

# 交互式视频质量 评价方法研究进展

## A Review of Interactive Video Quality Assessment Methods



李继龙 /LI Jilong<sup>1</sup>, 赵雪 /ZHAO Xue<sup>2</sup>, 杨铀 /YANG You<sup>3</sup>

1. 国家广播电视总局广播电视科学研究院, 中国 北京 100086;
  2. 武汉理工大学, 中国 武汉 430070;
  3. 华中科技大学, 中国 武汉 430074)
- (1. Academy of Broadcasting Science, National Radio and Television Administration, Beijing 100086, China;  
2. Wuhan University of Technology, Wuhan 430070, China;  
3. Huazhong University of Science and Technology, Wuhan 430074, China)

**摘要:** 在交互式视频应用快速发展的同时, 如何评价视频质量成为当前亟待解决的挑战性难题, 其成果对整个多媒体通信系统的各环节技术发展具有关键作用。从主观质量评价、客观质量评价两个角度综述了当前交互式视频质量评价的研究与应用现状, 其中主观质量评价方法包括主观视频质量评价数据库、主观视频质量评价打分与计算机制, 客观质量评价方法则包括视觉信号处理与分析、深度学习机制下的评价与建模方法等。在总结上述研究方法与成果的基础上, 展望了本领域的研究与发展。

**关键词:** 交互式视频; 视频质量评价; 主观质量评价; 客观质量评价

**Abstract:** With the rapid development of interactive video applications, the research on interactive video quality assessment becomes an urgent challenge to the community, because it is helpful to the development of other modules in the multimedia communication system. The researches on interactive video quality assessment via both objective and subjective methodologies are surveyed. Extant methods are then reviewed, including databases of subjective video quality assessment, score and computation mechanism of subjective video quality assessment, visual signal processing and analysis of objective video quality assessment, and deep learning based methods of objective video quality assessment. Based on the above surveys, future directions and open problems on the research of interactive video quality assessment are discussed.

**Keywords:** interactive video; video quality assessment; objective quality assessment; subjective quality assessment

DOI: 10.12142/ZTETJ.202101010  
网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210114.1456.004.html>

网络出版日期: 2021-01-15  
收稿日期: 2020-12-21

在视频通信系统中, 视频源与信道之间如同水源与水管的关系。如何克服信源与信道之间的各种矛盾, 从而给用户提供更加优质的视频服务, 一直以来都是业界追求的目标。然而,

自电视机诞生之后的 100 多年, 视频服务一直是被动接受的模式, 其发展变化无非只是从不同地点的同一时刻接受同一服务(广播电视), 变成在不同地点的不同时刻接受同一服务(基于互联网协议的互联网视频)。人们在观看视频的过程中始终无法主动改变正在播出的视频内容, 这使得该领域的研究与应用给人们所提供的想象

空间非常有限。近些年来, 在线视频服务开始从被动式向主动式转变, 出现了云端虚拟现实(VR)、云游戏等 Cloud VR 业务及以面向在线教育、在线会议的多视点视频业务等, 用户可以在终端通过“人-机-内容”交互的方式主动改变所看到的视频内容。在线视频服务有望在可预见的未来实现“千人千面”的特点。为此, 与这

基金项目: 国家重点研发计划(2017YFC0806202); 国家自然科学基金(61971203); 武汉市应用基础前沿(2019020701011422、2020020601012222)

类视频有关的质量评价问题开始涌现,成为传统视频质量评价研究领域中的新方向、新课题。

视频的质量评价主要面向终端用户,因此该环节位于多媒体通信系统的最末端,其目的在于为多媒体通信系统前端的采集、处理、编码等环节提供一个可供参考的评价依据,从而构成处理流程上的闭环。视频质量评价的研究对象总体而言可分为两个层面:一个是解决信道质量与显示质量之间的关系,主要考察的是用于描述信道质量的多种因素与视频重建客观质量之间的作用机制,一般称为关键质量指标(KQI);另一个是解决显示质量与用户感受之间的关系,主要考察的是用于描述图像重建质量的多种因素与人类视觉系统响应质量之间的作用机制,一般称为视频质量评价(VQA)。相比而言,由于涉及信道质量,因此关注通信终端应用的学者与企业比较重视对KQI的研究;而视频的信息失真与质量重建更多的是由有损压缩或视频处理的环节所带来的,因此涉及上述领域的学者和企业比较重视对VQA的研究。Cloud VR及多视点视频业务作为产业界中的新生事物,目前在KQI方面的研究较少,尚不构成体系;但是这两种视频形式在学术界的研究中已经历过较长的历程,因此在VQA方面的成果已具有一定规模,本文的工作也主要集中于此。

## 1 交互式视频质量的主观评价

如前所述,视频的最终接收者是用戶,因此视频质量的好坏应由人来决定。然而,终端用户因个人知识背景、观看环境,甚至观看时的情绪千差万别,其对视频质量优劣的反应也会各不相同,因此如何对视频质量进行有效的评价是一个极具挑战性的难题<sup>[1]</sup>。一般情况下,其研究可

分为主观、客观质量评价两个大的方向。视频主观质量评价采用“自顶向下”“以人为本”的研究模式,探索涉及人本体相关的因素与视频质量之间的联系;客观质量评价采用“自底向上”“以技术为本”的研究模式,探索和构建视频中的视觉信号与视频质量之间的映射关系。两种模式互为支撑,不可相互替代。

视频主观质量评价从技术手段上可通过邀请主观测试人员采取某种规定的打分机制,对具有不同失真类型、等级的视频进行打分,这涉及主观质量评价数据库、主观质量评价打分和计算机制等相关工作。在打分与计算机制方面,国际电信联盟无线电通信部门(ITU-R)和电信标准分局(ITU-T)制定了通用的主观质量打分与计算机制,如ITU-R BT.500-13<sup>[2]</sup>和ITU-T P.910<sup>[3]</sup>等。在打分的操作过程中,根据刺激方式的不同,主观质量评价方法可以分为单刺激、双刺激和多刺激的方式。单刺激即在一次打分过程中只播放失真视频,双刺激则在一次打分过程中随机播放参考、失真视频。在不同的标准机制中,操作流程略有不同。如ITU-R BT.500-13设计了单刺激连续质量估计方法(SSCQE)、双刺激失真分级方法(DSIS)、双刺激连续质量分级方法(DSCQS)、同时双刺激连续估计方法(SDSCE)等。ITU-T P.910设计了用于评价失真视频的打分方法,包括绝对类别打分法(ACR)、隐藏参考图绝对类别打分法(ACR-HR)、降质类别打分法(DCR)、匹配对比较法(PC)等。打分时可以采用百分制或等级打分制,其中较为常用的等级打分制提供了5个感受等级,即5(优秀)、4(良好)、3(一般)、2(差)、1(很差)。主观测试人员打分后,对异常数据进行处理,便可得到每个视频的平均主观意见得

分(MOS),然后再进一步通过计算失真图像与原始图像的MOS分数差得到差异平均主观意见得分(DMOS)。在绝大多数情况下,通过主观质量评价方法建立起来的主观数据库包含失真图像及其MOS/DMOS,为图像的客观质量评价方法提供了测试依据,而且人们一般也认为主观分数最接近图像的用户对视频质量的感知。目前,上述打分与计算机制是针对传统的非交互式的图像、视频业务的,并没有专门针对交互式视频设计与之相对应的打分与计算机制。虽然如此,大多数科研与工程技术人员认为上述打分与计算机制是与显示内容无关的,因此还可以将这些方法继续沿用至交互式视频的主观评价研究与应用中。在影响交互式视频主观质量的关键因素中,目前尚未有明确的研究成果,一些终端企业一方面参考了立体视频舒适度评价中的如眩晕、分辨率等因素,另一方面也站在企业自身的角度提出了包括黑边、交互延迟、卡顿等方面的因素<sup>[4]</sup>。这些工作为本领域未来的研究与发展提供了较好的思路。

主观质量评价数据库的建立是开展质量评价打分的前提,需要就应用过程中典型的情况进行表达,如分辨率、失真类型、失真等级等。目前针对交互式视频的主观质量评价所建立的数据库较少,其建立经历了从立体视频到交互式视频的发展过程。WANG X.等考虑了非对称失真特性对视觉感知质量的影响,建立了双目立体图像主观质量评价数据库<sup>[5]</sup>。该数据集包含4种不同的失真类型、10个场景共400组失真图像对。A. K. MOORTHY等针对对称失真,建立了包含20个场景共计365组失真图像对的LIVE-Phase-I数据集<sup>[6]</sup>。CHEN M. J.等同时考虑了对称和非对称失真特性的影响,建立了包含8个场景和360组失真图像对

的 LIVE-Phase-II 数据集<sup>[7]</sup>。针对立体图像质量评价的客观评价模型的建模需求, WANG J. H. 等建立的 Waterloo-IVC-3D 图像质量数据库<sup>[8]</sup>, 探索了信号失真分别对单目图像和立体图像视觉感知质量的影响。针对立体视频系统中的编码压缩方案对视觉感知质量的影响, WANG J. H. 等建立了 Waterloo-IVC-3D 视频质量数据库<sup>[9]</sup>。随着虚拟现实 (VR) 业务的广泛应用, 3D VR 内容的视觉质量评价得到了广泛关注。近期, CHEN M. 等建立了 LIVE-3D-VR 图像质量数据库<sup>[10]</sup>, 该数据库包含了 15 个 3D VR 场景、6 种失真类型, 共计 450 组失真图像的用户评分和眼动数据。前述工作主要针对自然场景内容, 未考虑交互视频中存在的虚拟视点绘制等过程对视觉感知质量的影响。在此基础上, YANG Y. 等以交互过程中所产生的虚拟视点为切入口, 建立了虚拟视点视频主观质量评价数据库<sup>[11-12]</sup>。该数据库主要考虑了多视点视频在彩色图、深度图压缩联合失真的情况下对虚拟视点图像绘制的相关影响, 重点考察了量化参数 (QP) 从 22 到 47, 且  $\Delta QP=5$  的条件下, 对 5 个不同分辨率的视频进行的失真处理。上述数据库的建立, 为本领域研究工作奠定了非常重要的基础。但是, 由于主观质量评价数据库的建立是一个极其耗费资源、投入大见效慢的工作, 受到各种外部因素的影响, 该方向的工作在近年来的推进相对迟缓。

## 2 交互式视频质量的客观评价

客观质量评价的目标在于克服主观质量评价对人本身的依赖, 仅依靠对视频信号的分析与计算即可实现视频质量的评价, 从而使得视频质量评价从分时、分空间的人为操作变成当时当地的自动计算, 这样可以大大提升多媒体通信系统的处理效率。

近年来, 交互式视频的客观质量评价以 360° VR 视频为主, 分别以该视频的球面映射 (SP)、等距柱面映射 (ERP)、立方体映射 (CMP) 等 3 种不同的方式为载体, 在其基础上提取视觉特征并加以建模, 来实现客观质量的评价。例如, 球面峰值信噪比 (S-PSNR)<sup>[13]</sup>、加权峰值信噪比 (WS-PSNR)<sup>[14]</sup> 等都是传统的峰值信噪比计算的基础上进行了微调, 以适应 VR 视频的应用。但是, 这些方法还是无法避免视频客观质量评价的典型问题, 即信号的失真不能代表视觉主观感受上的失真程度。为此, CHEN S. J. 在结构相似性度量的基础上提出了球面结构相似性 (S-SSIM) 度量模型, 能够取得比 WS-PSNR 更加贴近人眼主观感受的性能效果<sup>[15]</sup>。这种方法较为直观, 主要是将 SSIM 方法应用到了 SP 上, 因此研究人员认为应该还会有更好的处理模式来解决上述问题。在这种思路的影响下, 利用深度学习的方式进行视频质量评价是一种快速见效的研究手段。如 ZHANG L. 提出了综合局部描述子的图像质量评价方法 (IL-NIQE)<sup>[16]</sup>、LIU L. X. 提出了朝向梯度下的图像质量评价方法 (OG-IQA)<sup>[17]</sup>, 他们都通过反向传播神经网络来将图像的特征映射成为图像的客观质量。此外, 利用信号分析的方法进行 VQA 的也不在少数, 如 XUE W. 提出了用梯度幅值和高斯-拉普拉斯算子进行建模的方法<sup>[18]</sup>, A. MITTAL 提出了自然场景统计失真的方法<sup>[19]</sup>, YANG Y. 提出了基于 Counterlet 小波的方法<sup>[11, 20]</sup>、相似性评估法<sup>[21]</sup>等。这些方法虽然在计算效率上具有较好的性能, 但是它们对 SP 与二维图像之间相互转换时所具有的视觉失真缺乏有效的分析, 因此其最终的表现性能仍然有待提升。为此, H. T. LIM 提出了一种基于对抗生成网络的 VR 视频

质量评价方法, 将多种压缩失真、位置信息、视觉特征进行了融合, 取得了较好的计算效果<sup>[22-23]</sup>。

上述研究工作主要针对 SP 模式展开, 对 ERP 和 CMP 的模式研究较少。值得注意的是, SP 是一种平面与球面的相互映射过程, 虽然这种映射符合当前 VR 应用的工程需求, 但是存在着较多的几何失真。在这种本身就具有失真的图像上进行 VQA 计算, 是值得商榷的。相比于 ERP, SP 具有更小的失真, 而 CMP 的失真则几乎可以忽略不计。如何在这两种映射的基础上进行 VQA 的建模与计算, 并与 SP 进行有效关联, 是一个值得探索的方向。

## 3 交互式视频的发展趋势

在信息传递的各种形式中, 视听信号更容易让人们理解, 因此也成为了现实世界中信息的主要载体。自从视听业务以数字信号播出以来, 音视频信号在数字设备中的应用变得更加便利。这导致视听业务的表现形式越来越丰富, 人们对视听服务的需求不断激增, 这也倒逼着传统的用于承载音视频业务的通信方式不断发展。近些年来, 通信技术的不断发展, 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。信道越来越宽, 传输速率越来越快, 通信变得无处不在, 这些都使得信源与信道之间的抱团滚动式发展产生越来越大的影响力。自由视点电视的概念于 1996 年被提出, 它认为观众应该改变观看的视角, 从被动接收到主动改变所观看的内容, 形成千人千面的视觉效果<sup>[24]</sup>。虽然上述工作未能带来商业价值, 但是这个交互式媒体的思路与目前低时延、大带宽的通信技术相结合, 在近几年形成了 VR、云游戏、云主机的高交互视听业务, 它和在 2020 年新冠肺炎疫情期间发挥关键作用的在线教育、直

播连麦、在线会议、远程医疗等交互式视听业务模式一起开始逐渐被用户所接纳。未来媒体势必以千人千面为目标,朝着大数据量、大计算量、大通信量的方向发展。上述业务架构具有“云-边-端”协同计算特点,在未来一定会衍生出更丰富的媒体应用。

为了在这些关键应用中保障用户的体验,增强用户对交互式视频的粘滞度,无论是KQI还是VQA,仍有一些问题值得深入研究、探讨。

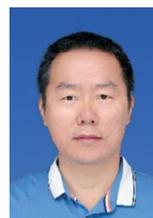
## 致谢

本文的工作得到深圳大学计算机学院王旭副教授的支持,在此特别表示感谢。

## 参考文献

- [1] HUYNH-THU Q, GARCIA M N, SPERANZA F, et al. Study of rating scales for subjective quality assessment of high-definition video [J]. IEEE transactions on broadcasting, 2011, 57(1): 1-14. DOI:10.1109/tbc.2010.2086750
- [2] ITU. Methodology for the subjective assessment of the quality of television pictures, Recommendation ITU-R BT.500-13 [EB/OL]. [2020-12-20]. [https://www.itu.int/dms\\_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-1!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-1!!PDF-E.pdf)
- [3] ITU. Subjective video quality assessment methods for multimedia applications, ITU-T P. 910 [EB/OL]. [2020-12-20]. <https://www.itu.int/rec/T-REC-P.910/en>
- [4] Cloud VR 用户体验与评测白皮书 [EB/OL]. [2020-12-20]. <https://www.huawei.com/minisite/static/cloud-vr-user-experience-evaluation-white-paper-cn.pdf>
- [5] WANG X, YU M, YANG Y, et al. Research on subjective stereoscopic image quality assessment [C]//Multimedia Content Access: Algorithms and Systems III. San Jose, CA, USA: SPIE, 2009: 18-22. DOI:10.1117/12.807641
- [6] MOORTHY A K, SU C C, MITTAL A, et al. Subjective evaluation of stereoscopic image quality [J]. Signal processing: image communication, 2013, 28(8): 870-883. DOI: 10.1016/j.image.2012.08.004
- [7] CHEN M J, SU C C, KWON D K, et al. Full-reference quality assessment of stereopairs accounting for rivalry [J]. Signal processing: image communication, 2013, 28(9): 1143-1155. DOI:10.1016/j.image.2013.05.006
- [8] WANG J H, REHMAN A, ZENG K, et al. Quality prediction of asymmetrically distorted stereoscopic 3D images [J]. IEEE transactions on image processing, 2015, 24(11): 3400-3414. DOI: 10.1109/tip.2015.2446942
- [9] WANG J H, WANG S Q, WANG Z. Asymmetrically compressed stereoscopic 3D videos: quality assessment and rate-distortion performance evaluation [J]. IEEE transactions on image processing, 2017, 26(3): 1330-1343. DOI: 10.1109/tip.2017.2651387
- [10] CHEN M, JIN Y, GOODALL Y, et al. Study of 3D virtual reality picture quality [J]. IEEE journal of selected topics in signal processing, 2020, 14(1):89-102
- [11] YANG Y, DAI Q. Contourlet-based image quality assessment for synthesised virtual image [J]. Electronics letters, 2010, 46(7): 492-494. DOI: 10.1049/el.2010.3522
- [12] YANG Y, WANG X, LIU Q, et al. User models of subjective image quality assessment on virtual viewpoint in free-viewpoint video system [J]. Multimedia tools and applications, 2016, 75(20): 12499-12519. DOI: 10.1007/s11042-014-2321-7
- [13] YU M, LAKSHMAN H, GIROD B. A framework to evaluate omnidirectional video coding schemes [C]//2015 IEEE International Symposium on Mixed and Augmented Reality. Fukuoka, Japan: IEEE, 2015: 31-36. DOI:10.1109/ismar.2015.12
- [14] SUN Y, LU A, YU L. AHG8: WS-PSNR for 360 video objective quality evaluation: ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0040 [S]. 2016
- [15] CHEN S J, ZHANG Y X, LI Y M, et al. Spherical structural similarity index for objective omnidirectional video quality assessment [C]//2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego, CA, USA: IEEE, 2018: 1-6. DOI:10.1109/icme.2018.8486584
- [16] LIN ZHANG, LEI ZHANG, BOVIK A C. A feature-enriched completely blind image quality evaluator [J]. IEEE transactions on image processing, 2015, 24(8): 2579-2591. DOI:10.1109/tip.2015.2426416
- [17] LIU L X, HUA Y, ZHAO Q J, et al. Blind image quality assessment by relative gradient statistics and adaboosting neural network [J]. Signal processing: image communication, 2016, 40: 1-15. DOI:10.1016/j.image.2015.10.005
- [18] XUE W, MOU X, ZHANG L, et al. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features [J]. IEEE transactions on image processing, 2014, 23(11): 4850-4862. DOI: 10.1109/tip.2014.2355716
- [19] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain [J]. IEEE transactions on image processing, 2012, 21(12): 4695-4708. DOI: 10.1109/tip.2012.2214050
- [20] 蒋刚毅, 王旭, 杨铀, 等. 基于Contourlet的质降图像质量评价模型 [J]. 光子学报, 2009, 20(5):1658-1662
- [21] 黄大江, 郁梅, 杨铀, 等. 基于相似度的立体图像对中心视点图像质量评价方法 [J]. 光子学报, 2008, 37(8):1673-1697
- [22] LIM H T, KIM H G, RA Y M. VR IQA Net: deep virtual reality image quality assessment using adversarial learning [C]//IEEE international conference on acoustics, speech and signal processing. Calgary, AB, Canada: IEEE, 2018: 6737-6741. DOI: 10.1109/ICASSP.2018.8461317
- [23] KIM H G, LIM H T, RO Y M. Deep virtual reality image quality assessment with human perception guider for omnidirectional image [J]. IEEE transactions on circuits and systems for video technology, 2020, 30(4): 917-928. DOI: 10.1109/tcsvt.2019.2898732
- [24] TANIMOTO M. FTV: free-viewpoint television [J]. Signal processing: image communication, 2012, 27(6): 555-570. DOI: 10.1016/j.image.2012.02.016

## 作者简介



**李继龙**, 国家广播电视总局广播电视科学研究院正高级工程师、学术带头人; 主要研究工作包括融合媒体、5G 广播电视、广播电视融合网、无线数字广播、信道编码和调制技术研究等; 曾参与多项国家、部委重要项目研发工作, 作为主要研究人员参与了有线/无线卫星融合网、卫星直播标准和数字音频广播标准的研究与制定; 曾获得广电总局“科技创新奖”一等奖一项、二等奖两项, “王选新闻科学技术奖”一等奖两项、二等奖一项; 发表论文 40 余篇, 出版著作 3 部, 获得授权国家发明专利 6 项。



**赵雪**, 武汉理工大学信息工程学院在读研究生; 主要从事机器学习、深度学习领域的研究工作。



**杨铀**, 华中科技大学电子信息与通信学院教授、博士生导师, 中国图象图形学会图象视频处理与通信专委会秘书长; 主要从事以视觉感知与计算为核心的计算机视觉、计算摄影学、立体视频系统等方面的研究工作; 2012 年获教育部高等学校科技成果转化一等奖, 2018 年当选英国国际工程技术学会会士 (IET Fellow), 2020 年获 TET 创新技术奖中“通信与信息技术”领域杰出创新奖; 主持和参与包括国家重点研发计划、国家自然科学基金面上项目、“863”项目、国家重大专项、国家重大科技成果转化等在内的项目 20 余项; 发表论文 80 余篇, 获得授权国家发明专利 24 项。