



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

中国图象图形学报

2021
06
VOL.26

ISSN1006-8961
CN11-3758/TB



中国图象图形学学会成立三十周年

图像图形学 发展 年度报告



第26卷第6期（总第302期）

2021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

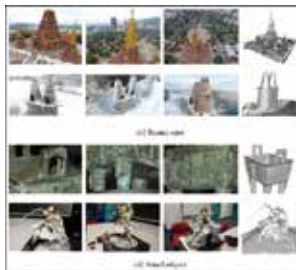
国内邮发代号 82-831

国内定价 60.00元

序言	王耀南
----------	-----



生物特征识别学科发展报告
(第1254页)



大规模室外图像3维重建技术
研究进展(第1429页)



虚实融合场景中的深度感知
研究综述(第1503页)

图像处理与通信技术

视频处理与压缩技术

贾川民, 马海川, 杨文瀚, 任文琦, 潘金山, 刘东, 刘家琰, 马思伟	1179
---	------

面向体验质量的多媒体计算通信

陶晓明, 杨铀, 徐迈, 段一平, 黄丹蓝, 刘文予	1201
----------------------------------	------

数字媒体取证技术综述

李晓龙, 俞能海, 张新鹏, 张卫明, 李斌, 卢伟, 王伟, 刘晓龙	1216
---	------

面向智慧城市的交通视频结构化分析前沿进展

赵耀, 田永鸿, 党建武, 付树军, 王恒友, 万军, 安高云, 杜卓然, 廖理心, 韦世奎	1227
--	------

生物特征识别学科发展报告

孙哲南, 赫然, 王亮, 阚美娜, 冯建江, 郑方, 郑伟诗, 左旺孟, 康文雄, 邓伟洪, 张杰, 韩璇, 山世光, 王云龙, 茹一伟, 朱宇豪, 刘云帆, 何勇	1254
--	------

自然场景文本检测与识别的深度学习方法

刘崇宇, 陈晓雪, 罗灿杰, 金连文, 薛洋, 刘禹良	1330
-----------------------------------	------

基于深度学习的跨模态检索综述

尹奇跃, 黄岩, 张俊格, 吴书, 王亮	1368
----------------------------	------

三维视觉和图形技术

三维视觉前沿进展

龙霄潇, 程新景, 朱昊, 张朋举, 刘浩敏, 李俊, 郑林涛, 胡庆拥, 刘浩, 曹汛, 杨睿刚, 吴毅红, 章国锋, 刘烨斌, 徐凯, 郭裕兰, 陈宝权	1389
--	------

大规模室外图像3维重建技术研究进展

颜深, 张茂军, 樊亚春, 谭小慧, 刘煜, 彭杨, 刘宇翔	1429
--------------------------------------	------

视觉传感成像技术与数据处理进展

王程, 陈峰, 汶德胜, 雷浩, 宋宗玺, 赵航芳	1450
---------------------------------	------

视觉—惯性导航定位技术研究进展

司书斌, 赵大伟, 徐婉莹, 张勇刚, 戴斌	1470
------------------------------	------

三维视觉测量技术及应用进展

张宗华, 刘巍, 刘国栋, 宋丽梅, 屈玉福, 李旭东, 魏振忠	1483
--	------

虚实融合场景中的深度感知研究综述

平佳敏, 刘越, 翁冬冬	1503
--------------------	------

可微绘制技术研究进展

许威威, 周漾, 吴鸿智, 过洁	1521
------------------------	------

沉浸式立体显示技术在临床医学领域中的应用

邵永航, 石俊生	1536
----------------	------

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Overview of biometrics research
(P1254)



Progress in the large-scale
outdoor image 3D reconstruction
(P1429)



Review of depth perception in
virtual and real fusion environment
(P1503)

Image Processing & Communication Technology

Video processing and compression technologies

Jia Chuanmin, Ma Haichuan, Yang Wenhan, Ren Wenqi, Pan Jinshan, Liu Dong, Liu Jiaying, Ma Siwei 1179

Multimedia computing communications

Tao Xiaoming, Yang You, Xu Mai, Duan Yiping, Huang Danlan, Liu Wenyu 1201

Overview of digital media forensics technology

Li Xiaolong, Yu Nenghai, Zhang Xinpeng, Zhang Weiming, Li Bin, Lu Wei, Wang Wei, Liu Xiaolong 1216

Frontiers of transportation video structural analysis in the smart city

Zhao Yao, Tian Yonghong, Dang Jianwu, Fu Shujun, Wang Hengyou, Wan Jun, An Gaoyun, Du Zhuoran, Liao Lixin, Wei Shikui 1227

Overview of biometrics research

Sun Zhenan, He Ran, Wang Liang, Kan Meina, Feng Jianjiang, Zheng Fang, Zheng Weishi, Zuo Wangmeng, Kang Wenxiong, Deng Weihong, Zhang Jie, Han Hu, Shan Shiguang, Wang Yunlong, Ru Yiwei, Zhu Yuhao, Liu Yunfan, He Yong 1254

Deep learning methods for scene text detection and recognition

Liu Chongyu, Chen Xiaoxue, Luo Canjie, Jin Lianwen, Xue Yang, Liu Yuliang 1330

3D Vision & Graphics Technology

Survey on deep learning based cross-modal retrieval

Yin Qiyue, Huang Yan, Zhang Junge, Wu Shu, Wang Liang 1368

Recent progress in 3D vision

Long Xiaoxiao, Cheng Xinjing, Zhu Hao, Zhang Pengju, Liu Haomin, Li Jun, Zheng Lintao, Hu Qingyong, Liu Hao, Cao Xun, Yang Ruigang, Wu Yihong, Zhang Guofeng, Liu Yebin, Xu Kai, Guo Yulan, Chen Baoquan 1389

Progress in the large-scale outdoor image 3D reconstruction

Yan Shen, Zhang Maojun, Fan Yachun, Tan Xiaohui, Liu Yu, Peng Yang, Liu Yuxiang 1429

Review on imaging and data processing of visual sensing

Wang Cheng, Chen Feng, Wen Desheng, Lei Hao, Song Zongxi, Zhao Hangfang 1450

Review on visual-inertial navigation and positioning technology

Si Shubin, Zhao Dawei, Xu Wanying, Zhang Yonggang, Dai Bin 1470

Overview of the development and application of 3D vision measurement technology

Zhang Zonghua, Liu Wei, Liu Guodong, Song Limei, Qu Yufu, Li Xudong, Wei Zhenzhong 1483

Review of depth perception in virtual and real fusion environment

Ping Jiamin, Liu Yue, Weng Dongdong 1503

Differential rendering: a survey

Xu Weiwei, Zhou Yang, Wu Hongzhi, Guo Jie 1521

Application of immersive 3D imaging technology in the clinic medical field

Tai Yonghang, Shi Junsheng 1536

中图法分类号: TN929.5 文献标识码: A 文章编号: 1006-8961(2021)06-1201-15

论文引用格式: Tao X M, Yang Y, Xu M, Duan Y P, Huang D L and Liu W Y. 2021. Quality of experience oriented multimedia computing communications. Journal of Image and Graphics, 26(06): 1201-1215 [陶晓明, 杨铀, 徐迈, 段一平, 黄丹蓝, 刘文予. 2021. 面向体验质量的多媒体计算通信. 中国图象图形学报, 26(06): 1201-1215] [DOI: 10.11834/jig.200864]

面向体验质量的多媒体计算通信

陶晓明¹, 杨铀², 徐迈³, 段一平¹, 黄丹蓝¹, 刘文予^{2*}

1. 清华大学, 北京 100084; 2. 华中科技大学, 武汉 430074; 3. 北京航空航天大学, 北京 100191

摘要: 随着移动互联网和通讯技术的发展, 多媒体通信技术成为国家信息产业发展的重大需求, 广泛地应用在视频会议、各类直播应用、远程医疗、远程监控和远程教育等方面。然而, 大容量多媒体通信业务面临着网络带宽的压力。本文将媒体计算引入通信系统, 建立新的多媒体通信研究范式, 从提升体验质量 (quality of experience, QoE) 的角度, 形成新的多媒体编码与传输方法, 从根本上降低网络带宽需求的压力。体验质量即信息接收者结合自身期望对客观信息载体的有关性能给出的主观评价, 是区别于服务质量 (quality of service, QoS) 的通信质量评价准则。本文介绍了 QoE 的评价准则, 分为基于用户的评价方法和基于客观参数的评价方法, 通过用户主观评分或对用户的相关生理、心理指标进行测量进而分析、推测用户的感受; 或者通过对业务客观指标的主观化修正实现体验质量的评价。本文综述了多媒体编码方法, 主要包括基于波形的编码和基于内容的编码方法。前者对任意视频信号进行有效编码而不需要分析视频内容, 如一系列视频编码标准; 后者识别视频序列中的物体和相关区域并对它们进行编码。本文阐述了 5G + AI (artificial intelligence) 时代的新型视频传输方法, 如多视点视频编码、4 K、8 K 视频编码、3D 立体视频、点云、光场、AR (augmented reality)、VR (virtual reality) 等视频业务。

关键词: 计算通信; 多媒体通信; 评价准则; 多媒体编码; 多媒体传输

Quality of experience oriented multimedia computing communications

Tao Xiaoming¹, Yang You², Xu Mai³, Duan Yiping¹, Huang Danlan¹, Liu Wenyu^{2*}

1. Tsinghua University, Beijing 100084, China; 2. Huazhong University of Science and Technology, Wuhan 430074, China;
3. Beihang University, Beijing 100191, China

Abstract: With the development of mobile Internet and communication technology, multimedia communication technology has become a major demand for the development of the national information industry, which is widely used in video conferencing, various live broadcast applications, telemedicine, remote monitoring, and remote education. However, large-capacity multimedia communication services face pressure on network bandwidth. Media computing is introduced into the communication system, a new multimedia communication research paradigm is established, and a new multimedia coding and transmission method is developed from the perspective of improving the quality of experience (QoE), which fundamentally reduces the pressure on network bandwidth requirements. The quality of experience refers to the subjective evaluation of the relevant performance of the objective information carrier by the information receiver in combination with its own expectations. It is a communication quality evaluation criterion that is different from the quality of service (QoS). This

收稿日期: 2020-12-31; 修回日期: 2021-02-18; 预印本日期: 2021-02-25

* 通信作者: 刘文予 liuwu@hust.edu.cn

基金项目: 国家重点研发计划项目 (2019YFB1803404, 2017YFC0806202); 国家自然科学基金项目 (61925105, 61801260, 61971203)

Supported by: National Key Research and Development Program of China (2019YFB1803404, 2017YFC0806202); National Natural Science Foundation of China (61925105, 61801260, 61971203)

article initially introduces the evaluation criteria of QoE, which are divided into user-based evaluation methods and objective parameter-based methods. The user's subjective score or the measurement of the user's relevant physiological and psychological indicators is used to analyze and infer the user's feelings. The subjective correction of business objective indicators realizes the evaluation of experience quality. Then, multimedia coding methods, including waveform-based coding and content-based coding methods are reviewed. The former effectively encodes any video signal without analyzing the video content, such as a series of video coding standards; the latter identifies objects and related areas in the video sequence and encodes them. Subsequently, this article describes new video transmission methods in the 5G + AI (artificial intelligence) era, such as multiview video coding, 4 K, 8 K video coding, 3D stereo video, point cloud, light field, AR (augmented reality), VR (virtual reality), and other video services. Video services and communication channels are the source and channel parts of the video communication system, respectively, and the relationship between them can usually be compared with the relationship between water sources and water pipes. The continuous development and changes in communication technology, especially the international competition of 5G technology and products, have attracted widespread attention. Channels continuously increase; hence, the transmission rates become fast. Communication becomes ubiquitous. Thus, the rolling development between the source and the channel becomes increasingly influential. Academia and industry mostly solve the high-definition and low-latency challenges faced by video communication systems from two aspects, as follows: video characterization and coding and video transmission. Video coding aims to find effective data compression techniques to reduce the bit rate of video sequences for real-time transmission on a given communication channel. Image compression coding uses the statistical characteristics of the image itself, as well as the user's visual physiology and psychology characteristics, to extract the effective information in the image and remove useless or less useful redundant information. Channel bandwidth varies with different applications and transmission media. Different types of video coding algorithms have been developed; they include effective coding of arbitrary video signals without the need to analyze video content and identify objects and related areas in the video sequence to encode them. The former approach is referred to as a waveform-based encoder, and the latter is a content-based video encoder. With the help of artificial intelligence and machine learning technology, the innovative exploration of information representation from "bit→structure" can effectively overcome the influence of noise and interference in the propagation environment and improve the reliability and efficiency of wireless communication. The amount of data transmitted is greatly reduced by introducing a priori knowledge. The audiovisual service breaks the traditional user-server (content server or content delivery network (CDN) server) video streaming push mode and adds edge computing nodes to satisfy the computing requirements of secondary encoding, virtual scene generation, and scene rendering, generated during the interaction. In the future, the media inevitably aims at "a thousand people with a thousand faces" and develops in the direction of large data volume, large calculation volume, and large communication volume. The business architecture has the characteristics of "cloud-side-end" collaborative computing, which derives richer media applications in the future. The quality of communication system is evaluated from the perspective of QoE. QoE reflects the subjective evaluation of the objective information carrier (voice, image, and video) after the information receiver perceives the objective information carrier's performance. The current QoE evaluation methods are mainly divided into two categories, namely, user-based evaluation methods and evaluation methods based on objective parameters. User-based evaluation methods include all evaluation methods that require user participation. Specific indicators or information about the QoE need to be obtained directly from users. The evaluation methods based on objective parameters realize the evaluation of experience quality through the subjective correction of objective indicators.

Key words: computing communication; multimedia communication; evaluation criteria; multimedia coding; multimedia transmission

0 引言

宽带无线通信的迅猛发展激发了人们对大容量

多媒体业务的需求。多媒体通信技术渗透到了人们生活的方方面面,广泛应用于视频监控、视频会议、视频直播、在线教育和自动驾驶等领域。随着互联网技术的发展,以视频为主体的移动多媒体数据急

剧增长,导致超高清视频业务面临挑战(陆建华, 2017)。Cisco(Cisco visual network index)报告指出(<http://www.199it.com/archives/804930.html>),2017—2022年的5年间,全球IP(internet protocol)流量将增长3倍;互联网流量繁忙时段将延长4.8倍,平均将延长3.7倍,视频、游戏和多媒体将占据数据流量的85%以上。与之对应的是,2022年全球固定宽带平均速度将从2017年的39 Mbps发展到75.4 Mbps,增长接近1倍,同时平均Wi-fi连接速度增长逾1倍(Cisco报告)。在目前或者可见的将来,信息产生的体量和对应的传输技术在增长速度上会产生一定程度的脱节。数据规模的不断扩张,使其与无线带宽资源之间的矛盾日益突出。在全球新冠疫情期间,多地下达封城令后,作为互联网基础设施最大消耗者的流媒体视频需求量激增,Netflix、YouTube、亚马逊旗下Prime Video和迪士尼旗下Disney+等流媒体视频服务被迫降低画质以减少带宽消耗(<https://tech.sina.com.cn/roll/2020-03-26/doc-iimxx-sth1786019.shtml>)。借助人工智能、机器学习技术,探索从“比特→结构”的信息表征方法可以有效克服传播环境中噪声与干扰的影响,提高无线通信的可靠性以及效率(陆建华,2017)。通过引入先验知识,大幅降低传输的数据量。

无线多媒体通信系统集成了图像采集处理、数据压缩、传输解析、编码解码及图像显示等多项功能。视频通信系统的实现步骤包括:1)发送端通过摄像头、麦克风等设备对声音、图像和文本等多媒体数据进行采集;2)对采集的数据进行压缩编码,打包后通过网络传输到接收端;3)接收端解析数据包,并将解码、转码后的多媒体数据进行显示播放。宽带视频通信业务与有限的带宽支持能力之间的矛盾尤为突出。解决的途径主要有两种:一是提升网络承载能力,二是提升视频的智能编码性能。5G将进一步推动视频通信领域的飞速发展。中国信息通信研究院预计,2020—2025年我国5G移动网络建设的总投资将达到1.2万亿元,因5G网络移动技术的商用而直接或间接地产生、带动的经济总产出为10.6万亿元(http://www.caict.ac.cn/kxyi/caictgd/201903/t20190305_195539.htm)。其中,高清视频是多媒体业务的主体,将是5G率先应用的领域。而5G的低时延高带宽也使得视频通信实现更加快速的发展。

本文将全面阐述无线多媒体通信的发展前沿,

从体验质量(quality of experience, QoE)的度量和计算方法、视频编码方法以及视频传输方法3个方面介绍国内外现状与前沿动态,并展望未来发展趋势。

1 国际研究现状

1.1 无线多媒体通信国际现状与发展前沿

随着多媒体通信技术的发展,视频通信技术已在视频会议、各类直播应用、远程医疗、远程监控和远程视频教育等方面取得了广泛应用。随着时代的进步,人们对于多媒体信息的要求日益增高,视频通信系统面临高清、低时延的挑战。

图像压缩编码利用图像本身的统计特性,以及用户的视觉、心理学特性,提取图像中的有效信息,去除无用或用处不大的冗余信息。目前视频编码技术H.261(Lei等,1994)、H.263(Wiegand等,1996)、H.264(Wiegand等,2003)、MPEG(moving picture experts group)-1、MPEG-2、MPEG-4(Watkinson,2004)和HEVC(high efficiency video coding)(Sullivan等,2012)多采用预测编码和变换编码相结合的基于块的混合编码模式。这种编码技术把每幅图像分成固定大小的块,第 k 帧的每个块用已经编码的若干帧为参考,通过其运动矢量所标记的相同尺寸的块合成得到。

对于人脸模型,已提出描述脸部表情的动作单元(action unit, AU)(Tong等,2007)。动作单元是由心理学者研发的脸部动作编码系统(facial action coding system, FACS)导出的。该系统的开发是为了“确定所有可能的视觉可以区分的脸部运动”。所用的测量单元是动作,而不是肌肉单元。例如,提动眉毛的额头肌肉被分为两个动作单元,取决于是内侧还是外侧的肌肉抬起内侧还是外侧的眼眉。几种动作单元可以同时运动,以便产生任意一种脸部表情。共有46种考虑脸部表情变化的动作单元,其中有12种粗略描述凝视方向变化和头的方位的动作单元。

现有研究工作采用机器学习方法,对视频中人的头部姿态和表情进行提取与编码,并在接收端由此重建逼真头像,无需将视频帧进行流式传输,大幅降低视频码率。例如,FreeNet(Zhang等,2020)提出将人脸表情从任意源人脸转移到目标人脸,实现基于动作单元的人脸驱动,生成说话人的视频。Za-

kharov 等人(2019)提出通过训练卷积神经网络来生成高度逼真的人头部图像,通过在大型数据集上进行元学习,实现对未见过头像的小样本学习(few shot learning)(Ravi 和 Larochelle, 2017)。实验表明,该方法可学习高度现实和个性化的新人头部模型,应用于人脸视频通信中。Watkinson(2004)提出单目视频序列(例如 Youtube 视频)的实时面部驱动方法 Face2Face,使用商业网络摄像头实时捕获演员的面部表情作为源序列,以此驱动目标视频序列,实现对 YouTube 视频的实时操作和逼真渲染。

NVIDIA 近日宣布推出一款名为 NVIDIA Maxine(Thies 等,2016)的视频会议平台,允许开发者利用 NVIDIA 的云端 AI 处理能力,提供诸如超分辨、噪声消除、面部优化和实时翻译等 AI 功能,并且所需带宽极低。对于 30 帧/s 的 H.264 视频而言,常规来说其所需的带宽为 2 918 KB/s,但在 NVIDIA 的平台下,所需的带宽被缩小至十分之一,即 292 KB/s。只需传输面部关键点,即可在视频端生成整个面部。

1.2 多媒体 QoE 度量和计算方法

无线通信传输技术以 5G 标准的演进为代表,新技术以更为密集的基站、更多天线、更高更宽频段和更为复杂的编码调制方式为主,虽然提升了频谱、功率效率,但是设备投入巨大,视频业务的进一步发展迫切需要变革性的新途径。探索以智能提升传输效能的方法成为国际前沿,美国弗吉尼亚理工、佐治亚理工、诺基亚—贝尔实验室等机构已经在数据驱动智能通信无线传输、数据感知资源管理等方面展开了研究;国内东南大学、华为公司等也提出利用人工智能改造现有视频通信系统。总体而言,现有研究多局限于将机器学习引入传输系统,多变场景下的适应性弱,训练对数据量的要求高且性能有限,需要发展网络条件下的高效感知、学习和资源利用新途径。研究表明,人眼视网膜有 10 亿个视杆、视锥细胞,相当于 10 亿像素的摄像机;而连接大脑初级视皮层的神经节细胞仅有 1 万个,相当于视网膜与大脑传输带宽仅为 8 Mbps;但人脑却可瞬间完成复杂的视觉认知。如果可以借鉴人类的感知、学习和归纳等视觉认知机理,寻求脑启发智能计算与通信协同,将计算和通信资源有效转化为人类认知 QoE 的提升,则可从根本上解决大容量业务对无线网络带宽需求的问题。

计算领域对于 QoE 的研究,特别是针对 QoE 的建模与度量的研究引起广泛的关注。Santos 等人(2012)将多个服务质量(quality of service, QoS)通信指标,如时延、带宽、阻塞率和丢包率等,作为变量映射到相应的 QoE 数值上,从而更全面地反映出通信系统整体状态与用户主观体验度之间的关系。已有研究(Hassan 等,2012;Feng 等,2012;Balachandran 等,2013)通过对用户在体验过程中简单的人体显式响应来衡量 QoE。例如:在语音通话中,检测用户要求对方重复说话内容的次数来衡量通话质量的好坏(Hassan 等,2012);在使用触摸屏电子设备时,依据用户手指滑动方式来分析其满意度(Feng 等,2012);在观看视频点播时,时间长短也能反映用户的体验质量(Balachandran 等,2013)。这类衡量和感知方法难以量化用户的满意程度,但却能直接体现用户的主观感受,为通信系统设计提供了新思路。

在通信领域,探索以用户 QoE 为中心的无线通信技术是当前的研究热点。Toni 等人(2012)根据数据包的重要性,给予不同质量的传输信道和不同等级的信道编码保护,使得用户体验质量有了明显提升。美国麻省理工学院(Massachusetts Institute of Technology, MIT)的 ParandehGheibi 等人(2011)研究了初始缓存大小和视频流畅度之间的折衷关系,为提升用户满意度奠定了理论基础,得到广泛引用。Dutta 等人(2012)通过合理调度资源使得用户观看视频的平均流畅程度最好。另外也有学者提出可以利用网络编码的方法来提升无线网络中用户的通信体验质量(Rezaee 等,2012)。为了适应未来移动互联网多媒体业务发展的需求,探索面向 QoE 的网络通信以提升网络服务能力成为最前沿的研究方向之一(Rezaee 等,2012)。目前在图像、视频的主观质量评价方面,有面向平面图像的 LIVE(Sheikh 等,2006),面向 3D 图像的 LIVE 3D(Chen 等,2013),面向 3D 视频(Yang 等,2016)以及面向虚拟视点(Yang 和 Dai,2010)的工作。

1.3 视频编码方法

多媒体通信技术的不断发展也让视频分辨率不断提高,高视频分辨率为人们带来更丰富的信息和更清晰的视觉体验。然而,提高分辨率意味着增加数据量,这对通信速率和存储空间都产生了很高的需求,限制了高清视频的进一步发展。

针对此类问题,国外的研究机构已经做出大量

工作,制定了多种压缩编码标准。例如国际电信联盟(International Telecommunications Union, ITU)提出的 H. 261(1988)(世界上第一个数字视频编码国际标准), H. 262(1995), H. 263(1996)和 H. 263+等。这些早期标准为数字视频的发展做出重要贡献。2003年,ITU 与国际标准化组织(International Standard Organization, ISO)联合提出 H. 264 标准,又称高级视频编码(advanced video coding, AVC)标准。与 H. 263 相比, H. 264 在保持同等视频质量的情况下码率平均降低约 50%(Wiegand 等, 2003),显著提高了视频压缩效率。随着时间的推移与多媒体技术的进步,高清视频迅速普及,用户对清晰度和画面质量的要求进一步提高,传统的编码标准再度遇到瓶颈。为了更有效地提高编码效率,2010年,由 ITU 与 ISO 组建成的视频编码联合组(Joint Collaborate Team on Video Coding, JCT-VC)提出高效率视频编码(high efficiency video coding, HEVC)方案。2013年, HEVC 正式成为国际标准。与 H. 264 相比, HEVC 在相同视频质量下能够进一步降低约 50%的码率(Sullivan 等, 2012)。2020年,多功能视频编码(VCC/H. 266)由联合视频专家组(Joint Video Experts Team, JVET)提出并确认为视频压缩标准,它是 HEVC 的后继标准。

视频编码标准发展到今天,需要同时考虑两个关键问题:

1)在保证视频质量的前提下,实现尽可能低的码率;

2)在保证效率(低码率)的前提下,尽可能降低编码复杂度。

针对这两个问题,国际上的学者提出了大量优化算法:

在码率控制领域,码率控制是在给定目标码率的前提下,通过在 GOP(group of picture)、帧、块等多种级别合理分配码率,达到优化视频质量的目的。国际学者针对不同的编码标准,提出了如 MPEG-4 标准的 VM8(Lee 和 Ebrahimi, 2012; Vetro 等, 1999), H. 263 标准的 TMN8(Ribas-Corbera 和 Lei, 1999), H. 264 标准下的 JVT-N046 等码率控制方法。这些码率控制算法可以划分为两类,即基于 R-Q 模型和基于 R-p 模型。R-Q 模型(Chiang 和 Zhang, 1997)可以用二次方程表征。在 H. 265 标准下, Choi 等人(2013)根据不同块大小下失真码率比

近似的特点,提出了一种像素级率量化模型(unified rate-quantization, URQ)(He 等, 2001),实现了低码率波动和高码率控制准确性。随着 H. 265 标准(Choi 等, 2013)下引入更加灵活的分割方式和预测模式, R-Q 模型很难准确预测码率与量化参数之间的关系,因为码率与量化参数之间不再是一一对应的关系,而是由所有的编码参数共同决定。

在感知编码部分, Wandell 和 Thomas(1997)的研究表明被测试者在观看视频的时候观察轨迹具有相似性,并且在一定程度上可预测。Wandell 和 Thomas(1997)在 H. 263 基础上,提出了利用眼动仪识别出人们观看视频的的关注点作为感兴趣区域(region of interest, ROI),而后再进行相关的编码资源分配。然而,这种方法需要利用到眼动仪设备,这在普通的视频编码时不太可行。因此,利用人类视觉系统(human visual system, HVS),并不借助眼动仪的自动 ROI 区域提取成为感知视频压缩的常用方法。在对话类视频中,公认的 ROI 区域便是人的脸部区域。Lee 和 Bovik(2003)提出了一种感知编码方法,在对话类场景里利用以块为单位的权重感知映射来减少肤色区域的量化变量(quantization parameter, QP),从而提升主观质量。此后,一种利用肤色的探测算法(Yang 等, 2005)被用于在对话类视频中提取出脸部区域,也即 ROI 区域,而后通过优化码率—失真—复杂度(rate-distortion-complexity, R-D-C)主观表现从而提升对话类视频的主观质量。

在视频编码复杂度优化方面, 2005—2010 年前后,许多研究针对 H. 264 标准提出了编码复杂度优化方案。其中大部分方案针对 H. 264 中的两种耗时过程进行简化,即运动估计(motion estimation, ME)(Hsu 等, 2002)和模式判决(mode decision, MD)(Huang 等, 2006; Kannangara 等, 2006)。Pan 等人(2005)则提出一种混合型方法,同时优化 ME 与 MD 过程。

HEVC 标准问世,得益于一系列先进的编码技术,例如高度灵活的编码树单元(coding tree unit, CTU)、丰富的帧内和帧间预测模式和可变尺寸块的离散变换等。在 HEVC 中,默认的 CTU 大小为 64×64 ,远大于 H. 264 中的宏块大小(16×16)。一个 CTU 既可以只包含一个编码单元(coding unit, CU),也可基于四叉树结构递归地分成多个 CU。一个 CU 可进一步分为若干个预测单元(prediction

unit, PU) 和变换单元(transform unit, TU)。在 PU 层面,可以选择一种帧内或帧间预测模式;在 TU 层面,预测之后的残差经离散块变换后进行编码。编码器需要衡量所有组合中的视频码率与失真,寻找最优的平衡点,此过程称为率失真优化(rate-distortion optimization, RDO)。在标准 HEVC 编码器中,需要递归检测所有可能的分割组合,并选取率失真代价最小的组合作为最优解,以实现 RDO。然而, RDO 中的递归检测和比较需要很高的计算量,这使得 HEVC 的编码复杂度大幅提高。针对此问题,国际学者亦提出众多复杂度优化方法。这些方法可分为两大类:启发式方法和基于机器学习的方法。

在启发式方法中部分研究(Wang 等,2007;Leng 等,2011;Kim 等,2016)通过降低 CU 分割的复杂度实现优化。具体而言,Wang 等人(2007)在帧级别决定 CU 分割的深度,主要思想是在决定当前帧 CU 深度时,跳过之前帧中较少出现的 CU 深度。Leng 等人(2011)根据每个 CU 中高频率关键点的数量判断是否分割该 CU。同样利用贝叶斯准则,Kim 等人(2016)则以完全率失真代价和低复杂度率失真代价为依据,提前结束 CU 分割中的递归检查过程,实现快速判决。除了优化 CU 分割过程,还可通过降低 PU 或 TU 分割复杂度实现优化。例如,Cho 和 Kim(2013)利用编码块标志(coding block flag, CBF)和已编码 PU 的率失真代价等特征,判断可能性最大的 PU 分割情况。Yoo 和 Suh(2013)根据视频内容,自适应地将较小的 PU 组合成更大的 PU,实现 PU 深度的快速判决。此外,Choi 和 Jang(2012)与 Vanne 等人(2014)优化了编码过程中其他操作的复杂度,例如多方向帧内预测模式选择和环路滤波操作等。

尽管能够有效降低编码复杂度,启发式方法存在一个普遍缺点:判决准则需要人为设定。例如,在 HEVC 帧内模式下,Choi 和 Jang(2012)将 CU 分割过程视为一个二分类问题,并利用逻辑回归算法构造分类器;在 HEVC 帧间模式下,Fu 等人(2019)利用数据挖掘技术,提出了 3 种能够简化 CTU 分割过程的策略。上述基于学习的方法同样利用了一些特征,如率失真损失、量化参数(quantization parameter, QP)和视频中的纹理复杂度等,但如何根据特征制定判决准则,包括判决门限的选取,不再需要人为设定,可通过训练模型来获得。可见,与启发式方法相

比,机器学习算法具有明显的优势。然而上述方法仍存在不足,即特征仍需要人工选取,这在很大程度上依赖先验知识,并可能错失一些在编码时不易发现的隐藏特征。

ITU 正在推出新一代的编码标准 VVC(versatile video coding),这一编码标准尚未最终定档,相关技术指标仍在研究和讨论阶段,相关研究领域的学者也在积极地递交提案。相比 AVC 和 HEVC, VVC 中采用了许多新的技术手段,例如四叉树和多类型树(quad-tree plus multiple-type-tree, QTMTT)的块分割方式、自适应多核变换(adaptive multiple transform, AMT)、双向光流(bi-directional optical flow, BIO)、局部照明补偿(local illumination compensation, LIC)等技术。这些新技术在使得 VVC 更加灵活、带来更高压缩性能的同时,也大大增加了编码复杂度。针对 VVC 中的复杂度优化问题,Correa 等人(2015)首先调研了两种不同分割情况时,对应得到的编码块选择分割模式的所占比例,发现无论是二叉树分割还是三叉树分割,当父编码块选择垂直分割后,子编码块选择垂直分割的可能性远远大于选择水平分割;父编码块选择水平分割的情况则正好相反。这是因为父子编码块共享了相同的纹理信息,因此在分割时会倾向于选择相同方向的分割方式。

Fu 等人(2019)提出了一种基于深度学习的面向下一代视频编码器帧内块分割的优化方法。HEVC 标准在原本只有四叉树分割的情况下,又增加了二叉树分割,提高了块分割的灵活度,进一步降低了码率。Fu 等人(2019)提出的方法可以支持这种更复杂的分割结构,并且可以确定最佳分割方式的子集,降低块分割过程的复杂度,最终实验表明该模型可以实现在几乎不影响编码效率的前提下提速 2 倍,在影响编码效率 1% 以内的前提下,提速 4 倍。

Galpin 等人(2019)提出了一种可调参数的轻量级机器学习方法,基于随机森林算法,通过对风险间隙大小的调整实现决策准确性的控制,模型的训练利用了视频序列的时间信息和空间信息,主要测试是在 VTM(VUC test model)和 HM(HEVC test model)之间的过渡参考软件 JEM 上,采用的配置是随机获取(random access, RA),后来直接移植到 VTM 5.0 进行再次测试,也获得了不错的效果,说明模型具有一定的泛化性。

1.4 视频传输方法

视频业务与通信渠道分别是视频通信系统中的信源与信道部分,它们之间的关系通常可比作水源与水管之间的关系。在人与人之间信息交互与传递的各种模态中,视觉与听觉模态更容易让人们理解与接受,因此也构成了目前现实世界中信息的主要载体。自从视听业务以数字信号形式开始播出以来,音视频信号在数字设备中的应用与编辑变得更加便利,也使得视听业务的表现形式越来越丰富和灵活,人们对视听服务的需求、依赖、甚至期待也开始激增,所有的这一切倒逼着传统的通信方式进行适配。近年来,通信模式与技术的不断发展与变化、特别是5G技术与产品的国际化竞争引起了人们的广泛关注。通信信道越来越宽、传输速率越来越快,通信变得无处不在,使得信源与信道之间的抱团滚动式发展在学术界、工业界和社会中产生越来越大的影响力(李继龙等,2021)。

视听业务正在形成由被动式向主动式的变化。从视听业务这个概念诞生以来,人们一直就是以被动接收视听服务为主要形式。一百多年来,这种形式的改变仅仅从所有接收者在不同地点的同一时刻接受同一服务(即广播式),变成在不同地点的不同时刻接受同一服务(即点播式)。自由视点电视的概念在2006年由日本的Masayuki Tanimoto提出(Tanimoto,2006),认为观众应该改变被动接收的局面,要像平常生活一样能够改变观看的视角,主动改变所观看的内容。通俗地说,就是以前看电视是导播怎么切换观众就怎么观看,以后应该是“千人千面”,不需要导播。这样的想法推动ISO/MPEG和ITU的联合视频标准组(joint video team, JVT)在2006年启动了多视点视频编码的专题,虽然相关工作一直延续至今,但是由于其数据量过于巨大、目前通信技术难以承载而导致其业务模式一直难以确定(Smolc,2011; Zhang和Yu,2012; Vetro等,2011)。虽然上述标准化工作尚未带来现实的商业价值,但是MPEG所提出的交互式媒体的发展思路与现有低延迟、大带宽的通信技术相结合,形成了虚拟现实(virtual reality, VR)、云游戏和云主机的高交互视听业务,以及在疫情期间发挥关键作用的在线教育、直播连麦、在线会议和远程医疗等互动式视听业务模式开始广泛地被用户所接纳(李继龙等,2021),一方面推动了虚拟现实所需的4K、8K视频编码技

术的发展,并促进了包括3D立体视频、光场、360°全景、点云和虚拟现实等为代表的MPEG-I标准的制定(Domański等,2017);另一方面也推动了屏幕编码技术的诞生以及超低延迟通信技术的应用(Xu等,2016; Zhu等,2014)。在这些业务的应用中,交互时延是关键指标,一般情况下,云游戏和云主机时延<100 ms、VR业务时延<200 ms、直播业务时延<400 ms等。为了达到这样的技术指标,传统IPTV(Internet protocol television)中所使用的UDP(user datagram protocol)、RTP(real-time transport protocol)协议已经很难满足,催生了DASH(dynamic adaptive streaming over HTTP)、HTTP-FLV、RTC(real-time communications)等协议和技术的广泛应用(Stockhammer,2011; Kua等,2017)。此外,由于VR视频所需的带宽较大,往往需要CDN(content delivery network)的支持与配合才能完成。多视点视听业务也随在线教育、互动体育的升温而开始出现,虽然不能达到Tanimoto(2006)所设想的自由视点效果,但是能够实现“坐在现场的不同位置就能看到不同角度的内容”等新颖的体验。这样的视听业务进一步打破了传统用户—服务器(内容服务器或CDN服务器)之间的视频流推送模式,加入了边缘计算节点以配合交互过程中产生的二次编码、虚拟场景生成和场景渲染等计算需求。未来媒体势必以“千人千面”为目标,朝着数据量大、计算量大和通信量大的方向发展,上述业务架构具有典型的“云—边—端”协同计算特点,在未来一定会衍生出形式更丰富、更具沉浸感的媒体应用。

视听业务也正在从文化、宣传领域向特殊极端应用延伸,这些特殊极端应用包括面向灾害救援的应急视频通信、面向深空探索的超远距离视频通信和面向车联网的灵活视频通信等。在面向文化、宣传的应用中,信道条件往往是优厚的,其变化往往来源于不同业务对同一信道资源的争夺。然而在特殊极端应用中,业务类型、数量相对较少,但是信道条件、组网条件极其恶劣,甚至很多信道资源在典型应用中往往不被用于视频业务,在组网和传输过程中没有适用的传输协议,因此不论其学术研究还是工程开发,往往在于克服信道上的恶劣条件并解决“有与无”的问题,较难在时延、交互等方面提出要求,在工程化过程中也往往采用了私有协议。

2 国内研究现状

2.1 无线多媒体通信国内现状与发展前沿

目前,学术界和产业界多从视频表征与编码、视频传输两方面解决视频通信系统面临的高清和低时延挑战。

通过挖掘多媒体中的语义信息,并对语义进行编码,可降低多媒体通信的带宽需求。国内学者提出用多模态数据(如声音、文本和图像)来重建说话人视频,实现低码率的视频通信(Suwajanakorn 等, 2017; Yu 等, 2021; Chen 等, 2020)。例如,给定任意语音剪辑或文本信息作为输入,旨在生成具有准确嘴唇同步的有声面部视频。利用多模态输入的互补性和视频帧间的相互依赖关系,大幅降低码率。在此情况下,仅需传输小数据量的语音、文本等,即可在接收端建视频。多媒体传输指的是多媒体信息借助通信网络的通信线路进行信息传输。通信网络多种多样,有数字数据网、综合业务数字网、公共电话网、专网和无线通道等。因此,提高终端对网络的适应性是关键问题。

2.2 多媒体 QoE 度量和计算方法

现有的视频质量评价方法通常以精确性指标 PSNR (peak signal to noise ratio)、SSIM (structural similarity) 等对无线视频通信质量进行衡量。QoE 是衡量视频通信服务能力的根本,视频感知质量是体现用户体验 QoE 的客观度量手段,因此,研究视频感知质量评价方法是必须要解决的核心问题。影响视频感知质量的因素种类各异,数量繁多,例如比特率、峰值速率、缓冲时延、缓冲次数和视频内容等可测参数,这些因素反映了视频的清晰度、流畅度和关注度。因此,创新性地以脑电响应为桥梁,建立可测参数与视频 QoE 之间的定量映射关系,实现视频 QoE 的精准稳定评测,建立了高清视频通信主观质量评价体系。

1) 基于脑电图 (electroencephalogram, EEG) 的视频清晰度评价。传统的视频清晰度主要是通过 PSNR 客观度量,以像素的微观精确性为前提,并不能反映用户的真实感受。借鉴脑神经信号测定方法,从人类数据中分析评价准则,首先利用 EEG 信号的时域特征区分不同的视频质量退化,进一步建立影响视频清晰度的可测参数与 EEG 的关联关系,

从而建立可测参数与视频清晰度之间的映射关系。

2) 基于 EEG 的视频流畅度评价。卡顿是影响视频流畅度的关键因素。然而通信网络中卡顿参数与用户体验之间的定量关系并不明确。视频流畅度是一种主观感受,脑电为用户主观感受的分析提供了一种有效手段。为此,拟提取脑电信号的频域特征区分不同的视频流畅度,建立卡顿等可测参数与 EEG 的关联关系,从而建立可测参数与视频流畅度之间的定量映射关系。

3) 基于 EEG 的视频关注度预测。现有关注度预测方法并未完全揭示影响人类视觉注意力的图像/视频特征,极大限制了视觉注意模型的预测精度。以视频集作为刺激材料并通过适当的实验设计将这些刺激呈现给多位被试者,通过设计滑窗掩膜,记录被试者的脑电信号,通过孪生网络,预测视频的关注度。

2.3 视频编码方法

在视频编码发展过程中,国内学者同样做出大量的贡献:

在码率控制领域,编码技术方面,针对 H. 265 标准, Li 等人 (2013, 2014) 提出了一种新的基于 R- λ 模型的码率控制算法。

在感知编码部分,在 H. 263 标准下, Yang 等人 (2005) 提出了一种感知码率控制方法,通过结合感官特征(亮度、纹理)和认知特征(肤色)获取视频的主观敏感区域,减小这些区域的量化参数来达到分配更多码率的效果。在 H. 264 标准下,针对会话类视频通信, Liu 等人 (2008) 提出了一种基于感兴趣区域的编码资源分配方法,通过帧间差异和肤色检测算法提取视频中的脸部区域,通过优化主观率失真复杂度提升视频的主观质量。

在视频编码的复杂度优化领域,2005—2010 年前后,国内许多研究针对 H. 264 标准提出了编码复杂度优化方案。主要包括对 H. 264 中的两种耗时过程进行简化,即运动估计 ME (Xu 和 He, 2008; Chen 等, 2006) 和模式判决 MD (Kua 等, 2017; Wu 等, 2005)。

针对 HEVC 标准,提出了两类方法:启发式方法和基于机器学习的方法。

1) 启发式方法。根据经验寻找特征,再以此为依据做出决策的方法。HEVC 编码的中间过程中存在一些特征,例如率失真损失、时间和空间相关性

等,这些特征和最终的CTU分割结果存在明显的相关性。因此,在CTU分割之前,可由这些特征预判分割结果,以节省递归检查的时间,实现复杂度优化。在这些方法中,Xiong等人(2014)和Shen等人(2012)通过降低CU分割的复杂度实现优化。Zhang和Ma(2014)基于锥形运动散度(pyramid motion divergence, PMD),利用K邻近分类法,进行CU深度的判决。Sullivan等人(2012)选取一些重要且易于计算的特征,如率失真代价和帧间 $2N \times 2N$ 模式的预测误差等,通过最小化贝叶斯风险准则,对CU分割做出快速判决。Wang等人(2006)通过优化编码过程中多方向帧内预测模式选择来降低编码的复杂度。

2) 基于机器学习的方法。能够从大量数据中自动学习并建立优化的模型。在HEVC帧内模式下,Liu等人(2016b, 2016c)将CU分割过程视为一个二分类问题,并利用逻辑回归算法构造分类器,而Watkinson(2004)则利用支持向量机(support vector machine, SVM)进行二分类。

随着机器学习技术的不断发展,一些深度学习方法已经能够自动提取特征,避免手动提取的缺陷。Weiland和Humayun(2008)提出了基于卷积神经网络(convolutional neural network, CNN)的CU深度预测方法,能够直接根据CU的图像信息训练CNN模型,并做出较为准确的预测。然而,Liu等人(2016a)方法的CNN结构较浅,只包含2个卷积层,分别含有6个和16个 3×3 的卷积核,可训练的参数也比较少。较浅的CNN结构和较少的参数数量,虽然可以在训练数据不足时避免过拟合,但同时也限制了预测精度,因此其性能有待提高。

Zhang等人(2015)提出了一种基于机器学习的HEVC编码单元深度快速决策方法,该方法旨在给定率失真指标下,优化编码单元的复杂度分配。文章首先将四叉树深度决策问题建模为3层二分类问题,之后在现有研究的基础上,改进并提出了一种灵活的编码单元深度决策结构,优化编码复杂度,由此设计出一种由不同二分类器组成的3输出联合分类器,用于控制预测错误的风险。最后在HEVC标准视频序列上进行测试,结果表明,相比HEVC的参考软件HM,这种算法可以将计算复杂度降低28.82%~70.93%,平均降低51.45%,但是码率的平均提升只有1.98%,通过有限码率的提升获得了

复杂度较大的下降,从整体来看优于当时最先进的方案。

Yang等人(2020)提出了一种针对VVC帧内模式的快速算法,分别优化了块分割和帧内模式选择这两个过程。文章首先对参考软件中这两个过程的输出结果进行统计分析,发现随着多分类树的引入,一方面提高了对局部纹理细节敏感度,但也大大增加了这一过程的复杂度;另一方面使得分割更加灵活,保留大块平坦的部分,而细节部分尺寸则更小,降低了码率,同时分割结果还具有一定的空间相关性,即相邻区域选择的分割模式较为一致。根据以上发现,使用决策树来降低块分割的复杂度,在决策树的训练中,筛选并使用3种(共9个)特征,其中包括全局纹理信息的块尺寸、归一化梯度、归一化最大梯度幅度、水平平均梯度、竖直平均梯度、局部纹理信息的上下和左右两部分,以及上下文信息中衡量QT深度和深度的两个指标;而在降低帧内模式选择方面,使用1维梯度下降算法。

2.4 视频传输方法

1) 交互式媒体的研究与应用。国内学者在面向下一代的交互式媒体及其传输方法的研究上做出了许多尝试。首先,交互式媒体本身寄托着用户对场景观察的无限需求,使得其内涵与外延至今仍然在不断扩充。以光场7维模型 $L=f(x, y, z, \lambda, \theta, \varphi, t)$ 为刻画模型,国内学者分别在改变角度(x, y)的多视角视频(杨铀等,2009)、自由视点视频(尤志翔等,2012)、全景视频(罗传飞等,2017)、光场多视角(Liu等,2016)、同时实现角度深度(x, y, z)多维变化的光线空间(蒋刚毅等,2009)、改变光照(θ, φ)的场景重光照(Yang和Liu, 2017; Li等, 2014)、2D与3D自由切换的沉浸式视频(Yang等, 2013)、视觉焦距可变的可调焦视频(Wu等, 2020a, b)等方面展开了大量研究工作,为交互式媒体当前及未来的应用奠定了基础。此外,国内学者也在国际期刊上组织了与交互式媒体相关的专刊,在国际学界引起了广泛的讨论(Lu等, 2016; Yang等, 2020)。

另一方面是传输方法随着媒体应用需求(带宽、延迟、交互和用户数量等)不断变化,所需要的技术手段和方式方法也在一直更新。目前情况下,人们很难给上述每一种交互式媒体应用形式研制对应的传输技术,因此一般统一将交互式媒体作为网

络上的虚拟现实视频业务需求来展开研究。但无论是哪一种形式,交互式媒体的传输带宽大、交互延迟低是其共性特征。为此,当面临大规模网络应用时,传输方法上就需要与传统媒体业务有所区别,比如,通过多路由、多点协同传输的方案(Ge等,2017; Yan,等,2019),基于用户意图发现的网络传输方案(李福亮等,2020),广播传输方案(孟祥昆等,2019)等。这些方法分别从不同的角度满足了特定网络条件下的交互式媒体业务需求。

2)广播式与定制式媒体业务。视听业务的发展不是通信技术发展的唯一原因,但可以从视听业务形式上的变化窥探通信技术发展的路线。首先是视听业务从广播式到定制式的变化。广播是音视频信号的播出方式中使用时间最长、技术最简单和成本最低的一种,从收音机、黑白电视机时代的模拟信号开始,到现在的有线电视网络(community antenna television, CATV)、地面电视广播(digital terrestrial multimedia broadcast, DTMB)和数字调频广播(Chinese depository receipt, CDR)等一直沿用至今。在这种模式中,电视台、广播台是内容生产的核心,也是内容播发的起点。只要在信号能够覆盖到的地方,人们就能够以无差别的方式进行接收。在通信技术的层面,广播式的播发模式不涉及地址问题(IP),更多地涉及调频、频点接入技术。这种播发模式的特点是初期建设投入非常巨大,需要打造完整的内容生产、内容播发和接收终端链条,因此相关产业更多地由政府、大财阀所掌握,一方面是意识形态相对较为单一;另一方面整个生产链条的高投入则给新国际标准的快速推行带来多重阻力。数字化之后,视听业务具备了通过互联网进行播发的能力,因此虽然传统的内容播发产业链条依然可以继续运行,并在新形势下产生像OTT(over-the-top)、交互式网络(IPTV)等高度定制化的视听业务,但人人媒体、自媒体时代以不可阻挡的姿态快速进入了发展轨道,诞生了大量以视听内容分享为载体的社交应用。在这些应用中,每个人都可以生产内容,并在通信信号可以覆盖到的任何地方实现内容播发,且成本较低。自媒体时代的到来,要求通信技术实现无死角覆盖,还需要提供更大的带宽、更低的时延和更稳定的信道质量。近20年来,以视听业务数字化为起点,人们在室外通信快速走过了从2G到5G的技术演进和基础建设,也完成了室内通信Wi-Fi从专

利到应用、从一代到六代的快速更新,并且越往后迭代更新的速度越快。目前,广播式与定制式的视听业务互相之间都不可替代,各自在其擅长的领域发挥着关键作用。如广播式视听业务在重大新闻实时播报、灾害与应急救援信息播发等条件下所具有的广域性、及时性是任何自媒体都无法比拟的,而分享生活、宣传自我的个性化需求也是自媒体所应对的核心。

3 国内外研究进展比较

如今,大量的对话类视频例如FaceTime正涌入人们的生活,不断填充人们的视频通信内容。此外,过去10年见证了视频分辨率和屏幕尺寸的飞速增长。特别是4K/8K超高清这类高分辨率的对话类视频的出现,给如今的带宽资源带来巨大的压力。鉴于视频数据的庞大性和信号带宽的有限性之间日益剧烈的矛盾,所以必须对高清视频进行高效率的压缩。

最早的视频编码标准是国际电信联盟ITU在1984年提出的H.120,然而由于其性能太差而在实践中不可用,因此在1988年ITU又提出了H.261标准,也是世界上第1个实用的视频编码标准。时至今日,当下国际常用的编码标准是H.265。这种编码标准提供了更为高效的编码方式,从而在一定程度上缓解了带宽受限的问题。由于H.265拥有灵活的图像分割规则、并行化视频编码以及其他的一些先进的方法,这种新兴的编码规则比上一代的编码标准H.264/AVC有了一个较大的提升。具体来讲,HEVC依然沿用MPEG系列和H.26x系列所采用的混合编码框架,但HEVC同时在原有的基础上进行了大量的改进和创新,比如基于四叉树结构的编码分割、多角度帧内预测等技术。目前,新一代的编码标准H.266/AVS3也在标准化的进程中,我国和JVET委员会分别于2019年和2020年发布AVS3第1版、VVC标准第1版。其中,VVC相比上一代的编码标准HEVC增加了30多种新的编码工具,覆盖了混合视频编解码系统框架中的每个模块。如VVC的编码单元最大可以覆盖到 128×128 像素亮度区域,而且其块划分除了支持四分树,也支持二分树和三分树,能够精确而灵活地描述物体等。

4 发展趋势与展望

视频通信的发展趋势表现在几个方面:

1) 视频分辨率不断提高。4 K 分辨率的超高清视频正在成为主流, 8 K 视频成为未来媒体的发展趋势。据悉, 2022 年北京冬奥会正在加速推进 AVS3 + 8 K + 5G 超高清视频布局。无线医疗、3 维全息视频会议、抖音和直播等新型应用, 展现出更为丰富和极致的功能体验, 多样化高清应用为视频通信带来了新的挑战。

2) 安全稳定运行。为了应接入设备源发的恶意流量激增引发的安全新问题, 立足通信构筑新的安全架构, 建立安全监测与恶意行为管控协同的新框架, 需解决“通信恶意行为的协同监控架构及其评估模型”这一科学问题, 攻克可信协议构建、恶意通信行为监控等关键技术(葛宁等, 2020)。

3) 物联网(internet of things, IoT)。物联网的广泛部署已经改变了网络服务的开发、部署和运营方式。最近, 大多数先进的物联网设备都配备了视觉传感器, 随后形成了所谓的视觉物联网(V-IoT)。由于需要视觉数据的感测和处理, V-IoT 严重依赖于计算机视觉处理技术。它们已广泛用于各种应用中, 例如工业生产、视频监控、智能交通、精确农业、娱乐以及自动地面和空中车辆。

V-IoT 旨在解决当前的 IoT 对象与复杂的多媒体处理之间的鸿沟, 将多媒体、信息学、通信和网络、统计和计算科学结合在一起。V-IoT 生态系统旨在通过有效地控制物理数据和事物, 通过智能识别、视觉信息分析、通信、监视、跟踪、预测和明智的管理, 有效地“看到”、“理解”和“连接”所有内容, 使网络连接更加可见和有价值。V-IoT 正在改善新兴应用程序和行业经验的可视化和智能水平, 解决了大量物联网设备和海量多媒体数据的负担, 涵盖了从使能技术到未来物联网通信类型的主题, 并研究了未来智能通信和多媒体计算技术。

参考文献 (References)

Balachandran A, Sekar V, Akella A, Seshan S, Stoica I and Zhang H. 2013. Developing a predictive model of quality of experience for internet video. *ACM SIGCOMM Computer Communication Review*,

43(4): 339-350 [DOI: 10.1145/2534169.2486025]

Chen L L, Cui G F, Liu C L, Li Z, Kou Z Y, Xu Y and Xu C L. 2020. Talking-head generation with rhythmic head motion//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, Scotland; Springer: 35-51 [DOI: 10.1007/978-3-030-58545-7_3]

Chen M J, Cormack L K and Bovik A C. 2013. No-reference quality assessment of natural stereopairs. *IEEE Transactions on Image Processing*, 22(9): 3379-3391 [DOI: 10.1109/TIP.2013.2267393]

Chen Z B, Xu J F, He Y and Zheng J L. 2006. Fast integer-pel and fractional-pel motion estimation for H. 264/AVC. *Journal of visual communication and image representation*, 17(2): 264-290 [DOI: 10.1016/j.jvcir.2004.12.002]

Chiang T and Zhang Y Q. 1997. A new rate control scheme using quadratic rate distortion model. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1): 246-250 [DOI: 10.1109/76.554439]

Cho S and Kim M. 2013. Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(9): 1555-1564 [DOI: 10.1109/TCSVT.2013.2249017]

Choi H, Yoo J, Nam J, Sim D and Bajić I V. 2013. Pixel-wise unified rate-quantization model for multi-level rate control. *IEEE Journal of Selected Topics in Signal Processing*, 7(6): 1112-1123 [DOI: 10.1109/JSTSP.2013.2272241]

Choi K and Jang E S. 2012. Early TU decision method for fast video encoding in high efficiency video coding. *Electronics Letters*, 48(12): 689-691 [DOI: 10.1049/el.2012.0277]

Correa G, Assuncao P A, Agostini L V and Da Silva Cruz L A. 2015. Fast HEVC encoding decisions using data mining. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(4): 660-673 [DOI: 10.1109/TCSVT.2014.2363753]

Domański M, Stankiewicz O, Wegner K and Grajek T. 2017. Immersive visual media—MPEG-I: 360 video, virtual navigation and beyond//*Proceedings of 2017 International Conference on Systems, Signals, and Image Processing*. Poznan, Poland; IEEE: 1-9 [DOI: 10.1109/IWSSIP.2017.7965623]

Dutta P, Seetharam A, Arya V, Chetlur M, Kalyanaraman S and Kurose J. 2012. On managing quality of experience of multiple video streams in wireless networks//*Proceedings of 2012 IEEE INFOCOM*. Orlando, USA; IEEE: 1242-1250 [DOI: 10.1109/INF-COM.2012.6195485]

Feng Y, Liu Z M and Li B C. 2012. GestureFlow: QoE-aware streaming of multi-touch gestures in interactive multimedia applications. *IEEE Journal on Selected Areas in Communications*, 30(7): 1281-1294 [DOI: 10.1109/JSAC.2012.120813]

Fu T, Zhang H, Mu F and Chen H B. 2019. Fast CU partitioning algorithm for H. 266/VVC intra-frame coding//*Proceedings of 2019 IEEE International Conference on Multimedia and Expo*. Shanghai, China; IEEE: 55-60 [DOI: 10.1109/ICME.2019.00018]

Galpin F, Racapé F, Jaiswal S, Bordes P, Le Léannec F and François E. 2019. CNN-based driving of block partitioning for intra slices en-

- coding//Proceedings of 2019 Data Compression Conference. Snowbird, USA; IEEE: 162-171 [DOI: 10.1109/DCC.2019.00024]
- Ge N, Chen X and Feng W. 2020. New network architecture for intelligent mobile communications. Bulletin of National Natural Science Foundation of China, 34(2): 150-153 (葛宁, 陈旭, 冯伟. 2020. 智能移动通信新架构探索. 中国科学基金, 34(2): 150-153) [DOI: 10.16262/j.cnki.1000-8217.2020.02.005]
- Ge X H, Pan L H, Li Q, Mao G Q and Tu S. 2017. Multipath cooperative communications networks for augmented and virtual reality transmission. IEEE Transactions on Multimedia, 19(10): 2345-2358 [DOI: 10.1109/TMM.2017.2733461]
- Hassan J A, Hassan M, Das S K and Ramer A. 2012. Managing quality of experience for wireless VOIP using noncooperative games. IEEE Journal on Selected Areas in Communications, 30(7): 1193-1204 [DOI: 10.1109/JSAC.2012.120805]
- He Z H, Kim Y K and Mitra S K. 2001. Low-delay rate control for DCT video coding via ρ -domain source modeling. IEEE Transactions on Circuits and Systems for Video Technology, 11(8): 928-940 [DOI: 10.1109/76.937431]
- Hsu R L, Abdel-Mottaleb M and Jain A K. 2002. Face detection in color images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5): 696-706 [DOI: 10.1109/34.1000242]
- Huang Y W, Hsieh B Y, Chien S Y, Ma S Y and Chen L G. 2006. Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology, 16(4): 507-522 [DOI: 10.1109/TCSVT.2006.872783]
- Jiang G Y, Fan L Z, Yu M and Chen K. 2009. Arbitrary viewpoint rendering based on ray-space interpolation. Acta Electronica Sinica, 37(8): 1799-1803 (蒋刚毅, 范良忠, 郁梅, 陈恳. 2009. 基于光线空间插值的任意视点绘制. 电子学报, 37(8): 1799-1803) [DOI: 10.3321/j.issn:0372-2112.2009.08.031]
- Kannangara C S, Richardson I E G, Bystrom M, Solera J R, Zhao Y F, MacLennan A and Cooney R. 2006. Low-complexity skip prediction for H.264 through Lagrangian cost estimation. IEEE Transactions on Circuits and Systems for Video Technology, 16(2): 202-208 [DOI: 10.1109/TCSVT.2005.859026]
- Kim N, Jeon S, Shim H J, Jeon B, Lim S C and Ko H. 2016. Adaptive keypoint-based CU depth decision for HEVC intra coding//Proceedings of 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Nara, Japan; IEEE: 1-3 [DOI: 10.1109/BMSB.2016.7521923]
- Kua J, Armitage G and Branch P. 2017. A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP. IEEE Communications Surveys and Tutorials, 19(3): 1842-1866 [DOI: 10.1109/COMST.2017.2685630]
- Lee J S and Ebrahimi T. 2012. Perceptual video compression: a survey. IEEE Journal of Selected Topics in Signal Processing, 6(6): 684-697 [DOI: 10.1109/JSTSP.2012.2215006]
- Lee S and Bovik A C. 2003. Fast algorithms for foveated video processing. IEEE Transactions on Circuits and Systems for Video Technology, 13(2): 149-162 [DOI: 10.1109/TCSVT.2002.808441]
- Lei S M, Chen T C and Sun M T. 1994. Video bridging based on H.261 standard. IEEE Transactions on Circuits and Systems for Video Technology, 4(4): 425-437 [DOI: 10.1109/76.313137]
- Leng J, Sun L, Ikenaga T and Sakaida S. 2011. Content based hierarchical fast coding unit decision algorithm for HEVC//Proceedings of 2011 International Conference on Multimedia and Signal Processing. Guilin, China; IEEE: 56-59 [DOI: 10.1109/CMSIP.2011.167]
- Li B, Li H and Li L. 2013. Adaptive bit allocation for R- λ model rate control in HM//Proceedings of the 13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG16 WP3 and ISO/IEC JTC1/SC. Incheon, Korea(South): [s. n.]
- Li B, Li H Q, Li L and Zhang J L. 2014. λ domain rate control algorithm for high efficiency video coding. IEEE Transactions on Image Processing, 23(9): 3841-3854 [DOI: 10.1109/TIP.2014.2336550]
- Li F L, Fan G Y, Wang X W, Liu S C, Xie K and Sun Q. 2020. State-of-the-art survey of intent-based networking. Journal of Software, 31(8): 2574-2587 (李福亮, 范广宇, 王兴伟, 刘树成, 谢坤, 孙琼. 2020. 基于意图的网络研究综述. 软件学报, 31(8): 2574-2587) [DOI: 10.13328/j.cnki.jos.006088]
- Li G N, Liu Y B and Dai Q H. 2014. Free-viewpoint video relighting from multi-view sequence under general illumination. Machine Vision and Applications, 25(7): 1737-1746 [DOI: 10.1007/s00138-013-0559-0]
- Li J L, Zhao X and Yang Y. 2021. Review of interactive video quality assessment methods. ZTE Technology Journal, 2021(1): 44-47 (李继龙, 赵雪, 杨铀. 2021. 交互式视频质量评价方法研究进展. 中兴通讯技术, 2021(1): 44-47)
- Liu D, Wang L Z, Li L, Xiong Z W, Wu F and Zeng W J. 2016a. Pseudo-sequence-based light field image compression//Proceedings of 2016 IEEE International Conference on Multimedia and Expo Workshops. Seattle, USA; IEEE: 1-4 [DOI: 10.1109/ICMEW.2016.7574674]
- Liu D Y, Liu X G and Li Y Y. 2016b. Fast CU size decisions for HEVC intra frame coding based on support vector machines//Proceedings of the 14th IEEE Intl Conference on Dependable, Autonomic, and Secure Computing, 14th Intl Conference on Pervasive Intelligence and Computing, 2nd Intl Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress. Auckland, New Zealand; IEEE: 594-597 [DOI: 10.1109/DASC-PICOM-DataCom-CyberSciTec.2016.168]
- Liu Y, Li Z G and Soh Y C. 2008. Region-of-interest based resource allocation for conversational video communication of H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology, 18(1): 134-139 [DOI: 10.1109/TCSVT.2007.913754]
- Liu Z Y, Yu X Y, Chen S L and Wang D S. 2016c. CNN oriented fast HEVC intra CU mode decision//Proceedings of 2016 IEEE International Symposium on Circuits and Systems. Montreal, Canada; IEEE: 2270-2273 [DOI: 10.1109/ISCAS.2016.7539036]
- Lu J H. 2017. Reflections on wireless communications. ZTE Technology

- Journal, 23(1): 2-5 (陆建华. 2017. 无线通信若干问题的思考. 中兴通讯技术, 23(1): 2-5) [DOI: 10.3969/j.issn.1009-6868.2017.01.001]
- Lu K, Yang Y and Zhen Y. 2016. Stereo data sensing, computation and perception. *Neurocomputing*, 215: 1-2 [DOI: 10.1016/j.neucom.2016.06.010]
- Luo C F, Kong D H, Liu X K, Xu K and Yang H. 2017. A business realization of VR panoramic video in smart home. *Telecommunications Science*, 33(10): 185-193 (罗传飞, 孔德辉, 刘翔凯, 徐科, 杨浩. 2017. 智慧家庭的VR全景视频业务实现. 电信科学, 33(10): 185-193 [DOI: 10.11959/j.issn.1000-0801.2017291])
- Meng X K, Wang Q, Wei N and Fu G T. 2019. Research on virtual reality omnidirectional video system for broadcasting network. *Radio and TV Broadcast Engineering*, 46(7): 78-81 (孟祥昆, 王强, 魏娜, 付光涛. 2019. 面向广播电视网络的虚拟现实全景视频系统方案研究. 广播与电视技术, 46(7): 78-81) [DOI: 10.16171/j.cnki.rtbte.20190007013]
- Pan F, Lin X, Rahardja S, Lim K P, Li Z G, Wu D J and Wu S. 2005. Fast mode decision algorithm for intraprediction in H.264/AVC video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(7): 813-822 [DOI: 10.1109/TCSVT.2005.848356]
- ParandehGheibi A, Médard, Ozdaglar A and Shakkottai S. 2011. Avoiding interruptions—a QoE reliability function for streaming media applications. *IEEE Journal on Selected Areas in Communications*, 29(5): 1064-1074 [DOI: 10.1109/JSAC.2011.110516]
- Ravi S and Larochelle H. 2017. Optimization as a model for few-shot learning//*Proceedings of International Conference on Learning Representations*. [s. l.]: [s. n.]
- Rezaee A, Du Pin Calmon F, Zeger L M and Medard M. 2012. Speeding multicast by acknowledgment reduction technique (SMART) enabling robustness of QoE to the number of users. *IEEE Journal on Selected Areas in Communications*, 30(7): 1270-1280 [DOI: 10.1109/JSAC.2012.120812]
- Ribas-Corbera J and Lei S. 1999. Rate control in DCT video coding for low-delay communications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(1): 172-185 [DOI: 10.1109/76.744284]
- Santos M A, Villalon J and Orozco-Barbosa L. 2012. QoE-Aware multicast mechanism for video communications over IEEE 802.11 WLANs. *IEEE Selected Areas in Communications*, 30(7): 1205-1214 [DOI: 10.1109/JSAC.2012.120806]
- Sheikh H R, Sabir M F and Bovik A C. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11): 3440-3451 [DOI: 10.1109/TIP.2006.881959]
- Shen X L, Yu L and Chen J. 2012. Fast coding unit size selection for HEVC based on Bayesian decision rule//2012 Picture Coding Symposium. Krakow, Poland: IEEE: 453-456 [DOI: 10.1109/PCS.2012.6213252]
- Smolic A. 2011. 3D video and free viewpoint video—from capture to display. *Pattern Recognition*, 44(9): 1958-1968 [DOI: 10.1016/j.patcog.2010.09.005]
- Stockhammer T. 2011. Dynamic adaptive streaming over HTTP—standards and design principles//*Proceedings of the 2nd Annual ACM Conference on Multimedia Systems*. San Jose, USA: ACM: 133-144 [DOI: 10.1145/1943552.1943572]
- Sullivan G J, Ohm J R, Han W J and Wiegand T. 2012. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12): 1649-1668 [DOI: 10.1109/TCSVT.2012.2221191]
- Suwajanakorn S, Seitz S M and Kemelmacher-Shlizerman I. 2017. Synthesizing Obama: learning lip sync from audio. *ACM Transactions on Graphics*, 36(4): 95 [DOI: 10.1145/3072959.3073640]
- Tanimoto M. 2006. Overview of free viewpoint television. *Signal Processing: Image Communication*, 21(6): 454-461 [DOI: 10.1016/j.image.2006.03.009]
- Thies J, Zollhöfer M, Stamminger M, Theobalt C and Nießner M. 2016. Face2Face: real-time face capture and reenactment of RGB videos//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 2387-2395 [DOI: 10.1109/CVPR.2016.262]
- Tong Y, Liao W H and Ji Q. 2007. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10): 1683-1699 [DOI: 10.1109/TPAMI.2007.1094]
- Toni L, Cosman P C and Milstein L B. 2012. Channel coding optimization based on slice visibility for transmission of compressed video over OFDM Channels. *IEEE Journal on Selected Areas in Communications*, 30(7): 1172-1183 [DOI: 10.1109/JSAC.2012.120803]
- Vanne J, Viitanen M and Hämläinen T D. 2014. Efficient mode decision schemes for HEVC inter prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(9): 1579-1593 [DOI: 10.1109/TCSVT.2014.2308453]
- Vetro A, Sun H F and Wang Y. 1999. MPEG-4 rate control for multiple video objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(1): 186-199 [DOI: 10.1109/76.744285]
- Vetro A, Wiegand T and Sullivan G J. 2011. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE*, 99(4): 626-642 [DOI: 10.1109/JPROC.2010.2098830]
- Wandell B and Thomas S. 1997. *Foundations of Vision*. USA: Sinauer Associates, Inc.
- Wang H L, Kwong S and Kok C W. 2007. An efficient mode decision algorithm for H.264/AVC encoding optimization. *IEEE Transactions on Multimedia*, 9(4): 882-888 [DOI: 10.1109/TMM.2007.893345]
- Wang J F, Wang J C, Chen J T, Tsai A C and Paul A. 2006. A novel fast algorithm for intra mode decision in H.264/AVC encoders//2006 IEEE International Symposium on Circuits and Systems. Kos: IEEE: 4 [DOI: 10.1109/ISCAS.2006.1693380]

- Watkinson J. 2004. The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4. Oxford: Elsevier/Focal Press
- Weiland J D and Humayun M S. 2008. Visual prosthesis. Proceedings of the IEEE, 96 (7): 1076-1084 [DOI: 10.1109/JPROC.2008.922589]
- Wiegand T, Lightstone M, Mukherjee D, Campbell T G and Mitra S K. 1996. Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard. IEEE Transactions on Circuits and Systems for Video Technology, 6 (2): 182-190 [DOI: 10.1109/76.488825]
- Wiegand T, Sullivan G J, Bjontegaard G and Luthra A. 2003. Overview of the H.264/AVC video coding standard. IEEE Transactions on Circuits and Systems for Video Technology, 13 (7): 560-576 [DOI: 10.1109/TCSVT.2003.815165]
- Wilson G M and Sasse M A. 2000. Do users always know what's good for them? Utilising physiological responses to assess media quality//McDonald S, Waern Y and Cockton G, eds. People and Computers XIV—Usability or Else! London: Springer: 327-339 [DOI: 10.1007/978-1-4471-0515-2_22]
- Wu D, Pan F, Lim K P, Wu S, Li Z G, Lin X, Rahardja S and Ko C C. 2005. Fast intermode decision in H.264/AVC video coding. IEEE Transactions on Circuits and Systems for Video Technology, 15 (7): 953-958 [DOI: 10.1109/TCSVT.2005.848304]
- Wu K J, Liu Q, Yin Y G and Yang Y. 2020a. Gaussian guided inter prediction for focal stack images compression//Proceedings of 2020 Data Compression Conference. Snowbird, USA: IEEE: 63-72 [DOI: 10.1109/DCC47342.2020.00014]
- Wu K J, Yang Y, Yu M and Liu Q. 2020b. Block-wise focal stack image representation for end-to-end applications. Optics Express, 28 (26): 40024-40043 [DOI: 10.1364/OE.413523]
- Xiong J, Li H L, Wu Q B and Meng F M. 2014. A fast HEVC inter CU selection method based on pyramid motion divergence. IEEE Transactions on Multimedia, 16 (2): 559-564 [DOI: 10.1109/TMM.2013.2291958]
- Xu J Z, Joshi R and Cohen R A. 2016. Overview of the emerging HEVC screen content coding extension. IEEE Transactions on Circuits and Systems for Video Technology, 26 (1): 50-62 [DOI: 10.1109/TCSVT.2015.2478706]
- Xu X Z and He Y. 2008. Improvements on fast motion estimation strategy for H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology, 18 (3): 285-293 [DOI: 10.1109/TCSVT.2008.918122]
- Yan J J, Wu D P, Wang H G and Wang R Y. 2019. Multipoint cooperative transmission for virtual reality in 5G new radio. IEEE MultiMedia, 26 (1): 51-58 [DOI: 10.1109/MMUL.2018.2879592]
- Yang H, Shen L Q, Dong X C, Ding Q, An P and Jiang G Y. 2020. Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding. IEEE Transactions on Circuits and Systems for Video Technology, 30 (6): 1668-1682 [DOI: 10.1109/TCSVT.2019.2904198]
- Yang X K, Lin W S, Lu Z K, Lin X, Rahardja S, Ong E and Yao S S. 2005. Rate control for videophone using local perceptual cues. IEEE Transactions on Circuits and Systems for Video Technology, 15 (4): 496-507 [DOI: 10.1109/TCSVT.2005.844458]
- Yang X K, Lin W S, Lu Z K, Lin X, Rahardja S, Ong E and Yao S S. 2005. Rate control for videophone using local perceptual cues. IEEE Transactions on Circuits and Systems for Video Technology, 15 (4): 496-507 [DOI: 10.1109/TCSVT.2005.844458]
- Yang Y and Dai Q. 2010. Contourlet-based image quality assessment for synthesised virtual image. Electronics Letters, 46 (7): 492-494 [DOI: 10.1049/el.2010.3522]
- Yang Y and Liu Q. 2017. Illumination attributes coding for virtual reality broadcasting system//Proceedings of 2017 Data Compression Conference. Snowbird, USA: IEEE: 469-469 [DOI: 10.1109/DCC.2017.11]
- Yang Y, Liu Q, Gao Y, Xiong B B, Yu L, Luan H B, Ji R R and Tian Q. 2013. Stereotime: a wireless 2D and 3D switchable video communication system//Proceedings of the 21st ACM international conference on Multimedia. Barcelona, Spain: ACM: 473-474 [DOI: 10.1145/2502081.2502275]
- Yang Y, Wang X, Liu Q, Xu M L and Wu W. 2016. User models of subjective image quality assessment on virtual viewpoint in free-viewpoint video system. Multimedia Tools and Applications, 75 (20): 12499-12519 [DOI: 10.1007/s11042-014-2321-7]
- Yang Y, Yang J Y and Adeli E. 2020. Guest editorial: AI-powered 3D vision. Journals and Magazines, 14 (12): 2627-2629 [DOI: 10.1049/iet-ipr.2020.1194]
- Yang Y, Yu M and Jiang G Y. 2009. Survey on interactive three-dimensional video systems. Journal of Computer-Aided Design and Computer Graphics, 21 (5): 569-578 (杨铀, 郁梅, 蒋刚毅. 2009. 交互式三维视频系统研究进展. 计算机辅助设计与图形学学报, 21 (5): 569-578)
- Yoo H M and Suh J W. 2013. Fast coding unit decision algorithm based on inter and intra prediction unit termination for HEVC//Proceedings of 2013 IEEE International Conference on Consumer Electronics. Las Vegas, USA: IEEE: 300-301 [DOI: 10.1109/ICCE.2013.6486903]
- You Z X, An P and Zhang Z Y. 2012. Key technologies of 3D auto-stereoscopic display systems and corresponding applications in China pavilion of 2010 expo. Video Engineering, 36 (2): 19-23 (尤志翔, 安平, 张兆杨. 2012. 3D 自由视点视频技术及其在中国馆中的应用. 电视技术, 36 (2): 19-23 [DOI: 10.3969/j.issn.1002-8692.2012.02.005])
- Yu L Y, Yu J, Li M Y and Ling Q. 2021. Multimodal inputs driven talking face generation with spatial-temporal dependency. IEEE Transactions on Circuits and Systems for Video Technology, 31 (1): 203-216 [DOI: 10.1109/TCSVT.2020.2973374]
- Zakharov E, Shysheya A, Burkov E and Lempitsky V. 2019. Few-shot adversarial learning of realistic neural talking head models//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9459-9468 [DOI: 10.1109/ICCV.2019.00955]

- Zhang H and Ma Z. 2014. Fast intra mode decision for high efficiency video coding (HEVC). IEEE Transactions on Circuits and Systems for Video Technology, 24(4): 660-668 [DOI: 10.1109/TCSVT.2013.2290578]
- Zhang J N, Zeng X F, Wang M M, Pan Y S, Liu L, Liu Y, Ding Y and Fan C J. 2020. FReeNet: multi-identity face reenactment//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5326-5335 [DOI: 10.1109/CVPR42600.2020.00537]
- Zhang Y and Yu L. 2012. Recent MPEG standardization activities on 3D video coding. ZTE Communications, 10(2): 9-12
- Zhang Y, Kwong S, Wang X, Yuan H, Pan Z Q and Xu L. 2015. Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding. IEEE Transactions on Image Processing, 24(7): 2225-2238 [DOI: 10.1109/TIP.2015.2417498]
- Zhu W J, Ding W P, Xu J Z, Shi Y H and Yin B C. 2014. Screen content coding based on HEVC framework. IEEE Transactions on Multimedia, 16(5): 1316-1326 [DOI: 10.1109/TMM.2014.2315782]

作者简介



陶晓明, 1981年生, 女, 教授, 博士生导师, 主要研究方向为无线多媒体通信理论及关键技术应用。

E-mail: taoxm@tsinghua.edu.cn



刘文予, 通信作者, 男, 教授, 博士生导师, 主要研究方向为人工智能、机器学习、计算机视觉、机器人。

E-mail: liuwu@hust.edu.cn

杨铀, 男, 教授, 博士生导师, 主要研究方向为多媒体通信、多媒体技术、计算摄像学。E-mail: yangyou@hust.edu.cn

徐迈, 男, 教授, 博士生导师, 主要研究方向为视频压缩、传输理论与关键技术。E-mail: maixu@buaa.edu.cn

段一平, 女, 助理研究员, 主要研究方向为无线多媒体通信。E-mail: yipingduan@mail.tsinghua.edu.cn

黄丹蓝, 女, 博士后, 主要研究方向为无线多媒体通信。E-mail: huangdl@mail.tsinghua.edu.cn