附件3

2022年度湖北省自然科学奖提名公示信息

|  |  |
| --- | --- |
| 项目名称 | 面向开放环境的视频人体动作识别理论与方法研究 |
| 提名单位 | 武汉大学 |
| 提名意见 | （不超过600字，如实对科学发现点的原创性、科学价值、国内外自然科学界公认度以及推动学科发展的作用进行概述）项目从视频图像预处理（滤波）、视频运动信息计算出发，以服务于人体行为理解高层视觉任务为目的，系统建立了视频人体动作识别的理论和方法，取得了多项创新成果：（1）提出了基于滤波引导的视频图像质量提升理论，解决了视频图像的运动模糊、滤波光晕问题，为视频运动估计和人体动作识别提供了坚实的基础；（2）发展了强抗干扰视频光流运动信息精确计算理论，解决了光流运动估计的噪声、大位移、运动模糊等难点问题；（3）建立并完善了基于多视觉模态的视频人体动作识别机制，启发性的提出了多视觉模态挖掘与融合理论，提高了开放环境下视频人体动作识别的精确性与鲁棒性。本研究部分成果获得了中国人工智能学会主办的ACAIT2020国际会议“最佳学生论文奖”（论文主题：视频骨骼人体动作识别）、ICCV2021‒MMVRAC挑战赛“鱼眼视频人体动作识别”、“骨骼人体动作识别”2个赛道的亚军。对照自然科学奖授奖条件，决定提名该项目为2022年度湖北省自然科学奖一等奖。 |
| 项目简介 | “人”作为视频图像的核心成分，由于其主观能动性，人体动作行为具有时空多样性；其次，人体动作行为发生的环境通常极为复杂。因此，视频人体动作识别是计算机视觉、视频理解领域公认的难题。在国家重点研发计划、国家自然科学基金、湖北省自然科学基金等项目的支持下，项目团队围绕“如何提高开放环境下视频人体动作识别的实用性”这一核心问题，在高性能视频人体动作识别模型研发、时空运动信息获取与利用、视频图像质量提升等方面取得了多项原创性成果。主要科学发现如下：1. 多视觉模态视频人体动作识别理论。人体动作识别不仅依赖人体表观信息，更需要整体考虑与动作相关的各种视觉模态信息及各模态的关联性。本研究率先提出了基于多视觉模态的视频人体动作识别概念与理论。建立了多视觉模态挖掘与计算、多视觉模态融合与优化的方法体系，实现了快速攫取丰富的视觉模态，增强了多视觉模态深度特征表达的可解释性，解决了动作识别的时空多变性难题。提升视频人体动作识别精度10%左右（从88.0%, 59.4%提升到95.6%, 71.4%）。2. 强抗干扰的视频光流运动信息精确计算理论。针对开放环境下的视频面临“噪声、大位移、遮挡、运动模糊”等多种干扰，导致光流运动估计精度不高，进而制约视频人体动作行为识别性能的问题，本研究提出了以变分能量函数和深度神经网络为载体的新型光流计算理论，建立了抗干扰性强的光流运动计算方案，为复杂场景下高效捕获视频（人体）目标的运动信息提供了坚实的理论基础与技术支撑，具有鲁棒、精细的优点（总体平均误差从7.2像素减少到4.7像素，精度提升34.7%）。3. 滤波引导的视频图像质量提升理论。针对任意方向的直线运动模糊视频图像，本研究提出了基于倒频谱的PSF 参数估计理论。进一步，为消除平滑图像过程中产生的光晕现象，提出了一种融合梯度信息的引导滤波方法，在结构相似性（SSIM）评价和峰值信噪比（PSNR）评价中分别取得最高约30%和15%左右的提升。五篇代表性论文发表在IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video Technology，中国图象图形学报（封面论文）等国内外计算机视觉与人工智能领域的顶级学术期刊上。总他引350多次，含ESI高引论文2篇。研究工作被5位中国工程院/美国发明家科学院/欧洲科学院院士、21位IAPR/IEEE Fellow等给予正面评价。如中国工程院高文院士，美国发明家科学院Alan Michaels院士，IEEE/IAPR Fellow英国萨里大学Josef Kittler教授等。基于本项目研究成果，与财富世界500强“深圳投资控股有限公司”旗下广东省最大安防类上市公司合作开发的视频人体动作行为识别产品在多个城市、多个重大项目中成功应用。其中，基于人体行为识别的智慧工地系统，应用在以中国天音大厦（深圳地标建筑）为代表的几十个建筑工地中，被评价为“减轻了管理者和安全员压力，降低了事故率”；高层建筑人体行为检测与识别系统，应用于深圳湾超级总部基地等，被评价为“…是目前国内关于高层建筑人员行为检测与识别最早的产品之一，填补了国内多项技术空白。实现了对高层动作行为…实时检测、预警”。 |
| 主要完成人（完成单位） | 涂志刚（武汉大学），谢伟（华中师范大学），李红艳（湖北经济学院） |
| 代表性论文（专著）目录 |
| 序号 | 论文（专著）名称/刊名/作者 | 年卷页码（xx年xx卷xx页） | 发表时间（ 年 月 日） | 通讯作者（含共同） | 第一作者（含共同） | 国内作者 | 他引总次数 | 检索数据库 | 论文署名单位是否包含国外单位 |
| 1 | Semantic Cues Enhanced Multi-modality Multi-Stream CNN for Action Recognition / IEEE Transactions on Circuits and Systems for Video Technology / Zhigang Tu, Wei Xie, Justin Dauwels, Baoxin Li, Junsong Yuan | 2019年29卷1423-1437页 | 2019-05-31 | 涂志刚 | 涂志刚 | 涂志刚，谢伟 | 45 | SCI | 是 |
| 2 | Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition / IEEE Transactions on Image Processing / Zhigang Tu, Hongyan Li, Dejun Zhang, Justin Dauwels, Baoxin Li, Junsong Yuan | 2019年28卷2799-2812页 | 2019-06-30 | 李红艳 | 涂志刚 | 涂志刚，李红艳，张德军 | 62 | SCI | 是 |
| 3 | A survey of variational and CNN-based optical flow techniques / Signal Processing: Image Communication / Zhigang Tu, Wei Xie, Dejun Zhang, Ronald Poppe, Remco C. Veltkamp, Baoxin Li, Junsong Yuan | 2019年72卷9-24页 | 2019-06-30 | 涂志刚、谢伟 | 涂志刚 | 涂志刚，谢伟，张德军 | 50 | SCI | 是 |
| 4 | 基于倒频谱的运动模糊图像PSF参数估计 / 武汉大学学报（信息科学版）/谢伟，秦前清 | 2008年33卷128-131页 | 2008-02  | 谢伟 | 谢伟 | 谢伟，秦前清 | 74 | CSCD | 否 |
| 5 | 融合梯度信息的改进引导滤波 / 中国图象图形学报 / 谢伟, 周玉钦, 游敏 | 2016年21卷1119-1126页 | 2016-09-16 | 谢伟 | 谢伟 | 谢伟, 周玉钦, 游敏 | 49 | CSCD | 否 |