图，并推广了事件 HDR 图像重建的思路。全景图的每个像素使用卡尔曼滤波器估计亮度梯度（基于式 4），再用泊松重建进行积分以获得绝对亮度。[238] 利用平台绕单轴旋转这一受约束的运动重建图像，再用这些图像估计立体深度。

随后，[117] 用正则化假设替代运动限制，使图像重建适用于一般的运动和场景。该工作通过含有多个惩罚项（数据拟合项，式 1；以及解的平滑性项）的变分框架，同时估计图像亮度与光流，以最好地解释离散为体素网格的时空事件体。它首次展示了在动态场景中由事件重建视频。后来，[135]、[235]、[236] 表明，甚至不估计运动也能重建图像。实现途径包括基于时间表面的变分图像去噪 [135]、[236]，或用从图像块学习的字典进行稀疏信号处理，将事件映射为图像梯度，再进行泊松积分 [235]。同期，[25]、[26] 的 VO 方法利用算出的场景深度和位姿，把 [24] 的图像重建技术扩展至 6 自由度相机运动：[25] 用稳健变分正则项减少噪声且提升重建图像的对比度，[26] 则将图像重建作为附带结果，因为实现 VO 并不需要它。近期，[62] 提出时间平滑滤波器以重建图像和持

续融合事件与帧。该滤波器在各像素独立工作，表明恢复亮度不需要在图像平面上作空间正则化，不过滤波器在自然地减少噪声与伪影的同时，也会牺牲一部分真实细节。更近的 [8]、[120] 提出深度学习方法，较既有方法有明显提升，并减轻了视觉伪影。回顾早期工作，曾用来实现图像重建的运动限制或人工设计的正则项，已被从自然场景取得的感知与数据驱动先验所取代，由此生成的图像看起来更自然。应注意，用于 VO 或 SLAM 的图像重建方法 [23]、[24]、[25] 假定场景静止；对运动做弱假设或不做假设的方法 [8]、[62]、[117]、[120]、[135]、[235]、[236] 则自然适用于重建任意场景（如动态场景）的视频。

除了从事件重建图像，另一类方法融合事件与帧（如来自 DAVIS [4]），以事件的高时间分辨率和 HDR 特性增强帧的亮度信息 [28]、[62]、[78]。这些方法也不依赖运动知识，最终都以式（2）为基础。[78] 在帧之间对事件直接逐像素积分。不过融合后的亮度很快受事件噪声污染（源于非理想效应、灵敏度失配、事件丢失等），所以每到一帧都重置融合。[62] 为减轻噪声，使用逐像素时间互补滤波器融合事件与帧：对事件作高通、对帧作低通。它考虑两种传感方式的互补性，因而高效：帧承载缓慢变化的亮度信息（时间低频），事件承载“变化”信息（高频）。[28] 的融合方法进一步借事件的高时间分辨率去除帧的运动模糊，从一幅模糊帧和事件生成高帧率、清晰的视频。它在优化框架中使用双重积分模型（一次积分恢复亮度，另一次积分去模糊）。上述方法的限制之一是仍有事件噪声所致伪影；与基于学习的方法 [8] 结合，或可缓解。

图像质量：重建质量直接受对比度阈值噪声影响（第 2.4 节）；该阈值因制造失配而随像素不同，也因入射光、时间等动态效应而变 [78]。多项研究 [8]、[62]、[118]、[120]、[236] 对图像质量做了量化；质量也受传感器空间分辨率影响。

应用：图像重建意味着原则上可以把事件转为亮度图像，再使用成熟的计算机视觉算法 [8]、[95]、[120]。这可能对事件视觉与帧视觉两个领域都产生较大影响。所得图像能捕捉高速运动和 HDR 场景，在某些应用中有利，但会增加计算代价、延迟和功耗。

虽然图像重建有助于识别 [235]、SLAM [25] 或光流估计 [117] 等任务，但 [113]、[119]、[128]、[153] 等研究也表明，完成这些任务不一定需要重建图像。图像重建最有价值的一点是，它提供比事件更不易受运动影响的场景表示（如外观图 [7]、[24]），并使建立事件对应更容易；事件对应是特征跟踪等某些事件处理任务面临的主要挑战之一 [64]。

4.7 运动分割

用静止事件相机观察时，分割运动目标比较简单：在照明恒定的假定下，事件仅来自目标运动 [122]、[124]、[180]。难点出现在相机运动时：图像平面上各处都可能触发事件 [13]、[132]、[155]（图 10），事件来自运动物体，也来自静态场景（其表观运动由相机自运动引起）；目标是推断每个事件的成因类别。但每个事件携带的信息很少，所以这样的逐事件分类很难。

图 10 运动相机下的目标分割与跟踪 / 原文第 18 页

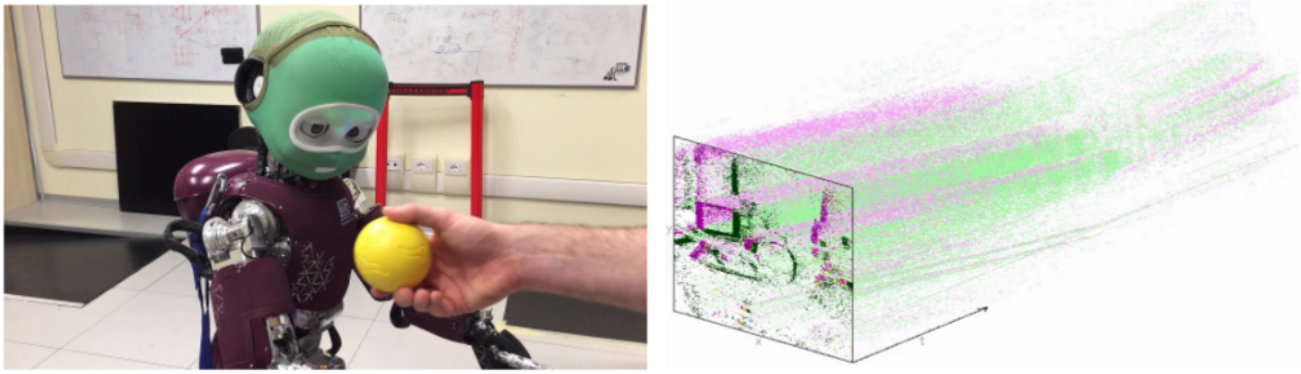


图 10. 意大利理工研究院（IIT）的 iCub 人形机器人在双眼各装有一台事件相机。此处它在头部运动产生的杂乱事件中分割、跟踪一个球。右图为图像帧上事件的时空可视化，按极性着色（正极性为绿色，负极性为红色）。图片由 [181] 提供。

图中术语：右图为时空事件；绿色正极性，红色负极性。

为克服这些挑战，研究者逐步减少用于求解的额外信息，研究复杂程度逐步提高的分割场景。额外信息可以是已知物体形状或已知运动，即算法知道“要找什么物体”或“预期什么运动”，通过检测观测与预期的一致或不一致来分割物体。提供的额外信息越少，问题就越接近无监督学习（如聚类）。在这种情况下，分割依赖一个关键认识：运动物体在图像平面上留下各自清晰可辨的事件迹线，可以推断产生这些迹线的物体轨迹，从而分割出物体 [155]。与聚类类似，这是一项关于物体运动参数（即“簇”）与事件-物体关联（即分割结果）的联合优化问题。

文献回顾：在物体形状已知的情形，[13] 给出一种在相机运动引起的杂乱事件中检测、跟踪圆形的方法。该方法基于霍夫变换，使用从事件时间窗口提取的光流信息。[181] 用粒子滤波器扩展该方法以提高跟踪稳健性，动态选取观测窗长度，适应目标加速引起的运动突变。[188] 用事件角点（第 4.1 节）作几何基元，检测并跟踪了更一般的物体形状。这种方法还要求知道控制相机运动的机器人关节信息。

[132]、[154]、[155] 在较弱的假设下用运动补偿事件图像 [127] 解决分割问题（第 3 节）。本质上，这项技术把按某个运动假设变换后可产生锐利边缘的事件归在一起。最简单的假设是线性运动模型（即恒定光流），但已足以描述：在短时间内，可把场景描述为由多个物体构成，各自生成符合不同线性运动模型的事件。上述分割算法就是要找到这样的场景描述。具体而言，[154] 先用线性运动补偿模型拟合占主导的事件，去掉这些事件，再用另一个线性模型拟合剩余事件，如此贪心迭代。因此它是按光流对事件聚类，产生具有锐利目标轮廓的运动补偿图像。类似地，[132] 将运动补偿模型拟合于主导事件（即背景），检测与模型不一致的部分（即运动目标），从而在杂乱事件中检出运动目标。研究者在标准相机难以应对的场景（HDR、高速）中测试了该方法，并发布数据集。[155] 提出迭代聚类算法，联合估计使运动补偿事件图像最锐利的事件-物体关联（即分割）与物体运动参数（即各簇）。该方法允许以一般的参数化运动模型 [128] 描述各目标，得到优于贪心方法 [132]、[154] 的结果。[133] 提出结合运动补偿的基于学习的分割方法，用 ANN 估计深度、自运动、独立运动物体的分割掩膜，以及物体的三维速度。它提供用于监督学习的事件数据集，含三维扫描物体的精确逐像素运动掩膜，即使在低光照和高速运动时也可靠。

分割是帧视觉的核心课题，在事件视觉中的探索却较少。随着研究转向更复杂的场景、事件视觉技术日益成熟，预计会有更多工作解决这一难题。

4.8 识别

算法：事件相机的识别算法日益复杂，从匹配简单形状的模板，演进到用基于人工特征的传统机器学习或现代深度学习方法，对任意边缘模式分类。这一发展旨在提高识别系统的表达能力（即近似能力），增强它抵抗数据失真的稳健性。

事件传感器的早期研究从用静止传感器跟踪运动物体开始。用事件驱动更新物体形状模型的位置，可检测、跟踪已知简单形状的物体，如斑团 [6]、圆 [53]、[239] 或直线 [54]。也可以与预先定义的模板匹配来检测简单形状，从而不必描述物体的几何形状。这种模板匹配方法在早期硬件上以卷积实现 [53]。

对于更复杂的物体，可用模板匹配低层特征而非物体整体，再用分类器基于观测到的特征分布作出判断 [109]。常用的是最近邻分类器，距离在特征空间计算。提高特征的不变性可提高准确率；逐层增加特征复杂性的层级模型能实现这一点。若选取了合适的特征，切换任务时只需重新训练最终分类器，这就引出应选择什么特征的问题。早期研究使用人工设计的方向特征，但从数据本身学习特征得到的结果好得多。最简单的做法是用单个样本得到各个模板，但这样会对样本噪声敏感 [15]。也可采用生成式方法，学习能精确重建输入的特征，例如 [138] 使用深度置信网络（DBN）。较新的工作用无监督学习获得特征：把事件数据聚类并把各簇中心作为特征 [109]。推断时，每个事件关联到离自己最近的特征，再由分类器处理观测到的特征分布。随着深度学习在帧式计算机视觉中兴起，许多研究尝试把深度学习工具用于事件识别，以反向传播学习特征。这种方法不需要在输出处另设分类器，缺点是训练需要多得多的标注数据。利用事件做图像识别还面临一个现实问题：事件域训练数据能否获得 [240]。[241] 采用虫洞学习，这是一种半监督方法，从 RGB 域检测器出发训练事件域检测器；到第二阶段，原来的教师成为学生，事件传感器对照明的一部分不变性转移到纯 RGB 检测器上。

多数基于学习的方法把事件或脉冲转成（稠密）张量，便于采用以图像为输入的层级模型，例如 ANN（图 11）。张量各元素的值可用不同方式计算（第 3.1 节）。简单方法使用时间表面或事件直方图帧；更稳健的方法使用具有指数衰减的时间表面 [109] 或采用平均时间戳的时间表面 [113]。也可使用图像重建方法（第 4.6 节）。有些识别方法在推断时把脉冲转成帧 [150]、[235]；另一些把训练好的 ANN 转成能直接在事件数据上工作的 SNN [137]。类似思路还可用于识别以外的任务 [22]、[108]。随着神经形态硬件发展（第 5.1 节），直接在 SNN 中进行学习 [143]、乃至直接在神经形态硬件上学习 [144]，越来越受到关注。

图 11 事件相机用于车辆和扑克牌识别 / 原文第 19 页

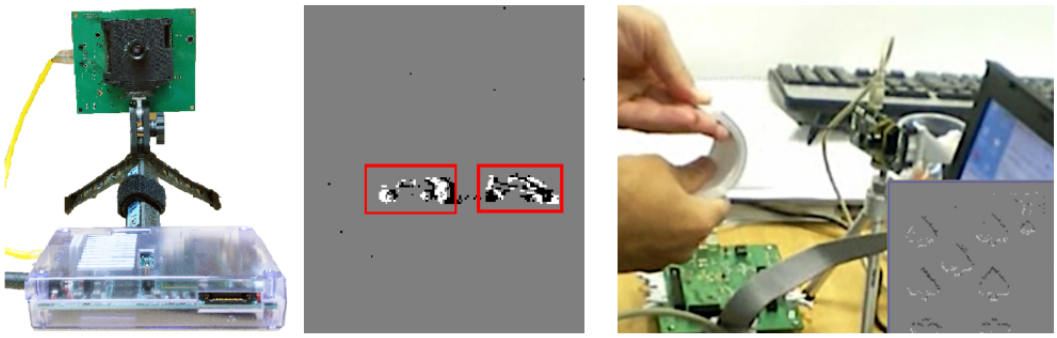


图 11. 运动物体识别。（a）接有 FPGA 的 DAVIS240C 传感器进行跟踪，并把事件区域发送至 IBM TrueNorth NS1e 评估平台做分类。街景结果用红框标出跟踪和分类的汽车。（b）[137] 使用事件驱动的卷积神经网络演示了极高速的目标识别（仅 0.65 s 内翻过整副 52 张牌）。

图中术语：Event camera and IBM TrueNorth = 事件相机和 IBM TrueNorth；Poker-DVS 为文中扑克牌实验。

任务：早期任务聚焦于用静止传感器检测是否存在某种简单形状（如圆形）[6]、[53]、[239]，随后很快转向对更复杂的形状分类，如扑克牌花色符号 [137]（图 11b）、印刷体字母 [15] 和人脸 [109]、[235]。手写数字分类一直是热门任务。受原始 MNIST 在帧式计算机视觉中作用的启发，研究者由它生成了若干事件版 MNIST 数据集 [58]、[242]。这些数据集仍是开发算法时很好的测试，成文时已有许多算法在此任务上达到 98 % 以上的准确率 [113]、[142]、[143]、[243]、[244]、[245]；但很少有人会把数字识别视为事件视觉的强项。更难的任务或涉及更难识别的目标，如 Caltech-101 和 Caltech-256 数据集（尽管两者在计算机视觉中仍被认为容易），或涉及更难的场景，如从移动载具上识别 [113]。迄今只有极少数工作处理这些任务，通常仍采取生成事件帧并用传统深度学习框架处理的方式。

识别的一个关键挑战是，事件相机响应场景中的相对运动（第 2.3 节），因而要求物体或相机处于运动中。因此，事件相机不太可能擅长识别静止或慢速物体，尽管将帧相机与事件相机的优点结合来处理这些应用的工作还很少。物体的事件外观高度依赖上述相对运动（图 5），因而可通过精确控制相机运动来辅助识别 [242]。

由于相机会响应动态信号，显而易见的应用包括按运动方式识别物体 [246]，或识别手势、动作等动态场景 [16]、[17]。这些任务因多了时间维，一般比静态物体识别更难，但这恰是事件相机的长处。

机会：事件相机有多种吸引人的特性，但事件识别要与现代帧方法竞争还有很长的路。比较两类方法很重要，同时也要记住各类传感器都有其长处。帧传感器的理想采集场景是传感器与物体均静止，这却是事件相机最不利的情形。事件识别要得到广泛采用，就需要找到能发挥自身长处的应用。这些应用不太可能类似于成熟、充分发挥帧传感器优点的计算机视觉识别任务；更可能是资源受限情况下的动态序列识别，或在运动平台上进行识别。在这些应用中找到并演示事件传感器的用途，仍是开放挑战。

尽管近年来事件数据集的质量有所改善，仍有提升空间。正采集的数据多得多，但标注仍然困难，目前并没有公认或标准的标注工具、格式。许多事件数据集源自帧视觉。虽在领域内发挥了重要作用，这些数据集本质上更适合发挥帧视觉的长处，因此不太可能促成新的事件传感器应用。数据采集和标

注费时费力，却不易得到认可；若能开发易用的事件数据采集与标注流程，将为领域作出重要贡献，特别是工具若能成熟到让非专业人员也可承接任务的程度。

4.9 神经形态控制

在生物体中，多数信息处理通过脉冲表示完成：脉冲编码感觉数据，完成计算，并传递执行器“命令”。因此，生物学表明，事件范式原则上不仅适用于感知和推理，也适用于控制。

神经形态视觉驱动的控制架构：这类架构（图 12）由神经形态传感器、事件估计器和传统控制器组成。估计器求出一个状态，控制器则根据状态计算控制量。控制器无需感知架构的异步性。

图 12 三种神经形态控制架构 / 原文第 20 页

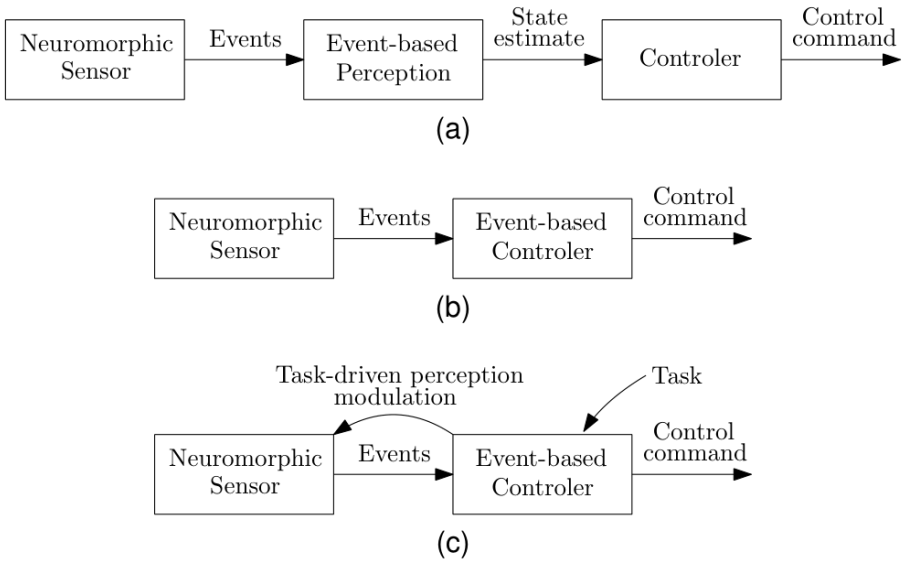


图 12. 基于神经形态事件的控制架构。在神经形态视觉驱动控制架构（a）中，神经形态传感器产生事件，事件感知系统生成状态估计，再异步调用传统控制器计算控制信号。在原生神经形态架构（b）中，事件直接引起控制量变化。（c）所示架构由任务指导所生成的事件。

图中术语：Neuromorphic Sensor = 神经形态传感器；Event-based Perception = 基于事件的感知；State estimate = 状态估计；Controller = 控制器；Control command = 控制命令；Task-driven perception modulation = 任务驱动的感知调节；Task = 任务。

从神经形态相机问世早期起，已有人演示神经形态视觉驱动的控制架构，证实了低延迟与高计算效率两项优势。最早的演示包括 CAVIAR 项目的基于脉冲卷积的目标跟踪 [53]，以及 [6]、[12] 的“机器人守门员”。另一个早期实例是平衡铅笔的机器人 [54]。演示中，两台 DVS 观测一根铅笔；铅笔作为倒立摆放在可移动的小车上。铅笔的三维状态估计延迟不到 1 ms。一个简单的手动调参 PID 控制器使铅笔保持直立平衡。该实验还在嵌入式系统上进行了演示，证明它能够在计算资源极其有限的条件下运行。

事件控制理论：从控制论和决策理论的角度，可以说明事件方法为何有意义。借用生物学的类比，事件控制可理解为经济学所说的“理性疏忽”[247]：更多信息有助于作出更好的决策，但若获取或处理信息有成本，仅凭可获得的部分信息作出决策就是理性的。

在事件控制中，控制信号异步改变 [248]。依“控制事件”的生成方式，这一概念有多种变体。一个重要区别是事件触发控制和自触发控制 [249]。在事件触发控制中，事件基于某些条件“外生”生成；例如，轨迹跟踪误差超过阈值时，可触发“重新计算控制量”的请求。自触发控制的控制器则根据

情况自行决定下一次应在何时被调用。例如，若状态接近目标，可让控制器“休眠”更久；若有需要，则更早重新计算控制信号。

事件控制的优点通常要从计算 / 通信成本与控制性能的权衡来理解。基本考虑是，虽然以无限高的频率重新计算控制量可以取得最佳控制性能（代价无限），但边际收益显著递减。控制理论有一个可靠原则：控制频率取决于被控对象和传感器的时间常数；控制量改变得远快于新信息的到达速度或执行器的响应速度并无意义。因此应选与被控对象动态特性相近、可因情况而调整的控制频率。例如，可以证明事件触发控制器仅用一部分计算就能达到相同性能；反过来，也可以用同样的计算量取得更好性能。在某些情形（标量线性高斯）下，可以得到这些权衡的闭式结果 [250]、[251]。（类似地，感知中的某些权衡也有闭式结果 [252]。）

遗憾的是，关于事件控制的大量文献对具身神经形态系统的用处有限。虽然表面上都处理“事件”，实际问题设置差别很大。例如，网络控制通常处理低维状态与偶发事件，重点是充分利用每一个事件。相比之下，装有事件相机的自主车辆通常需要从每秒可能数以百万计的事件中找到有用信号。将事件控制理论具体用于神经形态情形是相对较新的研究路径 [253]、[254]、[255]、[256]。主要挑战是处理视觉模态特有的非线性，这使其难以获得简洁的闭式结果。

神经形态控制的开放问题：最后介绍该领域的一些未解问题。

任务驱动的感知：在动物中，感知有价值是因为随后会行动，所收集的信息是有助于完成任务、可用于行动的信息。若控制器能依据任务和情境调节感知过程，将是一项重大进展。现有硬件虽能调节部分硬件偏置，但软件对感知过程的调节能力仍然有限。结合感兴趣区域机制、异质的相机偏置设置等，将提高灵活性，使控制在计算上更高效。

思考，快与慢：现有研究着重实现低延迟控制，但较少探讨如何将这一感觉运动层与智能体其余认知架构整合。再次借用仿生类比，并参考 Kahneman [257]，要将快速、本能、“情绪化”的系统与较慢、深思熟虑的系统结合起来。

5 事件系统与应用

5.1 神经形态计算

神经形态工程试图通过模仿大脑的结构与功能，获得其无与伦比的计算能力和效率的某些方面。这通常得到一种大规模并行的 SNN 硬件加速器（第 3.3 节），本文据此定义神经形态处理器。由于此类处理器内的神经元脉冲天然是异步的，神经形态处理器是事件相机最合适的计算搭档。它可直接处理事件相机送来的事件，无须转换；相比 CPU 等标准架构，其数据处理在时间和空间上的局部性更好，因而能构成低功耗、低延迟的计算机视觉系统。

神经形态处理器可按神经元模型的实现方式分类（表 4），主要分为模拟神经元（Neurogrid、BrainScaleS、ROLLS、DYNAP-se）、数字神经元（TrueNorth、Loihi、ODIN）和软件神经元（SpiNNaker）。有些架构也支持片上学习（Loihi、ODIN、DYNAP-1e）。为事件视觉系统评估神经形态处理器时，除功能和性能外，还应考虑以下准则：（i）软件开发生态：最小工具链包括用于构建和训练网络的 API、为硬件准备网络的编译器，以及将网络部署到硬件中的运行时库；（ii）事件视觉系统通常要求处理器作为独立系统使用，适合移动应用，而不能只托管在远程服务器上；（iii）神经形态处理器的可获得性。

表 04 神经形态处理器 / 原文第 22 页

Processor Reference	SpiNNaker [258]	TrueNorth [259]	Loihi [260]	DYNAP [261]	Braindrop [262]
Manufacturer	U. Manchester	IBM	Intel	aiCTX	Stanford U.
Year	2011	2014	2018	2017	2018
Neuron model	Software	Digital	Digital	Analog	Analog
On-chip learning	Yes	No	Yes	No	No
CMOS technol.	130 nm	28 nm	14 nm	180 nm	28 nm
Neurons/chip	4 k*	1024 k	131 k	1 k	4 k
Neurons/core	255*	256	1024	256	4096
Cores/chip	16*	4096	128	4	1
Synapses/chip	16 M	268 M	130 M	128 k	16 M
Boards	4- or 48-chip	1- or 16-chip	4- or 8-chip,	1-chip	1-chip
Software stack	sPyNNaker PACMAN	CPE/Eedn NSCP	Nengo Nx SDK	cAER libcAER	Nengo

表 4. 几种神经形态处理器的对比，按神经元模型类型排序。

表中术语：Processor = 处理器；Reference = 文献；Manufacturer = 厂商 / 研制机构；Year = 年份；Neuron model = 神经元模型（Software：软件；Digital：数字；Analog：模拟）；On-chip learning = 片上学习；CMOS technol. = CMOS 工艺；neurons/chip = 每芯片神经元数；neurons/core = 每核心神经元数；cores/chip = 每芯片核心数；synapses/chip = 每芯片突触数；Boards = 板卡（按所含芯片数计）；software stack = 软件栈。

要让这些处理器得到更广泛的使用，需要进一步发展：（i）构建更易用的生态，让用户更容易编写所需方法并部署到硬件；（ii）让硬件平台支持更多处理能力；（iii）扩大器件供应，不限于为特定合作伙伴提供的早期试用计划。

下列处理器（表 4）具有最成熟的开发流程，也是最容易获得的独立系统。更多细节见 [263]、[264]。

SpiNNaker（脉冲神经网络架构）用通用 ARM 核心模拟符合生物实际的人脑模型 [265]。其他处理器选择某些特定的简化神经元模型嵌入定制晶体管电路，而 SpiNNaker 则将神经元实现为在核心上运行的软件，牺牲硬件加速以最大限度提高模型灵活性。已有研究将 SpiNNaker 与事件相机结合用于立体深度估计 [165]、[266]、光流计算 [266]、[267]、目标跟踪 [268] 和识别 [269]。

TrueNorth 用数字神经元实时推断。每片芯片模拟 1 M（100 万）个神经元和 256 M（2.56 亿）个突触，分布在 4096 个神经突触核心中。它不支持片上学习，网络须在 GPU 等处理器上离线训练 [270]。

TrueNorth 已与事件相机结合构造端到端、低功耗和低延迟的事件视觉系统，用于手势识别 [17]、立体重建 [213] 和光流估计 [196]。

Loihi 用数字神经元进行实时推断和在线学习。每片芯片最多模拟 13.1 万个脉冲神经元和 130 M（1.3 亿）个突触。每个神经形态核心的学习引擎按包括 STDP 和强化学习在内的规则更新各突触 [260]。非脉冲网络可在 TensorFlow 中训练，再用 Applied Brain Research 的 Nengo Deep Learning 工具包将其近似为适于 Loihi 的脉冲网络 [271]。有计划用 768 片 Loihi 芯片构建名为 Pohoiki Springs 的更大系统（100 M〔1 亿〕个神经元和 1000 亿个突触），但和人脑的 800 万亿个突触相比，规模仍小。

DYNAP： aiCTX 的动态神经形态异步处理器（DYNAP）有两个版本，一个针对可扩展推断优化（Dynap-se），另一个针对在线学习优化（Dynap-le）。

Braindrop 是拥有 1 M（100 万）个神经元的 Brainstorm 系统 [262] 的一个核心的原型。它使用 Nengo [272] 编程并实现了神经工程框架 [273]。Braindrop 是斯坦福大学 Neurogrid 处理器的后续产品。

5.2 机载实时机器人应用

事件视觉传感器单位时间内产生的数据通常明显少于传统相机，因此可以设想多种应用：由直接连接传感器的简单计算系统实时提取相关视觉信息，无需 USB 连接。图 13 给出了这样的例子：一个运行在 200 MHz、板载 136 kB SRAM 的双核 ARM 微控制器实时读取和处理事件。此处，由传感器和微控制器组成的嵌入式系统操控一台简单的轮式机器人，执行循线、主动与被动目标跟踪、距离估计和简单建图等任务 [274]。

图 13 Pushbot 与下视相机四旋翼 / 原文第 22 页

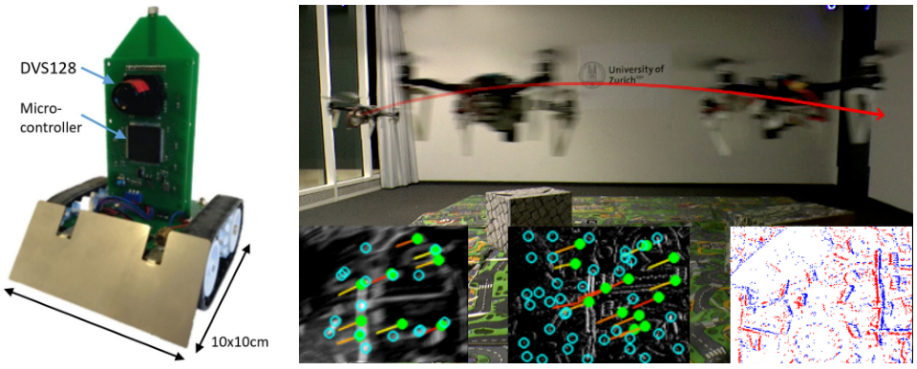


图 13.（a）Pushbot 上的嵌入式 DVS128，构成独立的闭环感知-计算-行动系统，用于导航和避障任务 [274]。（b）配备一台下视 DAVIS 的无人机，用于自主飞行 [27]。事件的高速度与高动态范围使其可在复杂光照下工作。同一视觉惯性里程计算法 [27] 也在高速场景中演示过，如把事件相机系在绳上旋转。

图中术语：Microcontroller = 微控制器；DVS128、Pushbot 与 DAVIS 为器件或平台名称。

另一个近传感器处理（“边缘计算”）的例子是 Speck 片上系统（SoC）[11]，它将 DVS 和 Dynap-se 神经形态 CNN 处理器集成在一起。其峰值功耗不到 1 mW，延迟不到 30 ms，适用领域包括低功耗连续目标检测、监视及车载系统。

原注 11 · https://www.speck.ai/

事件相机也用于计算资源有限的四旋翼飞行器，在有挑战的场景中实现自主着陆 [275] 或飞行 [27]（图 13b）。

6 资源

《事件视觉资源列表》[9] 由相关人士协作整理，收集了该领域的资料链接（论文、视频、组织、公司和研讨会），以及事件视觉所用的软件、驱动程序、代码、数据集、模拟器和其他重要工具。

6.1 软件

截至成文，还没有集成于 OpenCV、提供事件视觉算法的开源标准库。这样的资源很有价值，可以加速事件相机的采用。不过已有许多相当成熟的开源软件资源：

• jAER [276][12] 是一个面向事件传感器及事件处理的 Java 环境，支持降噪、特征提取、光流、用 IMU 消除旋转、CNN 和 RNN 推断等。虽说 Java 不太适合移动机器人，仍有数款非移动机器人 [6]、[12]、[54]、[277]，甚至一款移动 DVS 机器人 [150] 是在 jAER 中构建的。它提供桌面图形界面，便于录制、回放数据，也开放设备复杂的内部配置。它主要支持由神经信息学研究所（苏黎世大学与苏黎世联邦理工学院）开发、iniVation 经销的传感器。

原注 12 · https://jaerproject.org

• libcaer[13] 是一个精简的 C 库，用于访问、配置 iniVation 和 aiCTX 的神经形态传感器与处理器并获取其数据。它支持 DVS、DAVIS 相机及 Dynap-SE 神经形态处理器。

原注 13 · https://github.com/inilabs/libcaer

• ROS DVS 软件包[14] 由苏黎世大学与苏黎世联邦理工学院的机器人与感知研究组开发，基于 libcaer，为 DVS 和 DAVIS 提供 C++ 驱动。它集成到机器人操作系统（ROS）[278]，有便于录制、回放数据和连接其他传感器、执行器的高级工具，因而在机器人领域很流行。常用数据集 [98]、[197] 以这一格式提供。该软件包还提供支持内参标定和立体标定的工具。

原注 14 · https://github.com/uzh-rpg/rpg_dvs_ros

• 事件驱动的 YARP 项目[15] [279] 包括处理神经形态传感器的库（如装在 iCub 人形机器人上的 DVS），以及事件数据处理算法。它基于 Yet Another Robot Platform（YARP）中间件。

原注 15 · https://github.com/robotology/event-driven

• pyAER[16] 是由神经信息学研究所（苏黎世大学与苏黎世联邦理工学院）开发的 libcaer Python 封装，有望成为快速实验的常用工具。

原注 16 · https://github.com/duguyue100/pyaer

• DV[17] 是面向 iniVation DVS/DAVIS 的开源 C++ 软件，也是其官方 SDK 所用的软件。它支持快速、便捷地为下一代嵌入式和物联网应用部署先进的处理算法。

原注 17 · https://inivation.gitlab.io/dv/dv-docs/

其他开源软件工具和处理算法（用 Python、CUDA、Matlab 等编写）散见于网上从事事件视觉研究的团队主页 [9]。专有软件包括企业开发的软件开发工具包（SDK）。

6.2 数据集与模拟器

数据集和模拟器是推动事件驱动技术应用与研究的基本工具。它们有助于降低成本（截至成文，事件相机明显比标准相机昂贵），并通过量化基准跟踪进展（像传统计算机视觉中的 Middlebury、MPI Sintel、KITTI、EuRoC 等数据集一样）。

事件视觉数据集和模拟器的数量增长很快，[9] 按任务列出其中一部分。大致可分为面向运动估计或图像重建（回归）的、面向识别（分类）的，以及驾驶数据 [280] 等用于端到端任务的人工标注数据。第一组含光流、SLAM、目标跟踪、分割等数据集；第二组包括目标与动作识别数据集。

光流数据集包括 [22]、[190]、[281]。光流真值难以获得，因此 [190] 只考虑由 IMU 记录的纯旋转运动期间的光流，数据集因此没有平移（视差）运动引起的光流。[22]、[281] 的数据集以相机运动与场景深度（用 RGB-D 相机、立体相机对或 LiDAR 等测距传感器测得）在图像平面上诱导的运动场作为光流。自然地，光流真值会受到噪声及所用各传感器对齐和标定不准的影响。

位姿估计与 SLAM 数据集有 [225][18]、[98]、[197]、[281]、[282]。使用最多的一个见 [98]，已用于评估视觉里程计和视觉惯性里程计方法 [27]、[103]、[104]、[127]、[128]、[129]、[131]。该数据集也常用于评估角点检测器 [111]、[115] 和特征跟踪器 [64]、[126]。

原注 18 · http://ebvds.neurocomputing.systems

相比传统计算机视觉，事件识别数据集规模尚小。它们包括扑克牌（4 类）、人脸（7 类）、手写数字（36 类）、动态场景中的手势（石头、布、剪刀）、汽车等。通过类似扫视的运动，研究者得到常见帧式计算机视觉数据集的神经形态版本，如 MNIST 和 Caltech101 [242]、[283]。较新的数据集 [17]、[113]、[284]、[285] 在真实场景中采集，并非用帧数据生成。包括 [15]、[109]、[113]、[137]、[150]、[151] 在内的研究用这些数据集对事件识别算法进行基准测试。

[73]、[286] 的 DVS 仿真器与 [98] 的模拟器基于理想 DVS 像素的工作原理（式 2）。给定虚拟三维场景及 DAVIS 在其中运动的轨迹，模拟器生成相应的事件流、光强帧和深度图。[74] 扩展了这一模拟器；它采用自适应渲染方案，更接近真实照片，包含简单的事件噪声模型，并返回估计光流。v2e 工具 [40] 使用更真实的、含噪声和非理想效应的 DVS 像素模型，由视频生成事件，将建模扩展到低光照条件。现有事件相机的噪声及动态效应尚未得到全面表征，因此所用噪声模型目前有所简化。

6.3 研讨会

目前有两个每年举行的暑期学校，促进事件视觉等领域的研究：Telluride 神经形态认知工程研讨会（2020 年第 27 届，在美国）和 Capo Caccia 研讨会（2020 年第 12 届，在欧洲）。近年来，大型机器人学会议同期也举办了相关研讨会（IROS’15[19]、ICRA’17[20] 或 IROS’18[21]）。在顶级会议如 ISSCC’06、NIPS’09、CVPR’18、ECCV’18、ICRA’17、IROS’18、CVPR’19[22] 和多届 ISCAS 等，已有事件系统现场演示。随着事件视觉社群壮大，传统计算机视觉会议预计也会举办更多研讨会和现场演示。

原注 19 · http://innovative-sensing.mit.edu/

原注 20 · http://rpg.ifi.uzh.ch/ICRA17_event_vision_workshop.html

原注 21 · http://www.jmartel.net/irosws-home

原注 22 · http://rpg.ifi.uzh.ch/CVPR19_event_vision_workshop.html

7 讨论

事件视觉横跨计算机视觉、机器人学和神经形态工程等领域。不同社群着重利用事件范式的不同优势：有的看重低功耗，用于资源受限平台上的“常开”或嵌入式应用；有的看重低延迟，以实现反应迅速的系统；还有的更重视取得有助于感知环境的信息（高时间分辨率和 HDR），而对计算资源的限制较少。

在帧相机硬件和软件已相当成熟的时代，事件视觉是一项新兴技术。两者发展成熟度不同，因此比较在某些方面并不公平。不过事件相机已经显示出潜力，能够克服帧相机的一些局限，进入过去无法处理的场景。如文中多处研究机会指出，事件视觉在研究和开发上仍有相当大的提升空间。

由于很大程度上取决于应用，事件处理的最佳方法和表示尚无共识。这里要进行不同权衡，例如延迟与功耗、精度之间的权衡，或灵敏度与带宽、处理能力之间的权衡。例如，降低对比度阈值或提高分辨率（或同时采取两者）都会产生更多事件，但处理这些事件的算法和平台能力有限。一个具有挑战的研究方向是量化这些权衡，并开发可动态调整传感器参数或算法参数（或两者）、获得最佳性能的技术。

另一大挑战是开发从感知到控制及执行都原生采用事件的端到端仿生系统，使它们比同步帧系统更高效，并能作为长远方案。事件相机要求重新思考感知、控制和执行，尤其要重新考虑当前计算机视觉中占主流的深度学习方法：调整这些方法，或迁移其中的思想来处理事件，同时保持顶尖性能。主动视觉（将感知与控制结合起来）尤其适用于事件相机，因为事件显著依赖运动，而运动可以由机器人执行机构产生。

可把事件相机看作实现更高效近传感器处理的起点：只传输高层、非冗余信息，从而减少带宽、延迟和功耗。为此，可以把事件相机和处理硬件集成于同一传感器器件中（第 5.2 节的 Speck）；也可以使用其他仿生成像传感器，如单元处理器阵列 [287]、[288]，其中每个像素都有处理器，能对该像素及邻域的亮度执行多类计算。

8 结论

事件相机是一种革命性的传感器，相比传统帧相机拥有诸多优势，如低延迟、低功耗、高速度和高动态范围。因此，在当前传统相机无力应对的场景中，它在计算机视觉和机器人应用中具有很大的潜力。本文综述事件视觉，覆盖感知、计算和控制，重点介绍事件相机工作原理，以及为在特定应用中发挥其卓越特性而开发的算法，涉及从低层到高层视觉。神经形态感知与控制是新兴课题，正如全文所述，仍有许多机会。前路仍有诸多挑战；希望本文能为这一课题提供入门阐述，作为人类长期探索中的一步，朝着能以更高效、更符合生物启发的方式感知世界并与之交互的智能机器前进。

致谢

作者谨向所有为本文作出贡献的人致谢。感谢事件相机制造商提供表 1 数据，并讨论没有统一测试平台所造成的比较困难。特别感谢 Hyunsurk Eric Ryu（三星电子）、Chenghan Li（iniVation）、Davide Migliore（Prophesee）、Marc Osswald（Insightness）以及 Chen 教授（CelePixel）。也感谢各研究实验室的所有成员对本文初稿开展讨论并提供意见。还要感谢 IEEE TPAMI 的编辑和匿名审稿人，他们的建议帮助我们改进了论文。

参考文献

按原文编号与英文著录。

[1] M. Mahowald and C. Mead, “The silicon retina,” Scientific American, vol. 264, no. 5, pp. 76-83, May 1991.

[2] P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor,” IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 566-576, 2008.

[3] C. Posch, D. Matolin, and R. Wohlgenannt, “A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS,” IEEE J. Solid-State Circuits, vol. 46, no. 1, pp. 259-275, Jan. 2011.

[4] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, “A 240x180 130dB 3μs latency global shutter spatiotemporal vision sensor,” IEEE J. Solid-State Circuits, vol. 49, no. 10, pp. 2333-2341, 2014.

[5] B. Son, Y. Suh, S. Kim, H. Jung, J.-S. Kim, C. Shin, K. Park, K. Lee, J. Park, J. Woo, Y. Roh, H. Lee, Y. Wang, I. Ovsiannikov, and H. Ryu, “A 640x480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation,” in IEEE Intl. Solid-State Circuits Conf. (ISSCC), 2017.

[6] T. Delbruck and P. Lichtsteiner, “Fast sensory motor control based on event-based hybrid neuromorphic-procedural system,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2007, pp. 845-848.

[7] G. Gallego, J. E. A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, “Event-based, 6-DOF camera tracking from photometric depth maps,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 10, pp. 2402-2412, Oct. 2018.

[8] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “High speed and high dynamic range video with an event camera,” IEEE Trans. Pattern Anal. Mach. Intell., 2019.

[9] https://github.com/uzh-rpg/event-based_vision_resources.

[10] T. Delbruck, “Neuromorphic vision sensing and processing,” in Eur. Solid-State Device Research Conf. (ESSDERC), 2016, pp. 7-14.

[11] S.-C. Liu, B. Rueckauer, E. Ceolini, A. Huber, and T. Delbruck, “Event-Driven Sensing for Efficient Perception: Vision and audition algorithms,” IEEE Signal Process. Mag., vol. 36, no. 6, pp. 29-37, Nov. 2019.

[12] T. Delbruck and M. Lang, “Robotic goalie with 3ms reaction time at 4% CPU load using event-based dynamic vision sensor,” Front. Neurosci., vol. 7, p. 223, 2013.

[13] A. Glover and C. Bartolozzi, “Event-driven ball detection and gaze fixation in clutter,” in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2016.

[14] M. Litzenberger, A. N. Belbachir, N. Donath, G. Gritsch, H. Garn, B. Kohn, C. Posch, and S. Schraml, “Estimation of vehicle speed based on asynchronous data from a silicon retina optical sensor,” in IEEE Intell. Transp. Sys. Conf., 2006, pp. 653-658.

[15] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman, “HFirst: A temporal approach to object recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 10, pp. 2028-2040, 2015.

[16] J. H. Lee, T. Delbruck, M. Pfeiffer, P. K. Park, C.-W. Shin, H. Ryu, and B. C. Kang, “Real-time gesture interface based on event-driven processing from stereo silicon retinas,” IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 12, pp. 2250-2263, 2014.

[17] A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. D. Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, and D. Modha, “A low power, fully event-based gesture recognition system,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.

[18] P. Rogister, R. Benosman, S.-H. Ieng, P. Lichtsteiner, and T. Delbruck, “Asynchronous event-based binocular stereo matching,” IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 2, pp. 347-353, 2012.

[19] H. Rebecq, G. Gallego, E. Mueggler, and D. Scaramuzza, “EMVS: Event-based multi-view stereo-3D reconstruction with an event camera in real-time,” Int. J. Comput. Vis., vol. 126, no. 12, pp. 1394-1414, 2018.

[20] N. Matsuda, O. Cossairt, and M. Gupta, “MC3D: Motion contrast 3D scanning,” in IEEE Int. Conf. Comput. Photography (ICCP), 2015.

[21] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, “Event-based visual flow,” IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 2, pp. 407-417, 2014.

[22] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “EV-FlowNet: Self-supervised optical flow estimation for event-based cameras,” in Robotics: Science and Systems (RSS), 2018.

[23] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, “Interacting maps for fast visual interpretation,” in Int. Joint Conf. Neural Netw. (IJCNN), 2011, pp. 770-776.

[24] H. Kim, A. Handa, R. Benosman, S.-H. Ieng, and A. J. Davison, “Simultaneous mosaicing and tracking with an event camera,” in British Mach. Vis. Conf. (BMVC), 2014.

[25] H. Kim, S. Leutenegger, and A. J. Davison, “Real-time 3D reconstruction and 6-DoF tracking with an event camera,” in Eur. Conf. Comput. Vis. (ECCV), 2016.

[26] H. Rebecq, T. Horstschäfer, G. Gallego, and D. Scaramuzza, “EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time,” IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 593-600, 2017.

[27] A. Rosinol Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, “Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios,” IEEE Robot. Autom. Lett., vol. 3, no. 2, pp. 994-1001, Apr. 2018.

[28] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, “Bringing a blurry frame alive at high

frame-rate with an event camera," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[29] G. Cohen, S. Afshar, A. van Schaik, A. Wabnitz, T. Bessell, M. Rutten, and B. Morreale, "Event-based sensing for space situational awareness," in Proc. Advanced Maui Optical and Space Surveillance Technol. Conf. (AMOS), 2017.

[30] T. Chin, S. Bagchi, A. P. Eriksson, and A. van Schaik, "Star tracking using an event camera," in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.

[31] P. Lichtsteiner and T. Delbruck, "64x64 event-driven logarithmic temporal derivative silicon retina," in IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 2005, pp. 157-160.

[32] --, "A 64x64 AER logarithmic temporal derivative silicon retina," in Research in Microelectr. & Electron., PhD, vol. 2, 2005.

[33] P. Lichtsteiner, "An AER temporal contrast vision sensor," Ph. D. Thesis, ETH Zurich, Dept. of Physics (D-PHYS), Zurich, Switzerland, 2006.

[34] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128x128 120dB 30mW asynchronous vision sensor that responds to relative intensity change," in IEEE Intl. Solid-State Circuits Conf. (ISSCC), 2006, pp. 2060-2069.

[35] D. Neil, "Deep neural networks and hardware systems for event-driven data," Ph.D. dissertation, ETH-Zurich, Zurich, Switzerland, 2017.

[36] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output," Proc. IEEE, vol. 102, no. 10, pp. 1470-1484, Oct. 2014.

[37] K. A. Boahen, "A burst-mode word-serial address-event link-I: Transmitter design," IEEE Trans. Circuits Syst. I, vol. 51, no. 7, pp. 1269-1280, Jul. 2004.

[38] S.-C. Liu, T. Delbruck, G. Indiveri, A. Whatley, and R. Douglas, Event-Based Neuromorphic Systems. John Wiley & Sons, 2015.

[39] Y. Suh, S. Choi, M. Ito, J. Kim, Y. Lee, J. Seo, H. Jung, D.-H. Yeo, S. Namgung, J. Bong, J. seok Kim, P. K. J. Park, J. Kim, H. Ryu, and Y. Park, "A 1280x960 Dynamic Vision Sensor with a 4.95-μm pixel pitch and motion artifact minimization," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2020.

[40] T. Delbruck, Y. Hu, and Z. He, "V2E: From video frames to realistic DVS event camera streams," arXiv e-prints, 2020.

[41] M. Mahowald, "VLSI analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, California Institute of Technology, Pasadena, California, May 1992.

[42] D. il Dan Cho and T. jae Lee, "A review of bioinspired vision sensors and their applications," Sens. Mater., vol. 27, no. 6, pp. 447-463, 2015.

[43] L. Steffen, D. Reichard, J. Weinland, J. Kaiser, A. Rönnau, and R. Dillmann, "Neuromorphic stereo vision: A survey of bio-inspired sensors and algorithms," Front. Neurorobot., vol. 13, p. 28, 2019.

[44] T. Delbruck and C. A. Mead, "Time-derivative adaptive silicon photoreceptor array," in Proc. SPIE, Infrared sensors: Detectors, Electron., and Signal Process., vol. 1541, 1991, pp. 92-99.

[45] S.-C. Liu and T. Delbruck, "Neuromorphic sensory systems," Current Opinion in Neurobiology, vol. 20, no. 3, pp. 288-295, 2010.

[46] T. Delbruck, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2010, pp. 2426-2429.

[47] T. Delbruck, "Fun with asynchronous vision sensors and processing," in Eur. Conf. Comput. Vis. Workshops (ECCVW), 2012, pp. 506-515.

[48] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression," in IEEE Intl. Solid-State Circuits Conf. (ISSCC), 2010, pp. 400-401.

[49] E. Culurciello, R. Etienne-Cummings, and K. A. Boahen, "A biomorphic digital image sensor," IEEE J. Solid-State Circuits, vol. 38, no. 2, pp. 281-294, Feb. 2003.

[50] G. Orchard, D. Matolin, X. Lagorce, R. Benosman, and C. Posch, "Accelerated frame-free time-encoded multi-step imaging," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2014, pp. 2644-2647.

[51] R. Berner, C. Brandli, M. Yang, S.-C. Liu, and T. Delbruck, "A 240x180 10mw 12μs latency sparse-output vision sensor for mobile applications," in Proc. Symp. VLSI, 2013, pp. C186-C187.

[52] E. Fossum, "CMOS image sensors: electronic camera-on-a-chip," IEEE Trans. Electron Devices, vol. 44, no. 10, pp. 1689-1698, 1997.

[53] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, L. Camunas-Mesa, R. Berner, M. Rivas-Perez, T. Delbruck, S.-C. Liu, R. Douglas, P. Hafliger, G. Jimenez-Moreno, A. C. Ballcels, T. Serrano-Gotarredona, A. J. Acosta-Jimenez, and B. Linares-Barranco, "CAVIAR: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory-processing-learning-actuating system for high-speed visual object recognition and tracking," IEEE Trans. Neural Netw., vol. 20, no. 9, pp. 1417-1438, 2009.

[54] J. Conradt, M. Cook, R. Berner, P. Lichtsteiner, R. J. Douglas, and T. Delbruck, "A pencil balancing robot using a pair of AER dynamic vision sensors," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2009, pp. 781-784.

[55] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.

[56] Y. Nozaki and T. Delbruck, "Temperature and parasitic photocurrent effects in dynamic vision sensors," IEEE Trans. Electron Devices, vol. 64, no. 8, pp. 3239-3245, Aug. 2017.

[57] --, “Authors Reply to Comment on Temperature and Parasitic Photocurrent Effects in Dynamic Vision Sensors,” IEEE Trans. Electron Devices, vol. 65, no. 7, pp. 3083-3083, Jul. 2018.

[58] T. Serrano-Gotarredona and B. Linares-Barranco, “A 128x128 1.5% contrast sensitivity 0.9% FPN 3s latency 4mW asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers,” IEEE J. Solid-State Circuits, vol. 48, no. 3, pp. 827-838, Mar. 2013.

[59] M. Yang, S.-C. Liu, and T. Delbruck, “A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding,” IEEE J. Solid-State Circuits, vol. 50, no. 9, pp. 2149-2160, 2015.

[60] D. P. Moeys, F. Corradi, C. Li, S. A. Bamford, L. Longinotti, F. F. Voigt, S. Berry, G. Taverni, F. Helmchen, and T. Delbruck, “A sensitive dynamic and active pixel vision sensor for color or neural imaging applications,” IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 1, pp. 123-136, Feb. 2018.

[61] A. Rose, Vision: Human and Electronic. Plenum Press, New York, 1973.

[62] C. Scheerlinck, N. Barnes, and R. Mahony, “Continuous-time intensity estimation using event cameras,” in Asian Conf. Comput. Vis. (ACCV), 2018.

[63] --, “Asynchronous spatial image convolutions for event cameras,” IEEE Robot. Autom. Lett., vol. 4, no. 2, pp. 816-822, 2019.

[64] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, “EKLT: Asynchronous photometric feature tracking using events and frames,” Int. J. Comput. Vis., 2019.

[65] S. Bryner, G. Gallego, H. Rebecq, and D. Scaramuzza, “Event-based, direct camera tracking from a photometric 3D map using nonlinear optimization,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2019.

[66] G. Gallego, C. Forster, E. Mueggler, and D. Scaramuzza, “Event-based camera pose tracking using a generative event model,” 2015, arXiv:1510.01972.

[67] Prophesee Evaluation Kits, https://www.prophesee.ai/event-based-evk/, 2020.

[68] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, “A 1280x720 backilluminated stacked temporal contrast event-based vision sensor with 4.86µm pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline,” in IEEE Intl. Solid-State Circuits Conf. (ISSCC), 2020.

[69] H. E. Ryu, “Industrial DVS design; key features and applications,” http://rpg.ifi.uzh.ch/docs/CVPR19workshop/CVPRW19_Eric_Ryu_Samsung.pdf.

[70] M. Guo, J. Huang, and S. Chen, “Live demonstration: A 768 x 640 pixels 200meps dynamic vision sensor,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2017.

[71] S. Chen and M. Guo, “Live demonstration: CeleX-V: A 1M pixel multi-mode event-based sensor,” in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.

[72] Insightness Event-based Sensor Modules, http://www.insightness.com/technology/, 2020.

[73] M. L. Katz, K. Nikolic, and T. Delbruck, “Live demonstration: Behavioural emulation of event-based vision sensors,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2012, pp. 736-740.

[74] H. Rebecq, D. Gehrig, and D. Scaramuzza, “ESIM: an open event camera simulator,” in Conf. on Robotics Learning (CoRL), 2018.

[75] A. Censi and D. Scaramuzza, “Low-latency event-based visual odometry,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2014.

[76] M. Yang, S.-C. Liu, and T. Delbruck, “Analysis of encoding degradation in spiking sensors due to spike delay variation,” IEEE Trans. Circuits Syst. I, vol. 64, no. 1, pp. 145-155, Jan. 2017.

[77] C. Posch and D. Matolin, “Sensitivity and uniformity of a 0.18µm CMOS temporal contrast pixel array,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2011, pp. 1572-1575.

[78] C. Brandli, L. Muller, and T. Delbruck, “Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2014.

[79] M. Yang, S.-C. Liu, and T. Delbruck, “Comparison of spike encoding schemes in asynchronous vision sensors: Modeling and design,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2014, pp. 2632-2635.

[80] T. Delbruck, “Frame-free dynamic digital vision,” in Proc. Int. Symp. Secure-Life Electron., 2008, pp. 21-26.

[81] A. Khodamoradi and R. Kastner, “O(N)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors,” IEEE Trans. Emerg. Topics Comput., vol. PP, no. 99, pp. 1-1, 2018.

[82] D. Czech and G. Orchard, “Evaluating noise filtering for event-based asynchronous change detection image sensors,” in IEEE Int. Conf. Biomed. Robot. and Biomechatron. (BioRob), 2016, pp. 19-24.

[83] V. Padala, A. Basu, and G. Orchard, “A noise filtering algorithm for event-based asynchronous change detection image sensors on TrueNorth and its implementation on TrueNorth,” Front. Neurosci., vol. 12, 2018.

[84] T. Delbruck, V. Villanueva, and L. Longinotti, “Integration of dynamic vision sensor with inertial measurement unit for electronically stabilized event-based vision,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2014, pp. 2636-2639.

[85] R. Berner, “Highspeed USB2.0 AER interfaces,” Master’s thesis, ETH Zurich, Dept. of Electrical and Information Eng. (D-ITET), Zurich, Switzerland, 2006.

[86] iniVation, “Understandingthe performance of neuromorphic event-based vision sensors,” https://inivation.com/dvp/white-papers/, May 2020.

[87] G. Taverni, D. P. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. S. S. Bello, and T. Delbruck, “Front and back illuminated Dynamic and Active Pixel Vision Sensors

comparison," IEEE Trans. Circuits Syst. II, vol. 65, no. 5, pp. 677-681, 2018.

[88] D. B. Fasnacht and T. Delbruck, "Dichromatic spectral measurement circuit in vanilla CMOS," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2007, pp. 3091-3094.

[89] R. Berner, P. Lichtsteiner, and T. Delbruck, "Self-timed vertacolor dichromatic vision sensor for low power pattern detection," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2008, pp. 1032-1035.

[90] L. Farian, J. A. Leñero-Bardallo, and P. Häfliger, "A bio-inspired AER temporal tri-color differentiator pixel array," IEEE Trans. Biomed. Circuits Syst., vol. 9, no. 5, pp. 686-698, Oct. 2015.

[91] R. B. Merrill, "Color separation in an active pixel cell imaging array using a triple-well structure," US Patent US5 965 875A, 1999.

[92] R. F. Lyon and P. M. Hubel, "Eyeing the camera: Into the next century," in Proc. IS&T/SID 10th Color Imaging Conf., 2002, pp. 349-355.

[93] C. Li, C. Brandli, R. Berner, H. Liu, M. Yang, S.-C. Liu, and T. Delbruck, "Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2015.

[94] D. P. Moeys, C. Li, J. N. P. Martel, S. Bamford, L. Longinotti, V. Motsnyi, D. S. S. Bello, and T. Delbruck, "Color temporal contrast sensitivity in dynamic vision sensors," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2017, pp. 1-4.

[95] C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. Mahony, and D. Scaramuzza, "CED: color event camera dataset," in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.

[96] A. Marcireau, S.-H. Ieng, C. Simon-Chane, and R. B. Benosman, "Event-based color segmentation with a high dynamic range sensor," Front. Neurosci., vol. 12, 2018.

[97] C. Posch, D. Matolin, R. Wohlgenannt, T. Maier, and M. Litzenberger, "A microbolometer asynchronous Dynamic Vision Sensor for LWIR," IEEE Sensors J., vol. 9, no. 6, pp. 654-664, Jun. 2009.

[98] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," Int. J. Robot. Research, vol. 36, no. 2, pp. 142-149, 2017.

[99] G. Gallego, M. Gehrig, and D. Scaramuzza, "Focus is all you need: Loss functions for event-based vision," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[100] D. Gehrig, A. Loquercio, K. G. Derpanis, and D. Scaramuzza, "End-to-end learning of representations for asynchronous event-based data," in Int. Conf. Comput. Vis. (ICCV), 2019.

[101] D. Weikersdorfer and J. Conradt, "Event-based particle filtering for robot self-localization," in IEEE Int. Conf. Robot. Biomimetics (ROBIO), 2012.

[102] F. Paredes-Valles, K. Y. W. Scheper, and G. C. H. E. de Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," IEEE Trans. Pattern Anal. Mach. Intell., Mar. 2019.

[103] C. Reinbacher, G. Munda, and T. Pock, "Real-time panoramic tracking for event cameras," in IEEE Int. Conf. Comput. Photography (ICCP), 2017, pp. 1-9.

[104] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," IEEE Trans. Robot., vol. 34, no. 6, pp. 1425-1440, Dec. 2018.

[105] M. Liu and T. Delbruck, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors," in British Mach. Vis. Conf. (BMVC), 2018.

[106] A. Aimar, H. Mostafa, E. Calabrese, A. Rios-Navarro, R. Tapiador-Morales, I. Lungu, M. B. Milde, F. Corradi, A. Linares-Barranco, S. Liu, and T. Delbrück, "NullHop: A flexible convolutional neural network accelerator based on sparse representations of feature maps," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 3, pp. 644-656, Mar. 2019.

[107] J. Kogler, C. Sulzbachner, and W. Kubinger, "Bio-inspired stereo vision system with silicon retina imagers," in Int. Conf. Comput. Vis. Syst. (ICVS), 2009, pp. 174-183.

[108] A. I. Maqueda, A. Loquercio, G. Gallego, N. García, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018.

[109] X. Lagorce, G. Orchard, F. Gallupi, B. E. Shi, and R. Benosman, "HOTS: A hierarchy of event-based time-surfaces for pattern recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 7, pp. 1346-1359, 2017.

[110] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, "Motion history image: its variants and applications," Mach. Vis. and Applications, vol. 23, no. 2, pp. 255-281, Mar. 2012.

[111] I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 3177-3184, Oct. 2018.

[112] J. Manderscheid, A. Sironi, N. Bourdis, D. Migliore, and V. Lepetit, "Speed invariant time surface for learning to detect corner points with event-based cameras," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[113] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: Histograms of averaged time surfaces for robust event-based object classification," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018.

[114] V. Vasco, A. Glover, and C. Bartolozzi, "Fast event-based Harris corner detection exploiting the advantages of event-driven cameras," in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2016.

[115] E. Mueggler, C. Bartolozzi, and D. Scaramuzza, "Fast event-based corner detection," in British Mach. Vis. Conf. (BMVC), 2017.

[116] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, “Semi-dense 3D reconstruction with a stereo event camera,” in Eur. Conf. Comput. Vis. (ECCV), 2018.

[117] P. Bardow, A. J. Davison, and S. Leutenegger, “Simultaneous optical flow and intensity estimation from an event camera,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2016.

[118] S. Mostafavi I., L. Wang, Y.-S. Ho, and K.-J. Y. Yoon, “Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[119] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “Unsupervised event-based learning of optical flow, depth, and egomotion,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[120] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “Events-to-video: Bringing modern computer vision to event cameras,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[121] Y. Sekikawa, K. Hara, and H. Saito, “EventNet: Asynchronous recursive event processing,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

[122] M. Litzenberger, C. Posch, D. Bauer, A. N. Belbachir, P. Schön, B. Kohn, and H. Garn, “Embedded vision system for real-time object tracking using an asynchronous transient vision sensor,” in Digital Signal Processing Workshop, 2006, pp. 173-178.

[123] Z. Ni, A. Bolopion, J. Agnus, R. Benosman, and S. Régnier, “Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics,” IEEE Trans. Robot., vol. 28, no. 5, pp. 1081-1089, 2012.

[124] Z. Ni, S.-H. Ieng, C. Posch, S. Régnier, and R. Benosman, “Visual tracking using neuromorphic asynchronous event-based cameras,” Neural Computation, vol. 27, no. 4, pp. 925-953, 2015.

[125] D. Tedaldi, G. Gallego, E. Mueggler, and D. Scaramuzza, “Feature detection and tracking with the dynamic and active-pixel vision sensor (DAVIS),” in Int. Conf. Event-Based Control, Comm. Signal Proc. (EBCCSP), 2016.

[126] B. Kueng, E. Mueggler, G. Gallego, and D. Scaramuzza, “Low-latency visual odometry using event-based feature tracks,” in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2016.

[127] G. Gallego and D. Scaramuzza, “Accurate angular velocity estimation with an event camera,” IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 632-639, 2017.

[128] G. Gallego, H. Rebecq, and D. Scaramuzza, “A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018.

[129] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, “Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization,” in British Mach. Vis. Conf. (BMVC), 2017.

[130] A. Z. Zhu, N. Atanasov, and K. Daniilidis, “Event-based feature tracking with probabilistic data association,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2017.

[131] --, “Event-based visual inertial odometry,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.

[132] A. Mitrokhin, C. Fermuller, C. Parameshwara, and Y. Aloimonos, “Event-based moving object detection and tracking,” in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2018.

[133] A. Mitrokhin, C. Ye, C. Fermuller, Y. Aloimonos, and T. Delbruck, “EV-IMO: Motion segmentation dataset and learning pipeline for event cameras,” in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2019.

[134] T. Brosch, S. Tschechne, and H. Neumann, “On event-based optical flow detection,” Front. Neurosci., vol. 9, Apr. 2015.

[135] C. Reinbacher, G. Graber, and T. Pock, “Real-time intensity-image reconstruction for event cameras using manifold regularisation,” in British Mach. Vis. Conf. (BMVC), 2016.

[136] H. Akolkar, S. Panzeri, and C. Bartolozzi, “Spike time based unsupervised learning of receptive fields for event-driven vision,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2015.

[137] J. A. Perez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, and B. Linares-Barranco, “Mapping from frame-driven to frame-free event-driven vision systems by lowrate rate coding and coincidence processing-application to feed-forward ConvNets,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2706-2719, Nov. 2013.

[138] P. O’Connor, D. Neil, S.-C. Liu, T. Delbruck, and M. Pfeiffer, “Real-time classification and sensor fusion with a spiking deep belief network,” Front. Neurosci., vol. 7, p. 178, 2013.

[139] P. U. Diehl, D. Neil, J. Binas, M. Cook, S.-C. Liu, and M. Pfeiffer, “Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing,” in Int. Joint Conf. Neural Netw. (IJCNN), vol. 4, 2015, pp. 2933-2940.

[140] S. K. Esser, P. A. Merolla, J. V. Arthur, A. S. Cassidy, R. Appuswamy, A. Andreopoulos, D. J. Berg, J. L. McKinstry, T. Melano, D. R. Barch, C. di Nolfo, P. Datta, A. Amir, B. Taba, M. D. Flickner, and D. S. Modha, “Convolutional networks for fast, energy-efficient neuromorphic computing,” Proc. National Academy of Sciences, vol. 113, no. 41, pp. 11 441-11 446, 2016.

[141] B. Rueckauer, I.-A. Lungu, Y. Hu, M. Pfeiffer, and S.-C. Liu, “Conversion of continuous-valued deep networks to efficient event-driven networks for image classification,” Front. Neurosci., vol. 11, p. 682, 2017.

[142] S. B. Shrestha and G. Orchard, “SLAYER: Spike layer error reassignment in time,” in Conf. Neural Inf. Process. Syst. (NIPS), 2018.

[143] J. H. Lee, T. Delbruck, and M. Pfeiffer, “Training deep spiking neural networks using backpropagation,” Front. Neurosci., vol. 10, p. 508, 2016.

[144] E. Neftci, “Data and power efficient intelligence with neuromorphic learning machines,” iScience, 2018.

[145] J. Kogler, C. Sulzbachner, M. Humenberger, and F. Eibensteiner, “Address-event based stereo vision with bio-inspired silicon retina imagers,” in Advances in Theory and Applications of Stereo Vision. InTech, 2011, pp. 165-188.

[146] H. Li, G. Li, and L. Shi, “Classification of spatiotemporal events based on random forest,” in Advances in Brain Inspired Cognitive Systems (BICS), 2016.

[147] A. Nguyen, T. Do, D. G. Caldwell, and N. G. Tsagarakis, “Real-time 6DOF pose relocalization for event cameras with stacked spatial LSTM networks,” in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.

[148] E. Mueggler, C. Forster, N. Baumli, G. Gallego, and D. Scaramuzza, “Lifetime estimation of events from dynamic vision sensors,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2015.

[149] S.-H. Ieng, J. Carneiro, M. Osswald, and R. Benosman, “Neuromorphic event-based generalized time-based stereovision,” Front. Neurosci., vol. 12, p. 442, 2018.

[150] D. P. Moeys, F. Corradi, E. Kerr, P. Vance, G. Das, D. Neil, D. Kerr, and T. Delbruck, “Steering a predator robot using a mixed frame/event-driven convolutional neural network,” in Int. Conf. Event-Based Control, Comm. Signal Proc. (EBCCSP), 2016.

[151] I.-A. Lungu, F. Corradi, and T. Delbruck, “Live demonstration: Convolutional neural network driven by dynamic vision sensor playing RoShamBo,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2017.

[152] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, “DDD17: End-to-end DAVIS driving dataset,” in ICML Workshop on Machine Learning for Autonomous Vehicles, 2017.

[153] C. Ye, A. Mitrokhin, C. Parameshwara, C. Fermüller, J. A. Yorke, and Y. Aloimonos, “Unsupervised learning of dense optical flow and depth from sparse event data,” arXiv e-prints, 2018, 1809.08625.

[154] T. Stoffregen and L. Kleeman, “Simultaneous optical flow and segmentation (SOFAS) using Dynamic Vision Sensor,” in Australasian Conf. Robot. Autom. (ACRA), 2017.

[155] T. Stoffregen, G. Gallego, T. Drummond, L. Kleeman, and D. Scaramuzza, “Event-based motion segmentation by motion compensation,” in Int. Conf. Comput. Vis. (ICCV), 2019.

[156] L. A. Camunas-Mesa, T. Serrano-Gotarredona, and B. Linares-Barranco, “Event-driven sensing and processing for high-speed robotic vision,” in IEEE Biomed. Circuits Syst. Conf. (BioCAS), 2014.

[157] G. Orchard, R. Benosman, R. Etienne-Cummings, and N. V. Thakor, “A spiking neural network architecture for visual motion estimation,” in IEEE Biomed. Circuits Syst. Conf. (BioCAS), 2013.

[158] E. Chicca, P. Lichtsteiner, T. Delbruck, G. Indiveri, and R. Douglas, “Modeling orientation selectivity using a neuromorphic multi-chip system,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2006.

[159] R. L. D. Valois, N. P. Cottaris, L. E. Mahon, S. D. Elfar, and J. Wilson, “Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity,” Vision Research, vol. 40, no. 27, pp. 3685-3702, 2000.

[160] F. Rea, G. Metta, and C. Bartolozzi, “Event-driven visual attention for the humanoid robot iCub,” Front. Neurosci., vol. 7, 2013.

[161] L. Itti and C. Koch, “Computational modelling of visual attention,” Nature Reviews Neuroscience, vol. 2, no. 3, pp. 194-203, 2001.

[162] D. Marr and T. Poggio, “Cooperative computation of stereo disparity,” Science, vol. 194, no. 4262, pp. 283-287, 1976.

[163] M. Mahowald, The Silicon Retina. Boston, MA: Springer US, 1994, pp. 4-65.

[164] M. Osswald, S.-H. Ieng, R. Benosman, and G. Indiveri, “A spiking neural network model of 3D perception for event-based neuromorphic stereo vision systems,” Sci Rep, vol. 7, no. 1, Jan. 2017.

[165] G. Dikov, M. Firouzi, F. Röhrbein, J. Conradt, and C. Richter, “Spiking cooperative stereo-matching at 2ms latency with neuromorphic hardware,” in Conf. Biomimetic and Biohybrid Systems, 2017, pp. 119-137.

[166] E. Piatkowska, A. N. Belbachir, and M. Gelautz, “Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach,” in Int. Conf. Comput. Vis. Workshops (ICCVW), 2013.

[167] V. Vasco, A. Glover, Y. Tirupachuri, F. Solari, M. Chessa, and C. Bartolozzi, “Vergence control with a neuromorphic iCub,” in IEEE-RAS Int. Conf. Humanoid Robots (Humanoids), 2016.

[168] M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” Nature Neuroscience, vol. 2, no. 11, pp. 1019-1025, Nov. 1999.

[169] H. Akolkar, C. Meyer, X. Clady, O. Marre, C. Bartolozzi, S. Panzeri, and R. Benosman, “What can neuromorphic event-driven precise timing add to spike-based pattern recognition?” Neural Computation, vol. 27, no. 3, pp. 561-593, Mar. 2015.

[170] L. A. Camunas-Mesa, T. Serrano-Gotarredona, S. H. Ieng, R. B. Benosman, and B. Linares-Barranco, “On the use of orientation filters for 3D reconstruction in event-driven stereo vision,” Front. Neurosci., vol. 8, p. 48, 2014.

[171] M. B. Milde, D. Neil, A. Aimar, T. Delbrück, and G. Indiveri, “ADaPTION: Toolbox and benchmark for training convolutional neural networks with reduced numerical precision weights and activation,” arXiv e-prints, Nov. 2017.

[172] E. Stromatias, M. Soto, T. Serrano-Gotarredona, and B. Linares-Barranco, “An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data,” Front. Neurosci., vol. 11, Jun. 2017.

[173] E. Neftci, C. Augustine, S. Paul, and G. Detorakis, “Event-driven random back-propagation: Enabling neuromorphic deep learning machines,” Front. Neurosci., vol. 11, p. 324, 2017.

[174] M. B. Milde, O. J. N. Bertrand, H. Ramachandran, M. Egelhaaf, and E. Chicca, “Spiking elementary motion detector in neuromorphic systems,” Neural Computation, vol. 30, no. 9, pp. 2384-2417, Sep. 2018.

[175] H. Blum, A. Dietmller, M. Milde, J. Conradt, G. Indiveri, and Y. Sandamirskaya, “A neuromorphic controller for a robotic vehicle equipped with a dynamic vision sensor,” in Robotics: Science and Systems (RSS), 2017.

[176] L. Salt, D. Howard, G. Indiveri, and Y. Sandamirskaya, “Differential evolution and bayesian optimisation for hyper-parameter selection in mixed-signal neuromorphic circuits applied to UAV obstacle avoidance,” arXiv e-prints, 2017, 1704.04853.

[177] D. Drazen, P. Lichtsteiner, P. Häfliger, T. Delbrück, and A. Jensen, “Toward real-time particle tracking using an event-based dynamic vision sensor,” Experim. Fluids, vol. 51, no. 5, pp. 1465-1469, 2011.

[178] Z. Ni, C. Pacoret, R. Benosman, S.-H. Ieng, and S. Régnier, “Asynchronous event-based high speed vision for microparticle tracking,” J. Microscopy, vol. 245, no. 3, pp. 236-244, 2012.

[179] E. Piatkowska, A. N. Belbachir, S. Schraml, and M. Gelautz, “Spatiotemporal multiple persons tracking using dynamic vision sensor,” in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2012.

[180] X. Lagorce, C. Meyer, S.-H. Ieng, D. Filliat, and R. Benosman, “Asynchronous event-based multikernel algorithm for high-speed visual features tracking,” IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 8, pp. 1710-1720, 2015.

[181] A. Glover and C. Bartolozzi, “Robust visual tracking with a freely-moving event camera,” in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2017.

[182] D. R. Valeiras, X. Lagorce, X. Clady, C. Bartolozzi, S.-H. Ieng, and R. Benosman, “An asynchronous neuromorphic event-driven visual part-based shape tracking,” IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 12, pp. 3045-3059, Dec. 2015.

[183] M. A. Fischler and R. A. Elschlager, “The representation and matching of pictorial structures,” IEEE Trans. Comput., vol. C-22, no. 1, pp. 67-92, Jan. 1973.

[184] C. Harris and M. Stephens, “A combined corner and edge detector,” in Proc. Fourth Alvey Vision Conf., vol. 15, 1988, pp. 147-151.

[185] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in Int. Joint Conf. Artificial Intell. (IJCAI), 1981.

[186] X. Clady, S.-H. Ieng, and R. Benosman, “Asynchronous event-based corner detection and matching,” Neural Netw., vol. 66, pp. 91-106, 2015.

[187] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in Eur. Conf. Comput. Vis. (ECCV), 2006.

[188] V. Vasco, A. Glover, E. Mueggler, D. Scaramuzza, L. Natale, and C. Bartolozzi, “Independent motion detection with event-driven cameras,” in IEEE Int. Conf. Adv. Robot. (ICAR), 2017.

[189] Y. Hu, H. Liu, M. Pfeiffer, and T. Delbruck, “DVS benchmark datasets for object tracking, action recognition, and object recognition,” Front. Neurosci., vol. 10, p. 405, 2016.

[190] B. Rueckauer and T. Delbruck, “Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor,” Front. Neurosci., vol. 10, no. 176, 2016.

[191] R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, “Asynchronous frameless event-based optical flow,” Neural Netw., vol. 27, pp. 32-37, 2012.

[192] F. Barranco, C. Fermuller, and Y. Aloimonos, “Contour motion estimation for asynchronous event-driven cameras,” Proc. IEEE, vol. 102, no. 10, pp. 1537-1556, Oct. 2014.

[193] --, “Bio-inspired motion estimation with event-driven sensors,” in Int. Work-Conf. Artificial Neural Netw. (IWANN), Advances in Comput. Intell., 2015, pp. 309-321.

[194] J. Conradt, “On-board real-time optic-flow for miniature event-based vision sensors,” in IEEE Int. Conf. Robot. Biomimetics (ROBIO), 2015.

[195] S. Tschechne, R. Sailer, and H. Neumann, “Bio-inspired optic flow from event-based neuromorphic sensor input,” in Artificial Neural Networks in Pattern Recognition (ANNPR), 2014, pp. 171-182.

[196] G. Haessig, A. Cassidy, R. Alvarez-Icaza, R. Benosman, and G. Orchard, “Spiking optical flow for event-based sensors using IBM’s truenorth neurosynaptic system,” IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 4, pp. 860-870, Aug. 2018.

[197] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, “The multivehicle stereo event camera dataset: An event camera dataset for 3D perception,” IEEE Robot. Autom. Lett., vol. 3, no. 3, pp. 2032-2039, Jul. 2018.

[198] D. J. Fleet and A. D. Jepson, “Computation of component image velocity from local phase information,” Int. J. Comput. Vis., vol. 5, no. 1, pp. 77-104, 1990.

[199] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2003, 2nd Edition.

[200] S. Schraml, A. N. Belbachir, N. Milosevic, and P. Schön, “Dynamic stereo vision system for real-time tracking,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2010, pp. 1409-1412.

[201] J. Kogler, M. Humenberger, and C. Sulzbachner, “Event-based stereo matching approaches for frameless address event stereo data,” in Int. Symp. Adv. Vis. Comput. (ISVC), 2011, pp. 674-685.

[202] J. Lee, T. Delbruck, P. K. J. Park, M. Pfeiffer, C.-W. Shin, H. Ryu, and B. C. Kang, “Live demonstration: Gesture-based remote control using stereo pair of

dynamic vision sensors," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2012.

[203] E. Piatkowska, A. N. Belbachir, and M. Gelautz, "Cooperative and asynchronous stereo vision for dynamic vision sensors," Meas. Sci. Technol., vol. 25, no. 5, p. 055108, Apr. 2014.

[204] R. Benosman, S.-H. Ieng, P. Rogister, and C. Posch, "Asynchronous event-based Hebbian epipolar geometry," IEEE Trans. Neural Netw., vol. 22, no. 11, pp. 1723-1734, 2011.

[205] J. Carneiro, S.-H. Ieng, C. Posch, and R. Benosman, "Event-based 3D reconstruction from neuromorphic retinas," Neural Netw., vol. 45, pp. 27-38, 2013.

[206] D. Zou, P. Guo, Q. Wang, X. Wang, G. Shao, F. Shi, J. Li, and P.-K. J. Park, "Context-aware event-driven stereo matching," in IEEE Int. Conf. Image Process. (ICIP), 2016, pp. 1076-1080.

[207] D. Zou, F. Shi, W. Liu, J. Li, Q. Wang, P.-K. J. Park, C.-W. Shi, Y. J. Roh, and H. E. Ryu, "Robust dense depth map estimation from sparse DVS stereos," in British Mach. Vis. Conf. (BMVC), 2017.

[208] M. Firouzi and J. Conradt, "Asynchronous event-based cooperative stereo matching using neuromorphic silicon retinas," Neural Proc. Lett., vol. 43, no. 2, pp. 311-326, 2016.

[209] J. Kogler, F. Eibensteiner, M. Humenberger, C. Sulzbachner, M. Gelautz, and J. Scharinger, "Enhancement of sparse silicon retina-based stereo matching using belief propagation and twostage postfiltering," J. Electron. Imag., vol. 23, no. 4, pp. 1-15, 2014.

[210] Z. Xie, S. Chen, and G. Orchard, "Event-based stereo depth estimation using belief propagation," Front. Neurosci., vol. 11, 2017.

[211] Z. Xie, J. Zhang, and P. Wang, "Event-based stereo matching using semiglobal matching," Int. J. Advanced Robotic Systems, vol. 15, no. 1, 2018.

[212] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 328-341, Feb. 2008.

[213] A. Andreopoulos, H. J. Kashyap, T. K. Nayak, A. Amir, and M. D. Flickner, "A low power, high throughput, fully event-based stereo system," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018.

[214] A. Z. Zhu, Y. Chen, and K. Daniilidis, "Realtime time synchronized event-based stereo," in Eur. Conf. Comput. Vis. (ECCV), 2018.

[215] R. Szeliski, Computer Vision: Algorithms and Applications, ser. Texts in Computer Science. Springer, 2010.

[216] S. Schraml, A. N. Belbachir, and H. Bischof, "An event-driven stereo system for real-time 3-D 360 panoramic vision," IEEE Trans. Ind. Electron., vol. 63, no. 1, pp. 418-428, Jan. 2016.

[217] --, "Event-driven stereo matching for real-time 3D panoramic vision," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2015.

[218] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robustperception age," IEEE Trans. Robot., vol. 32, no. 6, pp. 1309-1332, 2016.

[219] R. T. Collins, "A space-sweep approach to true multi-image matching," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 1996.

[220] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in Int. Conf. Comput. Vis. (ICCV), 2011.

[221] C. Brandli, T. Mantel, M. Hutter, M. Höpflinger, R. Berner, R. Siegwart, and T. Delbruck, "Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor," Front. Neurosci., vol. 7, p. 275, 2014.

[222] J. N. P. Martel, J. Müller, J. Conradt, and Y. Sandamirskaya, "An active approach to solving the stereo matching problem using event-based sensors," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2018.

[223] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. The MIT Press, Cambridge, MA, 2005.

[224] D. Weikersdorfer, R. Hoffmann, and J. Conradt, "Simultaneous localization and mapping for event-based vision systems," in Int. Conf. Comput. Vis. Syst. (ICVS), 2013.

[225] D. Weikersdorfer, D. B. Adrian, D. Cremers, and J. Conradt, "Event-based 3D SLAM with a depth-augmented dynamic vision sensor," in IEEE Int. Conf. Robot. Autom. (ICRA), 2014.

[226] E. Mueggler, B. Huber, and D. Scaramuzza, "Event-based, 6-DOF pose tracking for high-speed maneuvers," in IEEE Int. Conf. Intell. Robot. Syst. (IROS), 2014, pp. 2761-2768.

[227] M. Milford, H. Kim, S. Leutenegger, and A. Davison, "Towards visual SLAM with event-based cameras," in The Problem of Mobile Sensors Workshop in conjunction with RSS, 2015.

[228] E. Mueggler, G. Gallego, and D. Scaramuzza, "Continuous-time trajectory estimation for event-based vision sensors," in Robotics: Science and Systems (RSS), 2015.

[229] Y. Zhou, G. Gallego, and S. Shen, "Event-based stereo visual odometry," arXiv preprint arXiv:2007.15548, 2020.

[230] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "Onmanifold preintegration for real-time visual-inertial odometry," IEEE Trans. Robot., vol. 33, no. 1, pp. 1-21, 2017.

[231] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in IEEE Int. Conf. Robot. Autom. (ICRA), 2007.

[232] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial SLAM using nonlinear optimization," Int. J. Robot. Research, 2015.

[233] A. Patron-Perez, S. Lovegrove, and G. Sibley, "A spline-based trajectory representation for sensor

fusion and rolling shutter cameras," Int. J. Comput. Vis., vol. 113, no. 3, pp. 208-219, 2015.

[234] L. von Stumberg, V. Usenko, and D. Cremers, "Direct sparse visual-inertial odometry using dynamic marginalization," in IEEE Int. Conf. Robot. Autom. (ICRA), 2018.

[235] S. Barua, Y. Miyatani, and A. Veeraraghavan, "Direct face detection and video reconstruction from event cameras," in IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2016, pp. 1-9.

[236] G. Munda, C. Reinbacher, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," Int. J. Comput. Vis., vol. 126, no. 12, pp. 1381-1393, 2018.

[237] C. Scheerlinck, H. Rebecq, D. Gehrig, N. Barnes, R. Mahony, and D. Scaramuzza, "Fast image reconstruction with an event camera," in IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2020.

[238] A. N. Belbachir, S. Schraml, M. Mayerhofer, and M. Hofstaetter, "A novel HDR depth camera for real-time 3D 360-degree panoramic vision," in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2014.

[239] G. Wiesmann, S. Schraml, M. Litzenberger, A. N. Belbachir, M. Hofstatter, and C. Bartolozzi, "Event-driven embodied system for feature extraction and object recognition in robotic applications," in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2012.

[240] D. Gehrig, M. Gehrig, J. Hidalgo-Carrió, and D. Scaramuzza, "Video to Events: Recycling video datasets for event cameras," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2020.

[241] A. Zanardi, A. J. Aumiller, J. Zilly, A. Censi, and E. Frazzoli, "Cross-modal learning filters for rgb-neuromorphic wormhole learning," in Proceedings of Robotics: Science and Systems, FreiburgimBreisgau, Germany, June 2019.

[242] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," Front. Neurosci., vol. 9, p. 437, 2015.

[243] D. Neil and S.-C. Liu, "Effective sensor fusion with event-based sensors and deep network architectures," in IEEE Int. Symp. Circuits Syst. (ISCAS), 2016, pp. 2282-2285.

[244] Y. Wu, L. Deng, G. Li, J. Zhu, and L. Shi, "Spatio-temporal backpropagation for training high-performance spiking neural networks," Front. Neurosci., vol. 12, 2018.

[245] A. Yousefzadeh, G. Orchard, T. Serrano-Gotarredona, and B. Linares-Barranco, "Active perception with dynamic vision sensors. minimum saccades with optimum recognition," IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 4, pp. 927-939, Aug. 2018.

[246] X. Clady, J.-M. Maro, S. Barré, and R. B. Benosman, "A motion-based feature for event-based pattern recognition," Front. Neurosci., vol. 10, Jan. 2017.

[247] C. A. Sims, "Implications of rational inattention," J. Monetary Economics, vol. 50, pp. 665-690, 2003.

[248] M. Miskowicz, Event-Based Control and Signal Processing, ser. Embedded Systems. CRC Press, 2018.

[249] P. Tabuada, "An introduction to event-triggered and self-triggered control," in IEEE Conf. Decision Control (CDC), 2012.

[250] K. J. Aström, Event Based Control. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 127-147.

[251] B. Wang and M. Fu, "Comparison of periodic and event-based sampling for linear state estimation," IFAC Proc. Volumes (IFAC-PapersOnline), vol. 19, pp. 5508-5513, 2014.

[252] A. Censi, E. Mueller, E. Frazzoli, and S. Soatto, "A powerperformance approach to comparing sensor families, with application to comparing neuromorphic to traditional vision sensors," in IEEE Int. Conf. Robot. Autom. (ICRA), 2015.

[253] E. Mueller, A. Censi, and E. Frazzoli, "Low-latency heading feedback control with neuromorphic vision sensors using efficient approximated incremental inference," in IEEE Conf. Decision Control (CDC), 2015.

[254] P. Singh, S. Z. Yong, J. Gregoire, A. Censi, and E. Frazzoli, "Stabilization of linear continuous-time systems using neuromorphic vision sensors," in IEEE Conf. Decision Control (CDC), 2016.

[255] A. Censi, "Efficient neuromorphic optomotor heading regulation," in IEEE Am. Control Conf. (ACC), 2015.

[256] E. Mueller, A. Censi, and E. Frazzoli, "Efficient high speed signal estimation with neuromorphic vision sensors," in Int. Conf. Event-Based Control, Comm. Signal Proc. (EBCCSP), 2015.

[257] D. Kahneman, Thinking, fast and slow. Farrar, Straus & Giroux, 2011.

[258] S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, and A. D. Brown, "Overview of the SpiNNaker system architecture," IEEE Trans. Comput., vol. 62, no. 12, pp. 2454-2467, 2013.

[259] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G.-J. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, and D. S. Modha, "TrueNorth: Design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 34, no. 10, pp. 1537-1557, 2015.

[260] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain et al., "Loihi: A neuromorphic manycore processor with on-chip learning," IEEE Micro, vol. 38, no. 1, pp. 82-99, 2018.

[261] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)," IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 1, pp. 106-122, 2018.

[262] A. S. Neckar, "Braindrop: a mixed signal neuromorphic architecture with a dynamical systems-based programming model," Ph.D.

dissertation, Stanford University, Stanford, CA, Jun. 2018.

[263] L. A. Camunas-Mesa, B. Linares-Barranco, and T. Serrano-Gotarredona, “Neuromorphic spiking neural networks and their memristor-CMOS hardware implementations,” Materials, vol. 12, no. 17, p. 2745, Aug. 2019.

[264] B. Rajendran, A. Sebastian, M. Schmuker, N. Srinivasa, and E. Eleftheriou, “Low-power neuromorphic hardware for signal processing applications: A review of architectural and system-level design approaches,” IEEE Signal Process. Mag., vol. 36, no. 6, pp. 97-110, Nov. 2019.

[265] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, “The SpiNNaker Project,” Proc. IEEE, vol. 102, no. 5, pp. 652-665, 2014.

[266] G. Haessig, F. Galluppi, X. Lagorce, and R. Benosman, “Neuromorphic networks on the SpiNNaker platform,” in IEEE Int. Conf. Artificial Intell. Circuits Syst. (AICAS), 2019.

[267] C. Richter, F. Röhrbein, and J. Conradt, “Bio inspired optic flow detection using neuromorphic hardware,” in Bernstein Conf., 2014.

[268] A. Glover, A. B. Stokes, S. Furber, and C. Bartolozzi, “ATIS + SpiNNaker: a fully event-based visual tracking demonstration,” in IEEE/RSJ Int. Conf. Intell. Robot. Syst. Workshops (IROSW), 2018.

[269] T. Serrano-Gotarredona, B. Linares-Barranco, F. Galluppi, L. Plana, and S. Furber, “ConvNets experiments on SpiNNaker,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2015.

[270] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura et al., “A million spiking-neuron integrated circuit with a scalable communication network and interface,” Science, vol. 345, no. 6197, pp. 668-673, 2014.

[271] P. Blouw, X. Choo, E. Hunsberger, and C. Eliasmith, “Benchmarking Keyword Spotting Efficiency on Neuromorphic Hardware,” arXiv e-prints, p. arXiv:1812.01739, Dec. 2018.

[272] T. Bekolay, J. Bergstra, E. Hunsberger, T. DeWolf, T. Stewart, D. Rasmussen, X. Choo, A. Voelker, and C. Eliasmith, “Nengo: a Python tool for building large-scale functional brain models,” Front. Neuroinf., vol. 7, p. 48, 2014.

[273] C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, Y. Tang, and D. Rasmussen, “A large-scale model of the functioning brain,” Science, vol. 338, no. 6111, pp. 1202-1205, 2012.

[274] N. Waniek, J. Biedermann, and J. Conradt, “Cooperative SLAM on small mobile robots,” in IEEE Int. Conf. Robot. Biomimetics (ROBIO), 2015.

[275] B. J. P. Hordijk, K. Y. Scheper, and G. C. D. Croon, “Vertical landing for micro air vehicles using event-based optical flow,” J. Field Robot., vol. 35, no. 1, pp. 69-90, Jan. 2017.

[276] jAER Project, https://github.com/SensorsINI/jaer, 2007.

[277] T. Delbruck, M. Pfeiffer, R. Juston, G. Orchard, E. Müggler, A. Linares-Barranco, and M. W. Tilden, “Human vs. computer slot car racing using an event and frame-based DAVIS vision sensor,” in IEEE Int. Symp. Circuits Syst. (ISCAS), 2015.

[278] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “ROS: an open-source Robot Operating System,” in ICRA Workshop Open Source Softw., vol. 3, no. 2, 2009, p. 5.

[279] A. Glover, V. Vasco, M. Iacono, and C. Bartolozzi, “The event-driven software library for YARP-with algorithms and iCub applications,” Front. Robotics and AI, vol. 4, p. 73, 2018.

[280] Y. Hu, J. Binas, D. Nei, S.-C. Liu, and T. Delbruck, “DDD20 Endto-End event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction,” in IEEE Int. Conf. Intell. Transportation Systems (ITSC), 2020.

[281] F. Barranco, C. Fermuller, Y. Aloimonos, and T. Delbruck, “A dataset for visual navigation with neuromorphic methods,” Front. Neurosci., vol. 10, p. 49, 2016.

[282] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, “Are we ready for autonomous drone racing? the UZH-FPV drone racing dataset,” in IEEE Int. Conf. Robot. Autom. (ICRA), 2019.

[283] C. Tan, S. Lallee, and G. Orchard, “Benchmarking neuromorphic vision: lessons learnt from computer vision,” Front. Neurosci., vol. 9, p. 374, 2015.

[284] S. Miao, G. Chen, X. Ning, Y. Zi, K. Ren, Z. Bing, and A. Knoll, “Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection,” Front. Neurorobot., 2019.

[285] E. Calabrese, G. Taverni, C. Awai Easthope, S. Skriabine, F. Corradi, L. Longinotti, K. Eng, and T. Delbruck, “DHP19: Dynamic vision sensor 3D human pose dataset,” in IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW), 2019.

[286] G. P. Garcia, P. Camilleri, Q. Liu, and S. B. Furber, “pyDVS: An extensible, real-time dynamic vision sensor emulator using off-the-shelf hardware,” in IEEE Symp. Series Comput. Intell. (SSCI), 2016.

[287] S. J. Carey, A. Lopich, D. R. Barr, B. Wang, and P. Dudek, “A 100,000 fps vision sensor with embedded 535 GOPS/W 256x256 SIMD processor array,” in VLSI Circuits Symp., 2013.

[288] J. N. P. Martel, “Unconventional processing with unconventional visual sensing: Parallel, distributed and event based vision algorithms & systems,” Ph.D. dissertation, ETH-Zurich, Zurich, Switzerland, 2019.