

顶会观察

# ACM Multimedia 2021

电子科技大学 徐行

ACM International Conference on Multimedia (简称 ACM Multimedia) 是世界多媒体领域最重要的顶级会议，也是中国计算机学会推荐的该领域唯一的 A 类国际学术会议。自 1993 年首次召开以来，ACM Multimedia 已成功举办了二十八届。2021 年第 29 届 ACM Multimedia 会议已于 2021 年 10 月 21 日至 25 日在中国成都举办，此次会议是第二次在中国、首次在大西南地区举办。

## 一、 论文录用情况

本次会议吸引了来自中国、美国、德国、澳大利亚、瑞典、法国、日本等 19 个国家和地区约 1100 余名人员注册参会，其中 ACM/SIGMM Member 有 396 名，Non-ACM/SIGMM Member 有 377 名，Student-ACM 有 109 名，Student-Non ACM 有 124 名；共收到了来自近 40 个国家共 2544 篇投稿论文，有效投稿 1942 篇，共有 219 名 Area Chair 以及 1233 名审稿人，最终录用了 542 篇。

本次 ACM Multimedia 内容覆盖了 Engaging Users with Multimedia、Experience、Multimedia Systems 和 Understanding Multimedia Content 四个主题。在被录用的论文中，占比最多的是 Deep Learning for Multimedia，占比 36%；其次是 Vision and Language，占比 14%；排名第三的是 Emerging Multimedia Applications，占比 8%；另外 Multimedia Search and Recommendation 和 Media Interpretation 均占比为 7%；Multimodal Fusion and Embedding 占比为 6%；Multimodal

Analysis and Description 占比为 5%；Emotional and Social Signals in Multimedia 和 Multimedia HCI and Quality of Experience 均只占 3%，而 Multimedia Art, Entertainment and Culture 占比只有 2%。

## 二、 会议日程概要

根据国内外疫情形势，本次会议采用独特的线上-线下混合会议模式。来自国内的参会者大多通过线下参会的方式来到会议现场参会，而在海外的参会者或者国内中高风险地区的参会者则通过线上参会方式参与会议。本次会议也针对性地提供了不同的线上参会软件和直播平台，方便线上参会者参与会议并进行在线交流。为期 5 天的大会学术报告共进行了 6 场 Keynote、13 场 Workshop、8 场 Tutorial、2 场 Panel 和 37 场 Session，充分展现了多媒体领域全球最新研究成果和前沿动态，以及对该领域未来发展方向创新引领。会议的第 1 和第 5 个会议日安排了 Workshop 和 Tutorial。第 2 到 4 个会议日则安排了主会期间的各个主旨报告，口头发表 session，特定主题 session 以及海报展示环节。各组织者，演讲者以及论文发表者通过现场参会或在线报告的方式介绍论文内容，参会者通过线下参会或加入线上会议的方式听报告或交流。本次会议每天中场休息和中午休息期间为现场参会者提供了茶歇和冷餐食，方便参会者的现场交流。从现场来看，交流氛围十分热烈，大家也十分珍惜疫情之下难得的线下交流机会。

会议最引人瞩目的四项多媒体领域大奖在大会晚

宴 Banquet 的热烈氛围中揭晓，包括最佳演示奖“ViDA-MAN: Visual Dialog with Digital Humans”、最佳开源奖“X-modaler: A Versatile and High-performance Codebase for Cross-modal Analytics”、最佳学生论文奖“aBio: Active Bi-Olfactory Display Using Subwoofers for Virtual Reality”，及最佳论文奖“Video Background Music Generation with Controllable Music Transformer”。

### 三、会议主旨演讲

今年大会主旨演讲邀请到了多媒体领域世界级的专家学者作为重量级嘉宾，分别带来了涵盖视频编码技术、多模态模型与学习、AI 教育等主题的 6 场演讲，从多个角度见证了多媒体技术的巨大潜力。

中国工程院院士，北京大学教授，鹏城实验室主任高文院士带来了“面向机器的视频编码技术”报告，剖析了传统视频编码技术在智能机器时代的缺陷，提出了社会发展对新一代针对智能机器的视频编码技术的迫切需要，并介绍了鹏城实验室在此方向上的最新实践与成果。密歇根大学安娜堡分校的 H. V. Jagadish 教授带来了“语义媒体转换”主题的演讲，分析了智能机器时代不同媒介的多媒体信息相互转换越发灵活便利的最新发展趋势，介绍了通过结构化的数据表来高保真地转换多媒体信息的最新技术。

法国 INRIA 和 Google 的研究员 Cordelia Schmid 带来了针对多模态视频的大规模训练的最新研究发展，介绍了 VideoBert, ViViT 等最新基于 Transformer 架构的多模态神经网络模型，并展示了跨模态监督训练范式的强大潜力。蚂蚁集团副总裁周靖人博士介绍了阿里巴巴和蚂蚁集团研发的最新大规模多模态预训练模型 M6 及其在大量下游学习任务上的卓越表现，以及一种最新的基于预训练对抗生成网络的图像编辑技术。

北卡罗莱纳州立大学的 James Lester 教授分享了 AI 技术对社会带来的机遇与挑战，介绍了计算机视觉、自然语言处理、机器学习等 AI 技术给人类教育事业带来的变革与发展。腾讯 AI 实验室、腾讯 Robotics

X 实验的主任张正友博士提出了“虚实集成世界”的新概念，介绍了虚拟现实、增强现实等技术带来的物理-虚拟世界融合的未来发展方向，以及其中涉及的人工智能、多媒体关键技术。

### 四、会议亮点 Sessions

Best Paper Session: 本次会议的 Best Paper Session 安排在主会第二天上午开场进行，共有 5 篇最佳论文提名的论文，内容涉及 Co-Speech 手势生成研究<sup>[1]</sup>、视频背景音乐生成<sup>[2]</sup>、场景文字识别<sup>[3]</sup>、多模态语音增强<sup>[4]</sup>、虚拟现实<sup>[5]</sup>，分别来自马里兰大学<sup>[1]</sup>、北京航空航天大学<sup>[2]</sup>、中国科学院信息工程研究所<sup>[3]</sup>、赫尔辛基大学<sup>[4]</sup>、台湾大学<sup>[5]</sup>。报告者通过线上或者现场的方式对论文内容进行介绍。最终，北京航空航天大学“Video Background Music Generation with Controllable Music Transformer”胜出获得了最佳论文奖。在该篇论文中，作者创造性地提出了视频背景音乐生成问题，能够依据视频播放内容自动生成合适的背景音乐。此外，该团队成员还在现场针对视频内容生成的音乐进行了实时演奏，体现了丰富的多媒体交互形式和视听觉体验。

Brave New Ideas Session: 此 Session 一直是作为 ACM Multimedia 会议的代表性的会议议程，探索未来多媒体领域最具研究和应用潜力的主题。本次 session 共包含 5 篇论文，来自中国科学技术大学<sup>[6]</sup>，纽约州立大学布法罗分校<sup>[7]</sup>，上海交通大学<sup>[8]</sup>，清华大学<sup>[9]</sup>和香港中文大学<sup>[10]</sup>。内容涉及图片质量评估、视频运动公式学习、用于抑郁症检测的音频编码、多媒体计算的新范式以及基于元宇宙(Metaverse)构建的虚拟大学原型。其中来自香港中文大学的 Haihan Duan 通过在元宇宙构建一个大学原型为例，介绍了元宇宙的概念和雏形，让大家对未来的元宇宙有了一个初步的认识。

Industrial Track Session: ACM Multimedia 会议自去年开始引入此 session 的形式，旨在介绍工业界所研究的前沿热点多媒体问题及产业应用情况，同时加强多媒体领域工业和学术领域研究者的交流。此次 session 共包含两个特邀报告和四篇口头发表论文。特邀报告中，来自阿里巴巴的 Xiansheng Hua 通过在线

的方式介绍了阿里巴巴公司内部中所使用的视觉搜索技术及典型应用案例；来自 OPPO 的 Yandong Guo 以风趣成熟的演讲风格给大家介绍了在 Metaverse 时代，针对智能移动设备的多媒体技术，重点对基于知识图谱的大规模视觉理解进行介绍，让大家对如何针对大规模多媒体数据构建知识图谱有了一个清晰的认识。四篇口头发表论文分别来自 OPPO、微软亚洲研究院、阿里巴巴、香港科技大学、京东 AI Lab 以及腾讯。论文内容涵盖多媒体算法如何应用在具体的工业场景中，例如，如何大规模视频数据自监督，对于语音识别的评分机制，通过知识蒸馏来提高行人检索精度和对遮挡文本的重建方法等。

## 参考文献

- [1] Uttaran Bhattacharya, Elizabeth Childs, Nicholas Rewkowski, Dinesh Manocha. Speech2affectivegestures: Synthesizing co-speech gestures with generative adversarial affective expression learning. 2021. In ACM Multimedia (ACM MM), Chengdu, China, 20-24 October, 2021, 2027—2036.
- [2] Shangzhe Di, Zeren Jiang, Si Liu, Zhaokai Wang, Leyan Zhu, Zexin He, Hongming Liu, Shuicheng Yan. Video Background Music Generation with Controllable Music Transformer. ACM MM 2021, 2037—2045.
- [3] Zhi Qiao, Yu Zhou, Jin Wei, Wei Wang, Yuan Zhang, Ning Jiang, Hongbin Wang, Weiping Wang. PIMNet: a parallel, iterative and mimicking network for scene text recognition. ACM MM 2021, 2046—2055.
- [4] Abhishek Kumar, Tristan Braud, Lik Hang Lee, Pan Hui. Theophany: Multimodal speech augmentation in instantaneous privacy channels. ACM MM 2021, 2056—2064.
- [5] You-Yang Hu, Yao-Fu Jan, Kuan-Wei Tseng, You-Shin Tsai, Hung-Ming Sung, Jin-Yao Lin, Yi-Ping Hung. aBio: Active Bi-Olfactory Display Using Subwoofers for Virtual Reality. ACM MM 2021, 2065—2073.
- [6] Zhu Y, Ma H, Peng J, et al. Recycling Discriminator: Towards Opinion-Unaware Image Quality Assessment Using Wasserstein GAN. ACM MM 2021.
- [7] Song L, Liu S, Liu C, et al. Learning Kinematic Formulas from Multiple View Videos. ACM MM 2021.
- [8] Zhang P, Wu M, Dinkel H, et al. Depa: Self-supervised audio embedding for depression detection. ACM MM 2021.
- [9] Kang Z, Li J, Zhu L, et al. Retinomorphic Sensing: A Novel Paradigm for Future Multimedia Computing. ACM MM 2021.
- [10] Duan H, Li J, Fan S, et al. Metaverse for social good: A university campus prototype. ACM MM 2021.

## 五、总结与展望

在 ACM Multimedia 会议的线上-线下混合举办期间，还有多个丰富多彩的 sessions，例如 Workshops、Tutorials、Art & Culture、Best Demo、Open Source Competition、Posters、艺术展、赞助企业展、午餐会等。笔者的理解仅代表个人观点，如有偏差还请各位读者指出。2022 年的 ACM Multimedia 会议将于 10 月 10 日至 14 日在葡萄牙里斯本举行，期待疫情显著缓解，全世界各地多媒体领域研究者能通过线下参会的方式相聚，共同探讨多媒体领域未来的发展。

推荐委员 金鑫 责任编辑 王金甲



## 徐行

电子科技大学副教授。主要研究方向为多模态信息感知与计算、跨媒体智能分析，已在多媒体、计算机视觉及人工智能相关领域的国内外学术期刊和会议累计发表论文 100 余篇，获得包含 2017 年国际多媒体大会 ACM MM (CCF-A) 最佳论文奖，2017 年国际多媒体展览会 ICME (CCF-B) 的最佳会议论文铂金奖等国际会议奖项 6 项。

Email: xing.xu@uestc.edu.cn