

# 基于事件相机的定位与建图算法: 综述

马艳阳<sup>1,2</sup>叶梓豪<sup>1,2</sup>刘坤华<sup>1,2</sup>陈龙<sup>1,2</sup>

**摘要** 事件相机是一种新兴的视觉传感器, 通过检测单个像素点光照强度的变化来产生“事件”。基于其工作原理, 事件相机拥有传统相机所不具备的低延迟、高动态范围等优良特性。而如何应用事件相机来完成机器人的定位与建图则是目前视觉定位与建图领域新的研究方向。本文从事件相机本身出发, 介绍事件相机的工作原理、现有的定位与建图算法以及事件相机相关的开源数据集。其中, 本文着重对现有的、基于事件相机的定位与建图算法进行详细的介绍和优缺点分析。

**关键词** 事件相机, 低延迟, 位姿估计, 定位与建图

**引用格式** 马艳阳, 叶梓豪, 刘坤华, 陈龙. 基于事件相机的定位与建图算法: 综述. 自动化学报, 2020, 46(x): 1-11

**DOI** 10.16383/j.aas.c190550

## Event-based Visual Localization and Mapping Algorithms: A Survey

MA Yan-Yang<sup>1,2</sup>YE Zi-Hao<sup>1,2</sup>LIU Kun-Hua<sup>1,2</sup>CHEN Long<sup>1,2</sup>

**Abstract** Event-based camera is a new type of visual sensor, which activates “events” by monitoring the changes of lighting intensity. Event camera offers low-latency output and tolerates high dynamic range, therefore, it is a new topic in the SLAM area applying event-based camera to robot localization and mapping. This work introduces the working principle of event-based camera and reviews present algorithms and dataset of event-based localization and mapping. Emphasis of this work is introducing and analysing event-based localization and mapping algorithms.

**Key words** Event camera, low latency, pose estimation, localization and mapping

**Citation** Ma Yan-Yang, Ye Zi-Hao, Liu Kun-Hua, Chen Long. Event-based visual localization and mapping algorithms: a survey. *Acta Automatica Sinica*, 2020, 46(x): 1-11

## 1 简介

感知自身在空间中的位置、方向、速度以及周围环境信息, 是自动驾驶系统、移动机器人等无人系统进行自主导航、路径规划<sup>[1]</sup>等任务的前提。因此, 同时定位与建图 (Simultaneous Localization and Mapping, SLAM) 技术<sup>[2]</sup>被广泛应用于自动驾驶、移动机器人、无人机、增强现实<sup>[3]</sup>等领域。在 SLAM 算法中常用的传感器包括全球导航卫星系统 (Global Navigation Satellite System, GNSS)、惯性导航系统 (Inertial Navigation System, INS)、激光雷达 (Light Detection and Ranging, Lidar)、毫米波雷达 (Radio Detection and Ranging, Ra-

dar)、相机等。相较于其他传感器, 相机具有体积小、成本低、易部署、低功耗、能够提供丰富的信息等优点; 但由于其视野受限、容易受光照影响, 同时无法直接获得深度信息, 基于视觉的定位与建图<sup>[4-6]</sup>方法比其它传感器的方法更复杂。因此, 视觉 SLAM 一直是 SLAM 领域里备受关注的研究方向。

单目相机无法获得深度信息, 这不仅使单目 SLAM 算法更复杂, 而且导致通过单目 SLAM 算法得到的相机位姿和地图具有尺度不确定性。而利用多个相机之间的内参 (Intrinsic) 和外参 (Extrinsic), 则可以将多个相机组成双目或多目相机系统; 然后根据多目相机系统对同一点的组合观测, 再利用三角测量的原理便可以计算出该点的深度。早在 2003 年, Olson 等人<sup>[7]</sup>便使用这种方法进行相机的位姿估计。然而由于计算量较大, 三角测量通常只能针对稀疏的特征点; 对所有像素进行三角测量和深度恢复, 需要有额外的设备和算法来支持。随着传感器技术的发展, 以微软的 Kinect<sup>[8]</sup>为代表的 RGB-D 相机可以通过红外结构光或飞行时间法 (Time of Flight, TOF) 来直接获取场景中的三维信息, 使得 RGB-D 相机成为视觉 SLAM 中代替双目相机的可选方案<sup>[9]</sup>。除此之外, 一些研究者结合

收稿日期 2019-07-25 录用日期 2019-12-15

Manuscript received July 25, 2019; accepted December 15, 2019

国家重点研发计划 (2018YFB1305002), 国家自然科学基金 (61773414) 资助

Supported by National Key Research and Development Program of China (2018YFB1305002), National Natural Science Foundation of China (61773414)

本文责任编辑 张哲

Recommended by Associate Editor ZHANG Zhe

1. 中山大学数据科学与计算机学院 广州 510006 2. 中山大学无人系统研究所 广州 510006

1. School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006 2. Institute of Unmanned Systems, Sun Yat-sen University, Guangzhou 510006

相机和 IMU 两种传感器来构建视觉 - 惯性系统 (Visual-Inertial System, VIS)<sup>[10-11]</sup>, 利用视觉信息修正 IMU 的累积误差, 同时利用 IMU 提供的线加速度和角速度获得相机的绝对位姿, 为视觉里程计提供良好的位姿初值, 并且可以为图像进行运动矫正. 视觉 SLAM 技术发展至今, 领域内已经有许多经典的算法, 其中基于纯视觉的代表性算法有 PTAM<sup>[12]</sup>、ORB-SLAM<sup>[13-14]</sup>、SVO<sup>[15]</sup>、LSD-SLAM<sup>[16-17]</sup> 等, 基于视觉 - 惯性系统的代表性工作有 MSCKF<sup>[18]</sup>、OKVIS<sup>[19]</sup> 以及香港科技大学研发的 VINS<sup>[20]</sup> 等.

尽管相机凭借其自身的优势, 在 SLAM 领域中被广泛地应用, 但是传统的 CMOS<sup>[21]</sup> 相机 (Active Pixel Sensor, APS) 存在着不少缺陷: 每个像素需要统一时间曝光, 快速运动时图像容易产生运动模糊; 相机获取信息的频率受曝光时间的限制; 在高动态范围 (High Dynamic Range, HDR) 的场景下容易出现图像部分过曝或部分欠曝的现象, 导致场景细节丢失. 这些缺陷的存在限制了视觉 SLAM 的使用场景, 因此一类新型的基于事件的动态视觉相机进入了 SLAM 研究者的视野. 基于事件的相机 (简称事件相机) 是一类异步 (asynchronous) 的相机<sup>[22]</sup>, 它的诞生是受到了生物视网膜的启发. 事件相机不再像传统的相机那样需要外部信号来触发像素统一曝光而获得完整的图像, 而是每个像素单独检测相对光照变化, 并输出光照变化的像素地址-事件 (Address-Event, AE) 流. 图 1<sup>[47]</sup> 为带一个黑色圆点的匀速旋转圆盘在事件相机中产生地址-事件流的示意图. 其中 X、Y 轴张成像素地址空间, 不同时刻被激发的像素点在时空上形成一条螺旋线.

传统相机的曝光时间是固定的, 即便某个像素上的光照没有发生变化, 它也会重复曝光. 这种工作原理导致传统相机输出的图像有高延迟、高冗余

的缺点. 而事件相机只输出由光照变化所触发的“事件”, 因此其输出的事件流是稀疏的、低延迟以及低冗余的. 此外, 事件相机在高动态范围的场景下也能良好地工作. 有着上述的优势, 事件相机在运动物体追踪<sup>[23-25]</sup>、光流估计<sup>[26-29]</sup>、高动态范围图像重建<sup>[30]</sup> 等问题上都得到成功的应用.

本文将对事件相机及其在 SLAM 领域的算法做详细的介绍, 文章结构如下: 第二节将介绍不同类型的事件相机; 第三节将介绍与分析基于事件相机的定位与建图算法; 用于定位与建图算法的事件相机数据集将在第四节介绍; 最后, 第五节是总结与展望.

## 2 不同种类的事件相机及其工作原理

### 2.1 事件相机概况

事件相机的出现, 可以追溯到 1992 年, Misha Mahowald 在他的博士论文<sup>[31]</sup> 中提出了一种新型视觉传感器“硅视网膜” (“Silicon Retina”). 这是首个输出地址 - 事件流的视觉传感器, 但它只是事件相机的雏形; 像素面积过大导致它无法被实际使用. 在随后几十年的发展中, 越来越多像素面积更小, 填充因数 (Fill Factor) 更大的事件相机被研发出来<sup>[32]</sup>. 其中被广泛接受并使用的可分为以下三类: DVS (Dynamic Vision Sensor)、ATIS (Asynchronous Time Based Image Sensor)、DAVIS (Dynamic and Active Pixel Vision Sensor). DVS 是输出只有事件流的事件相机, 而 ATIS 和 DAVIS 除了能够输出事件流, 还能够输出灰度信息. 上述三种事件相机已经有商用产品的版本, 包括 iniVation 公司的 DV-S128<sup>[34]</sup>、DAVIS240<sup>[39]</sup>, Prophesee 公司的 ATIS<sup>[37]</sup>, CelePixel 公司的 CeleX-IV<sup>[40]</sup>. 除此之外, 考虑到无论是自然界动物的视觉还是传统的 APS 相机都可以获取颜色信息, 研究者们尝试利用滤镜或分色棱镜, 将颜色信息融入到事件相机中<sup>[41-43]</sup>. 不过这方面的工作仍处在学术研究的阶段. 本章节的后续部分将对 DVS、ATIS、DAVIS 这三种事件相机的工作原理进行介绍.

### 2.2 DVS

DVS 是只输出地址 - 事件流 (简称事件流) 的一类事件相机<sup>[33-35]</sup>, 它的单个像素结构和工作原理简化图分别如图 2<sup>[34]</sup> 以及图 3<sup>[34]</sup> 所示. 其工作原理与传统的 APS 相机截然不同: 每个像素单独地检测照射到该像素上的光强对数 (Log Intensity) 的变化, 并根据光强对数的变化异步地输出变亮 (“ON”)

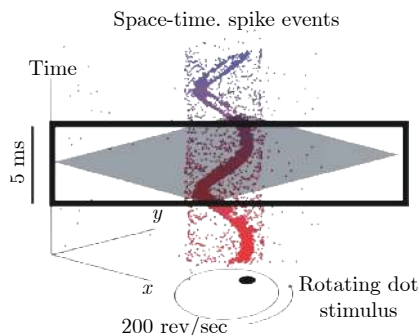
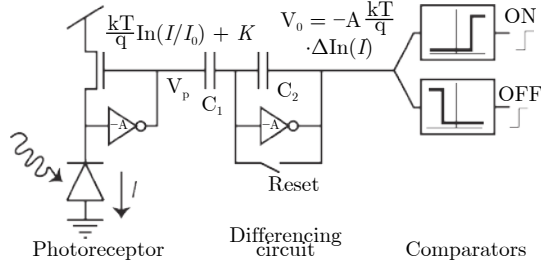
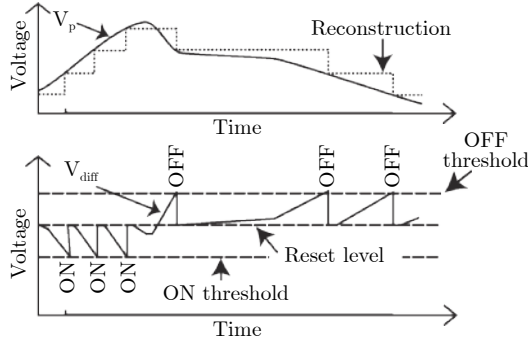


图 1 事件相机输出的地址 - 事件流<sup>[47]</sup>

Fig. 1 Address-event stream output by event-based camera<sup>[47]</sup>

图2 DVS像素结构原理图<sup>[34]</sup>Fig. 2 Abstracted DVS pixel core schematic<sup>[34]</sup>图3 DVS工作原理图<sup>[34]</sup>Fig. 3 Principle of DVS operation<sup>[34]</sup>

或者变暗 (“OFF”) 事件. 更具体地说, 当某个像素激发一个变亮或变暗事件时, 它会通过重置操作 (Reset) 来记录当前时刻的光强对数, 并持续检测光强对数的变化. 一旦某个时刻的光强对数相较于记录中的光强对数的增大或减小量超过一定的阈值, 该像素会激发一个新的变亮或变暗事件, 并记录新的光强对数. 由像素阵列激发的事件, 经过外围器件的处理后便可以发送给上位机.

通常, DVS 输出的事件  $e_k$  包含该事件发生的像素地址、时间和事件的类型 (极性). 即  $e_k = (\mathbf{x}_k, t_k, p_k)$ , 其中  $\mathbf{x}_k = (x_k, y_k)^T$ , 表示像素地址;  $t_k$  表示该事件发生的时间;  $p_k \in \{+1, -1\}$  表示事件的极性,  $p_k = +1$  为变亮事件, 反之则为变暗事件. 规定在像素  $x_k$  在  $t_k$  时刻的光强对数表示为  $L(\mathbf{x}_k, t_k) = \log(I(\mathbf{x}_k, t_k))$ , 其中  $I$  表示光照强度. 事件  $e_k$  被激发, 意味着像素  $x_k$  上的光强对数与该像素上一次激发事件时的光强对数之差, 达到了预先设置的阈值  $\pm C$ , 即:

$$\Delta L(\mathbf{x}_k, t_k) = p_k C \quad (1)$$

其中,

$$\Delta L(\mathbf{x}_k, t_k) = L(\mathbf{x}_k, t_k) - L(\mathbf{x}_k, t_k - \Delta t_k) \quad (2)$$

$\Delta t_k$  表示从像素  $x_k$  上次激发事件开始经过的时间.

上述的事件生成模型, 仅考虑了没有噪声的理想情况. 实际上, 任何相机都会受到噪声的影响, 包括来自感光器的外部噪声以及内部电路产生的噪声, DVS 也不例外. 通过设置阈值  $C$ , 可以调节 DVS 对噪声的敏感度: 阈值  $C$  越大, DVS 对噪声越不敏感, 但是 DVS 捕捉到的事件也会相应地减少. 阈值  $C$  越小, DVS 捕捉到的事件越多, 但信噪比会严重降低.

在某个像素上的光照强度的改变, 通常是由两种情况造成的: 场景亮度的改变、场景中的物体或相机自身发生了移动. 并且, 当场景亮度变化越快、场景中的物体或相机自身移动越快时, 单位时间内 DVS 产生的事件就越多. 事件源源不断地从 DVS 中异步地输出, 形成事件流. 根据硬件和设计的不同, DVS 输出速率可高达 300 Meps (events per second, eps)<sup>[35]</sup>, 延迟可低至 3 微秒<sup>[39]</sup>, 动态范围可达 130 dB. 这样的特性让 DVS 在捕捉高速运动的物体、应对高动态范围的光照环境、低延迟控制等应用上有着巨大的优势.

### 2.3 ATIS

单纯的 DVS 输出的事件流, 只提供了事件的事件地址、事件和极性, 而 ATIS 不仅能输出上述信息, 还能输出发生事件的像素的灰度值<sup>[36-37]</sup>. ATIS 的像素结构可以划分为两个子像素 (以下简称甲、乙子像素): 甲子像素包含完整的 DVS 像素结构, 它可以感应光强对数的变化并激发相应的事件; 在甲子像素激发事件的同时, 它还会激发乙子像素进行 “曝光”. ATIS 像素中的乙子像素的曝光方式, 与传统 APS 相机像素的曝光方式不同. 传统相机的曝光方式是预置曝光时间, 通过测定曝光后电容两端的电压来确定像素的灰度值. 因此, 光照强度越大、曝光时间越长, 像素的灰度值越大. 而 ATIS 像素中的乙子像素采用一种基于时间的异步曝光方式: 甲子像素激发事件的同时, 乙子像素内的一个电容被重置为高电平. 随着乙子像素持续受到光照, 电容两端的电压下降, 其两端电压从高电平下降到低电平所需的时间决定了该像素的灰度值. 电压下降时间越短, 说明该像素上光照强度越强, 灰度值越高; 下降时间越长, 则说明灰度值越低. 像素的灰度信息是以独立的事件流的形式输出的. 这种由事件触发的基于时间的曝光方式, 使得 ATIS 相机在高动态范围的场景下也能获得较好的灰度图 (传统 APS 相机容易过曝或欠曝). 并且只有产生事件的像素会输出灰度信息, 减少了信息的时间冗余和空间冗余. 然而 ATIS 相机也有一定的



缺陷: 在亮度过低的情况下曝光时间过长. 当曝光时间超过了下一次时间触发的时间, 电容会被强行重置导致曝光异常终止, 从而导致信息丢失.

## 2.4 DAVIS

DAVIS<sup>[38-39]</sup> 相机也是一种能够输出事件流和灰度信息的时间相机, 它是 DVS 相机和传统的 APS 相机的结合体. DAVIS 像素也分为两个子结构, 其中一个子结构用于监测光照变化的事件, 而另一个子结构像传统的 APS 相机一样进行同步的曝光. 不同于 ATIS 的是, DAVIS 像素的两个子结构是共用同一个感光器的, 而 ATIS 像素的两个子像素有各自的感光器. 因此, DAVIS 相机的像素面积较 ATIS 相机的像素面积更小, 前者的填充因数较后者的更大. 通过 DAVIS 获取的灰度图, 和通过传统 APS 相机获取的灰度图一样, 具有时间冗余性和空间冗余性, 并且无法承受高动态范围的环境. 因此, 从功能上看, DAVIS 只是 DVS 和 APS 的简单结合体.

## 3 基于事件相机的定位与建图算法

从 2012 年开始, 机器人领域逐渐出现基于事件相机和事件流的定位与建图算法的研究. 它们对事件流的使用方式各不相同, 并且其中的部分算法结合了其它传感器数据, 取得了良好的效果. 本节将按照算法中使用的传感器和数据作为分类的依据, 对基于事件相机的定位与建图算法做详细的介绍. 最后, 我们将本节中介绍的算法概要汇总整理至表 1.

### 3.1 基于纯事件流

Weikersdorfer 等在 2012 年提出了一种基于事件流与粒子滤波的定位算法<sup>[44]</sup>. 常规的基于粒子滤波的定位算法迭代主要分为两个步骤, 即用运动模型预测位姿和用观测模型校正位姿.<sup>[44]</sup> 预测位姿所使用的运动模型为随机扩散模型, 而校正位姿使用的观测模型则是作者提出的指数递减模型. 传统相机所拍摄的图像帧之间是独立的, 而 DVS 所输出的事件流中, 每一个事件并不是独立的, 作者正是针对该特性提出了指数递减的事件观测模型. 此外, 算法每一次位姿预测都是基于单个事件的, 这种方式利用了 DVS 相机低延迟的特性, 使定位算法的延迟更低、响应更快. 在<sup>[44]</sup> 的实验中, 机器人在地面上进行二维运动, 并且相机固定朝向天花板, 天花板上有预先布置好的纹理, 机器人到天花板的距离需要预先设定, 同时需要为算法提供天花板纹理的地图.

2013 年, Weikersdorfer 等对<sup>[44]</sup> 中的文章进行拓展<sup>[45]</sup>, 在粒子滤波的框架外维护了一个动态地图. 在粒子滤波的每一次迭代中, 先用上一次迭代产生的地图进行位姿的更新, 再用更新后的位姿来更新地图. 这使得算法从一个单纯的定位算法拓展为同时定位与建图算法, 不再需要提供先验的地图, 移动机器人可以进行定位与建图. 在<sup>[45]</sup> 的基础上, 作者结合碰撞检测器搭建了机器人空间自主探索系统<sup>[46]</sup>. 系统利用<sup>[45]</sup> 中的算法进行定位和建图, 并利用碰撞开关来探测墙壁和障碍物, 再利用启发式的算法进行路径规划.<sup>[44-46]</sup> 中, 机器人都只在室内的地面上做 3 自由度的运动, 并且要求天花板上有特殊

表 1 文中叙述的部分基于事件相机的 SLAM 算法及应用

Table 1 Event-based SLAM algorithms and applications

相关文献	所使用传感器	维度	算法类型	是否需要输入地图	发表时间(年)
[44]	DVS	2D	定位	是	2012
[45]	DVS	2D	定位与建图	否	2013
[47]	DVS	3D	定位	是	2014
[48]	DVS	3D	定位与建图	否	2016
[49]	DVS	3D	定位与建图	否	2016
[51]	DVS	3D	定位	是	2019
[52]	DVS, 灰度相机	3D	定位	否	2014
[53]	DVS, RGB-D相机	3D	定位与建图	否	2014
[55]	DAVIS	3D	定位	否	2016
[56]	DAVIS(内置IMU)	3D	定位	否	2017
[59]	DAVIS(内置IMU)	3D	定位与建图	否	2017
[64]	DAVIS(内置IMU), RGB相机	3D	定位与建图	否	2018
[65]	DAVIS(内置IMU)	3D	定位	否	2018

的纹理作为算法的前置条件, 这些缺陷限制了算法在实际环境中的应用。

Mueggler 等在 2014 年提出了一种基于事件流的 6 自由度定位算法<sup>[47]</sup>, 该算法建立于一个假设: DVS 所检测到的事件是由地图中灰度变化强烈的边缘产生的。因此算法以预先建好的三维边缘地图为输入, 将地图中的边和它产生的事件联系起来。在初始化阶段, 算法累积一定的事件点形成初始事件图像帧, 并在该图像上用 Hough 变换进行直线检测, 根据检测到的线段将事件点与地图中的边缘形成关联。

在算法的位姿追踪阶段, 算法根据产生的事件持续更新事件与地图边缘的关联, 将地图中的边重投影到像素平面坐标系, 并最小化投影后的边与相关联的事件的平方和误差, 从而计算出相机的位姿  $P$ , 下文的公式 (3) 用数学语言概括了这个过程。其中,  $l$  代表事件相机像素平面某条线段, 它是地图中的线段  $L_l$  在像素平面的投影,  $e_{l,i}$  表示与线段  $l$  关联的第  $i$  个事件。

$$P^* = \arg \min_P \sum_{l=1}^M \sum_{i=1}^N \|d(\pi(L_l, P), e_{l,i})\|^2 \quad (3)$$

算法实验是在空中机器人 (无人机) 上进行的: 实验中机器人在空中进行快速的翻转, 而算法能很好地跟踪机器人的位姿。然而, 文章实验中用到的地图较为简单: 在白墙上贴了一个黑色的正方形, 而且机器人一直面朝这个正方形。这种形状简单, 边缘明显的场景为算法的运行提供了便利, 作者并没有验证算法在复杂场景下的鲁棒性。

Hanme Kim 等提出了一种基于事件流和拓展卡尔曼滤波的定位与建图算法<sup>[48]</sup>。文章分别使用了三个不同的拓展卡尔曼滤波器 (EKF) 来估计相机的六自由度位姿、场景的灰度信息以及场景的深度信息。三个滤波器交错运行, 每一个滤波器输出的结果都会被用在其它两个滤波器中。算法仅使用 DVS 事件流, 对相机位姿的估计, 场景灰度的估计以及深度的估计都有着良好的效果。但在文章的实验中, 事件相机都是在一个很小的范围内移动的 (在几十厘米内来回摆动), 而且相机的朝向也没有很大的变化, 对场景深度和灰度的估计不仅需要对场景同一区域进行多次重复观测, 还需要较长时间才能收敛。这些限制可能导致算法无法在相机进行长距离运动的情况下运行。

2017 年, Rebecq 等将事件流累积成事件图像帧, 提出了基于事件图像帧的定位与建图算法 EVO<sup>[49]</sup>。在算法的建图部分, Rebecq 等使用了他们

在 2016 年提出的基于事件的空间扫描法<sup>[50]</sup>。基于事件的空间扫描法将传统图像三维重建中常用的空间扫描法拓展到事件相机中, 算法可以总结为三个步骤:

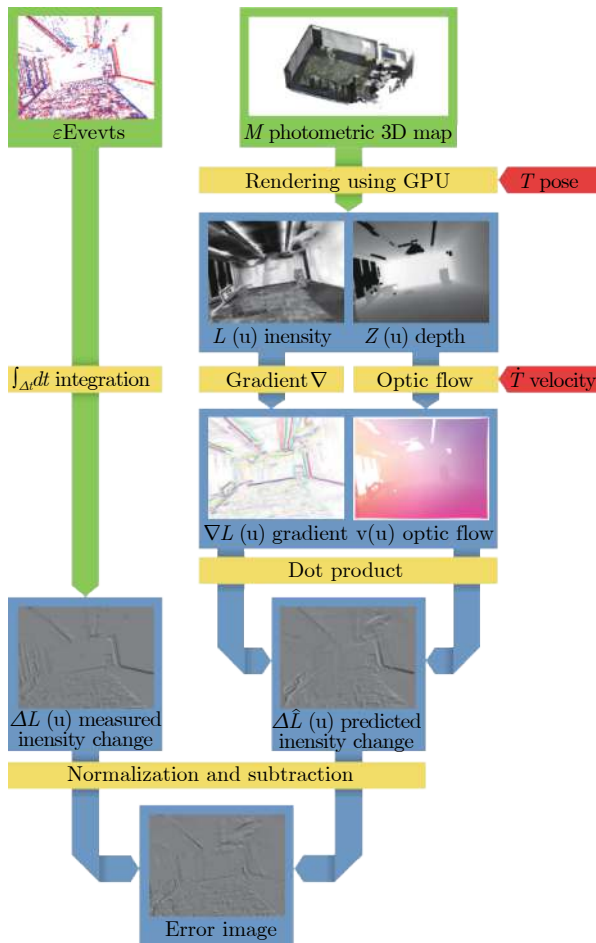
1. 根据相机的运动轨迹和记录的对应时刻产生的事件, 利用相机的内参, 将对应的事件投影到空间中, 形成一条经过相机光心和成像平面上该事件位置的射线。
2. 选取一个参考视角, 基于参考视角建立视差空间图 (DSI), 统计 DSI 中的每个体素被所有射线经过的次数。
3. 根据 DSI 的统计结果来确定某个体素中是否存在物体, 从而完成三维重建。

算法<sup>[49]</sup>中定位的部分, 则是将一定个数的事件累积成事件图像帧, 并用事件图像帧与当前已经建出的三维半稠密地图作匹配, 从而计算出相机的当前位姿。文章中的实验表明, 在室内外环境、高动态范围环境以及相机快速运动的情况下, 算法都具有良好的表现。作者用开关灯模拟亮度剧烈变化, EVO 也能良好地进行位姿追踪。唯一的问题在于, 将事件累积成事件图像帧的过程, 会在一定程度上增加算法的延迟, 这就浪费了 DVS 低延迟的特性。

2019 年, Bryner 等人提出了一种新的基于非线性优化的定位算法<sup>[51]</sup>。与传统的优化特征点重投影误差的思路不同的是, Bryner 等设计了一种基于光强变化图像 (Intensity-Change Image) 的误差函数。如图 4 所示, 算法根据输入的三维光度地图和当前估计的相机位姿, 用 GPU 计算出当前相机位姿下的灰度图和深度图, 然后对灰度图提取边缘, 同时根据估计相机速度求出对深度图求出光流图, 再结合边缘图和光流图构造估计的光强变化图  $\Delta \hat{L}(\mathbf{u})$ 。算法根据事件生成模型, 对事件流进行积分获得测量的光强变化图  $\Delta L(\mathbf{u})$ 。对  $\Delta \hat{L}(\mathbf{u})$  和  $\Delta L(\mathbf{u})$  归一化后求差值, 便得到了光强变化误差。最后, 算法根据光强变化误差来优化估计的相机位姿以及相机速度, 直到优化收敛。算法在实验中的精度表现十分优异: 在合成数据下算法的角度误差低于  $0.52^\circ$ , 在真实数据下算法的角度误差低于  $3.84^\circ$ 。然而由于优化的过程复杂, 算法无法做到实时运行, 甚至比实时慢了 2-3 个数量级。

### 3.2 基于事件流与传统图像

Censi 等结合使用 DVS 和普通相机, 提出了一种基于滤波的视觉里程计<sup>[52]</sup>。文章提出了一种结合了普通相机输出的灰度图和 DVS 输出的事件流的观测模型。其中心思想是, DVS 输出的事件流是由场景中灰度梯度较大的区域产生的, 而这些区域在

图4 Bryner 算法工作流程<sup>[51]</sup>Fig. 4 The workflow of Bryner's algorithm<sup>[51]</sup>

灰度图中能直接获取. 因此, 算法利用事件流与作者提出的观测模型相结合, 追踪两个普通图像帧之间的相机位姿, 使定位算法的延迟比普通的基于图像的视觉里程计更低. 经过作者的实验, 该算法在姿态估计上有较好的表现, 但对位移的估计表现不佳.

Weikersdorfer 等基于他们在<sup>[45]</sup>中的工作, 结合使用 DVS 和 RGB-D 相机, 提出了一种 6 自由度的定位与建图算法<sup>[53]</sup>. 与前面的工作一致, 该算法的核心也是粒子滤波算法, 不过粒子中位姿由原先的二维位姿拓展为三维位姿. 对 RGB-D 相机和 DVS 相机标定内外参后, 利用 RGB-D 相机提供的深度图构建深度映射, 为 DVS 输出的事件提供深度, 从而将滤波器中粒子的状态从二维空间提升到三维空间. 类似<sup>[45]</sup>, 算法也维护了一个动态地图, 地图由三维体素网格构成, 采用和<sup>[45]</sup>中类似的更新方式. 在实验中, 该算法的最低误差达到 3.1 cm(RMSE); 在不使用 GPU 加速地情况下, 算法运行速度也能达到实时的 20 倍. 算法的不足之处在于, 相比起低延迟、高响应速度的 DVS, RGB-D 提供的深度图

是有延迟的; 这就意味着在快速运动的情况下, 根据深度映射计算出的事件的深度是不够准确的.

Tedaldi 等在 2016 年提出了一种结合图像和事件流的特征点检测与追踪算法<sup>[54]</sup>. 该算法使用 DAVIS 相机作为传感器, 在灰度图上进行特征点检测, 并利用事件流来对特征点进行追踪. 在特征点提取阶段, 算法对相机输出的原始灰度图使用 Canny 算子进行边缘提取, 形成边缘图; 同时, 对原始灰度图提取 Harris 角点, 作为特征点; 然后对以特征点为中心的方形区块内的边缘图进行二值化, 形成模型点集 (model point set), 作为特征点的描述子以供后续的特征点追踪. 在特征点追踪阶段, 算法将特征点的描述子区块内发生的事件累积形成数据点集 (data point set), 然后将数据点集和描述子的模型点集使用 ICP 算法进行几何上的配准, 解算出特征点移动.

Kueng 等的工作<sup>[55]</sup>中使用了<sup>[54]</sup>中的算法进行特征点提取和追踪, 然后和常规的基于优化的视觉里程计类似, 用深度滤波器来估计特征点的深度, 并且用最小化重投影误差的方式来解算相机的位姿. 文中的特征点追踪算法有着很高的时空分辨率, 使得定位延迟低. 不过限于 DAVIS 中 APS 部分的自身缺陷, 在特征点提取和边缘提取的过程中还是会受到运动模糊的影响, 并且 APS 部分所能承受的动态范围仍然有限, 这限制了算法的适用场景.

### 3.3 基于事件与 IMU

Alex 等在 2017 年提出基于特征和卡尔曼滤波的 VIO 方法<sup>[56]</sup>, 在传统的 VIO 算法上进行了改进, 从事件流中提取特征点在图像中的移动轨迹并使用滤波方法将该轨迹与 IMU 的测量数据融合, 完成相机的位姿估计. 在<sup>[56]</sup>中, 系统将单个特征点和多个与其空间相邻的事件关联起来并完成追踪<sup>[57]</sup>. 得到特征点在图像中的移动轨迹后, 使用多状态约束的卡尔曼滤波器<sup>[58]</sup>将特征点轨迹与 IMU 信息相结合, 最终得到估计的相机位姿模型, 并不断通过新获取的 IMU 数据以及高频的事件流对位姿模型进行更新和修正. 文章中的实验表明, 使用事件相机来进行特征追踪较传统视觉相机而言拥有更小的漂移误差, 在长距离追踪中拥有更好的效果. 但由于实验采用了迭代的 EM 算法来进行特征的追踪, 开销较大, 因此算法无法做到在常规算力平台 (如笔记本电脑) 及载具中实时运行.

使用滤波方法的 SLAM 系统的定位和建图的精度会因为线性化过程导致的误差的累积而迅速降低. 针对这个现象, Rebecq 等在 2017 年提出基于



事件帧和非线性优化的定位与建图算法<sup>[59]</sup>. 这个算法的主要思想是将连续的事件流通过给定长度的时间窗口划分成事件帧, 基于事件帧进行非线性优化完成位姿追踪. 首先, 将从事件相机得到的连续事件流拆分成一组重叠的时间窗口, 将同窗口内的事件流累计成一个事件帧, 并通过 IMU 采集到的数据对得到的事件帧进行运动补偿<sup>[60]</sup>. 对于修正完成的事件帧, 使用经典的 FAST 角点检测方法<sup>[61]</sup> 和 Lucas-Kanade 追踪方法<sup>[62]</sup> 提取和追踪其中的特征点, 得到用于相机位姿估计的特征点移动轨迹. 最后, 使用非线性优化方法<sup>[63]</sup> 将由特征点的移动轨迹与来自 IMU 的数据结合, 将 IMU 误差项以完全概率的方式与特征重投影误差紧耦合, 优化联合了重投影误差和 IMU 误差的非线性代价函数, 周期性地改进相机的轨迹和 3D 界标的位置. 文章中通过实验验证了其方法可在普通笔记本上完成实时的 6 自由度位姿追踪以及 3D 地标的稀疏图还原.

在 2018 年, Rebecq 等继续对他们在<sup>[59]</sup> 中的工作进行拓展, 首次提出了将事件流、标准图像帧以及 IMU 测量数据三者紧耦合融合的状态估计方法<sup>[64]</sup>. 方法同时对来自事件相机的事件帧和来自标准相机的图像帧进行了特征点的追踪, 并将二者的追踪结果都作为非线性优化方法的输入并完成相机位姿的估计. 在<sup>[64]</sup> 中, 作者在四旋翼无人机上搭载了事件相机以及标准相机, 在光源高频切换、高速旋转的几种极端情况下对提出的特征点追踪方法进行了测试, 并实验证明了论文提出的方法在多种极端情况下 (高动态光照场景和高速运动) 均具有优秀的性能, 并且能够在算力有限的平台上完成实时计算. 同时, 该方法也拥有更高的准确度, 与仅基于事件帧以及 IMU 的方法<sup>[59]</sup> 相比提升了 130 %, 与仅基于标准图像帧以及 IMU 的方法相比提升了 85 %.

当前主流的 VIO 算法大多是基于特征的, 即从事件中提取特征点的移动轨迹并与 IMU 的数据融合完成位姿的追踪. Mueggler 等则在 2018 年提出了一个不依赖特征点的新方法: 使用连续的时间框架描述基于事件相机的 VIO 过程<sup>[65]</sup>. 与基于特征的传统 VIO 算法不同, <sup>[65]</sup> 中应用了连续时间框架<sup>[66-67]</sup>, 直接应用事件流中的时间和位置属性, 将事件和相机位姿建立对应关系, 使用平滑参数模型来描述相机的位姿变化情况. 最后, 使用三次样条插值、视觉-惯性束调整等优化方法来完成模型的优化, 减少存储开销, 提高计算效率. 文章中的实验显示在普通算力平台下方法无法实现实时运行, 需要通过使用高算力平台 (如多 GPU) 来达到实时运行的结果.

## 4 事件相机数据集

在硬件方面, 事件相机目前在市场上并不普及, 种类少且价格昂贵. 因此, 关于事件相机的开源数据集是算法开发和测试的重要工具. 好的数据集不仅可以帮助研究者避免硬件系统搭建的麻烦, 还可以帮助研究者对算法进行客观准确的评估. 本节总结了用于定位与建图的事件相机数据集, 并对其进行简要介绍, 最后将数据集的概要归纳至节末的表 2 中.

<sup>[53]</sup> 的工作中, 作者开放了实验用的数据集. 数据集中包含了 eDVS 相机和 RGB-D 相机输出的事件流、彩色图像以及深度图像. 相机在室内环境做 6 自由度的运动, 相机位姿的真值则是通过运动采集系统获取的.

<sup>[68]</sup> 中使用内置 IMU 的 DAVIS 相机, 采集了在室内环境下相机做纯旋转运动的数据, 相机姿态的真值是通过陀螺仪的输出进行角度积分获取的.

数据集<sup>[69]</sup> 中, 作者使用一个地面机器人和一个

表 2 DVS 公开数据集  
Table 2 Dataset provided by event cammera

相关文献	所使用传感器	相机运动自由度	数据采集场景	载具	是否提供真值	发表时间 (年)
[53]	eDVS相机, RGB-D相机	6DOF	室内	手持	是	2014
[68]	DAVIS(内置IMU)	3DOF(纯旋转)	室内, 仿真	旋转基座	是	2016
[69]	DAVIS, RGB-D相机	4DOF	室内, 仿真	地面机器人和云台	是	2016
[70]	DAVIS(内置IMU)	6DOF	室内 室外 仿真	手持	室内: 是 室外: 否 仿真: 是	2016
[71]	DAVIS	6DOF	室外	汽车	是	2017
[72]	2*DAVIS(内置IMU) 2*RGB相机(内置IMU) 16线激光雷达	6DOF	室内 室外 室内 到室外	四轴飞行器 摩托车 汽车 手持	是	2018
[73]	2*DAVIS(内置IMU) RGB-D相机	3DOF	室内	3*地面机器人	是	2018
[74]	DAVIS	6DOF	室内	手持	是	2019
[51]	DAVIS, IMU	6DOF	室内, 仿真	手持	是	2019

云台搭载了 DAVIS 和 RGB-D 相机, 提供了机器人在室内运动采集的数据以及作者合成的数据. 由于云台只能进行偏航角和俯仰角方向的旋转, 相机的运动只有 4 自由度. 相机位姿的真值是结合机器人的里程计以及云台的角度计算出来的.

数据集<sup>[70]</sup> 提供了多场景的 DAVIS 及其内置 IMU 的数据, 数据集包括室内数据、室外数据以及合成数据. 室内数据的真值来自运动采集系统, 而室外数据部分作者并未提供真值.

数据集<sup>[71]</sup> 提供了长距离室外道路场景的 DAVIS 数据, 由于该数据集原本目的是自动驾驶的车辆控制, 数据集提供了多种车辆控制信息, 但并未提供准确的 6 自由度位姿真值, 相机位置只能通过车载 GPS 提供的经纬度信息来获取.

数据集<sup>[72]</sup> 中, 作者搭建了多传感器的数据采集系统, 该系统包含了一对内置 IMU 的 DAVIS 相机、一对内置 IMU 的 RGB 相机、一个 16 线的激光雷达以及一个 GPS 传感器. 作者采集了多种场景的数据, 包括室内、室外、室内外切换、白天和黑夜. 除此之外, 数据采集系统搭载于不同的载具: 四轴飞行器、汽车、摩托车以及手持. 作者利用室内运动采集系统和室外运动采集系统等方法提供了真值. 该数据集不仅满足了第 3 节中所包含的类型的算法研究需求, 还满足了双目事件视觉算法研究的需求. 同时, 多场景和多载具的数据也大大方便了对算法适应性和鲁棒性的验证.

数据集<sup>[73]</sup> 使用了三个地面机器人, 每个地面机器人搭载了一对内置 IMU 的 DAVIS 相机, 以及一个 RGB-D 相机. 三个地面机器人在室内地面上同时运动, 并由运动采集系统采集机器人的位姿.

文章<sup>[51][74]</sup> 的作者也将他们实验中所使用的 DAVIS 数据开放出来.

上述数据集都提供了相机运动的真值, 可以用于对基于事件相机的定位与建图算法进行评估. 值得注意的是, 数据集<sup>[68]</sup> 中相机只做三维空间中的旋转运动; 数据集<sup>[69]</sup> 中的相机运动缺少竖直方向的位移和滚转旋转; 数据集<sup>[73]</sup> 中相机只有水平位移和偏航旋转. 这三个数据集相机运动均未达到 6 自由度, 不能够全面地评估基于事件相机定位与建图算法的性能. 数据集<sup>[71]</sup> 虽然相机运动达到了 6 自由度, 但数据集提供的真值只包含其中两个自由度的信息. 研究者在选用以上四个数据集时, 须注意与其它数据集搭配使用. 此外, 数据集<sup>[73]</sup> 为同一场景下, 多机器人同时采集的数据, 可以用于基于事件相机的多智体协同定位与建图的研究. 除了用于对基于事件相机的定位与建图算法的研究, 上述数据集中部分

也可以用于对基于事件相机的深度估计、基于事件相机的运动分割等方面的研究. 例如, 数据集<sup>[72][74]</sup> 给出了深度图的真值, 可以用于研究基于事件相机的深度估计算法. 数据集<sup>[69]</sup> 给出了二维图像运动场真值, 数据集<sup>[74]</sup> 给出了精确的场景内物体运动轨迹, 可以用于研究基于事件相机的运动分割算法.

## 5 总结与展望

本文以基于事件相机的定位与建图算法为中心, 介绍了事件相机的种类及其工作原理, 介绍了一些具有代表性的、基于事件相机的定位与建图算法, 以及与事件相机定位与建图相关的数据集. 尽管已经有许多基于事件相机的定位与建图算法被提出来, 但它们仍存在问题. 比如说, <sup>[55]</sup> 中的算法仍然受到了 APS 相机的限制, 无法应对快速运动; <sup>[49]</sup> 中的算法要把事件流累积成事件图像帧, 浪费了事件相机低延迟的特性; <sup>[48]</sup> 中的算法被局限于小范围的相机运动; <sup>[51]</sup> 中的算法需要很长时间才能收敛, 并且需要预先建立精确的三维光度地图.

总的来说, 基于纯事件流的定位建图算法存在的问题主要在于, 微小时间内事件相机获取的信息较少, 这是由事件相机的分辨率、信噪比以及带宽所决定的. 这导致基于纯事件流的定位建图算法或要求场景或相机运动简单, 或需要通过累积事件流形成事件图像帧. 前者使得算法的实用性下降, 而后者则牺牲了事件相机低延迟的优良特性. 基于事件流与传统图像的定位建图算法存在的问题在于, 对传统 APS 相机的依赖导致算法仍会受到相机延迟、运动模糊、高动态范围环境的影响. 而基于事件相机和 IMU 的算法是目前表现最佳的一类算法, 既保持了 VIO 低延迟、高频输出的优势, 又结合了事件相机可以应对低光照和高动态范围环境的优良特性. 然而 IMU 的使用会为算法带来额外的成本.

从本文介绍的算法可以看出, 事件相机可以应用于实现低延迟的定位与建图算法, 可以帮助克服视觉传感器受运动模糊和高动态范围环境的影响, 可以提升视觉定位与建图算法的精度. 但是对于如何利用事件相机进行回环检测, 降低累积误差, 并没有很好的研究工作. 在未来的研究中, 有望对事件相机做进一步的挖掘, 实现更加完善鲁棒的定位与建图算法.

## References

- 1 Burri M, Oleynikova H, Achtelik M W, Siegwart R. Realtime visual-inertial mapping, re-localization and planning onboard MAVs in unknown environments. In: *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hamburg, Germany: IEEE, 2015. 1872-1878



- 2 Chatila R, Laumond J P. Position referencing and consistent world modeling for mobile robots. **In: Proceedings of the 1985 IEEE International Conference on Robotics and Automation.** Louis, Missouri, USA: IEEE, 1985. Vol. 2: 138–145.
- 3 Chatzopoulos D, Bermejo C, Huang Z, P Hui. Mobile augmented reality survey: From where we are to where we go. *Ieee Access*, 2017, **5**: 6917–6950
- 4 Taketomi T, Uchiyama H, Ikeda S. Visual SLAM algorithms: a survey from 2010 to 2016. *Transactions on Computer Vision and Applications*, 2017, **9**(1): 16
- 5 Strasdat H, Montiel J M M, Davison A J. Visual SLAM: why filter?. *Image and Vision Computing*, 2012, **30**(2): 65–77
- 6 Younes G, Asmar D, Shammas E, J Zelek. Keyframe-based monocular SLAM: design, survey, and future directions. *Robotics and Autonomous Systems*, 2017, **98**: 67–88
- 7 Olson C F, Matthies L H, Schoppers M, Maimore M W. Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, 2003, **43**(4): 215–229
- 8 Zhang Z. Microsoft kinect sensor and its effect. *IEEE multimedia*, 2012, **19**(2): 4–10
- 9 Huang A S, Bachrach A, Henry P, et al. *Visual odometry and mapping for autonomous flight using an RGB-D camera*. Robotics Research. Springer, Cham, 2017: 235–252
- 10 Jones E S, Soatto S. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 2011, **30**(4): 407–430
- 11 Martinelli A. Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Transactions on Robotics*, 2011, **28**(1): 44–60
- 12 Klein G, Murray D. Parallel tracking and mapping for small AR workspaces **In: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality.** Nara, Japan: IEEE, 2007. 1–10
- 13 Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 2015, **31**(5): 1147–1163
- 14 Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 2017, **33**(5): 1255–1262
- 15 Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry. **In: Proceedings of the 2014 IEEE international conference on robotics and automation (ICRA).** Hong Kong, China: IEEE, 2014. 15–22
- 16 Engel J, Schops T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM. **In: Proceedings of the 2014 European conference on computer vision.** Zurich, Switzerland: Springer, 2014. 834–849
- 17 Engel J, Stöckler J, Cremers D. Large-scale direct SLAM with stereo cameras. **In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).** Hamburg, Germany: IEEE, 2015. 1935–1942
- 18 Li M, Mourikis A I. High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research*, 2013, **32**(6): 690–711
- 19 Leutenegger S, Lynen S, Bosse M, Siegwart R, Furgale P. Keyframe-based visual inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 2015, **34**(3): 314–334
- 20 Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 2018, **34**(4): 1004–1020
- 21 Fossum E R. CMOS image sensors: Electronic camera-on-a-chip. *IEEE transactions on electron devices*, 1997, **44**(10): 1689–1698
- 22 Delbruck T. Neuromorphic vision sensing and processing. **In: Proceedings of the 2016 46th European SolidState Device Research Conference (ESSDERC).** Lansanne, Switzerland: IEEE, 2016. 7–14
- 23 Delbruck T, Lichtsteiner P. Fast sensory motor control based on event-based hybrid neuromorphic-procedural system. **In: Proceedings of the IEEE international symposium on circuits and systems.** New Orleans, USA: IEEE, 2007. 845–848
- 24 Delbruck T, Lang M. Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor. *Frontiers in neuroscience*, 2013, **7**: 223
- 25 Glover A, Bartolozzi C. Event-driven ball detection and gaze fixation in clutter. **In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).** Daejeon, Korea: IEEE, 2016. 2203–2208
- 26 Benosman R, Ieng S H, Clercq C, Bartolozzi C, Srinivasan M. Asynchronous frameless event-based optical flow. *Neural Networks*, 2012, **27**: 32–37
- 27 Benosman R, Clercq C, Lagorce X, Ieng S H, Bartolozzi C. Event-based visual flow. *IEEE transactions on neural networks and learning systems*, 2013, **25**(2): 407–417
- 28 Rueckauer B, Delbruck T. Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor. *Frontiers in neuroscience*, 2016, **10**: 176
- 29 Bardow P, Davison A J, Leutenegger S. Simultaneous optical flow and intensity estimation from an event camera. **In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.** LAS VEGAS, USA: IEEE, 2016. 884–892
- 30 Reinbacher C, Graber G, Pock T. Real-time intensity image reconstruction for event cameras using manifold regularisation. *International Journal of Computer Vision*, 2018, **126**(12): 1381–1393
- 31 Mahowald M. VLSI analogs of neuronal visual processing: a synthesis of form and function. California Institute of Technology, 1992.
- 32 Posch C, Serrano-Gotarredona T, Linares-Barranco B, Delbruck T. Retinomorph event-based vision sensors: bioinspired cameras with spiking output. *Proceedings of the IEEE*, 2014, **102**(10): 1470–1484
- 33 Lichtsteiner P, Posch C, Delbruck T. A 128 X 128 120 db 30 mw asynchronous vision sensor that responds to relative intensity change. **In: Proceedings of the 2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers.** San Francisco, CA, USA: IEEE, 2006. 2060–2069
- 34 Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 2008, **43**(2): 566–576
- 35 Son B, Suh Y, Kim S, et al. 4. 1 A 640×480 dynamic vision sensor with a 9 μm pixel and 300 Meps address-event representation. **In: Proceedings of the 2017 IEEE International Solid-State Circuits Conference (ISSCC).** San Francisco, CA, USA: IEEE, 2017. 66–67
- 36 Posch C, Matolin D, Wohlgenannt R. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE Journal*

- al of Solid-State Circuits*, 2010, **46**(1): 259–275
- 37 Posch C, Matolin D, Wohlgenannt R. A QVGA 143 dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression. **In: Proceedings of the 2010 IEEE International Solid-State Circuits Conference-(ISSCC)**. San Francisco, CA, USA: IEEE, 2010. 400–401
  - 38 Berner R, Brandli C, Yang M, Liu S C, Delbruck T. A 240x180 120 db 10 mw 12 us-latency sparse output vision sensor for mobile applications. **In: Proceedings of the International Image Sensors Workshop**. Snowbird, Utah, USA: IEEE, 2013. 41–44
  - 39 Brandli C, Berner R, Yang M, Liu S C, Delbruck T. A 240×180 130 db 3  $\mu$ s latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 2014, **49**(10): 2333–2341
  - 40 Guo M, Huang J, Chen S. Live demonstration: A 768×640 pixels 200 Meps dynamic vision sensor. **In: Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS)**. Baltimore, Maryland, USA: IEEE, 2017. 1–1
  - 41 Li C, Brandli C, Berner R, et al. Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor. **In: Proceedings of the 2015 IEEE International Symposium on Circuits and Systems (ISCAS)**. Liston, Portugal: IEEE, 2015. 718–721
  - 42 Moeys D P, Li C, Martel J N P, et al. Color temporal contrast sensitivity in dynamic vision sensors. **In: Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS)**. Baltimore, Maryland, USA: IEEE, 2017. 1–4
  - 43 Marcireau A, Ieng S H, Simon-Chane C, Benosman R B. Event-based color segmentation with a high dynamic range sensor. *Frontiers in neuroscience*, 2018, **12**: 135
  - 44 Weikersdorfer D, Conradt J. Event-based particle filtering for robot self-localization. **In: Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)**. Guangzhou, China: IEEE, 2012. 866–870
  - 45 Weikersdorfer D, Hoffmann R, Conradt J. Simultaneous localization and mapping for event-based vision systems. **In: Proceedings of the 2013 International Conference on Computer Vision Systems**. St. Petersburg, Russia: Springer, 2013. 133–142
  - 46 Hoffmann R, Weikersdorfer D, Conradt J. Autonomous indoor exploration with an event-based visual SLAM system. **In: Proceedings of the 2013 European Conference on Mobile Robots**. Barcelona, Catalonia, Spain: IEEE, 2013. 38–43
  - 47 Mueggler E, Huber B, Scaramuzza D. Event-based, 6-DOF pose tracking for high-speed maneuvers. **In: Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems**. Chicago, USA: IEEE, 2014. 2761–2768
  - 48 Kim H, Leutenegger S, Davison A J. Real-time 3D reconstruction and 6-DoF tracking with an event camera. **In: Proceedings of the 2016 European Conference on Computer Vision**. Amsterdam, The Netherlands: Springer, 2016. 349–364
  - 49 Rebecq H, Horstschaefer T, Gallego G, Scaramuzza D. EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2016, **2**(2): 593–600
  - 50 Rebecq H, Gallego G, Scaramuzza D. EMVS: Event-based multi-view stereo. **In: Proceedings of the 2016 British machine vision conference (BMVC)**. York, UK: Springer, 2016(CONF).
  - 51 Bryner S, Gallego G, Rebecq H, Scaramuzza D. Eventbased, direct camera tracking from a photometric 3D map using nonlinear optimization. **In: the 2019 International Conference on Robotics and Automation (ICRA)**. Montreal, Canada: IEEE, 2019. 2
  - 52 Censi A, Scaramuzza D. Low-latency event-based visual odometry. **In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)**. Hong Kong, China: IEEE, 2014. 703–710
  - 53 Weikersdorfer D, Adrian D B, Cremers D, Conradt J. Event-based 3D SLAM with a depth-augmented dynamic vision sensor. **In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)**. Hong Kong, China: IEEE, 2014. 359–364
  - 54 Tedaldi D, Gallego G, Mueggler E, Scaramuzza D. Feature detection and tracking with the dynamic and active-pixel vision sensor (DAVIS). **In: Proceedings of the 2016 Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)**. Krakow, Poland: IEEE, 2016. 1–7
  - 55 Kueng B, Mueggler E, Gallego G, Scaramuzza D. Lowlatency visual odometry using event-based feature tracks. **In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**. Daejeon, Korea: IEEE, 2016. 16–23
  - 56 Zhu A Z, Atanasov N, Daniilidis K. Event-based visual inertial odometry. **In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, Hawaii, USA: IEEE, 2017. 5816–5824
  - 57 Zhu A Z, Atanasov N, Daniilidis K. Event-based feature tracking with probabilistic data association. **In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)**. Marina Bay, Singapore: IEEE, 2017. 4465–4470
  - 58 Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation. **In: Proceedings of the 2007 IEEE International Conference on Robotics and Automation (ICRA)**. Roma, Italy: IEEE, 2007. 3565–3572
  - 59 Rebecq H, Horstschaefer T, Scaramuzza D. Real-time Visual-Inertial Odometry for Event Cameras using Keyframe-based Nonlinear Optimization. **In: Proceedings of the 2017 British machine vision conference (BMVC)**. London, UK: Springer, 2017(CONF).
  - 60 Gallego G, Scaramuzza D. Accurate angular velocity estimation with an event cameras. *IEEE Robotics and Automation Letters*, 2017, **2**(2): 632–639
  - 61 Rosten E, Drummond T. Machine learning for high-speed corner detection. **In: Proceedings of the 2006 European conference on computer vision**. Graz, Austria: Springer, 2006. 430–443
  - 62 Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision. 1981. 121–130
  - 63 Leutenegger S, Furgale P, Rabaud V, et al. Keyframe-based visual-inertial slam using nonlinear optimization. **In: Proceedings of the 2013 Robotis Science and Systems (RSS)**. Berlin, German, 2013.
  - 64 Vidal A R, Rebecq H, Horstschaefer T, Scaramuzza D. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. *IEEE Robotics and Automation Letters*, 2018, **3**(2): 994–1001
  - 65 Mueggler E, Gallego G, Rebecq H, Scaramuzza D. Continuous-time visual-inertial odometry for event cameras. *IEEE Transactions on Robotics*, 2018, **34**(6): 1425–1440
  - 66 Mueggler E, Gallego G, Scaramuzza D. Continuous-time trajectory estimation for event-based vision sensors[R]. 2015
  - 67 Patron-Perez A, Lovegrove S, Sibley G. A spline-based trajectory

ory representation for sensor fusion and rolling shutter cameras. *International Journal of Computer Vision*, 2015, **113**(3): 208–219

- 68 Rueckauer B, Delbruck T. Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor. *Frontiers in neuroscience*, 2016, **10**: 176
- 69 Barranco F, Fermuller C, Aloimonos Y, Delbruck T. A dataset for visual navigation with neuromorphic methods. *Frontiers in neuroscience*, 2016, **10**: 49
- 70 Mueggler E, Rebecq H, Gallego G, Delbruck T, Scaramuzza D. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *The International Journal of Robotics Research*, 2017, **36**(2): 142–149
- 71 Binas J, Neil D, Liu S C, Delbruck T. DDD17: End-to-end DAVIS driving dataset. arXiv: 1711. 01458, 2017
- 72 Zhu A Z, Thakur D, Ozaslan T, Pfommer B, Kumar V, Daniilidis K. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception. *IEEE Robotics and Automation Letters*, 2018, **3**(3): 2032–2039
- 73 Leung S, Shamwell E J, Maxey C, Nothwang W D. Toward a large-scale multimodal event-based dataset for neuromorphic deep learning applications. **In: Proceedings of the 2018 Micro- and Nanotechnology Sensors, Systems, and Applications X.** International Society for Optics and Photonics. Orlando, Florida, USA: SPIE, 2018. 10639: 106391T
- 74 Mitrokhin A, Ye C, Fermuller C, Aloimonos Y, Delbruck T. EV-IMO: Motion segmentation dataset and learning pipeline for event cameras. arXiv: 1903. 07520, 2019



**马艳阳** 中山大学数据科学与计算机学院硕士研究生. 2014 年获得中山大学计算机科学与技术学士学位. 主要研究方向为机器人定位与建图技术.

E-mail: mayany3@mail2.sysu.edu.cn

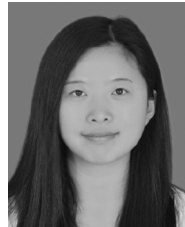
(**MA Yan-Yang** Master student at the School of Data and Computer Science, Sun Yat-sen University. He received his B.Sc. from Sun Yat-sen University in 2014. His research interest covers robot localization and mapping.)



**叶梓豪** 中山大学数据科学与计算机学院本科生. 主要研究方向为机器人定位与建图技术.

E-mail: yezh9@mail2.sysu.edu.cn

(**YE Zi-Hao** Bachelor at the School of Data and Computer Science, Sun Yatsen University. His research interest covers robot localization and mapping.)



**刘坤华** 中山大学数据科学与计算机学院博士后. 2019 年获得山东科技大学机电工程学院博士学位. 主要研究方向为自动驾驶环境感知.

E-mail: lkhzyf@163.com

(**LIU Kun-Hua** Postdoctoral at the School of Data and Computer Science, Sun Yat-sen University. She received her Ph.D. degree from the Mechanical and Electrical Engineering Institute, Shandong University of Science and Technology. Her research interest covers automatic driving environment perception.)



**陈 龙** 中山大学数据科学与计算机学院副教授. 于 2007 年、2013 年获得武汉大学学士、博士学位. 主要研究方向为自动驾驶, 机器人, 人工智能. 本文通信作者.

E-mail: chenl46@mail.sysu.edu.cn

(**CHEN Long** Associate Professor at the School of Data and Computer Science, Sun Yat-sen University. He received his B.Sc. degree and Ph.D. degree from Wuhan University in 2007 and 2013. His research interest covers autonomous driving, robotics and artificial intelligence. Corresponding author of this paper.)