

异步数据流中关键目标提取及分类研究*

张 姝, 杜从洋, 吴金建, 石光明, 谢雪梅



(西安电子科技大学 人工智能学院, 陕西 西安 710071)

通讯作者: 吴金建, E-mail: jinjian.wu@mail.xidian.edu.cn

摘 要: 事件相机因其生物视觉的启发渊源, 打破了计算机视觉领域的常规数据获取方式, 直击计算机视觉领域中 RGB 图像的痛点, 带来了二维图像传感器无法比拟的优势, 引起了广大研究者的密切关注. 事件相机带来去除冗余信息、快速感知能力、高动态范围的感光能力和低功耗特性等优势的同时, 其异步的事件数据无法直接应用于现有的计算机视觉处理模式. 因此, 利用基于关键事件点的分类方法对事件相机的数据流进行分类. 该方法检测带有重要信息的角点事件, 并只对角点事件进行特征提取. 在保留事件重要特征和凝练提取事件流特征的同时, 有效地减少了对其他事件的运算量. 对预设手势进行识别, 以此验证该方法的有效性, 实现了 97.86% 的准确率.

关键词: 事件相机; 关键目标; 异步数据流; 目标分类; 角点事件

中文引用格式: 张姝, 杜从洋, 吴金建, 石光明, 谢雪梅. 异步数据流中关键目标提取及分类研究. 软件学报, 2019, 30(Suppl. (2)): 9-16. <http://www.jos.org.cn/1000-9825/19013.htm>

英文引用格式: Zhang S, Du CY, Wu JJ, Shi GM, Xie XM. Research on key object extraction and classification in asynchronous data stream. Ruan Jian Xue Bao/Journal of Software, 2019, 30(Suppl. (2)): 9-16 (in Chinese). <http://www.jos.org.cn/1000-9825/19013.htm>

Research on Key Object Extraction and Classification in Asynchronous Data Stream

ZHANG Shu, DU Cong-Yang, WU Jin-Jian, SHI Guang-Ming, XIE Xue-Mei

(School of Artificial Intelligence, Xidian University, Xi'an 710071, China)

Abstract: Event camera has attracted the attention of the majority of researchers due to the inspiration of biological vision, breaks the way of regular data acquisition in the field of computer vision, directly hits the pain point of RGB images, and brings the advantages that 2D image sensors cannot match. Event Camera brings the advantages of removing redundant information, fast sensing capability, high dynamic range sensitivity and low power consumption, while its asynchronous event data cannot be directly applied to existing computer vision processing modes. Therefore, this paper classify the data stream using the key event based classification method. This method detects corner events with important information and only extracts features of corner events. While retaining the important features of event and condensing the extraction of event stream features, the amount of computation for other events is effectively reduced. The preset gesture is recognized to verify the validity of this method, achieving an accuracy of 97.86%.

Key words: event camera; key object; asynchronous data stream; object classification; corner event

近年来, 计算机视觉领域中涌现了大量优异的研究方法使用相机和电脑代替人眼对目标进行检测、识别和跟踪. 尽管研究方法的不断更迭和计算力的不断增强使得精确程度和处理速度有了很大的改善, 但是目前的视觉数据获取方式仍有一定的局限性. 现有的视频序列是普遍使用帧率为 30fps 的传统相机进行拍摄得到的, 这类相机的成像模式抓拍快速运动得到的图像序列会出现模糊化现象, 并且连续图像帧中包含大量的冗余未变化信息. 这些局限性的存在对计算机的运算速度和存储容量都有较高要求, 需要耗费大量的计算资源和计算

* 基金项目: 高分辨率对地观测系统重大专项青年创新基金; 装备预研教育部联合基金(6141A020336)

Foundation item: Young Fund for High Resolution Earth Observation Conference; Joint Foundation of Ministry of Education for Equipment Advanced Research (6141A020336)

收稿时间: 2019-08-17; 采用时间: 2019-11-04

时间.

事件相机(event camera)^[1-4]是一种新型的仿生视觉传感器,采用仿视网膜的动态像素点探测技术来感知并编码视觉信息.事件相机的每个像素点在光照变化刺激下独立产生异步事件.通过这种方式产生的事件流消除了传统相机以固定帧率输出的重复的静态冗余数据,同时其具有微秒级的时间分辨率,不受曝光时间和帧率的限制,可以捕获快速运动的连续异步数据,确保了运动的连续性.此外,事件相机还具有高动态范围特性,在光线过亮或过暗情况下也能很好工作.因此,事件相机具有极大的研究和应用价值.

事件相机带来其特有优势的同时,也带来了数据处理方面的挑战.为了充分利用事件相机的优势和对事件数据异步特性更合理的利用,对每个事件都应采用事件驱动(event-driven)的处理方式.因此,本文从上述问题出发,提出了一种基于事件信号的识别方法.由于事件流数据是一种三维的点云数据,与传统的图像帧有着本质的区别,基于此,本文将传统的基于图像帧的 Harris 角点检测和 SURF 特征描述应用在了事件流数据上,实现了事件驱动识别方法.并在一个特定动态手势上验证了该方法,得到了 97.86%的识别准确率.

1 事件相机

近年来,多种旨在自适应编码场景信息的事件相机投入了研究使用,如 DVS^[1],ATIS^[2],DAVIS^[4].事件相机每个像素检测光照强度随着时间的变化是否超过用户设定的阈值,当且仅当光照变化大于等于所设定的阈值时,才产生相应的事件信号.此外,还将光照由弱变强和由强变弱这两种情况下的信号输出分别称为正极性(ON)事件和负极性(OFF)事件.事件相机的光照强度响应如图 1 所示.

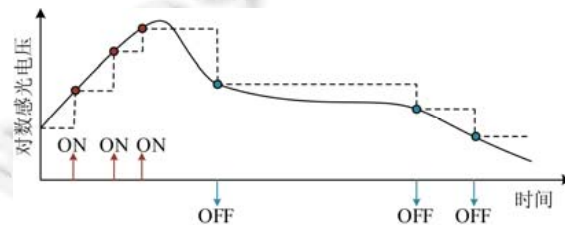


Fig.1 Illumination intensity response

图 1 光照强度响应

通常,利用以下符号表示一个事件:

$$e = (x, y, t, p) \quad (1)$$

其中,事件 e 表示在 t 时刻传感器像素阵列的 (x, y) 位置处感知到了场景光照变化而产生的事件.同时 p 表示光照变化属性, p 的取值是 $[-1, 1]$,当 $p=1$ 时,该事件为 ON 事件;当 $p=-1$ 时,该事件为 OFF 事件.

不同事件相机具有的主要优势因不同设计及工艺略有不同:(1) 低延时和高时间分辨率.DVS、ATIS 和 DAVIS 的延时分别为 $15\mu\text{s}$ 、 $4\mu\text{s}$ 和 $3\mu\text{s}$.对于帧率为 30fps 的传统相机而言,其延时高达 33ms,可见事件相机能更快地控制电路进行数据输出.(2) 高动态范围.DVS、ATIS 和 DAVIS 的动态范围分别为 120dB、125dB 和 130dB,而传统相机通常只能达到 60dB 的动态范围.(3) 低功耗.DVS、ATIS 和 DAVIS 的功耗分别为 24mW、50mW-175mW 和 5mW-14mW.

本文使用的数据采集设备是瑞士 INIVATION 公司生产的具体型号为 DAVIS240C 的事件相机设备.

2 相关工作

为了对事件流进行准确识别和分类,研究人员相继提出了不同的研究方法.对事件流数据不同的处理方法可以分为两大类:基于事件图像帧的方法和基于事件的方法.

基于事件图像帧的方法:通过某种转换方式将事件流转换为图像帧.Rivera-Acosta 等人^[5]将 50ms 内的事件进行堆叠得到图像帧,并对图像帧进行去噪和轮廓提取,最后利用神经网络进行分类.Park 等人^[6]将空间金字塔核变换应用在 10ms 内事件堆叠的图像上,统计不同金字塔层级内子区域的边缘方向,以此作为识别特征.

有工作^[7,8]将堆叠的图像直接输入进卷积神经网络(CNN)进行分类.Li 等人^[9]利用泄露整合放电(LIF)神经元控制时域编码层进行事件流可视化,并用支持向量机对 CNN 提取的图像特征矢量进行分类.此外,Peng 等人^[10]利用运动标志检测模块对事件流进行分段,堆叠每段事件流得到事件频率图,随后将乘以权重的事件频率图拉成一维矢量,得到最终的 BOE 特征用于分类.将事件流转换为图像帧使其能应用于现有计算机视觉处理模式,但是这种方法丢失了事件的时间属性,不可逆性操作削弱了事件的异步特性.

基于事件的方法:在事件层面上提取特征,而非使用转换的图像帧.很多工作^[11,12]利用脉冲神经网络(SNN)对事件流数据进行识别研究.然而,具有离散特性的事件数据使用梯度下降来训练 SNN 是有难度的.后续研究^[13-15]改变了 SNN 的权重初始化方案.同时,也有一些工作^[16,17]利用 CNN 网络进行研究.此外,还有其他基于事件的特征提取方法.Lagorce 等人^[18]利用时间面(time surfaces)特征建立了分层识别模型.随后,Sironi 等人^[19]在时间面基础上进行了改进,引入平均时间面直方图特征表示及一个新的学习架构.以上方法对事件流中所有事件数据都进行特征提取,并未筛选出具有特征性质的事件.

事件相机以异步形式输出场景的结构化信息,擅于捕捉运动物体轮廓的光照变化.角点作为物体轮廓中具有特定特征的点,选择事件流中的角点事件提高事件的信息含量有利于后续识别和分类.检测带有重要信息的事件,并只对角点事件进行特征提取,在保留事件重要特征和提取事件流特征的同时,有效地减少了对其他非角点事件的运算量.Harris 角点检测^[20]作为经典角点检测算法之一,具有计算简单、较高的稳定性和鲁棒性的优点.而 SURF 特征描述^[21]用更为高效的方式得到特征向量,具有计算速度快、鲁棒性好的优点.本文借鉴传统的 Harris 角点检测和 SURF 特征描述的实现了事件驱动的识别方法.

3 角点事件检测和特征提取

3.1 角点事件检测

传统的角点检测是对整个图像进行操作,在图像空间中寻找候选特征点.对于传入的每个事件而言,在整个事件相机的空间分辨率上进行计算是不可行的.因此对于每个事件,利用其时空邻域中过去事件进行有效辅助计算.

假设传入的事件为 $e_i = (x_i, y_i, p_i, t_i)$,将以 (x_i, y_i) 为中心的 $n \times n$ 空间邻域中 10ms 内已发生的事件进行叠加,得到事件 e_i 时空邻域的激活频率矩阵 A .计算激活频率矩阵的梯度 ∇A ,并利用梯度值计算结构张量 M_e :

$$\nabla A = (\partial A / \partial x, \partial A / \partial y) \quad (2)$$

$$M_e = \sum_{x,y} w(x,y) \begin{bmatrix} (\partial A / \partial x)^2 & (\partial A / \partial x)(\partial A / \partial y) \\ (\partial A / \partial x)(\partial A / \partial y) & (\partial A / \partial y)^2 \end{bmatrix} \quad (3)$$

其中 $w(x,y)$ 为高斯加权函数.

根据矩阵的特征值判断事件 e_i 的邻域属性,利用矩阵的迹和行列式进行判断的角点响应函数:

$$R_e = |M_e| - D_s Tr(M_e) \quad (4)$$

其中, $|M_e|$ 表示计算矩阵的行列式, D_s 为可调参数,用于调整角点事件检测的灵敏度, $Tr(M_e)$ 表示计算矩阵 M_e 的迹.当 $R_e > 0$ 时,当前事件为角点事件.

3.2 特征提取

利用 SURF 特征描述,对于每个角点事件,依然利用角点事件时空邻域中过去事件进行辅助计算.

假设现有角点事件 $e_j = (x_j, y_j, p_j, t_j)$,在以 (x_j, y_j) 为中心和 $6s$ 为半径的空间邻域中已发生的所有事件进行叠加,得到角点事件时空邻域的主方向矩阵 B ,其中, s 为尺度值.计算 B 中元素值在 x 和 y 方向的 Haar 小波响应并将小波响应值加权求和,选取滑动扇形区域内求和最长矢量,该矢量定义了角点事件的主方向.

为了描述以角点事件为中心的时空邻域,得到角点事件的主方向后,沿着选取的主方向,选取 $20s \times 20s$ 的方形区域,将该区域内已发生的所有事件进行累加得到该角点事件的描述矩阵 C .将 C 划分为 4×4 个子区域,在每

个子区域中利用 x 和 y 方向的 Haar 小波算子提取特征,并得到如下的向量描述:

$$V_e = (\sum dx, \sum dy, \sum |dx|, \sum |dy|) \quad (5)$$

由此,从 4×4 个子区域得到了该角点事件的 64 维联合向量.

4 基于手势识别的实验示例

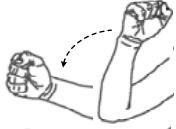

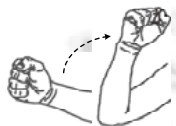
本文在“石头剪刀布”手势上进行实验,验证了角点事件检测和事件特征提取能够有效地进行识别.

4.1 数据采集

“石头剪刀布”手势的一个完整运动,见表 1,该手势既有手的形状变化也有空间位置的变化.

Table 1 Gestures

表 1 手势动作

运动阶段	运动说明
	首先紧握拳头,手臂向下运动
	在手臂到达最低点之前决定出石头、剪刀或布手势,到达最低点后以当前手形作为最终出手势,稍作停留
	手臂向上收回,并将手势慢慢变回拳头状的准备状态

本文采集了 15 个实验人员的手势数据,每个实验人员打手势过程持续大约 90s.事件流数据中带有部分噪声,因此,本文使用基于密度连接区域的密度聚类(DBSCAN)对任意一个 10ms 的事件流进行去噪,去噪后的事件流数据如图 2 所示,其中洋红色点表示 ON 事件,青蓝色点表示 OFF 事件.需要对该手势分割出关键手势事件流,然后对关键手势事件流数据进行识别.

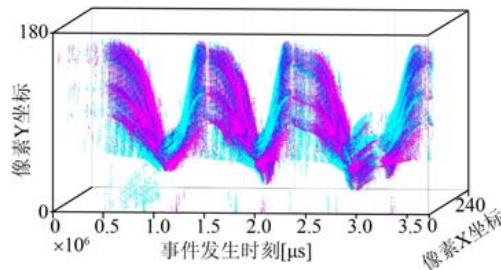


Fig.2 Denoised event stream

图 2 去噪后事件流

4.2 关键手势事件流提取

虽然事件相机只编码场景中的光照变化,但是优异的时间分辨率隐式地带来了运动信息.手势运动速度的快慢和手形的变化都会引起不同的事件数量.因此,可以通过事件数量的变化,提取关键手势的事件数据.

将手势事件流数据以 10ms 为间隔进行划分得到多个事件流子序列,统计每个事件流子序列中的事件数量,得到事件数量变化趋势.该趋势中事件数量从增加到减少的转折过程表示了关键手势和收回手势两种手势.先对事件流数量变化趋势进行平滑处理,再用一阶差分、二阶差分和事件数量阈值筛选出候选转折点.利用波峰跨度方法有效地筛选出真实转折点,波峰跨度阈值分别为 δ_1 和 δ_2 .最终提取的真实转折点如图 3 所示,转折点中

既包含关键手势,也包含收回手势,因此需要关键手势提取.

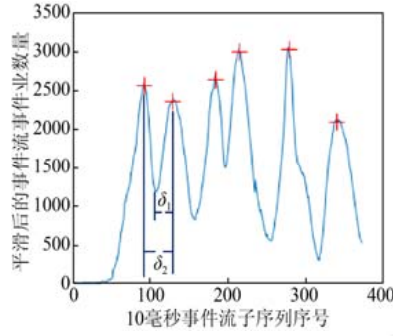


Fig.3 Real turning point

图3 真实转折点

本文利用角点事件检测选择带有重要信息事件的同时,也利用角点事件提取关键手势.角点事件检测算法中空间邻域设置为 5×5 ,同时,角点响应函数 R_e 的值大于等于 5 的事件被认为是角点事件.角点事件检测结果二值化图如图 4 所示,图中白色像素点表示非角点事件,红色像素点表示角点事件.从图 4 中可以看出,借鉴传统 Harris 角点检测方法的角点事件检测具有一定的抗噪能力,对后续操作提供了良好基础.

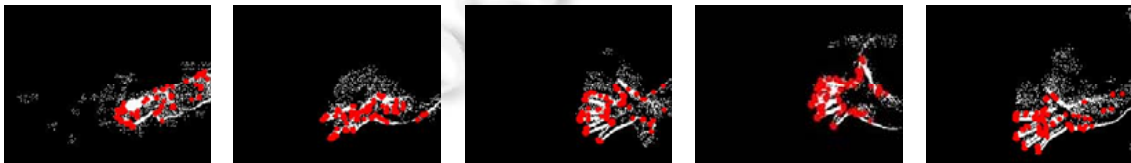


Fig.4 Viewable images of corner events

图4 角点事件可视图

本文利用两个相邻候选关键手势事件流特征事件的方向向量存在的关系选取真实关键手势事件流.首先将角点事件映射至二维空间,分析二维空间的特征事件分布,得到方向向量.利用两个相邻候选关键手势事件流特征事件的方向向量的向量积符号判断关键手势事件流,如图 5 所示.若向量积符号为正,则表示当前的候选关键手势事件流为收回阶段的手势事件流;若向量积符号为负,则表示当前的候选关键手势事件流为行动阶段的手势事件流,即为关键手势事件流.

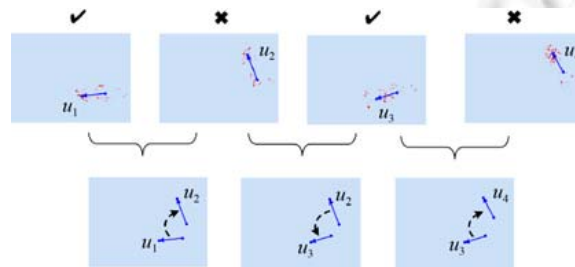


Fig.5 Schematic diagram of selecting key gesture

图5 关键手势筛选示意图

在 15 个实验人员 1 484 次出拳手势中提取到了 1 483 个关键手势事件流,同时也得到了 1 483 个关键手势相应的角点事件.

4.3 特征提取及手势识别结果

利用第 3.2 节描述的角度特征提取方法对关键手势事件流中的角点事件进行特征提取.本文使用 Kmeans 算法得到 k 个聚类中心,利用得到的聚类中心量化 SURF 特征,将该 SURF 特征归入最近的聚类中心所在的特征簇中.直至一个关键手势事件流中的每个 SURF 特征完成了上述映射过程,就得到了一个关键手势事件流对应的 k 维特征向量,便于选择合适的分类器进行手势分类.

利用肘部方法和 Kmeans++ 的初始化方法判断 k 值的选取.聚类指标为平均畸变程度,将畸变程度定义为所有类的平方误差之和.不同 k 值的平均畸变程度曲线如图 6 所示,该变化曲线并没有明显的拐点,初步确定 k 的取值范围在 100–500 之间.接着以支持向量机(SVM)最终的准确率确定 k 值.本文利用线性核函数完成搜索实验.从 15 个实验人员数据中,选取 3 个实验人员数据作为测试数据,为了避免人为指定测试数据对最终结果的影响,所以随机选取了 5 组测试数据,进一步为了避免 Kmeans 随机初始化的影响,每组实验重复运行 30 次.

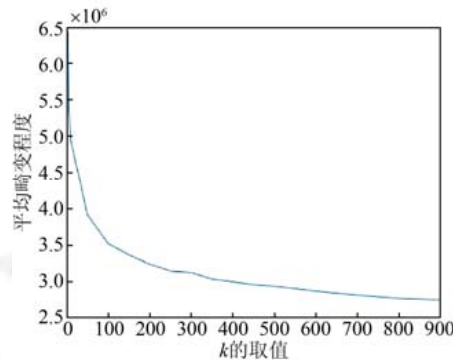


Fig.6 Average distortion degree curve of different k values

图 6 不同 k 值的平均畸变程度曲线

5 组实验的平均准确率如图 7 所示,从图 7 (a)中可以看出,在 k 取值为 20–800 时,准确率在 90%以上,而在 k 为 900 时,准确率急剧下降至 90%以下.其中 k 取值为 20–800 之间的准确率变化如图 7(b)所示,从中可以看出 k 取值为 140–600 之间的准确率在 96%以上.进一步细化区间得到图(c)中的变化曲线,可以从图(c)中可以看出在 k 取值为 360 时,最大平均准确率为 97.86%,如图 7(c)中圆形标识符所示.

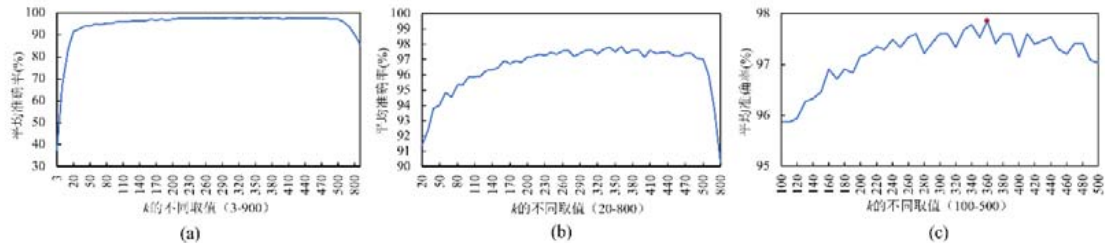


Fig.7 Accuracy variation of different cluster centers

图 7 不同聚类中心的准确率变化

从以上实验结果可以看出,根据事件数量变化转折点提取关键手势事件流的方法能够灵活可靠地提取关键手势事件流.应用于事件数据的 Harris 算法检测事件重要性的同时,进一步弱化了噪声的影响.对角点事件的时空邻域提取 SURF 描述特征,利用 SVM 完成最终的关键手势事件流识别,得到了较好的结果.

5 结论

二维图像因其局部区域中包含较多结构信息,所以计算机视觉、机器学习等领域根据该特性得到了快速发

展.但由于二维图像采集设备成像原理的限制,使得其越来越不能满足当前视觉任务的需求.受生物视网膜中神经元对视觉信息的处理机制启发,模拟生物视觉系统数据驱动特性的事件相机引起了广泛关注.为了对事件相机的事件数据异步特性更合理的利用,本文将传统的基于图像帧的Harris角点检测和SURF特征描述应用在了事件流数据上,实现了事件驱动的认识方法.该方法检测带有重要信息的角点事件,并只对角点事件进行特征提取,在保留事件重要特征和凝练提取事件流特征的同时,有效地减少了对其他事件的运算量.并且本文在一个特定动态手势上验证了该方法,提取关键手势,对关键手势基于角点事件的特征提取,得到了97.86%的识别准确率.

References:

- [1] Lichtsteiner P, Posch C, Delbruck T. A 128×128 120dB 15μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 2008,43(2):566–576.
- [2] Posch C, Matolin D, Wohlgenannt R. A QVGA 143dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE Journal of Solid-State Circuits*, 2011,46(1):259–275.
- [3] Leñero-Bardallo JA, Serrano-Gotarredona T, Linares-Barranco B. A 3.6 μs latency asynchronous frame-free event-driven dynamic-vision- sensor. *IEEE Journal of Solid-State Circuits*, 2011,46(6):1443–1455.
- [4] Brandli C, Berner R, Yang M, *et al.* A 240×180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 2014,49(10):2333–2341.3
- [5] Rivera-Acosta M, Ortega-Cisneros S, Rivera J, *et al.* American sign language alphabet recognition using a neuromorphic sensor and an artificial neural network. *Sensors*, 2017,17(10):2176.
- [6] Park PKJ, Lee K, Lee JH, *et al.* Computationally efficient, real-time motion recognition based on bio-inspired visual and cognitive processing. In: *Proc. of the IEEE Int'l Conf. on Image Processing (ICIP)*. 2015. 932–935.
- [7] Lungu IA, Corradi F, Delbrück T. Live demonstration: Convolutional neural network driven by dynamic vision sensor playing RoShamBo. In: *Proc. of the IEEE Int'l Symp. on Circuits and Systems (ISCAS)*. 2017.
- [8] Amir A, Taba B, Berg D, *et al.* A low power, fully event-based gesture recognition system. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 7243–7252.
- [9] Li H, Li G, Ji X, *et al.* Deep representation via convolutional neural network for classification of spatiotemporal event streams. *Neurocomputing*, 2018,299:1–9.
- [10] Peng X, Zhao B, Yan R, *et al.* Bag of events: an efficient probability-based feature extraction method for AER image sensors. *IEEE Trans. on Neural Networks and Learning Systems*, 2017,28(4):791–803.
- [11] Diehl PU, Neil D, Binas J, *et al.* Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In: *Proc. of the Int'l Joint Conf. on Neural Networks (IJCNN)*. 2015. 1–8.
- [12] Neil D, Pfeiffer M, Liu SC. Phased lstm: Accelerating recurrent network training for long or event-based sequences. *Advances in neural information processing systems*. 2016. 3882–3890.
- [13] Cao Y, Chen Y, Khosla D. Spiking deep convolutional neural networks for energy-efficient object recognition. *Int'l Journal of Computer Vision*, 2015,113(1):54–66.
- [14] Orchard G, Meyer C, Etienne-Cummings R, *et al.* HFirst: A temporal approach to object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2015,37(10):2028–2040.
- [15] Rueckauer B, Lungu IA, Hu Y, *et al.* Theory and tools for the conversion of analog to spiking convolutional neural networks. *arXiv preprint arXiv:1612.04052*, 2016.
- [16] Wang Q, Zhang Y, Yuan J, *et al.* Space-time event clouds for gesture recognition: from RGB cameras to event cameras. *IEEE Winter Conf. on Applications of Computer Vision (WACV)*. IEEE, 2019. 1826–1835.
- [17] Pérez-Carrasco JA, Zhao B, Serrano C, *et al.* Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing—application to feedforward ConvNets. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013,35(11):2706–2719.
- [18] Lagorce X, Orchard G, Galluppi F, *et al.* Hots: a hierarchy of event-based time-surfaces for pattern recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017,39(7):1346–1359.

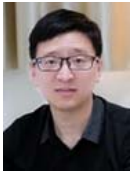
- [19] Sironi A, Brambilla M, Bourdis N, *et al.* HATS: Histograms of averaged time surfaces for robust event-based object classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 1731–1740.
- [20] Harris CG, Stephens M. A combined corner and edge detector. In: Proc. of the 4th Alvey Vision Conf. 1988. 189–192.
- [21] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. European conference on computer vision. Berlin: Springer-Verlag, 2006. 404–417.



张姝(1995—),女,四川南充人,硕士,主要研究领域为模式识别,机器学习.



杜从洋(1996—),男,硕士生,主要研究领域为模式识别,机器学习.



吴金建(1986—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为视觉显著性检测,恰可识别误差估计,图像质量评价.



石光明(1965—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为遥感探测与成像,立体深度成像,图像处理与压缩编码系统研发,智能信号处理算法.



谢雪梅(1967—),女,博士,教授,博士生导师,CCF 专业会员,主要研究领域为行为识别及目标检测,场景理解与视频分析,深度学习特征表示.