融合多层特征的窗口 6DoF 合成视频质量评价

唐婷琰[1],邹文辉[2],彭宗举[1,2*],陈 芬[1,2],金充充[2]
(1. 重庆理工大学电气与电子工程学院,重庆 400054;2. 宁波大学信息科学与工程学院,浙江宁波 315211)

摘 要: 六自由度(Six Degrees of Freedom, 6DoF)视频允许用户从全方位、任意视角身临其境体验场景,是下一代沉浸式视频产业的发展方向. 部分自由度受限的窗口 6DoF 视频近年来成为研究热点,本文提出面向窗口 6DoF 合成视频的主观数据库和客观质量评价方法. 在主观数据库方面,构建了包含两种交互路径不适性失真、四种绘制失真和四种压缩失真的窗口 6DoF 合成视频主观质量数据库 Windowed-6DoF,并开展主观质量测试及结果分析. 在客观质量评价方法方面,设计了一种融合多层特征的窗口 6DoF 合成视频无参考客观质量评价方法. 采用切比雪夫矩提取视频时域切片上的底层形状特征;采用 Resnet-50 网络提取视频的时域、空域高层语义特征并进行降维处理;最后采用随机森林将底层形状特征和高层语义特征进行融合,且训练得到窗口 6DoF 合成视频的客观质量评价模型. 在提出的数据库 Windowed-6DoF 和公共数据库 IRCCyN/IVC DIBR 的测试结果表明,本文提出的客观质量评价方法预测分数的皮尔逊线性相关系数分别达到 0.932 7 和 0.858 1,与主观评价分数具有较好的一致性.

关键词: 视频质量评价;窗口六自由度视频;交互路径;语义特征
基金项目: 国家自然科学基金(No.62371081);重庆市自然科学基金(No.cstc2021jcyj-msxmX0411,No.CSTB2022NSCQ-MSX0873)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)01-0193-16
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20240101

Quality Assessment for Windowed-6DoF Synthesized Video Based on Multilayer Features Fusion

TANG Ting-yan[1], ZOU Wen-hui[2], PENG Zong-ju[1,2*], CHEN Fen[1,2], JIN Chong-chong[2]
(1. *College of Electrical and Electronic Engineering*, *Chongqing University of Technology*, *Chongqing* 400054, *China*;
2. *Faculty of Information Science and Engineering*, *Ningbo University*, *Ningbo*, *Zhejiang* 315211, *China*)

Abstract: Six degrees of freedom (6DoF) video, allowing users to experience the scene from omnidirectional and arbitrary perspective, is the development direction of the next-generation immersive video system. The windowed 6DoF video with limited degrees of freedom is a hot research topic in recent years. This paper proposes a subjective database and an objective quality assessment method for the windowed 6DoF synthesized video. For subjective database, we build a subjective quality database called Windowed-6DoF. The database contains 128 windowed 6DoF synthesized videos which involve discomfort caused by two viewpoint switching paths, distortions caused by four rendering schemes, and four levels of compression. Then subjective quality tests are conducted on the database and the test results are analyzed. For objective quality assessment, we design a no reference quality assessment method for windowed 6DoF synthesized video which fuses multilayer features. Tchebichef moment is used to extract the low layer shape features of temporal video slices. Resnet-50 network is used to extract the high-level semantic features of video in temporal and spatial domains, and consequently reduce the dimensionality of features. Finally, the random forest is used to fuse the low layer shape features and high layer semantic features, and train the quality assessment model of windowed 6DoF synthesized video. We respectively test the method on the proposed Windowed-6DoF database and IRCCyN/IVC DIBR database. The experimental results show that the Pearson linear correlation coefficient of the proposed method are 0.932 7 and 0.858 1, respectively. The predicted scores of the objective method are consistent with the subjective assessment scores.

Key words: video quality assessment; windowed six degrees of freedom video; interaction path; semantic features

收稿日期:2024-01-26;修回日期:2024-08-08;责任编辑:赵克
*通讯作者:彭宗举

Foundation Item(s): National Natural Science Foundation of China (No.62371081); National Natural Science Foundation of Chongqing (No.cstc2021jcyj-msxmX0411, No.CSTB2022NSCQ-MSX0873)

1 引言

随着5G、超高清和虚拟现实等新兴技术的发展,沉浸式视频逐渐应用到智慧教育、展览展陈、文化旅游、医疗卫生和工业制造等领域中[1,2]. 六自由度(Six Degrees of Freedom,6DoF)视频系统允许用户从全方位、任意视角身临其境体验场景,是下一代沉浸式视频产业的发展方向[3]. 6DoF视频包括3个旋转自由度和3个平移自由度,在探索6DoF视频过程中,国际标准化组织MPEG提出了部分自由度受限窗口6DoF视频的概念[4]. 基于深度图绘制(Depth Image Based Rendering,DIBR)是实现窗口6DoF视频的主流方法之一,DIBR能合成虚拟视点以实现用户交互浏览时视点和视角的无缝切换. 在窗口6DoF视频系统中,压缩编码、视点切换和虚拟视点绘制等环节引起的视频失真显著降低了用户的视觉体验质量. 因此,研究各种失真对用户观看质量的影响,提出窗口6DoF合成视频的主客观质量评价方法,对于优化沉浸式视频系统的各个环节具有重要意义.

目前,已有如基于模块化无参考视频质量评价[5]、基于时空特征解析无参考视频质量评价(Blind Video Quality Assessment based on Spatio-Temporal Feature Resolver,STFR-BVQA)[6]、基于各种预训练模型的高效视频质量评价[7]、基于片段采样的高效端到端视频质量评价[8]、基于邻里代表片段抽样的高效端到端视频质量评价(Fragment spatial-temporal Video Quality Assessment,Faster-VQA)[9]和基于美学技术角度的视频质量评价[10]等面向传统视频质量评价方法被提出. 但传统视频与合成视频之间存在较大差距,其质量评价方法没有考虑到合成视频失真的特殊性. 虚拟视点合成图像/视频质量评价(Synthesized Image/Video Quality Assessment,S-I/VQA)方法都是面向自由视点视频(Free Viewpoint Videos,FVV)系统中的虚拟视点合成图像/视频,能有效评价由于几何失真、压缩失真等引起的质量变化. 与FVV系统相比,窗口6DoF视频系统参考视点更多、用户交互路径更加复杂多样. 因此,窗口6DoF合成视频存在新的失真类型,现有面向FVV的S-I/VQA方法不能有效评价这些新失真. 当前,鲜有针对窗口6DoF虚拟视点合成视频质量评价的研究. 为解决上述问题,本文提出一种融合多层特征的无参考窗口6DoF合成视频质量评价方法,贡献如下:

(1)面向用户不同的交互路径,构建了1个窗口6DoF合成视频主观质量数据库Windowed-6DoF,具体包含128个窗口6DoF合成视频,涉及2种交互路径不适性失真、4种绘制失真和4种压缩失真.

(2)提出了融合底层和高层特征的6DoF合成视频客观质量评价方法. 采用切比雪夫矩提取视频时域切片上的底层形状特征;采用深度学习的方法分别提取视频的空域、时域高层语义特征并进行降维处理;采用随机森林将底层形状特征和高层语义特征进行融合,并训练得到窗口6DoF合成视频的客观质量评价模型.

2 相关工作

2.1 主观质量评价方法

为了开展虚拟视点合成图像/视频主观质量评价工作,国内外研究团队建立了虚拟视点合成图像/视频数据库. 虚拟视点合成图像数据库主要考虑空域模糊、拉伸、扭曲和空洞等失真. 现有公开虚拟视点合成图像数据库主要有IRCCyN/IVC DIBR图像数据库[11]、MCL-3D图像数据库[12]、IVY图像数据库[13]和IETR DIBR图像数据库[14]. 这些数据库均包含原始图像、失真图像及其主观分数. 虚拟视点合成视频数据库中的失真,除了包括与虚拟视点合成图像相似的空域失真以外,还包括视频播放过程中的时域闪烁失真. 现有公开虚拟视点合成视频数据库主要有IRCCyN/IVC DIBR视频数据库[15]和SIAT视频数据库[16]. 其中,IRCCyN/IVC DIBR视频数据库包含3个场景序列、分辨率为1 024×768的102个视频,主要考虑了几何失真,对压缩失真考虑较少. SIAT视频数据库着重考虑了视频中的纹理/深度压缩失真,包含10个场景序列、分辨率为1 024×768和1 920×1 088的140个视频. 除这2个主流公共数据库外,Wang等人[17]建立了VRTS DIBR合成视频数据库,在考虑几何失真的基础上还考虑了类离散余弦变换失真;Peng等人[18]建立了合成立体视频数据库,涉及到FVV系统中的多重组合失真.

然而,以上虚拟视点合成图像/视频数据库是针对FVV系统提出的,视点仅局限在水平基线上切换,用户对场景的沉浸式体验受限. 此外,这些数据库中的视频通过较早的绘制方法生成,存在着一些如空洞等严重的陈旧性几何失真. 这类失真在现有先进绘制技术下已得到明显改善. 因此,这些数据库难以用于研究窗口六自由度中合成视频质量评价. 为解决上述问题,本文建立了包含压缩失真、绘制失真和交互路径不适性失真的窗口6DoF合成视频主观质量数据库. 所提数据库充分考虑了用户在观看窗口6DoF视频时的视觉不适性,进一步拓展了用户体验场景的自由度,有助于推动新一代6DoF沉浸式视频系统的发展.

2.2 客观质量评价方法

现有的I/VQA方法分为全参考(Full Reference, FR)、半参考和无参考(No Reference, NR)3大类[19]. 对于FR方法,Sadbhawna等人[20]提出了一种S-IQA方法,通过卷积神经网络(Convolutional Neural Network, CNN)提取特征,利用参考图像与失真图像之间的余弦相似性计算质量评分. Zhang等人[21]首先通过离散小波变换提取多尺度结构失真,然后整合梯度幅度相似性突出失真特征,并通过形态学操作和中值滤波来排除不敏感的特征,利用小波尺度和子带失真特征图的标准差池化得到合成图像的质量分数. Thakur等人[22]提出了一种基于上下文区域的合成IQA方法,通过计算参考图像深度能量图与失真图像之间的相关性,提取失真局部区域的离散余弦变化系数,从而得到质量评分. Zhang等人[23]提出了一种基于稀疏表示的S-VQA方法,通过字典学习和稀疏表示来度量视频中的闪烁失真. Zhang等人[24]提出了一种基于时空域失真的S-VQA方法,并运用该方法提升3D视频编码的性能. 实际应用系统中,虚拟视点合成视频不存在相应的参考视频,上述FR方法不能直接用于窗口6DoF合成视频质量评价.

在NR S-IQA方法方面,Jakhetiya等人[25]提出了一种基于核岭回归的IQA方法,通过估计具有几何失真的完整失真表面,来预测合成图像质量评分. Tian等人[26]提出一种无参考合成视点IQA(No-reference Image Quality assessment of Synthesized Views, NIQSV)方法,主要对虚拟视点图像的模糊和扭曲失真进行评价. Tian等人[27]在NIQSV的基础上,进一步增加了对空洞和拉伸失真的度量,得到改进后的NIQSV+方法. Gu等人[28]提出一种自回归结合阈值(Autoregression-Plus Thresholding, APT)的质量评价方法,通过获取图像像素值与预测值之间的误差来提取图像失真,并利用视觉显著性排除非几何失真区域. Gu等人[29]提出一种多尺度的自然场景统计(Multiscale Natural Scene Statistical, MNSS)模型,根据自然场景图像的自相似性和主结构一致性提取局部和全局统计特征. Wang等人[30]提出一种基于离散小波变换的S-IQA方法,在小波域对几何失真和全局清晰度失真进行度量,并提取图像复杂度特征,通过池化得到最终质量分数. Sadbhawna等人[31]采用深度学习识别拉伸失真(Stretching Identification using Deep Learning, SI-DL),并对3D合成视频进行质量评价. Ling等人[32]提出一种基于生成对抗网络的S-IQA方法. 上述方法集中在度量图像空域上的局部几何失真强度,以便预测虚拟视点合成图像的质量.

与S-IQA方法不同,S-VQA方法还需进一步考虑时间域上的失真. Gu等人[29]提出的MNSS方法中,在IQA的基础上进一步加入时间池化策略来整合单个视频帧的质量,从而预测虚拟视点合成视频的质量. 该方法无法度量时间域上的视频闪烁失真. Kim等人[33]提出了一种基于时间临界不一致性(Critical Temporal Inconsistency, CTI)的S-VQA方法,通过测量相邻两帧闪烁区域的结构相似性来评价虚拟视点视频的质量. 然而该方法忽略了空域失真对视频质量的影响. Wang等人[34]提出一种基于时空域特征的S-VQA方法,在空域中估计每帧的高频能量,在时域中采用光流法估计相邻帧之间的运动场,然后计算相邻光流场的结构相似性. Wang等人[35]提出一种基于时空纹理不一致性的S-VQA方法,在空域中提取每帧的纹理图,再利用有向光流直方图提取相邻帧的纹理信息变化. Sandić-Stanković等人[36]通过高-高小波子带中的选定区域来估计视频质量,并采用阈值来选择失真敏感区域. Sandić-Stanković等人[37]定义了1个面向形态学的闭合差算子(Difference of Closings, DoC),并在不同尺度分辨率下使用DOC和高斯差分算子(Difference of Gaussians, DoG)提取特征,然后通过广义回归神经网络(General Regression Neural Network, GRNN)建立特征与主观分数之间的关系得到合成I/VQA方法模型DoC-DoG-GRNN. Jin等人[38]将2种模拟人脑的学习方法相结合,提出一种基于多模态的无参考S-VQA方法(Multi-Model Learning based Blind Synthesized Video Quality Assessment, MML-BSVQA). 为了平衡模型性能和效率,Yan等人[39]提出了一种基于稀疏采样的VQA方法. Jia等人[40]提出一种由空间特征感知模块、时间运动特征感知模块和质量得分融合模块组成的端到端S-VQA方法.

上述S-I/VQA方法主要针对FVV系统提出的,能较好地评价合成图像/视频中的失真. 这些方法无法有效衡量窗口6DoF合成视频中的新型失真,比如交互路径不适性失真. 目前还鲜见针对窗口6DoF视频系统中S-I/VQA方法的报道. 因此,本文面向窗口6DoF合成视频中的新型失真,提出一种融合多层特征的窗口6DoF合成视频无参考客观质量评价方法.

3 窗口6DoF合成视频主观质量数据库构建及失真分析

图1是窗口6DoF视频的示意图. 其中,用户观看场景内容时,围绕y、z轴旋转受限,沿x轴平移受限[41]. 本文首先构建了1个窗口6DoF合成视频主观数据库Windowed-6DoF,然后提出了针对窗口6DoF合成视频的客观质量评价方法,并在提出的Windowed-6DoF数据库和IRCCyN/IVC DIBR数据库上分别测试该客观质

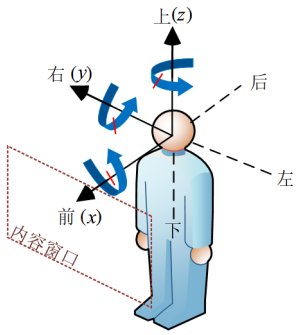


图1 窗口6DoF可视化

量评价方法预测分数与主观评价分数的一致性.

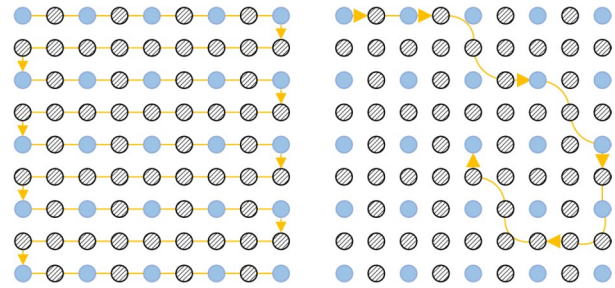
3.1 主观质量数据库构建

本文构建了1个探索性的窗口6DoF合成视频主观质量数据库,主要考虑了3种类型失真:不同程度的压缩失真、不同虚拟视点绘制方案导致的失真和不同视点/视角切换导致的交互路径不适性失真. 数据库的构建过程主要包括视频压缩、虚拟视点绘制和交互路径设计等环节.

首先,选择4个稀疏相机阵列采集的序列ETRI-Chef[42]、TechnicolorPainter[43]、OrangeKitchen[44]和OrangeShaman[44],采用3种量化参数(Quantization Parameter, QP)对(25, 34)、(35, 42)和(45, 48),通过3D-HEVC的HTM-16.2参考软件压缩得到三种压缩等级的失真视频. QP对中,前后2个数字分别对应彩色视频和深度视频的QP.

与FVV系统中仅采用基线上的参考视点来绘制虚拟视点不同,窗口6DoF视频系统可采用平面分布的多个参考视点来绘制虚拟视点,绘制方案更加复杂. 具体地,本数据库针对原始视频和3种压缩等级的失真视频,选取拟绘制视点最邻近的1~4个视点作为参考视点,基于VSRS4.3平台采用DIBR技术进行绘制得到窗口6DoF合成视点. 参考视点数目为1、2、3和4个的4种绘制方案分别用S1、S2、S3和S4来表示. 此外,由于人类观影时,场景内容切换的最佳舒适距离为4~6个像素[45],本文根据相邻视点间的舒适距离来确定各视频序列拟绘制的窗口6DoF合成视点的数量. 因原始序列的相机间距存在一定的差异性,本数据库各序列所绘制得到的合成视点数量并不相同.

在窗口6DoF视频系统中,用户欣赏场景时,视点和视角会发生变化. 本文将该变化过程定义为交互路径. 人眼在观察场景时会自动聚焦于一些显著性的物体,以追求更好的交互体验,不恰当的交互路径会给用户带来视觉不适感. 因此,本文主要探讨了S形固定交互路径和显著性引导交互路径对人类视觉感知质量的影响. 图2(*a*)为S形固定交互路径,用户从左到右、从上到下切换视点;采用S形交互路径,用户难以欣赏到场景显著性内容,视觉体验质量较低. 图2(*b*)为显著性引导交互路径,用户可根据场景内容的变化动态地调整观看视点和视角,使显著区域位于图像中心,能获得更舒适的体验质量;不同序列具有不同的显著性引导交互路径.



(*a*) 固定交互路径 (*b*) 显著性引导交互路径

图2 窗口6DoF合成视频2种交互路径

显著性引导交互路径的计算方法:采用时空显著性算法得到视频序列的时空显著图,并计算面积最大的显著性区域的质心,再沿着质心方向移动视点;随着视频内容的不断变化,每20帧更新一次质心计算,且相邻视点的切换距离控制在4~6个像素之内,以消除2个视点之间因切换距离过大造成的视觉不适性. 最终构建的Windowed-6DoF主观质量数据库采用了4个序列,考虑了4种压缩状态(不压缩和3种等级压缩)、4种绘制方案和2条交互路径,合计生成了128个含有混合失真的合成视频. 其详细属性如表1所示.

本文参考ITU-R BT.500-11[46]的建议,采用单刺激法对数据库进行主观质量测试. 显示设备为55英寸超高清显示器(CHiQ_55Q3R_02DB),显示分辨率为3 840×2 160,峰值亮度调整为50 cd/m^2,观察距离为显示器高度的3倍. 邀请了34名视力正常或者矫正后视力正常的观察者(19男15女,年龄22~34岁)参与实验,且观察者均从事图像处理相关工作. 主观实验过程如下:在室内正常照明条件下,观察者首先观看视频说明;播放1次被随机打乱的待测试合成视频;最后观察者根据表2所示的五级质量和损伤量表对观看的各视频质量进行打分. 观察者的连续主观测试时间均控制在30 min内. 因此,对128个视频的测试实验被随机分成了两部分,两部分测试之间观察者需休息3~5 min以避免视觉疲劳对观察者主观感知的影响.

在数据处理阶段,本方法根据协议进一步筛除异常数据,保留95%的评分数据作为主观分数置信区间. 因此,移除每个视频的2名观察者打出的最低和最高分数,剩下的32个分数为该视频的有效分数. 然后计算每个视频的平均意见得分(Mean Opinion Score, MOS)值,作为该视频的主观质量评价分数,如式(1)所示.

表1　窗口6DoF合成视频序列详细属性

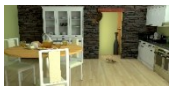
序列名称	序列预览图	分辨率	帧率	相机间距	视点分布	QP	绘制方法	交互路径
ETRIChef		1 920×1 080	30	10×10	5×5	NO,(25,34) (35,42) (45,48)	S1,S2,S3,S4	tra,sal
TechnicolorPainter		2 048×1 088	30	7×7	4×4	NO,(25,34) (35,42) (45,48)	S1,S2,S3,S4	tra,sal
OrangeKitchen		1 920×1 080	30	11.5×11.5	5×5	NO,(25,34) (35,42) (45,48)	S1,S2,S3,S4	tra,sal
OrangeShaman		1 920×1 080	30	10×10	5×5	NO,(25,34) (35,42) (45,48)	S1,S2,S3,S4	tra,sal

表2　五级质量和损伤量表

主观分数	质量	损伤
1	劣	很讨厌
2	差	讨厌
3	中	稍微察觉
4	良	可察觉,但不讨厌
5	优	不可察觉

$$\mathrm{MOS}_a=\frac{1}{L}\sum_{b=1}^{L}V_{ab} \tag{1}$$

其中,V_{ab}表示b号观察者给视频a打出分数,L表示观察者数量,MOS_a表示视频a的主观质量分数. 本文通过分析组间一致性验证每个视频有效分数个数选取的有效性. 将每个视频的所有观察者评分随机平均分为2组,分别计算每组分数的均值. 然后计算2组分数的皮尔逊线性相关性系数(Pearson Linear Correlation Coefficient,PLCC)值. 该过程重复100次,取中值结果作为组间一致性值. 图3为组间一致性与实验人员数量的关系曲线. 可以看出,在当实验人员数量达到28人时,组间一致性趋于饱和. 因此,本文对每个视频选择32个有效分数是合理的.

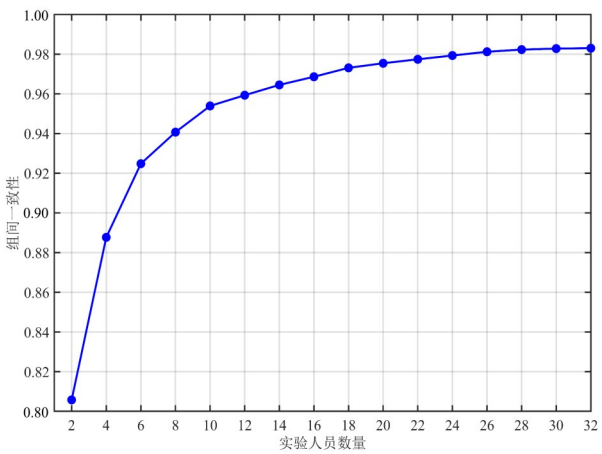


图3　组间一致性与实验人员数量关系曲线

图4给出了序列OrangeShanman的MOS值直方图,该值越大表示视频质量越好. 其中,sal表示显著性引导交互路径,tra表示传统固定交互路径. 横坐标表示了绘制方法和压缩等级,例如S1_25表示采用S1方法绘制,QP对为(25,34)的视频. 可见,观察者对显著性引导交互路径、QP值更小和更多参考视点绘制的窗口6DoF合成视频的视觉感知质量评分更高. 图5(a)、图5(b)和图5(c)分别为Windowed-6DoF数据库中交互路径不适性失真、绘制失真和压缩失真的级别的示例以及MOS值.

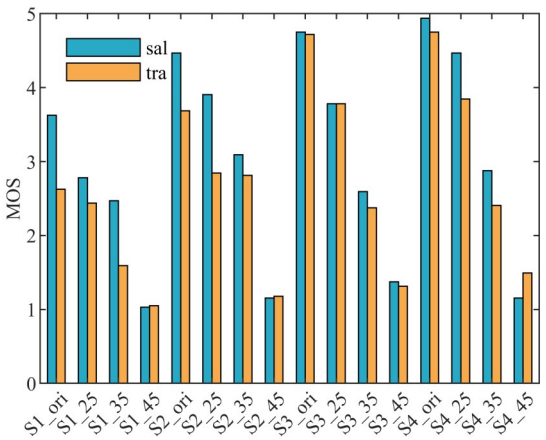


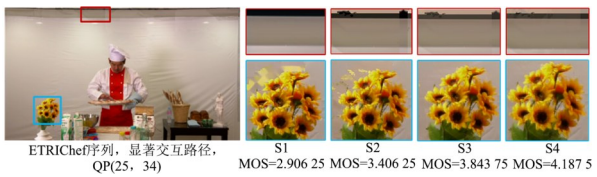
图4　OrangeShanman序列MOS值

3.2　失真分析

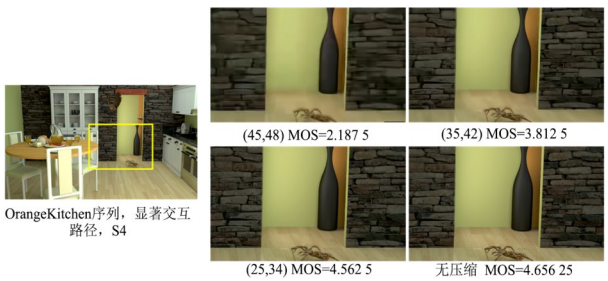
窗口6DoF合成视频的失真可分为空域和时域失真. 其中,空域失真为压缩编码引起的模糊失真;时域失真为交互路径和虚拟视点绘制引起的交互路径不适性失真和闪烁失真. 图4表明显著性引导交互路径的MOS值普遍高于S形传统固定交互路径,可见显著性引导交互路径更符合人类视觉机制.

在6DoF合成视频绘制过程中,S1和S2方案采用了较少的参考视点,合成视频中物体边缘存在严重的像素突变和局部几何失真,进而产生时域上的不一致性

(*a*) 交互路径不适性失真及MOS值



(*b*) 绘制失真及MOS值

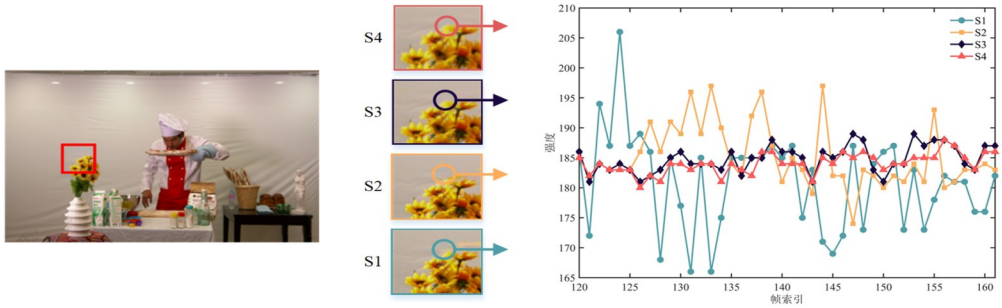


(*c*) 压缩失真及MOS值

图5 Windowed-6DoF数据库中不同失真类型及MOS值

即闪烁效应. 图6(*a*)为窗口6DoF视频ETRIchef序列的第120帧,图6(*b*)为4种绘制方案绘制的视频帧中黄色花朵的局部放大图. 图6(*c*)展示了坐标为(532,500)处像素强度随时间的变化情况. 可见,随着绘制方案中参考视点的增加,几何失真越来越小;S3、S4绘制的窗口6DoF合成视频随时间变化平缓,而S1、S2绘制的合成视频在相同坐标处强度随时间变化更剧烈. 因此,较少参考视点绘制出的合成视频会带来更加严重的闪烁效应.

总之,压缩失真、交互路径不适性失真和绘制失真不仅会引起窗口6DoF合成视频空域上高层语义信息的改变,还会引起时域上高层语义和底层形状信息的改变,最终降低人眼视觉的感知质量. 由于窗口6DoF合成视频相关数据库样本少且不易获得,预训练的CNN模型可用来提取窗口6DoF合成视频的时空域中的绘制失真和压缩失真. 另外,现有的数据库中含有交互路径不适性失真的视频数量较少,无法满足端到端深度模型需要大量数据训练的要求. 因此,本文提出基于底层形状特征和高层语义特征的无参考窗口6DoF合成视频客观质量评价方法. 该方法采用传统方法对交互路径不适性失真进行度量,采用预训练CNN模型对绘制和压缩失真进行度量,设计了传统加深度特征的混合模型.



(*a*) 视频ETRIchef序列第120帧 (*b*) 4种绘制方案的视频帧中黄色花朵的局部放大图 (*c*) 坐标为(532,500)处像素强度随时间的变化曲线

图6 窗口6DoF合成视频绘制失真分析

4 窗口6DoF合成视频客观质量评价方法

本文提出的窗口6DoF合成视频客观质量评价方法总流程如图7所示,主要包括时域切片底层形状特征提取模块和CNN高层语义特征提取模块. 首先,将输入视频进行时空域转换,得到输入视频XT-Y时域切片和XY-T空域切片. 然后,采用底层形状特征提取模块在二值化后的XT-Y时域切片上进行形状描述,从而得到时域切片底层形状特征. 其次,对XY-T空域切片,考虑视频相邻帧间语义特征随时间的变化特性,采用CNN高层语义特征提取模块分别提取时域、空域高层语义特征并进行降维处理. 最后,采用随机森林将底层形状特征和高层时空域语义特征进行融合,且训练得到质量评价模型,从而获得窗口6DoF合成视频的客观质量分数.

4.1 时域切片底层形状特征

通常视频可用由宽度W、高度H和帧数量T构成的3维阵列来表示,通过对视频的3元素W、H和T进行不

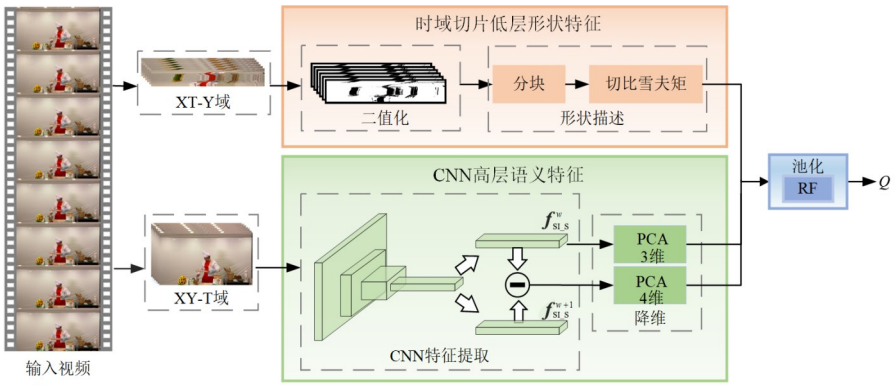
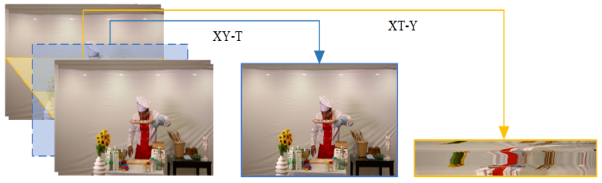


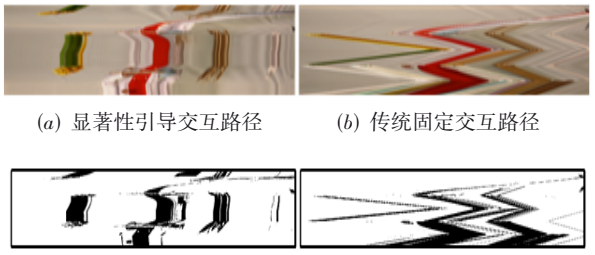
图7　客观质量评价方法总流程图

同的组合能够构成不同的时空域. 图8直观地展示了窗口6DoF合成视频的XY-T空间域切片和XT-Y时间域切片. 由于人眼观看视野的水平延伸大于垂直延伸,因此,不同的交互路径在水平切片XT-Y域切片上有着更显著的表现[47]. 图9(*a*)为显著性引导交互路径视频的XY-T域切片,图9(*b*)为传统固定交互路径视频的XT-Y域切片. 可见,不同交互路径的视频在XT-Y域中呈现出明显不同的形状信息. 本文通过对XT-Y域中物体所呈现出的形状进行描述,进而对因交互路径引起的失真进行度量. 之前研究已证实切比雪夫矩为有效的形状描述算子[48],因此本文采用切比雪夫矩来描述窗口6DoF合成视频的时域切片底层形状特征.



(*a*) 视频序列　(*b*) XY-T空间域切　(*c*) XT-Y时间域切片

图8　窗口6DoF合成视频时空域转换



(*a*) 显著性引导交互路径　(*b*) 传统固定交互路径

(*c*) 图(*a*)二值化图片　(*d*) 图(*b*)二值化图片

图9　不同交互路径在XT-Y域的轨迹分析

本文首先采用自适应二值化方法对XT-Y域切片进行二值化处理,除去彩色视频中的冗余信息,其次采用切比雪夫矩来描述XT-Y域切片中的形状信息. 具体地,将二值化XT-Y域切片图像划分为$k\times k$的小块,再计算每个小块的[(*k*–1)+(*k*–1)]阶矩:

$$\boldsymbol{C}_{mn}=\begin{pmatrix} T_{00} & T_{01} & \cdots & T_{0(k-1)} \\ T_{10} & T_{11} & \cdots & T_{1(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ T_{(k-1)0} & T_{(k-1)1} & \cdots & T_{(k-1)(k-1)} \end{pmatrix} \tag{2}$$

其中,$\boldsymbol{C}_{mn}$表示XT-Y域第m幅切片图像中第n个图像块的[(*k*–1)+(*k*–1)]阶矩,T_{00}为直流分量,其他元素为交流分量. 本文首先用交流分量对视频图像的形状特征进行描述,并采用式(3)计算每个图像块的交流能量和:

$$E_{mn}=\sum_{i=0}^{(k-1)}\sum_{j=0}^{(k-1)}\left(T_{ij}\right)^2-\left(T_{00}\right)^2 \tag{3}$$

其中,E_{mn}表示表示XT-Y域第m幅切片图像中第n个图像块的交流能量和. 接着,计算XT-Y域中单个切片的能量:

$$E_m=\frac{1}{N}\sum_{p=1}^{N}E_{mp} \tag{4}$$

其中,E_m表示XT-Y域第m幅切片图像的能量,p表示图像块序号,N为XT-Y域切片中图像块的总数. 最后,采用XT-Y域切片特征的均值和标准差来表示窗口6DoF合成视频的时域切片底层形状特征值:

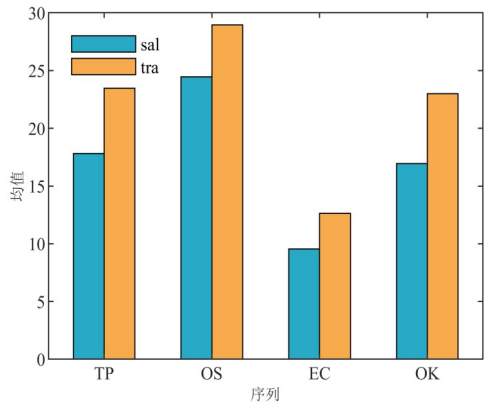
$$\bar{f}_{\text{vpd}}=\frac{1}{M}\sum_{q=1}^{M}E_q \tag{5}$$

$$\tilde{f}_{\text{vpd}}=\sqrt{\frac{1}{M}\sum_{q=1}^{M}\left(E_q-\bar{f}_{\text{vpd}}\right)^2} \tag{6}$$

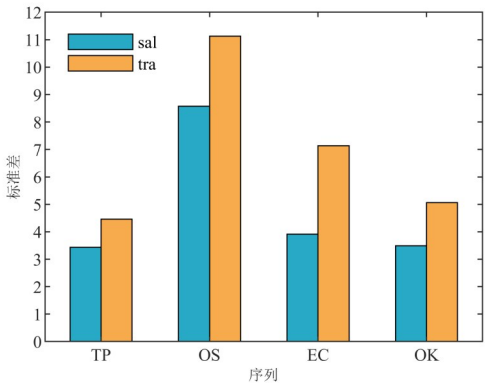
其中,$\bar{f}_{\text{vpd}}$表示XT-Y域切片特征均值,$\tilde{f}_{\text{vpd}}$表示XT-Y域切片特征标准差,M表示XT-Y域切片总数.

图10给出了在Windowed-6DoF数据库中不同交互路径下的XT-Y域切片特征的均值和标准差. 其中,sal表示显著性引导交互路径,tra表示传统固定交互路径. 横坐标TP、OS、EC和OK分别为场景视频Technicolor-Painter、OrangeShaman、ETRIChef和OrangeKitchen,且以上视频QP对和绘制方法相同. 可以明显地看出,当

视频场景、QP对和绘制方法相同时，显著性引导交互路径的均值和标准差均小于传统固定交互路径，这表明本文所提出的方法可有效提取时域切片的底层形状特征.



(*a*) 切片形状特征的均值



(*b*) 切片形状特征的标准差

图 10 切片形状特征的均值和标准差

最终，窗口 6DoF 合成视频时域切片底层形状特征 $\boldsymbol{F}_{\mathrm{vpd}}$ 为

$$\boldsymbol{F}_{\mathrm{vpd}}=\left(\overline{f}_{\mathrm{vpd}},\tilde{f}_{\mathrm{vpd}}\right) \tag{7}$$

4.2 CNN 高层语义特征

本文采用预训练 CNN 模型在 XY-T 域上对窗口 6DoF 合成视频的每帧图像进行高层语义特征的提取. 具体的，采用在 ImageNet 数据库上预训练的 Resnet-50 模型提取图像语义特征，首先输入大小为 224×224×3 大小的图像；然后对输入图像进行卷积、正则化、激活函数、最大池化的计算. 其中，网络包括了 4 个由 2 种 Bottleneck 结构[49]构成的 Block1-4. 经过 Block1-4 后输出大小为 $7\times7\times2\,048$ 的特征图，通过全局平均池化层（Global Average Pooling，GAP）转换成特征向量作为图像的高层语义特征图.

由于视频的时空域失真，所以本方法提取语义特征和相邻两帧的特征之差，为方便描述，将其分别称为空域高层语义特征和时域高层语义特征：

$$\boldsymbol{f}_{\mathrm{SI\_S}}^{(w)}=\mathrm{GAP}\left(\boldsymbol{I}_w\right) \tag{8}$$

$$\boldsymbol{f}_{\mathrm{SI\_T}}^{(t)}=\mathrm{ABS}\left(\boldsymbol{f}_{\mathrm{SI\_S}}^{(w+1)}-\boldsymbol{f}_{\mathrm{SI\_S}}^{(w)}\right) \tag{9}$$

其中，$\boldsymbol{f}_{\mathrm{SI\_S}}^{(w)}$ 表示第 w 幅图像的空域高层语义特征，$w\in[1,T]$，T 为视频的帧数. $\boldsymbol{f}_{\mathrm{SI\_T}}^{(t)}$ 表示 w+1 和 w 帧之间的时域高层语义特征，t 表示 w+1 和 w 帧的帧间序号，$t\in[1,T-1]$；GAP 表示全局平均池化，ABS 表示绝对值. 窗口 6DoF 合成视频的空域高层语义特征和时域高层语义特征分别为

$$\overline{\boldsymbol{F}}_{\mathrm{SI\_S}}=\frac{1}{T}\sum_{w=1}^{T}\boldsymbol{f}_{\mathrm{SI\_S}}^{(w)} \tag{10}$$

$$\overline{\boldsymbol{F}}_{\mathrm{SI\_T}}=\frac{1}{T-1}\sum_{t=1}^{T-1}\boldsymbol{f}_{\mathrm{SI\_T}}^{(t)} \tag{11}$$

其中，$\overline{\boldsymbol{F}}_{\mathrm{SI\_S}}$ 表示空域高层语义特征，$\overline{\boldsymbol{F}}_{\mathrm{SI\_T}}$ 表示时域高层语义特征.

由于时域切片底层形状特征 $\boldsymbol{F}_{\mathrm{vpd}}$ 和 CNN 高层语义特征输出维度不平衡，其中 $\boldsymbol{F}_{\mathrm{vpd}}$ 为 2 维，$\overline{\boldsymbol{F}}_{\mathrm{SI\_S}}$ 和 $\overline{\boldsymbol{F}}_{\mathrm{SI\_T}}$ 都为 2 048 维，如果直接将这 3 个特征串联起来，会导致 $\boldsymbol{F}_{\mathrm{vpd}}$ 的表达被抑制. 因此，本文对 $\overline{\boldsymbol{F}}_{\mathrm{SI\_S}}$ 和 $\overline{\boldsymbol{F}}_{\mathrm{SI\_T}}$ 分别采用主成分分析（Principal Component Analysis，PCA）[50]方法对特征降维，降维后的空域高层语义特征 $\boldsymbol{F}_{\mathrm{SI\_S}}$ 和时域高层语义特征 $\boldsymbol{F}_{\mathrm{SI\_T}}$ 为窗口 6DoF 合成视频的 CNN 高层语义特征.

4.3 质量分数预测

将提取到 $\boldsymbol{F}_{\mathrm{vpd}}$、$\boldsymbol{F}_{\mathrm{SI\_S}}$ 和 $\boldsymbol{F}_{\mathrm{SI\_T}}$ 构成窗口 6DoF 合成视频整体特征 $\boldsymbol{F}$，并输入随机森林[51]得到回归函数 $R(\cdot)$，从而通过 R 将特征映射到客观分数 Q：

$$Q=R(\boldsymbol{F}) \tag{12}$$

将实验数据库分为 80% 的训练集和 20% 的测试集，并进行 1 000 次测试训练，最终选取结果的中间值作为窗口 6DoF 合成视频的客观预测分数.

5 实验结果与分析

本文利用 Windowed-6DoF 和 IRCCyN/IVC DIBR 视频数据库，通过与现有的 IQA/VQA 方法进行对比实验，测试了提出的窗口 6DoF 合成视频客观质量评价方法的性能.

5.1 性能比较

本文采用 PLCC、斯皮尔曼排序相关性系数（Spearman Rank Order Correlation Coefficient，SROCC）和均方根误差（Root Mean Square Error，RMSE）来衡量方法的有效性. 一种较好的视频质量评价模型应获得较大的 PLCC、SROCC 和较小的 RMSE. 图 11 为在 Windowed-

6DoF 数据库上的 1 000 次训练测试的 PLCC、SROCC 和 RMSE 的箱型图,可以看出各指标的箱型形状较扁平,大部分数据聚集于中值附近,反映此模型具有较好的稳定性. 表 3 和 4 分别列出了传统图像/视频质量评价方法(Traditional Image/Video Quality Assessment, T-I/VQA)、S-I/VQA 和本文提出的窗口 6DoF 视频质量评价方法在 Windowed-6DoF 和 IRCCyN/IVC DIBR 视频数据库上的性能比较. 其中,粗体和下划线分别表示最优和次优的性能.

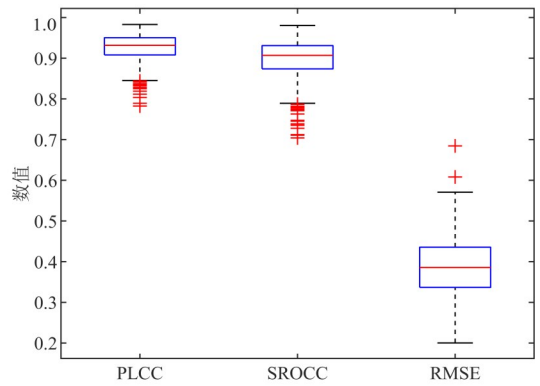


图 11　Windowed-6DoF 数据库 1 000 次实验各指标分布箱形图

从表 3 可以看出,在 Windowed-6DoF 数据库上本文方法 PLCC 性能最优. 具体的,在各种 T-I/VQA 方法中,Faster-VQA 在 SROCC 上优于本文方法说明了其预测分数在排序上的一致性优于本文方法,但本文的 PLCC 和 RMSE 皆优于 Faster-VQA,说明了本文方法的预测分数的精度与主观分数 MOS 值更加接近. 窗口 6DoF 视频中的失真与传统图像/视频失真的差别十分明显,传统图像/视频失真通常表现为全局分布,而窗口 6DoF 合成视频的几何失真表现为局部不规则分布. T-I/VQA 方法主要是针对传统图像/视频失真提出的,不能有效评价窗口 6DoF 合成视频中局部失真. 在各类 S-I/VQA 方法中,MML-BSVQA 方法性能较优,但 PLCC、SROCC 和 RMSE3 项性能均不及本文提出的方法. S-I/VQA 方法虽然解决了 T-I/VQA 方法无法有效度量的局部几何失真问题,但缺乏对交互路径不适性失真的度量. 总之,本文方法充分考虑了窗口 6DoF 合成视频的混合失真特性,且引入深度学习方法来提取失真视频的空域、时域高层语义特征,相较于现有方法具有一定优势.

为了更加直观地比较各方法性能,图 12 给出了几种代表性方法预测的客观质量分数与在 Windowed-6DoF 数据库上主观分数 MOS 值的散点图. 其中,横坐标表示客观方法预测的分数,纵坐标表示主观 MOS 值. 散点越靠近对角线表示客观分数与主观

表 3　现有质量评价方法在 Windowed-6DoF 数据库上的性能比较

方法	类型	Windowed 6DoF 数据库		
		PLCC	SROCC	RMSE
Wen[5]	T-VQA NR	0.931 4	0.908 2	0.376 2
STFR-BVQA[6]	T-VQA NR	0.916 7	0.922 8	**0.310 5**
Faster-VQA[9]	T-VQA NR	0.924 8	**0.957 2**	0.480 4
NIQSV[26]	S-IQA NR	0.361 0	0.227 0	1.033 2
NIQSV+[27]	S-IQA NR	0.464 9	0.426 7	0.966 2
APT[28]	S-IQA NR	0.270 2	0.134 1	1.050 7
MNSS[29]	S-IQA NR	0.588 1	0.601 7	0.882 6
Wang[30]	S-IQA NR	0.597 6	0.602 3	0.875 0
SI-DL[31]	S-IQA NR	0.526 6	0.478 6	0.927 7
DoC-DoG-GRNN[37]	S-VQA NR	0.854 5	0.809 9	0.552 9
MML-BSVQA[38]	S-VQA NR	0.920 8	0.893 1	0.413 5
LI[48]	T-IQA NR	0.624 6	0.502 1	0.852 2
BIQI[52]	T-IQA NR	0.564 4	0.555 6	0.900 8
BRISQUE[53]	T-IQA NR	0.618 3	0.558 5	0.857 7
NIQE[54]	T-IQA NR	0.608 7	0.562 0	0.865 8
VIIDEO[55]	T-VQA NR	0.529 3	0.514 2	0.925 9
VSFA[56]	T-VQA NR	0.728 1	0.670 4	0.755 2
Simple-VQA[57]	T-VQA NR	0.896 4	0.902 3	0.492 6
Dendi[58]	T-VQA NR	0.829 4	0.806 1	0.594 2
本文方法	Win-VQA NR	**0.932 7**	0.911 0	0.385 3

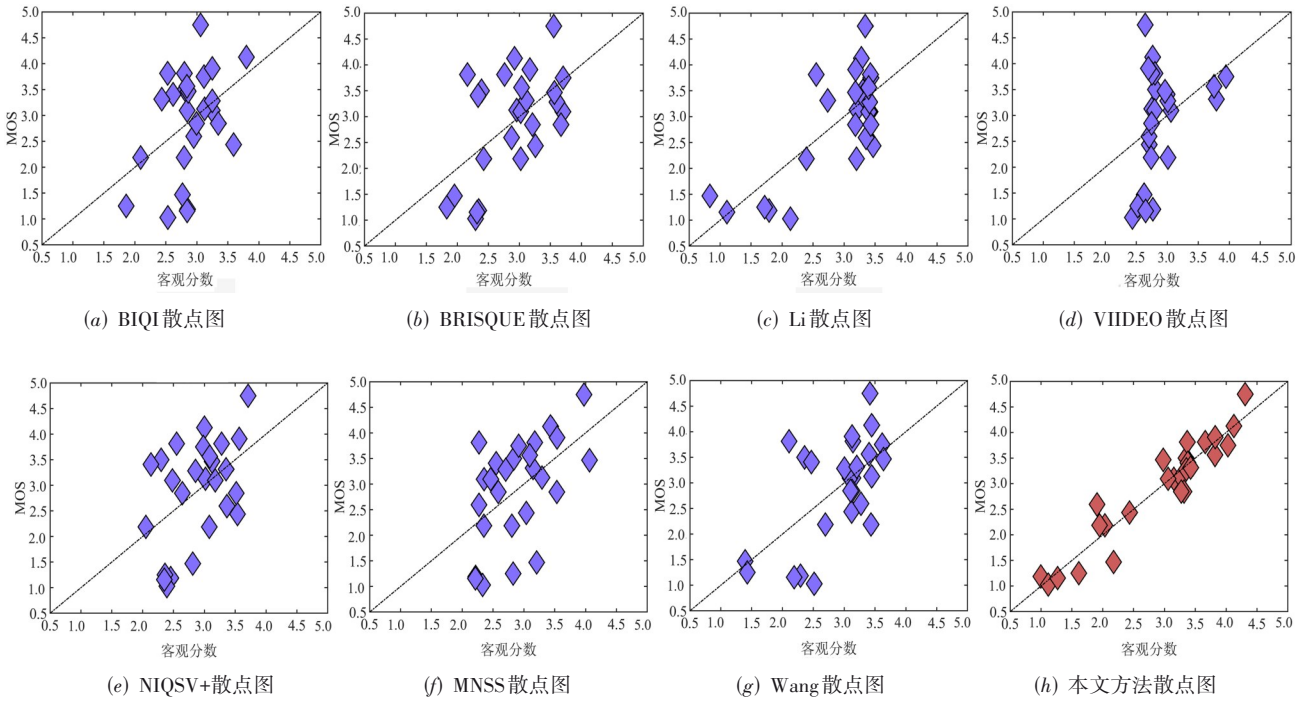
注:加粗和下划线数据分别表示性能最优和次优.

MOS 值越接近,对应评价方法性能更佳. 可见,本文提出的窗口 6DoF 合成视频质量评价方法的散点图呈现出更高的线性度和收敛性,其客观质量评价分数与主观质量分数保持着更高的一致性,更符合人眼视觉感知特性.

从表 4 中可以看出,本文方法在 IRCCyN/IVC DIBR 公开视频数据库上的 PLCC 和 SROCC 性能仅次于 MML-BSVQA 方法. 在 RMSE 上仅次于 STFR-BVQA 方法,但 PLCC 和 SROCC 优于 STFR-BVQA 方法. 由于 IRCCyN/IVC DIBR 视频数据库仅考虑了压缩和绘制失真,且合成视点的绘制方案仅采用 FVV 系统基线分布的参考视点来绘制. 本文方法针对窗口 6DoF 合成视频中的压缩失真、绘制失真和新型的交互路径不适性失真进行设计. IRCCyN/IVC DIBR 数据库中的失真类型与本文算法考虑的失真类型的差异是本文算法 PLCC 和 SROCC 性能稍微次于 MML-BSVQA 的原因.

5.2　统计显著性分析

为了进一步验证本文所提方法与现有方法的统计显著性,分别在 Windowed-6DoF 和 IRCCyN/IVC DIBR 视频数据库进行了 F 检验. 具体地,方法 X 和方法 Y 的 F 检验分数可以表示如下:



(*a*) BIQI散点图 (*b*) BRISQUE散点图 (*c*) Li散点图 (*d*) VIIDEO散点图

(*e*) NIQSV+散点图 (*f*) MNSS散点图 (*g*) Wang散点图 (*h*) 本文方法散点图

图12 Windowed-6DoF数据库上的预测分数与主观MOS的散点图

表4 现有的质量评价方法在IRCCyN/IVC DIBR视频数据库上的性能比较

方法	类型	IRCCyN/IVC DIBR数据库		
		PLCC	SROCC	RMSE
PSNR	T-IQA FR	0.675 9	0.542 1	0.551 0
Wen[5]	T-VQA NR	0.788 4	0.814 8	0.413 1
STFR-BVQA[6]	T-VQA NR	0.723 4	0.700 8	**0.355 0**
NIQSV[26]	S-IQA NR	0.551 6	0.493 6	0.623 6
NIQSV+[27]	S-IQA NR	0.523 7	0.385 5	0.636 9
APT[28]	S-IQA NR	0.590 2	0.583 5	0.603 6
MNSS[29]	S-IQA NR	0.457 0	0.455 3	0.665 0
Wang[30]	S-IQA NR	0.634 6	0.651 7	0.577 8
SI-DL[31]	S-IQA NR	0.463 8	0.424 2	0.662 4
CTI[33]	S-VQA NR	0.721 7	0.721 8	0.401 2
MML-BSVQA[38]	S-VQA NR	**0.891 4**	**0.873 4**	0.367 7
LI[48]	T-IQA NR	0.204 0	0.242 5	0.731 9
BIQI[52]	T-IQA NR	0.402 5	0.412 2	0.684 4
BRISQUE[53]	T-IQA NR	0.350 1	0.369 3	0.700 3
NIQE[54]	T-IQA NR	0.581 6	0.381 2	0.608 2
VIIDEO[55]	T-VQA NR	0.551 8	0.380 0	0.623 5
SSIM[59]	T-IQA FR	0.445 9	0.349 9	0.669 2
本文方法	Win-VQA NR	<u>0.858 1</u>	<u>0.816 5</u>	<u>0.366 2</u>

注:加粗和下划线分别表示最优和次优.

$$F_{\text{score}}=\left(\frac{\sigma_{\text{X}}}{\sigma_{\text{Y}}}\right)^2 \tag{13}$$

其中,F_{score}为F检验分数,σ为RMSE值,X、Y表示比较方法.通过MATLAB中的finv函数,置信水平设置为95%,得到2个数据库的阈值F_{critical}. 如果$F_{\text{score}}>F_{\text{critical}}$,表明Y方法的统计性能明显优于X方法;如果$F_{\text{score}}$位于$\left[1/F_{\text{critical}},F_{\text{critical}}\right]$,表明X和Y方法的统计性能相当;如果$F_{\text{score}}<1/F_{\text{critical}}$,表明Y方法的统计性能明显差于X方法.表5和表6分别列出了在Windowed-6DoF和IRCCyN/IVC DIBR视频数据库上X和Y这2种方法之间的统计显著性能对比结果.其中,“1”、“0”和“−1”分别表示Y方法的性能优于、相当于和差于X方法.在Windowed-6DoF数据库上,本文方法相当的统计性能与MML-BSVQA相当,优于其他对比方法;在IRCCyN/IVC DIBR视频数据库上,本文方法与CTI和MML-BSVQA相当,优于其他对比方法.

5.3 参数研究

本文在Windowed-6DoF数据库上测试了k阶切比雪夫矩对本文方法的影响,图13显示了不同k值下底层形状特征的评估性能结果.从图中可以看出,随着k值的增加,模型性能呈现出先上升后下降的趋势,当k值为16时性能达到最优,其PLCC和SROCC分别达到了0.724 4和0.670 7,因此本文将k值设置为16.

采用PCA方法对CNN高层语义特征进行降维处理时,主成分r的选取与累计特征贡献率有关:

表5　本文方法与对比方法在Windowed-6DoF数据库上的统计显著性

X方法 Y方法	文献[9]	文献[27]	文献[30]	文献[31]	文献[37]	文献[38]	文献[48]	文献[53]	文献[54]	文献[57]	文献[58]	本文方法
Faster-VQA[9]	0	1	1	1	0	−1	1	1	1	0	1	**−1**
NIQSV+[27]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
Wang[30]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
SI-DL[31]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
DoC-DoG-GRNN[37]	0	1	1	1	0	−1	1	1	1	0	0	**−1**
MML-BSVQA[38]	1	1	1	1	1	0	1	1	1	1	1	**0**
LI[48]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
BRISQUE[53]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
NIQE[54]	−1	0	0	0	−1	−1	0	0	0	−1	−1	**−1**
Simple-VQA[57]	0	1	1	1	0	−1	1	1	1	0	1	**−1**
Dendi[58]	−1	1	1	1	0	−1	1	1	1	−1	0	**−1**
本文方法	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**

注:加粗数据表示本文方法性能.

表6　本文方法与对比方法在IRCCyN/IVC DIBR视频数据库上的统计显著性能

X方法 Y方法	文献[27]	文献[30]	文献[31]	文献[33]	文献[38]	文献[48]	文献[52]	文献[54]	本文方法
NIQSV+[27]	0	0	0	−1	−1	0	0	0	**−1**
Wang[30]	0	0	0	−1	−1	1	0	0	**−1**
SI-DL[31]	0	0	0	−1	−1	0	0	0	**−1**
CTI[33]	1	1	1	0	0	1	1	1	**0**
MML-BSVQA[38]	1	1	1	0	0	1	1	1	**0**
LI[48]	0	−1	0	−1	−1	0	0	−1	**−1**
BIQI[52]	0	0	0	−1	−1	0	0	0	**−1**
NIQE[54]	0	0	0	−1	−1	1	0	0	**−1**
本文方法	**1**	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**0**

注:加粗数据表示本文方法性能.

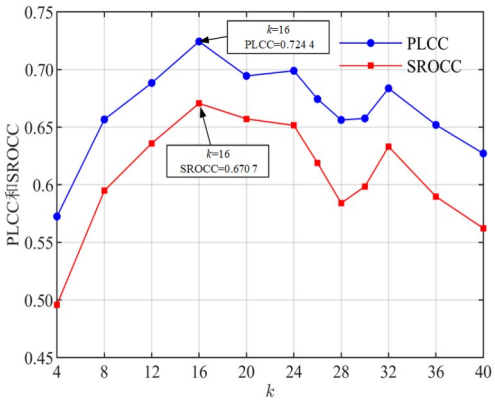


图13　不同k值下的模型性能

$$\beta_r=\frac{\sum_{v=1}^{r}\lambda_v}{\sum_{u=1}^{d}\lambda_u},\ (r\leqslant d) \tag{14}$$

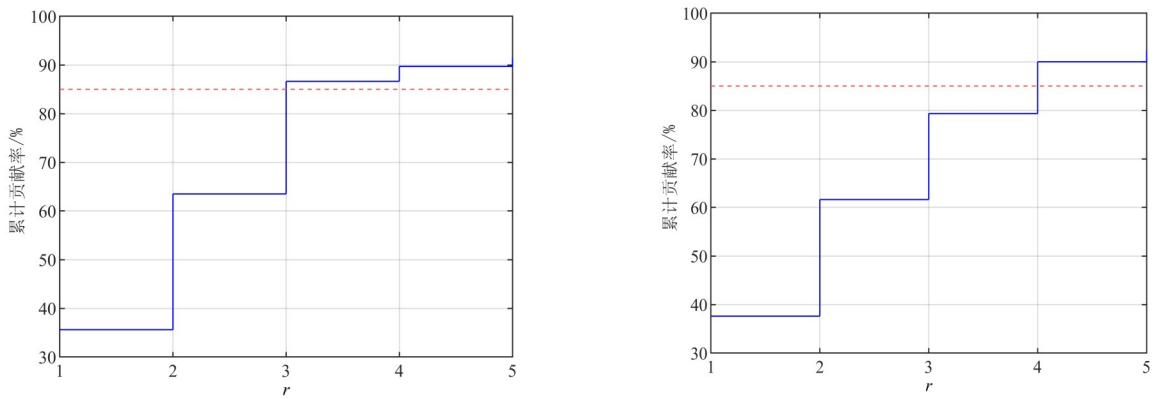
其中,β_r表示前r个主成分的累计贡献率,λ_u和λ_v分别为第u和v个特征值,d表示所有主成分的个数.

一般情况下,选取特征累计贡献率达85%的前r个主成分[60]. 如图14所示,当β取85%时用红色虚线表示. 图14(a)表示空间域高层语义特征,图14(b)表示时间域高层语义特征. 当空间域高层语义特征$\boldsymbol{F}_{\mathrm{SI_S}}$和时间域高层语义特征$\boldsymbol{F}_{\mathrm{SI_T}}$的$r$分别为3和4时,特征累计贡献率已达85%. 因此,本文将$\boldsymbol{F}_{\mathrm{SI_S}}$和$\boldsymbol{F}_{\mathrm{SI_T}}$的维数$r$分别设置为3和4.

5.4　跨数据库实验

由于窗口6DoF合成视频数据库较少,无法采用2个窗口6DoF合成视频数据库进行跨库实验. 因此,采用本文所提的数据库和IRCCyN/IVC DIBR视频数据库进行跨库实验. 具体跨数据库实验方法与对比方法MML-BSVQA中的跨数据库实验方法相同,在1个数据库中的80%样本上训练,然后在另1个数据库中随机选择20%样本进行测试. 实验结果如表7所示. 由于数据库中的失真类型不一致,跨库测试的PLCC、SROCC和RMSE并不理想,比在同一数据库中训练测试的性能有所降低. Windowed-6DoF视频数据库中包含IRCCyN/IVC DIBR视频数据库中不存在的交互路径不适性失真,且Windowed-6DoF数据库中的失真为复合型失真. 本文方法中含有针对交互路径不适性失真的度量,而对比方法是针对FVV系统提出的,且并没有对交互路径不适性失真的度量. 因此,训练和测试数据库分别为Windowed-6DoF和IRCCyN/IVC DIBR时,本文方法要优于对比方法.

训练和测试数据库分别为IRCCyN/IVC DIBR和Windowed-6DoF时,2种方法均不能学习到交互路径不适性失真特征. 然而,本文方法注重交互路径不适性失真,仅通过单一深度模型提取到压缩和绘制失真;



(*a*) 空间域高层语义特征 (*b*) 时间域高层语义特征

图14 不同r值对应的累计贡献率

表7 本文算法与MML-BSVQA方法的跨库性能对比

训练数据库	测试数据库	方法	PLCC	SROCC	RMSE
Windowed-6DoF	IRCCyN/IVC DIBR	MML-BSVQA	0.454 1	0.347 1	0.666 1
		Proposed	0.548 3	0.422 3	0.596 4
IRCCyN/IVC DIBR	Windowed-6DoF	MML-BSVQA	0.451 6	0.341 3	0.973 3
		Proposed	0.414 0	0.166 8	0.967 8

MML-BSVQA方法是针对压缩和绘制失真提出的多模态学习方法,能更好地学习到库中所包含的压缩和绘制失真.因此,本文方法在IRCCyN/IVC DIBR数据库中训练得到的模型在Windowed-6DoF库上测试时性能不如MML-BSVQA

5.5 训练集百分比研究

为了研究训练集百分比对本文所提出的VQA模型性能的影响,分别将训练集相对于整个数据库的比例设置为90%、80%、70%、60%、50%,并在Windowed-6DoF和IRCCyN/IVC DIBR视频数据库上进行测试.实验结果如表8所示,随着训练集相对整个数据库的比例下降,本文所提出的VQA模型性能也随之下降.在训练集的比例为50%的情况下,本文VQA方法在2个数据库上的PLCC仍能达到0.889 2和0.787 2,且优于大多数现有S-I/VQA方法.可见,本文所提出的VQA方法对训练集数量敏感度较低,即使仅采用少量的训练集,本文方法仍具有较好的性能.因此,本文将训练集百分比设置为80%.

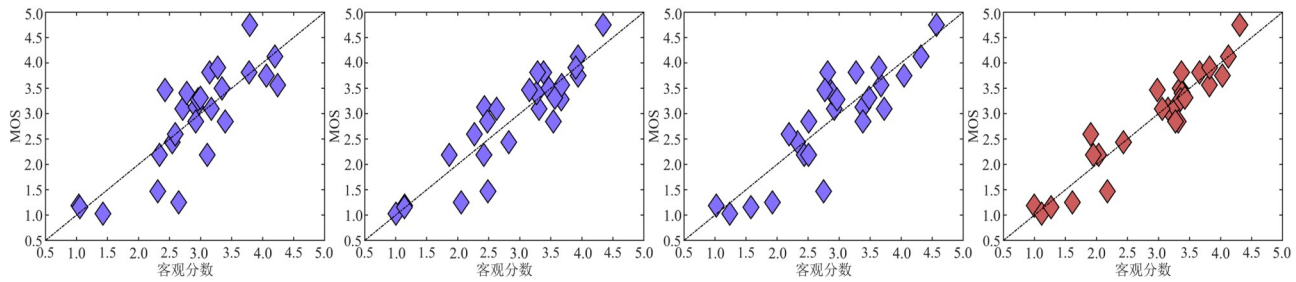
表8 本文所提出的客观方法在不同大小训练集下的结果

视频数据库	训练集百分比-测试集百分比/%	PLCC	SROCC	RMSE
Windowed-6DoF	50~50	0.889 2	0.871 4	0.496 5
	60~40	0.908 1	0.890 4	0.451 7
	70~30	0.921 9	0.904 0	0.418 9
	80~20	0.932 7	0.911 0	0.385 3
	90~10	0.949 5	0.912 1	0.323 1
IRCCyN/IVC DIBR	50~50	0.787 2	0.776 7	0.456 2
	60~40	0.810 1	0.791 6	0.429 0
	70~30	0.825 9	0.803 6	0.408 7
	80~20	0.858 1	0.816 5	0.366 2
	90~10	0.913 2	0.828 5	0.270 3

5.6 消融实验

为了研究不同特征分量对算法性能的影响,本文进行了消融实验.消融实验结果如表9所列.语义特征的性能高于时域切片底层形状特征的性能,说明CNN高层语义特征的贡献度更大.在Windowed-6DoF数据库中,单个特征都取得了不错的性能.当特征组件为2个时PLCC在0.888 4到0.928 8之间.当特征组件全部使用时,模型性能表现最

优.可见,随着特征组件数量的增加,模型性能逐渐提升.图15(*a*)、15(*b*)和15(*c*)分别直观地展示由单特征$\boldsymbol{F}_{\text{vpd}}$、$\boldsymbol{F}_{\text{SI_S}}$和$\boldsymbol{F}_{\text{SI_T}}$训练模型所得客观分数与主观MOS的散点图,图15(*d*)为特征组件全部使用时训练模型的散点图.可以看出,图15(*d*)中的点相较图15(*a*)、图15(*b*)和图15(*c*)中的点更加聚集在对角线上,即当特征组件全部使用时性能达到最优.这说明本文提出的底层形状特征和高层语义特征在描述窗口6DoF合成视频失真特性方面的有效性.对于IRCCyN/IVC DIBR视频数据库,其失真来源于压缩和绘制失真,这些失真会在不同程度上破坏高层语义特征信息;时域切片底层形状特征主要是对交互路径不适性引起的失真进行度量,而IRCCyN/IVC DIBR数据库中并不存在交互路径不适性失真.因此,在IRCCyN/IVC DIBR数据库中,采用空域高层语义特征能有效衡量失真,但融合底层形状特征后性能反而有所降低.



(*a*) $\boldsymbol{F}_{\text{vpd}}$　(*b*) $\boldsymbol{F}_{\text{SI\_S}}$　(*c*) $\boldsymbol{F}_{\text{SI\_T}}$　(*d*) Proposed

图15　不同特征训练模型所得客观分数与主观MOS的散点图

表9　消融实验结果

模型	特征			Windowed-6DoF			IRCCyN/IVC DIBR		
	$\boldsymbol{F}_{\text{vpd}}$	$\boldsymbol{F}_{\text{SI\_S}}$	$\boldsymbol{F}_{\text{SI\_T}}$	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
1	√	—	—	0.724 4	0.670 7	0.733 6	0.524 9	0.411 0	0.609 0
2	—	√	—	0.887 0	0.852 4	0.497 1	0.642 3	0.509 2	0.553 4
3	—	—	√	0.868 4	0.834 5	0.520 1	0.861 0	0.815 8	0.359 9
4	√	—	√	0.917 5	0.889 1	0.424 6	0.854 0	0.812 0	0.370 3
5	√	√	—	0.888 4	0.846 6	0.493 3	0.662 8	0.595 3	0.560 4
6	—	√	√	0.928 8	0.905 2	0.393 2	**0.865 0**	**0.820 5**	**0.358 3**
7	√	√	√	**0.932 7**	**0.911 0**	**0.385 3**	0.858 1	0.816 5	0.366 2

注:加粗表示性能最优.

5.7　算法复杂度分析

为了对本文的算法复杂度进行分析,本文在Windowed-6DoF数据库上测试并对比了部分评价性能相近的方法的平均运行时间和基于深度学习方法的模型参数量.这些方法包括Wen、Dendi和DoC-DoG-GRNN.所有方法均在i9-10900X CPU、64G内存和NVIDIA RTX 2080Ti GPU的计算机上进行测试,结果如表10所列.可以看出,在本文方法在运行时间上少于方法Dendi和DoC-DoG-GRNN,多于方法Wen.在模型参数方面,本文方法明显优于方法Wen.

表10　本文方法与对比方法的运行时间和参数量

方法	Wen	Dendi	DoC-DoG-GRNN	Proposed
时间/s	91.59	2 373.65	345.66	119.72
参数量/M	88.32	—	—	23.45

6　总结

窗口6DoF合成视频质量评价对促进窗口6DoF视频系统的发展和推广有着重要意义.本文提出了1个窗口6DoF合成视频主观质量视频数据库Windowed-6DoF,包含4种压缩质量、4种绘制方法和2条交互路径的共128个混合失真合成视频.同时,本文还提出了一种无参考窗口6DoF合成视频客观质量评价方法.首先在时域切片上提取底层形状特征,对交互路径不适性失真进行度量.然后分别提取空域、时域高层语义特征并进行降维处理.最后在本文数据库Windowed-6DoF和1个公开IRCCyN/IVC DIBR虚拟视点数据库上进行实验.实验结果表明,本文所提方法与现有方法相比具有一定的优越性,其客观预测分数与主观MOS值保持较高的一致性.

参考文献

[1] LIN T C, AOUIDIDI A, CHEN Z T, et al. VIRD: Immersive match video analysis for high-performance badminton coaching[J]. IEEE Transactions on Visualization and Computer Graphics, 2024, 30(1): 458-468.

[2] 丁颖, 刘延伟, 刘金霞, 等. 虚拟现实全景图像显著性检测研究进展综述[J]. 电子学报, 2019, 47(7): 1575-1583.
DING Y, LIU Y W, LIU J X, et al. An overview of research progress on saliency detection of panoramic VR images[J]. Acta Electronica Sinica, 2019, 47(7): 1575-1583. (in Chinese)

[3] 王旭, 刘琼, 彭宗举, 等. 6DoF视频技术研究进展[J]. 中国图象图形学报, 2023, 28(6): 1863-1890.
WANG X, LIU Q, PENG Z J, et al. Research progress of six degree of freedom (6DoF) video technology[J]. Journal of Image and Graphics, 2023, 28(6): 1863-1890. (in Chinese)

[4] JUNG J, KROON B, DORÉ R, et al. Common Test Conditions on 3DoF+ and Windowed 6DoF[R]. San Diego: MPEG, 2018.

[5] WEN W, LI M, ZHANG Y, et al. Modular blind video quality assessment[C]//Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 2763-2772.

[6] BI X D, HE X H, XIONG S H, et al. Blind video quality assessment based on spatio-temporal feature resolver[J]. Neurocomputing, 2024, 574: 127249.

[7] YUAN K, LIU H B, LI M D, et al. PTM-VQA: Efficient video quality assessment leveraging diverse PreTrained models from the wild[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 2835-2845.

[8] WU H N, CHEN C F, HOU J W, et al. FAST-VQA: Efficient end-to-end video quality assessment with fragment sampling[C]//[M]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 538-554.

[9] WU H N, CHEN C F, LIAO L, et al. Neighbourhood representative sampling for efficient end-to-end video quality assessment[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 15185-15202.

[10] WU H N, ZHANG E L, LIAO L, et al. Exploring video quality assessment on user generated contents from aesthetic and technical perspectives[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 20087-20097.

[11] BOSC E, PEPION R, LE CALLET P, et al. Towards a new quality metric for 3-D synthesized view assessment [J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(7): 1332-1343.

[12] SONG R, KO H, JAY KUO C C. MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source[J]. Journal of Information Science and Engineering, 2015, 31(5): 1593-1611.

[13] JUNG Y J, KIM H G, RO Y M. Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(7): 1201-1214.

[14] TIAN S S, ZHANG L, MORIN L, et al. A benchmark of DIBR synthesized view quality assessment metrics on a new database for immersive media applications[J]. IEEE Transactions on Multimedia, 2019, 21(5): 1235-1247.

[15] BOSC E, HANHART P, LE CALLET P, et al. A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data[C]//2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). Piscataway: IEEE, 2013: 100-105.

[16] LIU X K, ZHANG Y, HU S D, et al. Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4847-4861.

[17] WANG X C, WANG K, YANG B L, et al. Perceptual quality assessment on DIBR synthesized videos with composite distortions[C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2020: 186-190.

[18] PENG Z J, WANG S P, CHEN F, et al. Quality assessment of stereoscopic video in free viewpoint video system[J]. Journal of Visual Communication and Image Representation, 2019, 63: 102569.

[19] 高敏娟, 党宏社, 魏立力, 等. 全参考图像质量评价回顾与展望[J]. 电子学报, 2021, 49(11): 2261-2272.
GAO M J, DANG H S, WEI L L, et al. Review and prospect of full reference image quality assessment[J]. Acta Electronica Sinica, 2021, 49(11): 2261-2272. (in Chinese)

[20] Sadbhawna, JAKHETIYA V, CHAUDHARY S, et al. Perceptually unimportant information reduction and cosine similarity-based quality assessment of 3D-synthesized images[J]. IEEE Transactions on Image Processing, 2022, 31: 2027-2039.

[21] ZHANG H, ZHENG D S, ZHANG Y, et al. Quality assessment for DIBR-synthesized views based on wavelet transform and gradient magnitude similarity[J]. IEEE Transactions on Multimedia, 2024, 26: 6834-6847.

[22] THAKUR S, JAKHETIYA V, SUBUDHI B N, et al. Context region identification based quality assessment of 3D synthesized views[J]. IEEE Transactions on Multimedia, 2022, 25: 6183-6193.

[23] ZHANG Y, ZHANG H, YU M, et al. Sparse representation based video quality assessment for synthesized 3D videos[J]. IEEE Transactions on Image Processing, 2020,

29: 509-524.
[24] ZHANG Y, YANG X X, LIU X K, et al. High-efficiency 3D depth coding based on perceptual quality of synthesized video[J]. IEEE Transactions on Image Processing, 2016, 25(12): 5877-5891.
[25] JAKHETIYA V, GU K, JAISWAL S P, et al. Kernel-ridge regression-based quality measure and enhancement of three-dimensional-synthesized images[J]. IEEE Transactions on Industrial Electronics, 2021, 68(1): 423-433.
[26] TIAN S S, ZHANG L, MORIN L, et al. NIQSV: A no reference image quality assessment metric for 3D synthesized views[C]//2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2017: 1248-1252.
[27] TIAN S S, ZHANG L, MORIN L, et al. NIQSV+: A No-reference synthesized view quality assessment metric[J]. IEEE Transactions on Image Processing, 2018, 27(4): 1652-1664.
[28] GU K, JAKHETIYA V, QIAO J F, et al. Model-based referenceless quality metric of 3D synthesized images using local image description[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2018, 27(1): 394-405.
[29] GU K, QIAO J F, LEE S, et al. Multiscale natural scene statistical analysis for no-reference quality evaluation of DIBR-synthesized views[J]. IEEE Transactions on Broadcasting, 2020, 66(1): 127-139.
[30] WANG G, WANG Z, GU K, et al. Blind quality metric of DIBR-synthesized images in the discrete wavelet transform domain[J]. IEEE Transactions on Image Processing, 2020, 29: 1802-1814.
[31] Sadbhawna, JAKHETIYA V, MUMTAZ D, et al. Stretching artifacts identification for quality assessment of 3D-synthesized views[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 31: 1737-1750.
[32] LING S Y, LI J, CHE Z H, et al. Re-visiting discriminator for blind free-viewpoint image quality assessment[J]. IEEE Transactions on Multimedia, 2020, 23: 4245-4258.
[33] KIM H G, RO Y M. Measurement of critical temporal inconsistency for quality assessment of synthesized video[C]//2016 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2016: 1027-1031.
[34] WANG G C, WANG Z Y, GU K, et al. Reference-free DIBR-synthesized video quality metric in spatial and temporal domains[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1119-1132.
[35] WANG G C, SUN K Z, TANG L J. No-reference DIBR-synthesized video quality assessment based on spatio-temporal texture inconsistency measurement[C]//2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). Piscataway: IEEE, 2022: 1-4.
[36] SANDIC-STANKOVIC D D, KUKOLJ D D, LE CALLET P. Fast blind quality assessment of DIBR-synthesized video based on high-high wavelet subband[J]. IEEE Transactions on Image Processing, 2019, 28(11): 5524-5536.
[37] SANDIC-STANKOVIC D D, KUKOLJ D D, LE CALLET P. Quality assessment of DIBR-synthesized views based on sparsity of difference of closings and difference of Gaussians[J]. IEEE Transactions on Image Processing, 2022, 31: 1161-1175.
[38] JIN C C, PENG Z J, CHEN F, et al. Multi-modal learning-based blind video quality assessment metric for synthesized views[J]. IEEE Transactions on Broadcasting, 2024, 70(1): 208-222.
[39] YAN J B, LI J, FANG Y M, et al. Subjective and objective quality of experience of free viewpoint videos[J]. IEEE Transactions on Image Processing, 2022, 31: 3896-3907.
[40] JIA R L, ZHANG Y H, XU J, et al. Quality of experience assessment for free-viewpoint video[C]//2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Piscataway: IEEE, 2023: 1-6.
[41] WIEN M, BOYCE J M, STOCKHAMMER T, et al. Standardization status of immersive video coding[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2019, 9(1): 5-17.
[42] BAE S J, PARK S, KIM J W, et al. Camera Array Based Windowed 6-DOF Moving Picture Contents[R]. San Diego: MPEG, 2018.
[43] DOYEN D, BOISSON G, GENDROT R. EE_DEPTH: New Version of the Pseudo-Rectified Technicolor Painter Content[R]. Ljubljana: MPEG, 2018.
[44] BOISSONADE P, JUNG J. Proposition of New Sequences for Windowed-6DoF Experiments on Compression Synthesis and Depth Estimation[R]. Ljubljana: MPEG, 2018.
[45] JUNG J, BOISSONADE P, FOURNIER J, et al. Proposition of Navigation Paths and Subjective Evaluation Method for Windowed 6DoF Experiments on Compression, Synthesis, and Depth Estimation[R]. Ljubljana: MPEG, 2018.
[46] INSTALLATIONS T, LINE L. Subjective video quality assessment methods for multimedia applications[J]. Recommendation ITU-TP, 1999, 910(37): 5.
[47] CHA E Y, JALIL PIRAN M, SUH D Y. A gaze-based real-time and low complexity no-reference video quality assessment technique for video gaming[J]. Multimedia Tools and Applications, 2024, 83(7): 20889-20908.
[48] LI L D, LIN W S, WANG X S, et al. No-reference image

blur assessment based on discrete orthogonal moments[J]. IEEE Transactions on Cybernetics, 2016, 46(1): 39-50.
[49] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
[50] 胡钢, 徐翔, 张维明, 等. 基于主成分分析的网络节点重要性指标贡献评价[J]. 电子学报, 2019, 47(2): 358-365.
HU G, XU X, ZHANG W M, et al. Contribution analysis for assessing node importance indices with principal component analysis[J]. Acta Electronica Sinica, 2019, 47(2): 358-365. (in Chinese)
[51] CRIMINISI A, SHOTTON J, KONUKOGLU E. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning[J]. Foundations and Trends in Computer Graphics and Vision, 2012, 7(2-3): 81-227.
[52] MOORTHY A K, BOVIK A C. A two-step framework for constructing blind image quality indices[J]. IEEE Signal Processing Letters, 2010, 17(5): 513-516.
[53] 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.
[54] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a "completely blind" image quality analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.
[55] MITTAL A, SAAD M A, BOVIK A C. A completely blind video integrity oracle[J]. IEEE Transactions on Image Processing, 2016, 25(1): 289-300.
[56] LI D Q, JIANG T T, JIANG M, et al. Quality assessment of in-the-wild videos[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 2351-2359.
[57] SUN W, MIN X K, LU W, et al. A deep learning based no-reference quality assessment model for UGC videos[C]// Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 856-865.
[58] DENDI S V R, CHANNAPPAYYA S S. No-reference video quality assessment using natural spatiotemporal scene statistics[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2020: 29: 5612-5624.
[59] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
[60] 周程灏, 王治乐, 刘尚阔. 基于空间变化点扩展函数的图像直接复原方法[J]. 光学学报, 2017, 37(1): 110001.
ZHOU C H, WANG Z L, LIU S K. Method of image restoration directly based on spatial varied point spread function [J]. Acta Optica Sinica, 2017, 37(1): 110001. (in Chinese)

作者简介

唐婷琰 女, 2000 年 4 月生, 重庆江津人. 重庆理工大学电气与电子工程学院硕士研究生. 主要研究方向为视频质量评价.
E-mail: 1136854007@qq.com

陈 芬 女, 1973 年 3 月生, 四川邻水人. 重庆理工大学教授. 主要研究方向为视频图像处理.
E-mail: chenfen@cqut.edu.cn

邹文辉 女, 1979 年 11 月生, 四川广安人. 宁波大学与重庆理工大学联合培养博士研究生. 主要研究方向为多媒体信息处理、图像/视频质量评价.
E-mail: 37712814@qq.com

金充充 女, 1994 年 10 月生, 浙江绍兴人. 主要研究方向为视频图像处理.
E-mail: jinchongchong94@163.com

彭宗举 男, 1973 年 5 月生, 四川南部人. 重庆理工大学教授、博士生导师. 主要研究方向为多媒体信号处理与编码等.
E-mail: pengzongju@126.com