

JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象报 图形学报

2022
06
VOL.27

ISSN1006-8961
CN11-3758/TB



中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）



第27卷第6期（总第314期）
2022年6月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

序言王耀南 I



面向智慧交通的图像处理与边缘计算(第1743页)

视觉理解与计算成像

基于深度学习的视觉目标检测技术综述

曹家乐, 李亚利, 孙汉卿, 谢今, 黄凯奇, 庞彦伟 1697

面向复杂场景的人物视觉理解技术

马利庄, 吴飞, 毛启容, 王鹏杰, 陈玉珑 1723

面向智慧交通的图像处理与边缘计算

曹行健, 张志涛, 孙彦赞, 王平, 徐树公, 刘富强, 王超, 彭飞, 穆世义, 刘文予, 杨铀 ... 1743

视觉弱监督学习研究进展

任冬伟, 王旗龙, 魏云超, 孟德宇, 左旺孟 1768

智能遥感: AI赋能遥感技术

孙显, 孟瑜, 刁文辉, 黄丽佳, 张新, 骆剑承, 高连如, 王佩瑾, 闫志远, 郜丽静, 董文, 冯琪超, 李霁豪, 付琨 1799

脉冲视觉研究进展

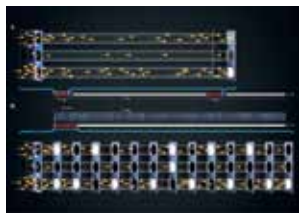
黄铁军, 余肇飞, 李源, 施柏鑫, 熊瑞勤, 马雷, 王威 1823

计算成像前沿进展

顿雄, 付强, 李浩天, 孙天成, 王建, 孙启霖 1840

移动在线实时绘制技术研究综述

刘畅, 霍宇驰, 张严辞, 张乾, 郑家祥, 唐睿, 余耿, 王锐, 贾金原 1877



脉冲视觉研究进展(第1823页)

数据挖掘和信息交互

表格识别技术研究进展

高良才, 李一博, 都林, 张新鹏, 朱子仪, 卢宁, 金连文, 黄永帅, 汤帆 1898

多媒体隐写研究进展

张卫明, 王宏霞, 李斌, 任延珍, 杨忠良, 陈可江, 李伟祥, 张新鹏, 俞能海 1918

大脑多模态成像技术定量研究进展

叶慧慧, 何宏建, 方静苑, 童琪琦, 周子涵, 刘华锋 1944

多模态人机交互综述

陶建华, 巫英才, 喻纯, 翁冬冬, 李冠君, 韩腾, 王运涛, 刘斌 1956

文化遗产活化关键技术研究进展

耿国华, 何雪磊, 王美丽, 李康, 贺小伟 1988

情感计算与理解研究发展概述

姚鸿勋, 邓伟洪, 刘洪海, 洪晓鹏, 王甦菁, 杨巨峰, 赵思成 2008

跨模态脑图谱数据融合研究进展

罗娜, 宋明, 杨正直, 蒋田仔 2036



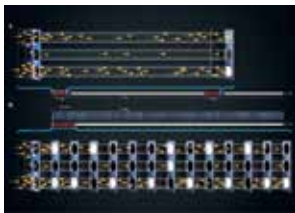
移动在线实时绘制技术研究综述(第1877页)

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



The review of image processing and edge computing for intelligent transportation system(P1743)



Advances in spike vision(P1823)



A review of real-time rendering technology based on mobile internet platforms(P1877)

Visual Understanding & Computational Imaging

A survey on deep learning based visual object detection

Cao Jiale, Li Yali, Sun Hanqing, Xie Jin, Huang Kaiqi, Pang Yanwei 1697

Visual recognition technologies for complex scenarios analysis

Ma Lizhuang, Wu Fei, Mao Qirong, Wang Pengjie, Chen Yulong 1723

The review of image processing and edge computing for intelligent transportation system

Cao Xingjian, Zhang Zhitao, Sun Yanzan, Wang Ping, Xu Shugong, Liu Fuqiang, Wang Chao, Peng Fei, Mu Shiyi, Liu Wenyu, Yang You 1743

Progress in weakly supervised learning for visual understanding

Ren Dongwei, Wang Qilong, Wei Yunchao, Meng Deyu, Zuo Wangmeng 1768

The review of AI-based intelligent remote sensing capabilities

Sun Xian, Meng Yu, Diao Wenhui, Huang Lijia, Zhang Xin, Luo Jiancheng, Gao Lianru, Wang Peijin, Yan Zhiyuan, Gao Lijing, Dong Wen, Feng Yingchao, Li Jihao, Fu Kun 1799

Advances in spike vision

Huang Tiejun, Yu Zhaofei, Li Yuan, Shi Boxin, Xiong Ruiqin, Ma Lei, Wang Wei 1823

Recent progress in computational imaging

Dun Xiong, Fu Qiang, Li Haotian, Sun Tiancheng, Wang Jian, Sun Qilin 1840

A review of real-time rendering technology based on mobile internet platforms

Liu Chang, Huo Yuchi, Zhang Yanchi, Zhang Qian, Zheng Jiexiang, Tang Rui, Yu Geng, Wang Rui, Jia Jinyuan 1877

Data Mining and Information Interaction

A survey on table recognition technology

Gao Liangcai, Li Yibo, Du Lin, Zhang Xinpeng, Zhu Ziyi, Lu Ning, Jin Lianwen, Huang Yongshuai, Tang Zhi 1898

Overview of steganography on multimedia

Zhang Weiming, Wang Hongxia, Li Bin, Ren Yanzhen, Yang Zhongliang, Chen Kejiang, Li Weixiang, Zhang Xinpeng, Yu Nenghai 1918

Research progress of quantitative multimodal brain imaging technology

Ye Huihui, He Hongjian, Fang Jingwan, Tong Qiqi, Zhou Zihan, Liu Huafeng 1944

A survey on multi-modal human-computer interaction

Tao Jianhua, Wu Yingcai, Yu Chun, Weng Dongdong, Li Guanjun, Han Teng, Wang Yuntao, Liu Bin 1956

Research progress on key technologies of cultural heritage activation

Geng Guohua, He Xuelei, Wang Meili, Li Kang, He Xiaowei 1988

An overview of research development of affective computing and understanding

Yao Hongxun, Deng Weihong, Liu Honghai, Hong Xiaopeng, Wang Sujing, Yang Jufeng, Zhao Sicheng 2008

Survey on cross-modal fusion based on brain atlas data

Luo Na, Song Ming, Yang Zhengyi, Jiang Tianzi 2036

中图法分类号: TP752 文献标识码: A 文章编号: 1006-8961(2022)06-1823-17

论文引用格式: Huang T J, Yu Z F, Li Y, Shi B X, Xiong R Q, Ma L and Wang W. 2022. Advances in spike vision. Journal of Image and Graphics, 27(06):1823-1839 (黄铁军, 余肇飞, 李源, 施柏鑫, 熊瑞勤, 马雷, 王威. 2022. 脉冲视觉研究进展. 中国图象图形学报, 27(06):1823-1839) [DOI:10.11834/jig.220175]

脉冲视觉研究进展

黄铁军^{1,2,3,*}, 余肇飞², 李源¹, 施柏鑫^{1,2}, 熊瑞勤¹, 马雷^{1,3}, 王威⁴

1. 北京大学计算机学院, 视频与视觉技术国家工程研究中心, 北京 100871; 2. 北京大学人工智能研究院, 北京 100871;
3. 北京智源人工智能研究院, 北京 100084; 4. 北京通用人工智能研究院, 北京 100080

摘要: 视频是视觉信息处理的基础概念, 传统视频的帧率只有几十 Hz, 不能记录光的高速变化过程, 成为限制机器视觉速度的天花板, 其根本原因在于视频概念脱胎于胶片成像, 未能发挥电子和数字技术的潜力。脉冲视觉模型通过感光器件捕获光子, 累积能量达到约定阈值时产生脉冲, 形成脉冲的时间越长, 表明收到的光信号越弱, 反之光信号越强, 据此可估计任意时刻的光强, 从而实现连续成像。采用普通器件, 研制了比影视视频快千倍的超高速成像芯片和相机, 进而基于脉冲神经网络实现了超高速目标检测、跟踪和识别, 打破了机器视觉提速依赖算力线性增长的传统范式。本文从脉冲视觉模型表达视觉信息的生物学基础和物理原理出发, 介绍了脉冲视觉原理的软件模拟器及其模拟真实世界光子传播的计算过程, 描述了基于脉冲视觉原理的高灵敏光电传感器器件及芯片的工作机理和结构设计、基于脉冲视觉的影像重建原理以及脉冲视觉信号与普通图像信号融合的计算摄像算法与计算摄像系统, 介绍了基于脉冲神经网络的超高速运动目标检测、跟踪与识别, 通过对比国际国内相关研究内容和发展现状, 展望了脉冲视觉的发展与演进方向。脉冲视觉芯片和系统在工业(高铁、电力和轮机等不停机监测, 智能制造高速监视等)、民用(高速相机、智能交通、辅助驾驶、司法取证和体育判罚等)以及国防(高速对抗)等领域都具有巨大应用潜力, 是未来值得重点关注和研究的一个重要方向。

关键词: 脉冲视觉; 脉冲神经网络; 视觉信息处理; 类脑视觉; 人工智能

Advances in spike vision

Huang Tiejun^{1,2,3,*}, Yu Zhaofei², Li Yuan¹, Shi Boxin^{1,2}, Xiong Ruiqin¹, Ma Lei^{1,3}, Wang Wei⁴

1. School of Computer Science, National Engineering Research Center of Visual Technology, Peking University, Beijing 100871, China;
2. Institute for Artificial Intelligence, Peking University, Beijing 100871, China; 3. Beijing Academy of Artificial Intelligence, Beijing 100084, China; 4. Beijing Institute for General Artificial Intelligence, Beijing 100080, China

Abstract: Video is the conceptual base of visual information processing technology. The frame rate of traditional video is tens of Hertz, which is incapable to represent the Ultra-high speed change process of light. It also constrains to the speed of machine vision. The concept of video is originated from film imaging, which can't unleash the full potential of electronic and digital technology. The spike vision model generates a spike when the photon energy captured by a photosensitive device reaches the predefined threshold. The longer the spike timing is, the weaker the received light signals are. The occurred light intensity can be inferred to achieve consistent imaging. Current research of spiking vision is focused on devel-

收稿日期: 2022-03-04; 修回日期: 2022-03-31; 预印本日期: 2022-04-07

* 通信作者: 黄铁军 tjhuang@pku.edu.cn

基金项目: 科技创新 2030——“新一代人工智能”重大项目(2021ZD0109802, 2021ZD0109803)

Supported by: National Key R&D Program of China (2021ZD0109802, 2021ZD0109803)

oping disruptive ultra-high speed imaging chips and cameras that are 1 000 times faster than videos. It is substituting traditional paradigm derived from machine vision speed linear growth of computing power. Based on the biological evidence and physical principle of spiking vision, our review analyzes the software simulator of spiking vision, the computational process of the photon propagation simulation, spiking vision based visual image reconstruction, and the detailed mechanism of photoelectric sensor structure and the chip design. Our review reveals the ultra-high speed target motion detection, and spiking neural networks based tracking and recognition. The evolution of spiking vision is forecasted. Future spiking vision based chip and system has multifaceted potentials to further harness the industrial tasks (e.g., the monitoring of high-speed train, power grid, turbine and intelligent manufacturing), civil applications (e.g., high-speed camera, intelligent transportation, auxiliary driving, judicial forensics and sports arbitration), and national defense construction (e.g., high-speed confrontation campaigns).

Key words: spiking vision; spiking neural networks; visual information processing; brain-inspired vision; artificial intelligence

0 引言

数码相机是真正数码吗?通常回答是肯定的,因为数码相机利用 CCD (charge-coupled device)/CMOS (complementary metal-oxide-semiconductor) 图像传感器和数字电路取代了胶片。然而,数码相机本质上仍停留在模拟时代,因为它无保留地继承了图像和视频这种表达视觉信息的形式,这种形式对于化学胶片记录是必需的,但对于纯数码系统并不必要。事实上,图像无法记录曝光时间内光线的变化过程,视频甚至会丢失相邻两次曝光之间的所有动态过程。此外,帧率只有几十 Hz 的相机无法拍摄超过其帧率的高速场景,高速像机帧率可以达到数千甚至数万,需要昂贵的光学系统、传感器和快门控制系统,但得到的每帧图像质量仍然不高,而且高帧率带来的巨大数据量对后续传输、存储和处理带来巨大压力。由此可见,图像和视频已经成为相机捕捉快速变化的光过程的最大障碍。

尽管如此,图像和视频却被默认继承实现机器视觉的基本范式被长期假定为“摄像头+计算机+算法”。1966年春,美国·麻省理工学院人工智能实验室主要筹备人 Papert 教授组织学生开展“暑期视觉项目”时提出:“计算机连上摄像头,‘描绘它看到什么’这个问题一个暑期项目就能搞定。在今年夏天结束时,我们将开发出电子眼”。这种假定和思维模式对机器视觉影响至今,很少有人质疑用图像序列表达视觉信息的合理性,更很少有人质疑用计算机算法进行视觉信息处理的合理性。事实上,在生物视觉系统中,根本就不存在图像序列,也没有机械的算法,传统机器视觉(计算机视觉)只是把成熟

的工业产品(摄像头和计算机)以及常用的信息处理手段(算法)生搬硬套到视觉这个问题上。以视频这个概念为例,两帧图像之间的运动变化过程已经丢失了,因此需要运动估计、光流等算法根据相邻帧的差异估算出运动,但这在生物视觉系统中既不存在,也无必要。

对生物视觉神经结构和机理的研究基本上是和计算机视觉并行的。20世纪50年代,生理学家就发现生物视网膜中存在对运动敏感的神经元。1953年英国神经科学家 Barlow (1953)发现青蛙视网膜有一种神经节细胞对运动斑点敏感,1959年美国麻省理工学院的 Lettvin 等人(1959)又在青蛙视网膜发现4种新功能的神神经节细胞,同年美国哈佛大学的 Hubel 和 Wiesel (1959)在猫的初级视皮层发现对移动边缘敏感的神经元。眼睛并非像照相机那样向大脑传送静态图像,而是采用异步脉冲序列方式向大脑报告光线变化,这已是生物视觉研究领域的常识,但是,脉冲序列编码光学刺激的精细机理,至今还未搞清楚。

研制模拟生物视网膜新型相机的努力起始于20世纪80年代末。1988年,美国加州理工大学 Mead 教授和其学生 Mahowald (1988)提出了模拟视网膜部分层次的硅电路方案(Mead 和 Mahowald, 1988),1991年提出表达视网膜事件输出的异步传输协议 AER (asynchronous event representation), 1994年设计出首款 AER 视觉芯片,可检测光强变化并输出事件脉冲。Mahowald 毕业后到瑞士苏黎世大学工作,所在团队后来开发出了动态视觉传感器(dynamic vision sensor, DVS)芯片和相机,称为事件相机(event camera)。事件相机感光单元仅在亮度变化超过阈值时才会产生事件,模拟的是低等动物(青蛙、蛇等)视网膜对动态目标的编码机制,捕

捉的是光的变化,在场景不变时不输出事件,优点是能够捕捉高速变化,缺点是不能像传统相机那样成像,因此不能替代传统视频相机。

脉冲视觉是北京大学黄铁军团队(中国专利 ZL201610045011)(黄铁军等,2016,2019)提出的一套能够替代且优于传统视频的视觉表达新体系,是受灵长类动物视网膜中央凹的神经回路结构和信息处理机制启发而提出的,其感光单元不断捕获光子,当累积强度超过给定阈值时产生脉冲。脉冲相机在保持时空关系的情况下直接将输入光子流映射为比特流,脉冲流具有清晰的物理意义,可以重建任何时刻的图像,因此能够完全替代传统相机。

与传统视频相机相比,事件相机和脉冲相机的输出都是脉冲流,可以进一步采用脉冲神经网络(spiking neural networks, SNNs)完成高速视觉任务。两者都可以完成高速目标检测跟踪,脉冲相机还能完成目标分类识别等传统视觉任务,因此能够全面替代传统视频相机。

脉冲相机对传统视频相机的颠覆主要在于“超速”和“全时”。“超速”人眼千倍,能够“看清”高速旋转叶片的文字。“全时”则是指从芯片采集的神经脉冲序列中可以重构出任意时刻的画面,就像人脑能够连续感知外部环境一样。

本文将归纳脉冲视觉原理方法、脉冲芯片与相机、脉冲成像与重构,以及脉冲视觉算法等几个方面的发展现状,分析该领域的关键技术,并展望未来发展空间。

事件相机相关研究已经30年,国内外相关论文很多(李家宁和田永鸿,2021;桑永胜等,2019;王程等,2021),本综述的国际进展部分,主要讨论事件相机。脉冲视觉模型是2015年提出的,最近相继获得中美日韩欧专利(中 ZL201610045011,美 10523972,日 6754434,韩 520000216573,欧 16595EPPC),脉冲相机芯片和相机已经研制成功,但尚未正式销售,因此相关论文主要是北京大学团队发表的,将在本综述的国内进展部分介绍,实际上这也是国际最新进展,只是囿于综述格式,才如此安排。

1 国际研究现状

1.1 事件相机

事件相机和传感器的相关研究最早可以追溯到

20世纪80年代,发展过程主要经历了两个大的标志性阶段:用脉冲进行事件表达及 AER 协议的提出;动态视觉传感器 DVS 的提出及其后续改进方案。

事件表达和 AER 协议的相关研究,最早由美国加州理工大学 Mead 和 Mahowald(1988)提出,该方案对视网膜细胞的部分功能进行了初步模拟,并使用芯片进行了验证,传感器电路的像素规模为 48×48 。为了减小数据维度,Meade 实验室的 Sivilotti(1991)于1991年提出了异步通信协议,将传感器常用的像素串行扫描输出协议替换为数据驱动的处理形式,即像素数据传输需要经历请求—握手过程,并设计了相应的仲裁机制应用于后续仿视网膜传感器的设计中。在后续的研究中,Mahowald(1992)先后于1992年和1994年提出了利用 AER 数据脉冲表示形式的异步感知架构和首款 AER 视觉传感器,具体描述了 AER 协议在2维仿视网膜传感器像素阵列中的实现形式。随后,Mahowald 毕业并加入瑞士苏黎世大学神经信息研究所工作,开启了欧洲 AER 和事件相机的相关研究。

美国宾夕法尼亚大学的 Boahen(2000)(此前在 Mead 实验室)在2000年讨论了在脉冲编码神经形态芯片中使用 AER 进行通信的方法,并给出了权衡带宽、排队时间和吞吐量之后的仲裁电路设计,随后西班牙塞维利亚微电子研究所于2004年设计了基于 AER 协议的积分发放神经形态像素,该电路使 AER 传感器具备了带符号权重计算的能力。

DVS 传感器和相机的提出始于2002年,由 Mahowald 在瑞士苏黎世大学神经信息研究所的同事 Kramer 教授设计并实现(Kramer,2002),该传感器采用对比度编码的方式实现了对连续变化的光强进行持续记录和瞬态记录两种模式,并模拟了视网膜中 ON 和 OFF 通道的部分功能。2004年,瑞士苏黎世大学的 Lichtsteiner 等人(2004)对 DVS 电路进行了改进,使用 $0.35 \mu\text{m}$ 工艺实现了 32×32 像素的传感器芯片。该像素电路具备 ON 和 OFF 通道的输出并实现了更加对称的正负输出,此方法通过对数压缩和时域滤波实现了冗余数据去除。

DVS 传感器主要以记录事件的输出为主,对于相对静止的像素并不输出信息,为了获取场景的全部信息,往往需要使用 DVS 传感器输出的事件信息与传统图像传感器输出的图像信息进行融合。相关

研究主要集中在 DVS 与传统图像传感器像素的融合设计方面,代表性工作包括 ATIS 传感器 (asynchronous time-based image sensor) 和 DAVIS 传感器 (dynamic and active pixel vision sensor)。

ATIS 传感器由 Posch 等人 (2010) 所在的奥地利技术研究所提出,在 2008 年到 2011 年间将数字像素传感器 (digital pixel sensor, DPS) 的部分结构与 DVS 结构融合,形成基于时间的异步图像传感器 ATIS,像素电路中包含变化检测器和光强测量装置,在检测到光强变化 (事件) 时开始测量事件的间隔,从而估算亮度信息并输出。由于该结构对静止像素不产生信息,因此无法完整地估计场景的纹理。

DAVIS 传感器由瑞士苏黎世大学的 Brandli 等人 (2014) 在 2014 年提出,为了解决 DVS 传感器无法输出绝对光强信息的问题,他们在 DVS 基础上增加有源像素传感器 (active pixel sensor, APS),设计出动态有源像素视觉传感器 DAVIS。DAVIS 通过异步输出 DVS 事件和同步输出全局曝光图像来满足绝对光强信息的需求。在 $0.18\ \mu\text{m}$ 工艺下实现了 240×180 像素的传感器芯片。由于 APS 部分输出的图像速率远低于事件流,因此无法做到很好的同步效果。目前 ATIS 和 DAVIS 传感器在事件表达和纹理重建方面,尚无法很好地融合给出场景的全部信息。

近几年关于事件相机的研究主要在 AER 和 DVS 的基础上结合先进结构和工艺持续提升性能。2015 年,西班牙塞维利亚微电子研究所的 Leñero-Bardallo 等人 (2015) 设计出的基于事件的 DVS 传感器同时具备强度模式和空间对比度模式,并可以与脉冲密度调制和第一时间脉冲输出模式结合。在 $0.35\ \mu\text{m}$ 工艺下实现 64×64 像素的传感器芯片,具有小于 $1\ \text{ms}$ 的延迟和 $19\ \text{mA}$ 的低功耗。2018 年,该团队提出了基于异步帧的图像传感器 (Leñero-Bardallo 等, 2018), 传感器像素具有对光照强度进行自适应调节的能力,并具有编程配置功能对灵敏度和输出数据速率进行调节,提高了传感器动态范围。2017 年,三星先进技术研究院的 Son 等人 (2017) 提出了改进的 G-AER (group address-event representation) 机制,采用字节序列形式提高事件输出能力,并使用 $90\ \text{nm}$ 背照工艺实现了 640×480 像素的 DVS,像素尺寸为 $9\ \mu\text{m} \times 9\ \mu\text{m}$,事件输出速率高达 $300\ \text{Meps}$ 。2020 年,索尼与 Prophesee 联合发

布了基于背照式 3D 堆叠技术的 DVS 图像传感器,事件输出速率进一步提升到 $1.066\ \text{Geps}$ 。

1.2 事件相机模拟器

事件相机的高速感知、无运动模糊、高动态范围等优势启发了新一代视觉应用,然而事件相机并未普及,搭建实验环境、采集数据比较困难,这制约了研究界的进展。在真实环境下进行视觉应用技术,具有实施复杂度高、实时反馈困难、训练成本巨大和风险控制管理困难等问题,无法模拟成本高、复杂性高和多样性高的应用场景。以自动驾驶行业为例,车辆碰撞、复杂交通情况下的行人车辆事故等场景,其试错成本无法估量,不被道德法律所允许。为了解决这些问题,事件相机模拟器应运而生。

在传统图像帧领域,仿真器已有多年研究历史,如自动驾驶的代表仿真环境 CARLA (car learning to act) (Dosovitskiy 等, 2017)、微软的飞行模拟器 Microsoft Airsim (Shah 等, 2018) 以及可用于视觉学习的高真实感互动环境 UnrealCV (Qiu 等, 2017) 等。这里重点介绍事件相机的相关模拟器,如 ESIM (event camera simulator) (Rebecq 等, 2018)、v2e (Hu 等, 2021b) 等。

ESIM 是一个开源的事件相机模拟器,它有两个主要模块:场景渲染引擎和事件模拟器单元。场景渲染引擎模拟了一个有多个物体和光照条件的真实场景,一个虚拟的相机按照给定的轨迹移动,用一系列的参数对这个场景进行渲染。在运行过程中,渲染引擎可以将一个时间戳作为输入,并生成相机在该时间绘制的图像。特别地,时间戳可以是任意精度的,ESIM 可以通过渲染引擎实现任意的时间采样。模拟器单元遵循事件相机的机制,它将两幅强度图像作为输入,计算每个像素的变化,并通过假设这两幅图像之间亮度的线性变换来生成事件。之后,它可以根据事件生成的结果调整下一个采样的时间戳,并使用新的时间戳进行下一轮的采样。

能够合成虚拟事件流的软件工具还包括 v2e 和 EventGAN (Zhu 等, 2021)。v2e 从视频到事件是基于一个叫做 Super SloMo (Jiang 等, 2018) 的超分辨率网络和 ESIM 的模拟单元。在运行时,它首先在传统相机拍摄视频的时间尺度上增强分辨率。然后,它使用 ESIM 的模拟器单元在新的图像系列上生成事件。EventGAN 训练一个网络,它将两幅图像作为输入,并在这两幅图像之间生成事件作为输出。

EventGAN 引入了 3 种损失, 以使网络快速而准确地收敛。

1.3 事件相机辅助的影像重建与光流估计

针对事件相机的影像重建, Kim 等人(2008)设计了一个扩展卡尔曼滤波框架来根据脉冲数据重建梯度图, 并在后续工作中进一步预测了 6 自由度相机运动。Bardow 等人(2016)提出利用原始对偶算法(primal-dual algorithm)来同时优化光流和图像光强估计的方法。Scheerlinck 等人(2019)提出了一种轻量级的网络来加速光强推断的过程。针对模拟数据和真实数据之间的差异所带来的网络泛化性受限问题, Stoffregen 等人(2020)提出一种较为新颖的策略, 用来消除模拟数据和真实数据之间的差异, 从而在多个数据集上提升了重建性能。

在影像重建过程中, 场景光流估计也是一个高度关联的问题, 因此事件相机的光流估计在近年获得了很多关注。Benosman 等人(2012)通过在时空域估计一个平面, 针对每个事件进行局部线性估计。Rueckauer 和 Delbruck(2016)的研究表明, 该算法对手工设计的离差值抑制要求较高, 难以与事件相机真实产生的脉冲相匹配。Gallego 等人(2018)提出对比度最大化算法, 通过最大化光流变形之后图像的对比度来估计事件脉冲的运动轨迹, 从而得到光流。为了应对深度学习时代大数据的要求, Zhu 等人(2018a)提出了 MVSEC(multi vehicle stereo event camera)数据集, 该数据集包含较大量的室内与室外场景的事件相机及其光流数据。在此基础上, Zhu 等人(2018b)提出采用类似 U-Net 形式编码—解码结构的 EV-FlowNet 算法。在该算法中, 神经网络的输入由脉冲数据同时进行脉冲计数与时间表征进行构建。此外, 该神经网络借助 DAVIS 相机提供的灰度图像作为辅助, 构建了自监督光照一致性损失函数来进行训练。该算法的升级版(Zhu 等, 2019a)提出评估经光流变形之后脉冲的模糊程度来构建损失函数, 而不使用灰度图像进行辅助, 并提出了基于时间戳进行时域插值的新型脉冲表征方案。除了深度学习方法之外, 脉冲神经网络(spiking neural networks, SNNs)由于其脉冲输入形式与事件相机的输出具有更好匹配性, 近年来在事件相机光流估计中也取得不少发展。Paredes-Valles 等人(2020)基于脉冲神经网络的稳态脉冲时序依赖可塑性(spike-timing-dependent plasticity, STDP)设计了无监督的

分层网络来实现事件相机光流估计。Spike-FlowNet(Lee 等, 2020)采用脉冲神经网络与人工神经网络的混合模型, 首先基于脉冲神经网络对输入的事件脉冲进行编码, 编码得到的特征由层数较深的人工神经网络进行进一步金字塔编码, 再由人工神经网络进行逐级解码得到光流。Gehrig 等人(2021)基于传统相机中的相关层(correlation layer), 设计了基于相关与特征变形的事件流序列光流估计算法, 该算法使用了循环神经网络进一步挖掘脉冲流的时域特性。

1.4 事件相机辅助的计算摄像

事件相机相较于传统图像传感器具有高动态、高速等特性, 但是由于其特殊的成像范式与图像/视频格式, 使得其难以兼容当下的计算摄像算法。然而其独特的性能优势如果通过合理设计相应的算法得以充分发挥, 可以解决很多在传统相机领域遇到瓶颈的计算摄像问题, 因而成为近年来计算摄像领域的研究热点。事件相机的高动态、高速的性质使其在计算摄像任务中主要用做辅助信号来源, 用以提升图像的帧率和质量等, 在视频插帧、图像增强和深度估计等问题上得到了较为广泛的应用。

在视频插帧方面, Tulyakov 等人(2021)提出“时间镜头”(Time Lens)的方法, 利用混合相机系统, 拍摄了在时空上对齐的彩色视频和事件信号流。通过 2 种方式来预测中间帧, 一是利用事件信号流直接预测; 二是将事件信号转换为光流再估计中间帧。对这 2 种方式所得结果加权得到最终的预测结果, 从而有效地挖掘事件信号中的时序信息来补偿低帧率的彩色视频。

在图像增强方面, Pan 等人(2019)提出事件重积分模型, 该模型利用事件信号构建了清晰图像和模糊图像之间的关系, 通过斐波那契查找求解单变量优化问题, 实现了从模糊低帧视频到清晰高帧视频的恢复。Mostafavi 等人(2020)首次提出利用事件信号流的高速信息来重构高分辨率图像, 同时进一步扩展利用事件信号来对灰度图像进行超分辨率。Jing 等人(2021)基于事件相机提出事件异步插值模块和视频超分辨率模块, 事件异步插值模块挖掘事件信号中的时序信息重构出高速视频中的高频信息, 随后利用视频超分辨率模块利用重构的高频信息实现彩色视频帧分辨率的提高。

在深度估计方面, Andreopoulos 等人(2018)基

于事件相机将深度估计算法嵌入到相机图像处理器中,实现了低功耗的深度估计。Zhu 等人(2018a)将事件信号表示成时序同步事件视差量(time synchronized event disparity volumes),解决了相机运动中出现的模糊问题,得到了更好的深度估计结果。Tulyakov 等人(2019)首次利用深度学习的方法从事件信号中估计深度,提出了事件信号的嵌入方法,使其从连续信号变成离散信号,并基于此设计了神经网络,得到密集深度估计结果。

在信号增强方面,事件相机在具有高速高动态特点的同时,对噪声信号也更加敏感,所以对事件信号进行去噪逐渐受到重视。Baldwin 等人(2020)搭建了第 1 个动态视觉传感器噪声标记数据集(DVS-NOISE20),提出事件概率掩码模型(event probability mask, EPM),并设计了事件去噪卷积神经网络(event denoising convolutional neural network, EDnCNN)对神经形态信号进行去噪。

国外对差分型事件相机在计算摄像领域的应用研究,为积分型脉冲相机的研究提供了宝贵的经验,同时也具有重要的启示意义。

1.5 基于事件相机的视觉算法

面向极端环境或高速运动的场景,传统相机由于低频率成像特点,易因运动模糊丢失或漏检目标,而事件相机高采样率、高动态范围的优势,使得基于事件相机的高速目标检测跟踪成为当下的研究热点。

基于脉冲信号直接进行检测跟踪的研究工作还很少,许多方法都是先将脉冲转为图像或时间表面,再在转化后的图像上进行后续算法处理。美国马里兰大学 Ye 等人(2020)利用四参数运动模型将脉冲流转化为时间灰度图像,并对转化后的图像进行全局最小化以分离不同运动目标,最后采用卡尔曼滤波的方法对高速运动目标稳定跟踪。瑞士苏黎世联邦理工大学 Rueckauer 等人(2016)采用了联合 DVS 和 APS 的方式进行运动目标检测,其将 DVS 脉冲流检测出目标候选区域并用卷积神经网络进行分类,再利用粒子滤波对目标定位与跟踪。此外,还有来自澳大利亚阿德莱德大学和昆士兰大学的 Chin 等人(2019)将 DVS 的脉冲流转为图像,并采用旋转评估跟踪天空恒星目标。

这种将脉冲转为图像或时间面的方法,并未充分挖掘异步脉冲信号的视觉时空特性,因此,也有一

些方法是直接在脉冲流的时空域上进行检测跟踪。Hinz 等人(2017)对脉冲信号进行在线聚类,应用于交通场景的车辆目标跟踪。Orchard 等人(2013)提出多核并行聚类的目标跟踪方法,适用于方向和尺度变化的多目标高速运动。此外,为解决高速运动多目标遮挡的问题,Benosman 等人(2012)还采用了立体视觉匹配聚类跟踪算法。此类方法在简单视觉场景中能有效跟踪多个目标,但在聚类重叠区域存在鲁棒性差的缺点。

除了对脉冲进行聚类的方式外,还有采用目标分割的方法,将脉冲流分成若干个区域并提取感兴趣目标的精确位置。Schraml 和 Belbachir(2010)将脉冲流进行时空聚类并构建立体视觉系统,可对多人进行实时分割与行为分析。Chen 等人(2018)采用“基于检测的跟踪(tracking-by-detection)”的模式在脉冲流上测试了 Mean shift、DBSCAN 和 Wave-Cluster 共 3 种聚类算法,并验证了 4 种多目标跟踪算法,最终搭建了对交通场景的车辆目标实时分割与跟踪的系统。Barranco 等人(2018)提出一种均值漂移的聚类方法,在操控机器人上实时对多目标进行分割与跟踪。Stoffregen 等人(2019)对脉冲流进行时空聚类来分割运动目标及背景,尤其适用于高速运动及极端光照条件的运动目标分割及背景建模。

2 国内研究进展

2.1 脉冲视觉模型

在传统图像和视频的成像模型中,对于相机视锥范围内高速射入的大量光子(如图 1(a)),相机按照预定帧率 f ,分别在时刻 t_1, t_2, t_3, \dots (时间间隔均为 $\frac{1}{f}$ s) 进行图像采集。图像采集时所有感光单元同时进行持续时长为 Δt (即曝光时间, $\Delta t < 1/f$) 的光子捕获,然后分别将累积的光强记录下来,按照感光单元空间排布而形成的光强分布就是图像,图像按等时间间隔排列的序列就是视频。从全光函数的角度看(Adelson 和 Bergen, 1991),图像记录的不是 t_1, t_2, t_3, \dots 时刻的状态,而是持续 Δt 的物理过程的累积。对于视频来说,采集每帧图像所用的累积时间 Δt 小于等于两帧图像之间的时间间隔 $1/f$,这意味着, $1/f - \Delta t$ 这段时间内的信息彻底

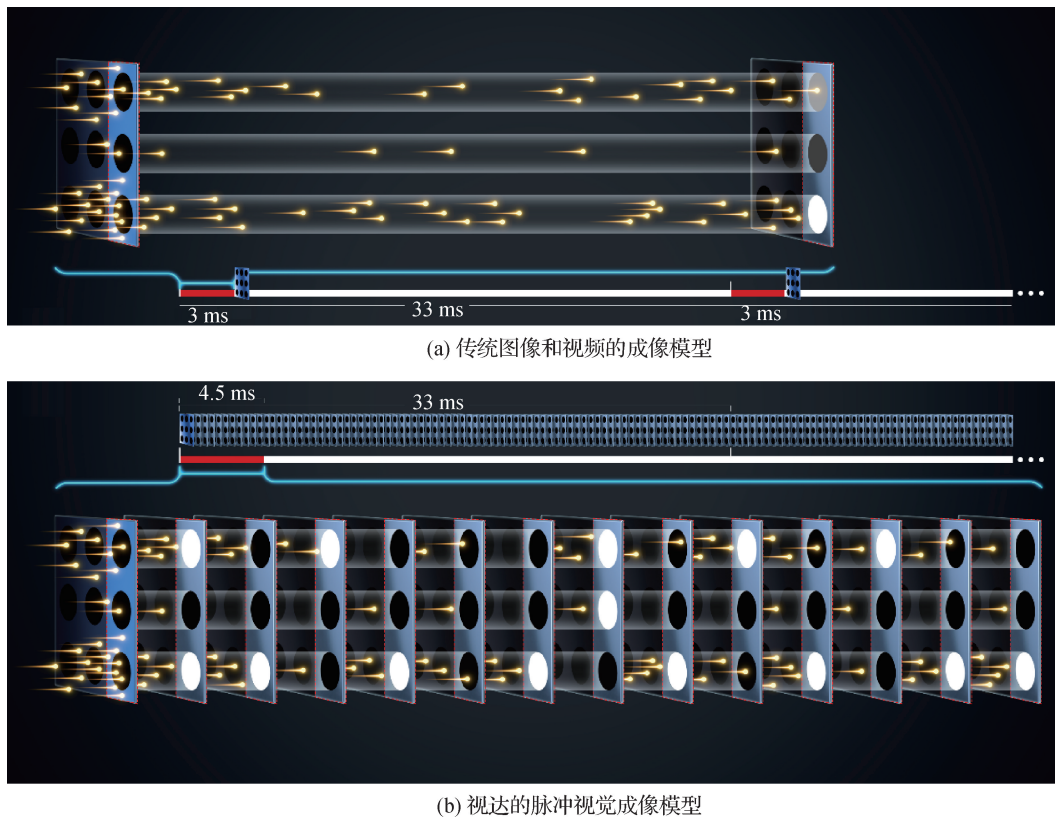


图 1 图像与视达在视觉信息表示方面的整体比较

Fig. 1 Comparison of imaging mechanism between conventional image/video and Vidar

((a) imaging mechanism of conventional image/video; (b) spiking imaging mechanism of Vidar)

丢失了, Δt 时段内的运动过程也被“压扁”到图像上而丢失, 因此视频的时域采样率不是对物理过程的一个完整采样。

图像和视频的同步曝光且曝光时间相同的设计已经成为数码相机捕捉快速变化的光子世界的最大障碍, 这样的设计是没有必要的。北京大学 2015 年 (2016 年申请专利, 相继获得中美日韩欧专利: 中 ZL201610045011, 美 10523972, 日 6754434, 韩 520000216573, 欧 16595EPPC) 提出了一种新的视觉模型——脉冲视觉模型 (spike vision model), 这种模型对光信息的表达称为视达 (Vidar, 以代替视频 Video)。

视达在空域采样方面与传统图像和视频并无二致, 脉冲相机可采用与普通相机完全相同的感光器件, 也就是大家熟知的 CMOS 和 CCD, 也采用空间排布的点阵表达空域信息 (如图 1(b))。视达和视频的根本不同是采用了新的时域采样方式, 所有感光器件并不同步按相同的曝光时长进行曝光, 而是持续捕获光子, 当收集的光强超过约定阈值就产生

一个脉冲, 这个脉冲及形成这个脉冲所持续的时长称为视元。每个感光器件产生的视元按照时间次序排成序列, 最简单的形式就是一个比特流, 1 表示在该时刻出现了一个脉冲, 0 表示处于累积状态, 比特 1 和它之前另一个 1 之间的所有 0 构成一个数码视元。每个感光器件产生一个脉冲流, 所有感光器件产生的脉冲流按照器件的空间排布组成脉冲流阵列, 这就是视达。

视达比视频的突出优势在于有效保留了各个采样位置光的时域变化过程。采用单光子敏感器件, 一个光子就能激发一个脉冲, 这时视达相机记录的就是精确完整的物理事实。普通感光器件捕获一组光子才能激发一个脉冲, 是对物理事实的一种粗糙表示, 但物理过程的时间关系仍然得到了最大程度的保留, 而不像视频那样通过人为规定将时间关系整齐划一地强制为数十 Hz。事实上, 今天广泛普及的 CMOS 感光器件, 时间灵敏度已经达到数十 ns, 采用视达这种新模型, 就可以实现千万 Hz 的高速时域采样, 记录极快的物理过程。北京大学开发的

第一款芯片设定的采样频率是4万 Hz,也就是比人类视觉和普通相机快1 000倍,已经可以清晰拍摄时速350 km的高铁和转速达每分钟7 200转的硬盘。视达记录了一定空间范围内各个位置光的精细变化过程,其物理意义十分清晰明确,因此,用视达可以生成传统的图像和视频。事实上,对于任意指定时刻,可以从覆盖该时刻的视元估计该位置的相对光强,更可以参考之前的和空间上邻近的更多视元估计出更精细的光强,从而得到任意时刻的精细图像。视达这种蕴含了任意时刻影像的能力称为全时成像(fulltime imaging),或者说连续成像(continuous imaging)。

2.2 脉冲相机模拟器

开发脉冲相机的北京大学黄铁军团队也是最早开发脉冲相机模拟器的研究队伍之一,该模拟器主要应用在光流估计、多尺度采样、脉冲检测跟踪算法和合成场景生成等方面。

在模拟器整体框架方面,北京大学马雷等人(2022a)提出了针对脉冲相机的模拟器,模拟了脉冲相机工作机理,进而帮助评估基于脉冲相机的检

测算法效能。通过紧密地结合渲染引擎与采样过程,可以搭建自己需要的场景并生成对应的脉冲数据,其中几个主要部分是传感器轨迹生成、渲染引擎和采样过程。传感器轨迹描述传感器(虚拟摄像机)的运动,默认情况下模拟器使用随机点拟合的样条曲线作为传感器轨迹,该模拟器也支持用户自定义轨迹输入。渲染引擎真实模拟了现实世界中采样的物理过程,可用于生成场景中虚拟摄像机采样的图像序列,该模拟器使用nvisii(Morriscal等,2021)和Blender作为引擎。采样过程包括渲染引擎和摄像机模拟之间的交互,更详细地说,模块首先将传感器轨迹发送到渲染引擎,在生成虚拟图像后将其转换为脉冲流,转换流程依据脉冲相机原理,即将图像像素强度视为光照强度,对光照强度按照时序进行累加作为累积值,在累积值到达预设阈值后,释放脉冲数据。

为了方便地生成脉冲数据,脉冲相机模拟器提供了一些方便的附加功能,通过使用这些附加功能,用户可以根据自身需求获得相关的数据。图2展示了脉冲相机模拟器的一些渲染结果。

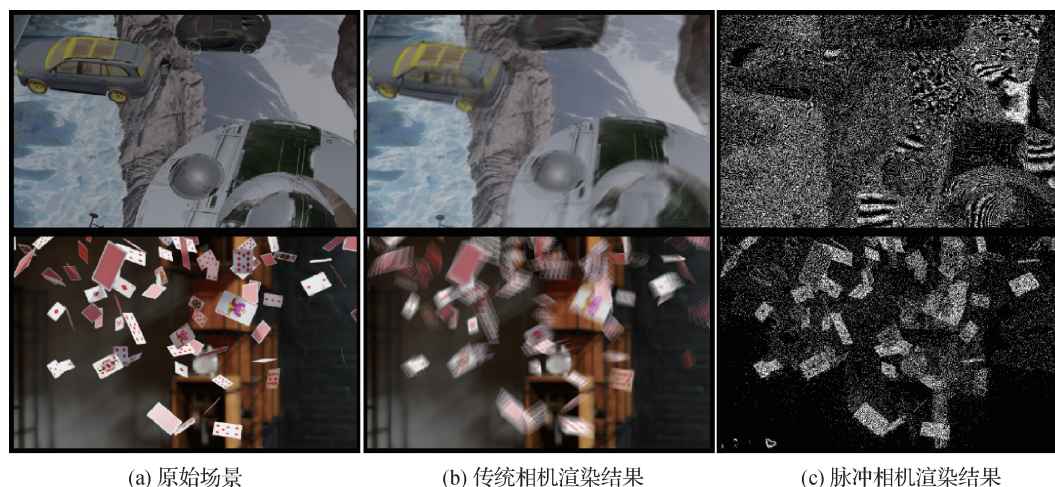


图2 脉冲相机模拟器的渲染结果

Fig. 2 Spike camera simulation results((a) corresponding virtual scenes;

(b) rendering results of conventional photorealistic rendering algorithm; (c) rendering results of spike camera simulator)

2.3 脉冲视觉芯片与相机

北京大学黄铁军团队从2015年起在脑认知与类脑计算专项中完成了视网膜仿真工作,初步积累了仿视网膜脉冲形态传感器的技术原理。2017年,该团队以大脑初级视觉系统结构仿真结果和

功能模拟机理为基础,研制完成高灵敏和高保真的脉冲视觉芯片及高速智能摄像系统,大幅度提高普通摄像头捕捉高速目标并进行精细识别的能力。

脉冲芯片版图如图3所示。该芯片采用异步像

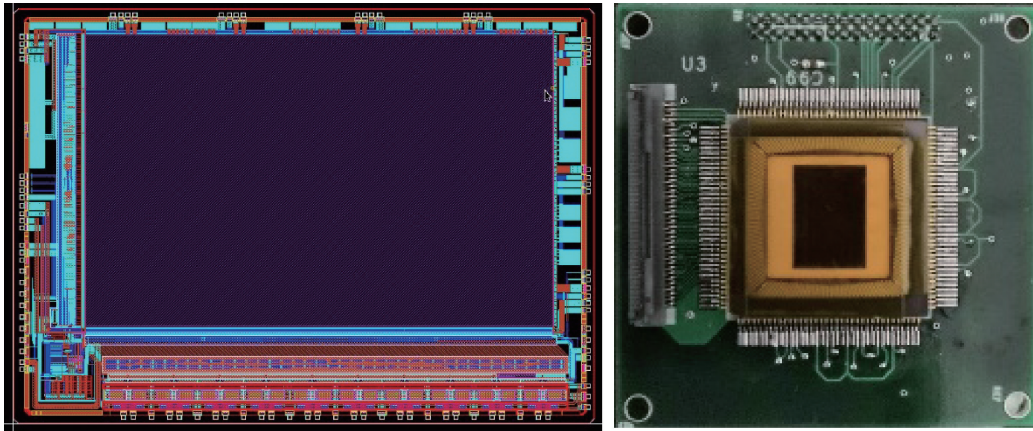


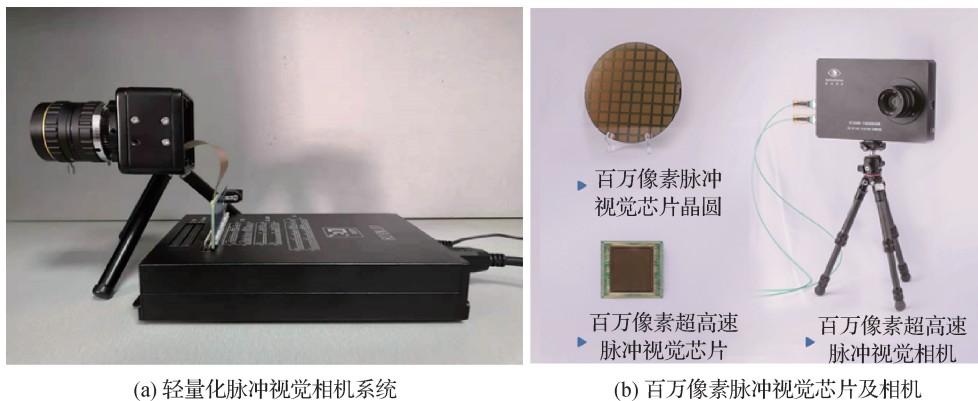
图3 脉冲视觉芯片版图及芯片图

Fig. 3 Layout and packaged chip of Vidar

素触发结构,分辨率为 400×250 像素,达到了 $40\,000$ Hz 的采样频率,动态范围超过 100 dB,结合高速传感模块和实时视觉计算模块组成的轻量化脉冲视觉相机系统,可实现对超高速目标的采集。

2021年黄铁军团队研制成功百万像素脉冲视觉芯片,分辨率达到 100 万像素,采样频率达到 $40\,000$ Hz,进一步缩小了单个像素的尺寸,提高了脉冲视觉采集的清晰度,如图4所示。该团队研

制的脉冲视觉芯片与相机可对高速运动物体实现清晰捕捉。在高铁示范应用中,能够对距离 2 m、相对时速达到 700 km/h (0.5 马赫) 的高铁实时检测跟踪,同时可以清晰分辨高铁在高速运动中形成的空气激波;在科研观测示范应用中,能够对 6 马赫的静风洞进行长期实时观测,可以对高速空气运动中产生的激波变化过程完成记录和实时检测处理。



(a) 轻量化脉冲视觉相机系统

(b) 百万像素脉冲视觉芯片及相机

图4 百万像素脉冲相机

Fig. 4 Mega-pixel Vidar camera((a) lightweight Vidar camera;(b) mega-pixel Vidar chip and camera)

单光子相机(single-photon camera)是面向极低光照情形提出的一种高灵敏度相机,通过采用单光子雪崩二极管(single photon avalanche diodes, SPAD),可以记录单个光子的到达,通过对光子进行计数,可以实现低光照或高动态范围条件下的更高质量成像。单光子相机可以视为一种特殊的脉冲相机,一个脉冲对应一个光子。针对脉冲相机开发的相关算法,也适用于单光子相机。

2.4 脉冲影像重建

国内学者近年来对脉冲相机等脉冲视觉传感器的影像重建进行了研究。对于脉冲相机而言,它以脉冲阵列的形式对动态视觉过程进行持续记录,在达到阈值时异步地发放脉冲,抛弃了曝光时间窗的概念,突破了传统相机的局限性。基于脉冲间隔法(texture from interval, TFI)和基于时间窗平均法(texture from window, TFW)是两种基础的脉冲相机

重建算法(Zhu等,2019b)。基于脉冲间隔的方法利用脉冲间隔随着光强增加而减小这一特性,反映的是极短时间内的瞬时光强。到达光子波动性对TFI的影响较大,通常呈现出明显噪声以及信号的时域波动。基于时间窗平均的重建方法,反映的是相对更长时段内的平均光照强度,对于静态场景产生的图像通常更稳定,但当物体高速运动时,该方法会引入显著的运动模糊。针对上述问题,Zhao等人(2020a)提出了对脉冲数据沿着运动轨迹进行时域滤波的方法,在提高图像信噪比的同时能够不引起运动模糊。Zhu等人(2020)提出了视网膜启发的脉冲神经网络模型,来解决脉冲相机的图像重建问题。该模型可以较好地提升图像的对比度。Zheng等人(2021)利用短时可塑性机制(short-term plasticity mechanism, STP)来解决脉冲相机的重建问题。该算法可以更加高效地解决脉冲相机重建问题。Zhao等人(2021a)考虑到当前脉冲相机的空间分辨率较为有限,基于脉冲相机的成像原理,提出了一种运动引导的脉冲相机超分辨率成像算法(motion-guided spike super resolution, MGSR)。该算法可以利用场景中物体与相机间的相对运动,从低分辨率的脉冲流中重建出高分辨率的重建图像序列。此外,受启发于深度学习强大的分析与建模能力,一些基于深度学习的脉冲相机影像重建框架涌现:例如Zhao等人(2021b)基于脉冲序列的特性设计了一个端到端的脉冲相机重建网络(spike to image network, Spk2ImgNet);Zhu等人(2021)提出将事件相机数据与脉冲相机数据相结合,从而进一步提升重建图像的动态范围。

国内学者近年来也对脉冲视觉传感器的光流估计进行了一些研究。Zhao等人(2020b)采用运动物体对应点在短时内的脉冲间距一致性假设,基于脉冲的运动轨迹对当前的自然光学场景进行表征,并在此基础上进行了脉冲数据流的光流估计。Hu等人(2021a)提出了SPIFT(spikingly flying things)与PHM(photo-realistic highspeed motion)数据集,用于脉冲相机光流估计的训练与评估,并且提出了用于脉冲流光流估计的神经网络SCFlow。该网络首先基于先验光流对脉冲流进行沿时间方向的卷积,然后基于类PWC-Net(image pyramid, warping and cost volume)(Sun等,2018)结构进行特征金字塔的编码—解码,并在解码过程中使用特征变换,实现了光流的渐进估计。

针对单光子相机的影像重建,Gyongy等人(2018)提出基于运动补偿来重新排列光子的方法,从而降低重建图像的模糊效应。Ma等人(2020)设计了一个Quanta burst photography框架来高效地对齐和融合二值图像,从而重建出高信噪比、高动态范围和低模糊的图像。Chi等人(2020)提出了一个学生教师框架(student-teacher framework)来解决噪声和高速运动之间的矛盾。Ingle等人(2021)通过分析光子到达的时间特性,提出了一种新的单光子相机成像算法,该算法极大地提升了重建图像的动态范围。Seets等人(2021)针对单光子相机提出了一种运动模糊去除算法,来提升其在动态场景中的重建性能。

2.5 脉冲计算摄像

北京大学团队是国内最早利用神经形态视觉信号研究计算摄像问题的研究队伍之一,其研究的主要应用体现在利用脉冲信号的高动态特性克服单幅图像高动态范围恢复的非适定性。

真实世界中有着100 000 : 1的动态范围(dynamic range, DR),但是普通相机拍摄的照片所能覆盖的动态范围非常有限,往往只有1 000 : 1,因此在采用传统摄像设备对动态范围较高的真实世界场景进行拍摄时,经常会由于过曝光、欠曝光导致图像亮处过饱和、暗处无细节,使得所拍摄的图像质量显著下降。这是因为传统相机所能拍摄的动态范围受到内部硬件固有的限制,从而无法完整记录真实场景的动态范围。

如何突破传统相机的高动态范围性能极限,是计算摄像的热点研究问题。对于高层计算机视觉任务来说,低动态范围(low dynamic range, LDR)的图像会使得特征识别准确率下降,使得系统在具有剧烈明暗变化的场景中无法正常工作。因此,高动态范围(high dynamic range, HDR)成像的有效性以及稳定性将直接影响高层计算机视觉任务的效率。脉冲相机的每个像素输出信号彼此独立,每个像素的脉冲信号按照时间先后顺序形成“脉冲序列”。由于其积分采样原理,不同强度的光强会产生稀疏程度不同的脉冲流,因此对场景中亮处和暗处的信息都能得到较好的保留,这为实现人类视觉友好的高动态范围成像提供了可能。

1)融合神经形态视觉相机和传统相机的高动态范围成像技术。利用神经形态视觉相机具有捕捉

高动态范围场景信息的能力, 可以将神经形态相机输出的脉冲流重建为一张包含高动态范围信息的灰度图。但是由于其空间分辨率较低 (400×250 像素), 且不能记录颜色信息, 因此将由脉冲重建的灰度图与传统基于帧的彩色相机拍摄的高分辨率彩色照片进行融合, 可以得到包含高动态范围信息的高分辨率彩色图像。Han 等人 (2020) 提出了一个包含 4 个步骤的融合流程, 这 4 个步骤分别是色彩空间转换、灰度图上采样、灰度空间融合和色彩信息补

偿, 每个步骤通过神经网络中相应设计的模块来实现。Han 等人 (2020) 还搭建了一个混合相机系统用于拍摄真实数据, 包含一部型号为 Point Grey Chameleon 3 的普通彩色工业相机和一部 Vidar 神经形态视觉相机, 二者通过分光镜获得相同的视场。通过将两部相机输出数据的融合, 利用上述算法能够恢复出工业相机缺失的高动态范围信息和脉冲相机缺失的色彩和高分辨率信息。该混合相机系统的示意图如图 5 所示。

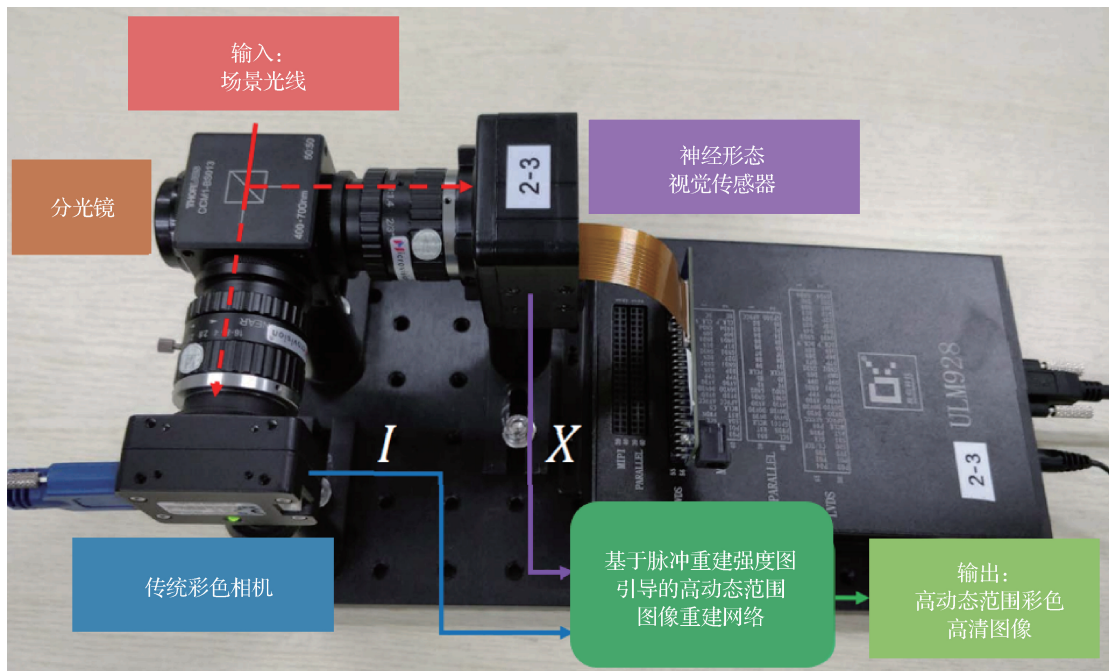


图 5 神经形态视觉相机和传统彩色相机的混合成像系统

Fig. 5 Hybrid imaging system fusing a neuromorphic camera and a conventional RGB camera

2) 基于神经形态视觉相机构造余数成像模型的高动态范围技术。Zhao 等人 (2015) 首次提出了余数成像系统 (modulo camera system) 这个概念, 可以获得理论上动态范围无上限 (不会饱和) 的单张图高动态范围成像。与传统摄像设备所记录的图像帧 (经过量化和动态范围裁剪) 不同, 余数成像系统所记录的是场景辐射强度对相机量化比特上限的余数值 (称为余数图像, modulo image), 并提出算法从余数图像来重建高动态范围图像。然而, 该原型采用数字像素焦平面阵列 (digital-pixel focal plane array, DFPA) 来搭建余数成像系统, 且采用基于马尔可夫随机场 (Markov random field, MRF) 的优化算法来完成余数图像的高动态范围重建, 由于其算法在非余数区域存在误解, 所以对于局部强对比度区域

以及彩色图像 RGB 三通道的颜色对齐处理欠佳。

为克服 Zhao 等人 (2015) 方法的缺陷, Zhou 等人 (2020) 对余数成像模型进行了改进, 其采用积分型神经形态视觉相机来构建系统原型, 并提出一种基于深度学习的方法来提高余数图像的高动态范围重建的算法鲁棒性, 为单帧 8 bit 高动态范围成像提供原理和性能上达到新高度的解决方案。与传统图像传感器的成像原理不同, 积分型神经形态视觉相机的感光器件捕获光子能量达到约定阈值而产生脉冲。此方法将脉冲相机配置为余数采集模式, 即设置一个脉冲数目门限值, 每个像素位置所收集到的脉冲数目达到此门限值则立刻清零重置, 最终得到由脉冲所积累的图像作为余数图像。不同于 Zhao 等人 (2015) 所采用的整数优化模型, 此方法采用概

率模型对余数图像的形成原理进行分析,将高动态范围图像拆分为余数图像和多个二元翻转掩膜(binary rollover mask)的求和,并建立余数图像和每一个二元翻转掩膜之间的概率关系,最终将余数图像到高动态范围图像的恢复过程当做一个迭代的二元标记问题来处理。

2.6 脉冲视觉算法

国内基于神经形态视觉传感器的目标检测跟踪方法大多数也是基于事件相机。Huang 等人(2018)提出了一种事件指导的支持向量机模型(event-guided support vector machine, ESVM),通过向传统的支持向量机模型中引入事件信息,提高对高速运动物体的检测跟踪性能。其中事件指导的方式有两种,一种是事件位置导向搜索定位,另一种是事件灰度导向采样补充。这两种方法都是采用其团队自主研发的 CeleX 相机产生的脉冲流,其可以记录灰度的变化信息。脉冲信号重构插帧来提高图像序列帧率,同时脉冲流可指导运动区域用于高速目标跟踪。Chen 等人(2019)提出了一种基于事件的检测跟踪方法(event-based tracking-by-detection, ETD),可获得运动物体所在位置的候选框,进而实现连续的跟踪。采用自适应时间面(adaptive time-surface with linear time decay, ATSLTD)将事件相机产生的脉冲流转为图像序列,然后用检测与跟踪一体化的方法对目标进行跟踪。除了将脉冲流转化为图像的方式,Chen 等人(2020)还提出使用端到端的视网膜运动表达神经网络(retinal motion regression network, RM-RNet)处理事件流数据,获得事件流中各个运动物体的 5-DoF(degree of freedom)物体运动参数。其提出的通过线性时间衰减实现同步时间平面的表达(time-surface with linear time decay, TSLTD),可将异步事件流的时空信息有效编码为具有清晰运动模式的 TSLTD 平面,进而实现在极端光照环境下的高速运动物体的检测跟踪。然而,这些方法本质上还是采用将事件流投影至一个时间平面再做后续处理的方式,没有发挥事件相机低延迟的优点。Yang 等人(2019)提出了一种结合脉冲神经网络和传统神经网络的模型(DashNet),使用脉冲神经网络处理事件传感器产生的脉冲流数据,使用人工神经网络处理 APS 产生的视频帧。这种新型的数据融合和网络融合的方式或许会成为将来处理复杂视觉任务的计算范式。

2.7 脉冲视觉应用

脉冲视觉主要应用在以下几个方面:

1) 工业领域。早期工作利用 ATIS 的高时间分辨率对工业微型机床的高速运动的抓手进行跟踪与系统反馈(Ni 等,2012)、采用 DVS 对高速运转平台上的工业部件进行实时识别(Belbachir 等,2011)。来自瑞士苏黎世大学和苏黎世联邦理工学院的 iniVation 是一家基于事件的动态视觉传感器(DVS)技术的公司,其研发的 Dynamic Vision 平台可进行高速 3D 基础设施扫描以进行预测性维护、高速生产检查等工业视觉应用。法国公司 Prophesee 最近推出了全面的评估套件(evaluation kit, EVK),帮助机器视觉系统开发人员全面评估基于事件的视觉传感器的性能。

2) 民用领域。DVS 公司与手机制造商合作开发一种兼具基于事件像素和正常像素的传感器,用户可以拍摄普通照片,但对于特定场景,用户可以采用基于事件的像素。韩国三星公司结合传统相机与 DVS 相机的特点,降低曝光过程中产生的运动模糊,快速感知动态变化,相关成果应用于手机的触屏唤醒。中国 CelePixel 公司开发了基于 CeleX V 的驾驶员疲劳检测系统(Guo 等,2016);国内北京大学田永鸿教授团队(Li 等,2019)将 DVS 和 APS 结合用于自动驾驶的目标联合检测,其性能在高速运动场景及极端光照条件下均有显著性的提升。

3) 国防领域。Rodriguez 等人(2021)将 DVS 应用到港口监控,实现对帆船、摩托艇和大型船只的运动监控,Falanga 等人(2020)将 DVS 作为无人机的视觉导航系统,相比传统相机在高速运动场景与极端光照下有更好的跟踪能力。

3 国内外研究进展比较

替代传统视频相机的早期工作主要由国外研究机构开展,其中最具代表性的是美国加州理工大学的早期探索和瑞士苏黎世大学研制的 DVS 事件相机,这些工作开辟了一个重要的研究方向,主要包括 AER 传输协议、DVS 系列相机及处理算法。与传统相机相比,基于 AER 的事件相机具备如下优势:可以实现连续记录、解耦快门速度和帧率以及高速记录物体变化。与此同时,基于 AER 的相机也存在一个较大的问题,即由于不携带绝对光强信号,因而无

法重建纹理图像。

北京大学黄铁军团队借鉴灵长类动物视网膜中央凹编码机理,直接使用脉冲流表达光子流,能够有效保留物理光流的时序和更为准确的时间信息。脉冲相机具有以下3方面的新特性:1)全时成像(full-time imaging),每个视元的持续时间和光强成反比,据此可以得到该像素位置任意时刻的光强,从视达中得到任意时刻图像,实现全时成像;2)自由动态范围成像(free dynamic range, FDR),如果时间窗口扩大到多个视元,则可以得到更高动态范围的图像,称为自由动态范围成像,单幅图像内实现不同亮度的目标均清晰;3)宽频成像(broad frequency imaging),利用相机高频能力,获得各像素的振动频率,从而区分不同目标(螺旋桨、抖动、海浪、云雾和建筑等)。脉冲相机具备了视频相机和事件相机的优点,而且有新的特点,因此有望全面替代两者。

在计算摄像方面,国外最具代表性的是瑞士苏黎世大学的传感器课题组以及机器感知课题组,其重点关注差分型事件相机。近年来,国内学者对差分型事件相机和积分型脉冲相机都有深入研究,并贡献了大量研究成果,华为等国内企业也不断加大在脉冲计算摄像算法理论研究方面的投入,与国内外研究机构合作,产出一系列高水平研究成果。

在视觉算法方面,由于对脉冲流数据处理方法尚处于摸索阶段,国内外所提出的方法大多数是在“脉冲流—时间平面”转化上进行方法创新,进而在转化所得的时间平面上进行运动物体的时空信息挖掘。这种处理模式没有发挥神经形态视觉传感器低延迟的优势,处理方法上仍需改进。近几年神经形态芯片技术的发展,清华大学施路平教授团队提出的融合脉冲神经网络和人工神经网络的方式,可直接处理脉冲流数据和图片数据,利用其团队研发的天极芯片,在保证较高的检测跟踪性能下,发挥事件流低延迟的优势,但事件相机与传统相机数据的时间对齐仍是数据融合时的一个难题,而脉冲相机具有脉冲流时空完备性优势,使得其与事件相机相比在高速成像方面具有不可替代的优势,发展潜力巨大。

4 发展趋势与展望

在脉冲视觉建模与相机模拟器方面,将继续开

展采样模式的探究,进一步探究多尺度采样的信号表征能力与稀疏编码能力。实际采样中噪声对脉冲相机的干扰非常明显,因此现阶段模拟器生成的脉冲数据和实际脉冲数据还存在差异,需要进一步探究采样噪声的精细模拟方法和理论。此外,除了光流与检测跟踪,其他视觉任务如深度图估计等也有很大创新机会。

在脉冲摄像器件方面,下一阶段的脉冲相机将以提升分辨率为主要发展方向,逐步弥补当前原型机较低的分辨率与相对较差的成像质量。与此同时,脉冲摄像传感器与传统视觉传感器的融合成像系统,作为一种兼具优势、取长补短的成像器件,也将是未来研究的重要方向之一。

在脉冲视觉芯片与系统方面,后续的研究将围绕以下几个方面展开:1)积分型像素与差分型像素的融合。将传感器记录光强绝对信息与记录光强变化信息的能力进行融合,面向不同应用场景提供更多纹理细节和更多变化信息的选择,同时在传输的信息和带宽之间实现平衡。2)脉冲原理传感器与传统视觉传感器的融合。充分发挥脉冲原理传感器与传统视觉传感器的优势,实现面向不同应用场景时提供更多时域信息和更多空域信息的选择,实现灵活高效的采样。3)脉冲感知芯片与脉冲处理芯片的融合。通过采用系统级封装或芯片堆叠的工艺形式,将脉冲感知芯片与脉冲处理芯片整合,实现高效数据传输和实时反馈调节,从而大幅提高传感器的感知能力。

在脉冲视觉算法方面,相关研究仍将保持以深度学习方法为主的发展态势,更多的新模型、新方法将被引入,以提升相关算法的输出准确性、实时性,满足更多现实任务的需求。此外,作为一种新兴的神经网络结构,脉冲神经网络是处理脉冲信号的理想手段,随着相关基础理论的研究推进,可以预见,脉冲神经网络在脉冲视觉领域更会大有作为。

致谢 本文由中国图象图形学学会类脑视觉专委会组织撰写,该专委会更多详情请见链接:<http://www.csig.org.cn/detail/3185>。

参考文献(References)

- Adelson E H and Bergen J R. 1991. The plenoptic function and the elements of early vision//Landy M and Movshon J A, eds. Computa-

- tional Models of Visual Processing. Cambridge: MIT Press; 3-20 [DOI: 10.7551/mitpress/2002.003.0004]
- Andreopoulos A, Kashyap H J, Nayak T K, Amir A and Flickner M D. 2018. A low power, high throughput, fully event-based stereo system//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE; 7532-7542 [DOI: 10.1109/CVPR.2018.00786]
- Baldwin R W, Almatrafi M, Asari V and Hirakawa K. 2020. Event probability mask (EPM) and event denoising convolutional neural network (EDnCNN) for neuromorphic cameras//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE; 1698-1707 [DOI: 10.1109/CVPR42600.2020.00177]
- Bardow P, Davison A J and Leutenegger S. 2016. Simultaneous optical flow and intensity estimation from an event camera//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE; 884-892 [DOI: 10.1109/CVPR.2016.102]
- Barlow H B. 1953. Summation and inhibition in the frog's retina. The Journal of Physiology, 119 (1): 69-88 [DOI: 10.1113/jphysiol.1953.sp004829]
- Barranco F, Fermuller C and Ros E. 2018. Real-time clustering and multi-target tracking using event-based sensors//Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid, Spain: IEEE; 5764-5769 [DOI: 10.1109/IROS.2018.8593380]
- Belbachir A N, Hofstätter M, Litzberger M and Schön P. 2011. High-speed embedded-object analysis using a dual-line timed-address-event temporal-contrast vision sensor. IEEE Transactions on Industrial Electronics, 58 (3): 770-783 [DOI: 10.1109/TIE.2010.2095390]
- Benosman R, Ieng S H, Clercq C, Bartolozzi C and Srinivasan M. 2012. Asynchronous frameless event-based optical flow. Neural Networks, 27: 32-37 [DOI: 10.1016/j.neunet.2011.11.001]
- Boahen K A. 2000. Point-to-point connectivity between neuromorphic chips using address events. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 47 (5): 416-434 [DOI: 10.1109/82.842110]
- Brandli C, Berner R, Yang M H, Liu S C and Delbruck T. 2014. A 240 × 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits, 49 (10): 2333-2341 [DOI: 10.1109/JSSC.2014.2342715]
- Chen G, Cao H, Aafaque M, Chen J N, Ye C B, Röhrbein F, Conrads J, Chen K, Bing Z S, Liu X B, Hinz G, Stechele W and Knoll A. 2018. Neuromorphic vision based multivehicle detection and tracking for intelligent transportation system. Journal of Advanced Transportation, 2018; #4815383 [DOI: 10.1155/2018/4815383]
- Chen H S, Suter D, Wu Q Q and Wang H Z. 2020. End-to-end learning of object motion estimation from retinal events for event-based object tracking//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI; 10534-10541 [DOI: 10.1609/aaai.v34i07.6625]
- Chen H S, Wu Q Q, Liang Y J, Gao X B and Wang H Z. 2019. Asynchronous tracking-by-detection on adaptive time surfaces for event-based object tracking//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM; 473-481 [DOI: 10.1145/3343031.3350975]
- Chi Y H, Gnanasambandam A, Koltun V and Chan S H. 2020. Dynamic low-light imaging with quanta image sensors//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer; 122-138
- Chin T J, Bagchi S, Eriksson A and van Schaik A. 2019. Star tracking using an event camera//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: IEEE; 1646-1655 [DOI: 10.1109/CVPRW.2019.00208]
- Dosovitskiy A, Ros G, Codevilla F, López A and Koltun V. 2017. CARLA: an open urban driving simulator//Proceedings of the 1st Annual Conference on Robot Learning. Mountain View, USA: PMLR; 1-16
- Falanga D, Kleber K and Scaramuzza D. 2020. Dynamic obstacle avoidance for quadrotors with event cameras. Science Robotics, 5 (40): #eaa9712 [DOI: 10.1126/scirobotics.aa9712]
- Gallego G, Rebecq H and Scaramuzza D. 2018. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE; 3867-3876 [DOI: 10.1109/CVPR.2018.00407]
- Gehrig M, Millhäusler M, Gehrig D and Scaramuzza D. 2021. E-RAFT: dense optical flow from event cameras [EB/OL]. [2022-02-15]. <https://arxiv.org/pdf/2108.10552.pdf>
- Guo M H, Ding R X and Chen S S. 2016. Live demonstration: a dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand//2016 IEEE International Symposium on Circuits and System. Montreal, Canada: IEEE; 456 [DOI: 10.1109/ISCAS.2016.7527274]
- Gyongy I, Dutton N A W and Henderson R K. 2018. Single-photon tracking for high-speed vision. Sensors, 18 (2): #323 [DOI: 10.3390/s18020323]
- Han J, Zhou C, Duan P Q, Tang Y H, Xu C, Xu C, Huang T J and Shi B X. 2020. Neuromorphic camera guided high dynamic range imaging//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE; 1727-1736 [DOI: 10.1109/CVPR42600.2020.00180]
- Hinz G, Chen G, Aafaque M, Röhrbein F, Conrads J, Bing Z S, Qu Z N, Stechele W and Knoll A. 2017. Online multi-object tracking-by-clustering for intelligent transportation system with neuromorphic vision sensor//Proceedings of the 40th Annual German Conference on AI. Dortmund, Germany: Springer; 142-154 [DOI: 10.1007/978-3-319-67190-1_11]
- Hu L W, Zhao R, Ding Z L, Ma L, Shi B X, Xiong R Q and Huang T J. 2021a. Optical flow estimation for spiking camera [EB/OL].

- [2022-02-15]. <https://arxiv.org/pdf/2110.03916.pdf>
- Hu Y H, Liu S C and Delbruck T. 2021b. V2e: from video frames to realistic DVS events//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA; IEEE; 1312-1321 [DOI: 10.1109/CVPRW53098.2021.00144]
- Huang J, Wang S Z, Guo M H and Chen S S. 2018. Event-guided structured output tracking of fast-moving objects using a CeleX sensor. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9): 2413-2417 [DOI: 10.1109/TCSVT.2018.2841516]
- Huang T J, Shi L P, Tang H J, Pan G, Chen Y J and Yu J Q. 2016. Research on multimedia technology 2015—advances and trend of brain-like computing. *Journal of Image and Graphics*, 21(11): 1411-1424 (黄铁军, 施路平, 唐华锦, 潘纲, 陈云霁, 于俊清. 2016. 多媒体技术研究: 2015——类脑计算的研究进展与发展趋势. *中国图象图形学报*, 21(11): 1411-1424) [DOI: 10.11834/jig.20161101]
- Huang T J, Yu Z F and Liu Y J. 2019. Brain-like machine: thought and architecture. *Journal of Computer Research and Development*, 56(6): 1135-1148 (黄铁军, 余肇飞, 刘怡俊. 2019. 类脑机的思想与体系结构综述. *计算机研究与发展*, 56(6): 1135-1148) [DOI: 10.7544/issn1000-1239.2019.20190240]
- Hubel D H and Wiesel T N. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3): 574-591 [DOI: 10.1113/jphysiol.1959.sp006308]
- Ingle A, Seets T, Buttafava M, Gupta S, Tosi A, Gupta M and Velten A. 2021. Passive inter-photon imaging//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE; 8581-8591 [DOI: 10.1109/CVPR46437.2021.00848]
- Jiang H Z, Sun D Q, Jampani V, Yang M H, Learned-Miller E and Kautz J. 2018. Super SloMo: high quality estimation of multiple intermediate frames for video interpolation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE; 9000-9008 [DOI: 10.1109/CVPR.2018.00938]
- Jing Y C, Yang Y D, Wang X C, Song M L and Tao D C. 2021. Turning frequency to resolution: video super-resolution via event cameras//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE; 7768-7777 [DOI: 10.1109/CVPR46437.2021.00768]
- Kim H, Handa A, Benosman R, Ieng S H and Davison A. 2008. Simultaneous mosaicing and tracking with an event camera//Proceedings British Machine Vision Conference 2014. Nottingham, UK; BMVA Press [DOI: 10.5244/C.28.26]
- Kramer J. 2002. An integrated optical transient sensor. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 49(9): 612-628 [DOI: 10.1109/TCSH.2002.807270]
- Lee C, Kosta A K, Zhu A Z, Chaney K, Daniilidis K and Roy K. 2020. Spike-FlowNet: event-based optical flow estimation with energy-efficient hybrid neural networks//Proceedings of the 16th European Conference on Computer Vision. Glasgow, Scotland; Springer; 366-382
- Lettvin J Y, Maturana H R, McCulloch W S and Pitts W H. 1959. What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47(11): 1940-1951 [DOI: 10.1109/JRPROC.1959.287207]
- Li J N, Dong S W, Yu Z F, Tian Y H and Huang T J. 2019. Event-based vision enhanced: a joint detection framework in autonomous driving//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Shanghai, China; IEEE; 1396-1401 [DOI: 10.1109/ICME.2019.00242]
- Li J N and Tian Y H. 2021. Recent advances in neuromorphic vision sensors: a survey. *Chinese Journal of Computers*, 44(6): 1258-1286 (李家宁, 田永鸿. 2021. 神经形态视觉传感器的研究进展及应用综述. *计算机学报*, 44(6): 1258-1286) [DOI: 10.11897/SP.J.1016.2021.01258]
- Leñero-Bardallo J A, Carmona-Galán R and Rodríguez-Vázquez Á. 2015. A high dynamic range image sensor with linear response based on asynchronous event detection//Proceedings of European Conference on Circuit Theory and Design. Trondheim, Norway; IEEE; 1-4 [DOI: 10.1109/ECCTD.2015.7300079]
- Leñero-Bardallo J A, Guerrero-Rodríguez J M, Carmona-Galan R and Rodríguez-Vázquez A. 2018. On the analysis and detection of flames with an asynchronous spiking image sensor. *IEEE Sensors Journal*, 18(16): 6588-6595 [DOI: 10.1109/JSEN.2018.2851063]
- Lichtsteiner P, Delbruck T and Kramer J. 2004. Improved ON/OFF temporally differentiating address-event imager//Proceedings of the 11th IEEE International Conference on Electronics, Circuits and Systems. Tel Aviv, Israel; IEEE; 211-214 [DOI: 10.1109/ICECS.2004.1399652]
- Ma S Z, Gupta S, Ulku A C, Bruschi C, Charbon E and Gupta M. 2020. Quanta burst photography. *ACM Transactions on Graphics*, 39(4): #79 [DOI: 10.1145/3386569.3392470]
- Mahowald M. 1992. VLSI Analogs of Neuronal Visual Processing: a Synthesis of form and Function. Pasadena; California Institute of Technology [DOI: 10.7907/4bdw-fg34]
- Mead C A and Mahowald M A. 1988. A silicon model of early visual processing. *Neural Networks*, 1(1): 91-97 [DOI: 10.1016/0893-6080(88)90024-X]
- Morrison N, Tremblay J, Lin Y, Tyree S, Birchfield S, Pascucci V and Wald I. 2021. Nvisii: a scriptable tool for photorealistic image generation [EB/OL]. [2022-02-15]. <https://arxiv.org/pdf/2105.13962.pdf>
- Mostafavi I S M, Choi J and Yoon K J. 2020. Learning to super resolve intensity images from events//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE; 2765-2773 [DOI: 10.1109/CVPR42600.2020.00284]
- Ni Z J, Bolopion A, Agnus J, Benosman R and Regnier S. 2012. Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics. *IEEE Transactions on Robotics*, 28(5): 1081-1089 [DOI: 10.1109/TRO.2012.2198930]

- Orchard G, Benosman R, Etienne-Cummings R and Thakor N V. 2013. A spiking neural network architecture for visual motion estimation// Proceedings of 2013 IEEE Biomedical Circuits and Systems Conference. Rotterdam, the Netherlands; IEEE: 298-301 [DOI: 10.1109/BioCAS.2013.6679698]
- Pan L Y, Scheerlinck C, Yu X, Hartley R, Liu M M and Dai Y C. 2019. Bringing a blurry frame alive at high frame-rate with an event camera//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 6813-6822 [DOI: 10.1109/CVPR.2019.00698]
- Paredes-Valles F, Scheper K Y W and de Croon G C H E. 2020. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation; from events to global motion perception. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8): 2051-2064 [DOI: 10.1109/TPAMI.2019.2903179]
- Posch C, Matolin D and Wohlgenannt R. 2010. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. IEEE Journal of Solid-State Circuits, 46(1), 259-275 [DOI: 10.1109/JSSC.2010.2085952]
- Pei J, Deng L, Song S, Zhao M, Zhang Y, Wu S, Wang G, Zou Z, Wu Z, He W and Chen F. 2019. Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature, 572(7767): 106-111 [DOI: 10.1038/s41586-019-1424-8]
- Qiu W C, Zhong F W, Zhang Y, Qiao S Y, Xiao Z H, Kim T S and Wang Y Z. 2017. UnrealCV: virtual worlds for computer vision// Proceedings of the 25th ACM international conference on Multimedia. Lisboa, the Portuguese Republic; ACM: 1221-1224 [DOI: 10.1145/3123266.3129396]
- Rodriguez J A, Mauger J, Parameswaran S, Zeller-Townson R and Cauble G. 2021. Dynamic vision sensor datasets in the maritime domain//Proceedings of SPIE 11870, Artificial Intelligence and Machine Learning in Defense Applications III. Spain; SPIE: 118700G [DOI: 10.1117/12.2600971]
- Rueckauer B and Delbruck T. 2016. Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor. Frontiers in Neuroscience, 10: #176 [DOI: 10.3389/fnins.2016.00176]
- Sang Y S, Li R H, Li Y Q, Wang Q W and Mao Y. 2019. Research on neuromorphic vision sensor and its applications. Chinese Journal on Internet of Things, 3(4): 63-71 (桑永胜, 李仁昊, 李耀仟, 王蔷薇, 毛耀. 2019. 神经形态视觉传感器及其应用研究. 物联网学报, 3(4): 63-71) [DOI: 10.11959/j.issn.2096-3750.2019.00133]
- Scheerlinck C, Barnes N and Mahony R. 2019. Continuous-time intensity estimation using event cameras//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia; Springer: 308-324 [DOI: 10.1007/978-3-030-20873-8_20]
- Son B, Suh Y, Kim S, Jung H, Kim J S, Shin C, Park K, Lee K, Park J, Woo J and Roh Y. 2017. 4.1 A 640 × 480 dynamic vision sensor with a 9 μm pixel and 300Meps address-event representation// Proceedings of the 2017 IEEE International Solid-State Circuits Conference. San Francisco, USA; IEEE: 66-67 [DOI: 10.1109/ISSCC.2017.7870263]
- Schraml S and Belbachir A N. 2010. A spatio-temporal clustering method using real-time motion analysis on event-based 3D vision//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, USA; IEEE: 57-63 [DOI: 10.1109/CVPRW.2010.5543810]
- Seets T, Ingle A, Laurenzis M and Velten A. 2021. Motion adaptive deblurring with single-photon cameras//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA; IEEE: 1944-1953 [DOI: 10.1109/WACV48630.2021.00199]
- Shah S, Dey D, Lovett C and Kapoor A. 2018. Airsim: high-fidelity visual and physical simulation for autonomous vehicles//Hutter M and Siegwart R, eds. Field and Service Robotics; Results of the 11th International Conference. Cham; Springer: 621-635 [DOI: 10.1007/978-3-319-67361-5_40]
- Sivilotti M A. 1991. Wiring Considerations in Analog VLSI Systems, with Application to Field-Programmable Networks. Pasadena, USA; California Institute of Technology [DOI: 10.7907/stj4-kh72]
- Stoffregen T, Gallego G, Drummond T, Kleeman L and Scaramuzza D. 2019. Event-based motion segmentation by motion compensation// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea(South); IEEE: 7243-7252 [DOI: 10.1109/ICCV.2019.00734]
- Stoffregen T, Scheerlinck C, Scaramuzza D, Drummond T, Barnes N, Kleeman L and Mahony R. 2020. Reducing the sim-to-real gap for event cameras//Proceedings of the 16th European Conference on Computer Vision. Glasgow, Scotland; Springer: 534-549 [DOI: 10.1007/978-3-030-58583-9_32]
- Sun D, Yang X, Liu M Y and Kautz J. 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Louisiana, USA; IEEE:8934-8943
- Tulyakov S, Fleuret F, Kiefel M, Gehler P and Hirsch M. 2019. Learning an event sequence embedding for dense event-based deep stereo// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea(South); IEEE: 1527-1537 [DOI: 10.1109/ICCV.2019.00161]
- Tulyakov S, Gehrig D, Georgoulis S, Erbach J, Gehrig M, Li Y Y and Scaramuzza D. 2021. Time Lens: Event-Based Video Frame Interpolation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual; IEEE:16155-16164
- Wang C, Chen F, Wen D S, Lei H, Song Z X and Zhao H F. 2021. Review on imaging and data processing of visual sensing. Journal of Image and Graphics, 26(6): 1450-1469 (王程, 陈峰, 汶德胜, 雷浩, 宋宗玺, 赵航芳. 2021. 视觉传感成像技术与数据处理进展. 中国图象图形学报, 26(6): 1450-1469) [DOI: 10.11834/jig.200852]
- Yang Z Y, Wu Y J, Wang G R, Yang Y K, Li G Q, Deng L, Zhu J and

- Shi L P. 2019. DashNet: a hybrid artificial and spiking neural network for high-speed object tracking [EB/OL]. [2022-02-15]. <https://arxiv.org/pdf/1909.12942.pdf>
- Ye C X, Mitrokhin A, Fermüller C, Yorke J A and Aloimonos Y. 2020. Unsupervised learning of dense optical flow, depth and egomotion with event-based sensors//Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA; IEEE: 5831-5838 [DOI: 10.1109/IROS45743.2020.9341224]
- Zhao H, Shi B X, Fernandez-Cull C, Yeung S K and Raskar R. 2015. Unbounded high dynamic range photography using a modulo camera//Proceedings of 2015 IEEE International Conference on Computational Photography. Houston, USA; IEEE: 1-10 [DOI: 10.1109/ICCPHOT.2015.7168378]
- Zhao J, Xie J Y, Xiong R Q, Zhang J, Yu Z F and Huang T J. 2021a. Super resolve dynamic scene from continuous spike streams//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada; IEEE: 2513-2522 [DOI: 10.1109/ICCV48922.2021.00253]
- Zhao J, Xiong R Q and Huang T J. 2020a. High-speed motion scene reconstruction for spike camera via motion aligned filtering//Proceedings of 2020 IEEE International Symposium on Circuits and Systems. Seville, Spain; IEEE: 1-5 [DOI: 10.1109/ISCAS45731.2020.9181055]
- Zhao J, Xiong R Q, Liu H F, Zhang J and Huang T J. 2021b. Spk2ImgNet: learning to reconstruct dynamic scene from continuous spike stream//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 11991-12000 [DOI: 10.1109/CVPR46437.2021.01182]
- Zhao J, Xiong R Q, Zhao R, Wang J, Ma S W and Huang T J. 2020b. Motion estimation for spike camera data sequence via spike interval analysis//Proceedings of 2020 IEEE International Conference on Visual Communications and Image Processing. Macau, China; IEEE: 371-374 [DOI: 10.1109/VCIP49819.2020.9301840]
- Zheng Y J, Zheng L X, Yu Z F, Shi B X, Tian Y H and Huang T J. 2021. High-speed image reconstruction through short-term plasticity for spiking cameras//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 6354-6363 [DOI: 10.1109/CVPR46437.2021.00629]
- Zhou C, Zhao H, Han J, Xu C, Xu C, Huang T J and Shi B X. 2020. UnModNet: learning to unwrap a modulo image for high dynamic range imaging//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada; MIT Press: 1559-1570
- Zhu A Z, Chen Y B and Daniilidis K. 2018a. Realtime time synchronized event-based stereo//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 438-452 [DOI: 10.1007/978-3-030-01231-1_27]
- Zhu A Z, Thakur D, Özarslan T, Pfommer B, Kumar V and Daniilidis K. 2018b. The multivehicle stereo event camera dataset: an event camera dataset for 3D perception. IEEE Robotics and Automation Letters, 3(3): 2032-2039 [DOI: 10.1109/LRA.2018.2800793]
- Zhu A Z, Wang Z Y, Khant K and Daniilidis K. 2021. EventGAN: leveraging large scale image datasets for event cameras//Proceedings of 2021 IEEE International Conference on Computational Photography. Haifa, Israel; IEEE: 1-11 [DOI: 10.1109/ICCP51581.2021.9466265]
- Zhu A Z, Yuan L Z, Chaney K and Daniilidis K. 2018c. EV-FlowNet: self-supervised optical flow estimation for event-based cameras [EB/OL]. [2022-02-15]. <https://arxiv.org/pdf/1802.06898.pdf>
- Zhu A Z, Yuan L Z, Chaney K and Daniilidis K. 2019a. Live demonstration: unsupervised event-based learning of optical flow, depth and egomotion//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA; IEEE: 1694-1694 [DOI: 10.1109/CVPRW.2019.00216]
- Zhu L, Dong S W, Huang T J and Tian Y H. 2019b. A retina-inspired sampling method for visual texture reconstruction//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Shanghai, China; IEEE: 1432-1437 [DOI: 10.1109/ICME.2019.00248]
- Zhu L, Dong S W, Li J N, Huang T J and Tian Y H. 2020. Retina-like visual image reconstruction via spiking neural model//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 1435-1443 [DOI: 10.1109/CVPR42600.2020.00151]
- Zhu L, Li J N, Wang X, Huang T J and Tian Y H. 2021. NeuSpikeNet: high speed video reconstruction via bio-inspired neuromorphic cameras//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada; IEEE: 2380-2389 [DOI: 10.1109/ICCV48922.2021.00240]

作者简介



黄铁军, 1970年生, 男, 教授, 主要研究方向为视觉信息处理与类脑视觉。

E-mail: tjhuang@pku.edu.cn

余肇飞, 男, 研究员, 主要研究方向为类脑视觉计算与脉冲神经网络。E-mail: yuzf12@pku.edu.cn

李源, 男, 高级工程师, 主要研究方向为视觉信息处理与类脑芯片设计。E-mail: yuanli@pku.edu.cn

施柏鑫, 男, 研究员, 主要研究方向为计算影像学、计算机视觉。E-mail: shiboxin@pku.edu.cn

熊瑞勤, 男, 研究员, 主要研究方向为视频图像信号处理。E-mail: rxqiong@pku.edu.cn

马雷, 男, 副研究员, 主要研究方向为类脑智能、实时真实感绘制、科学计算可视化。E-mail: lei.ma@pku.edu.cn

王威, 男, 研究员, 主要研究方向为计算机视觉、认知计算与常识推理。E-mail: wangwei@bigai.ai