

沉浸式视频质量评价综述

曾焕强^{1,2} 丁 瑞² 黄海靓² 陈 婧² 朱建清¹

(1. 华侨大学工学院, 福建泉州 362021; 2. 华侨大学信息科学与工程学院, 福建厦门 361021)

摘 要: 近年来,随着计算机视觉处理技术的深入研究和虚拟现实设备的不断普及,沉浸式媒体技术飞速发展并在各个领域得到初步应用。考虑到沉浸式视频产生的质量损耗会影响用户身临其境的沉浸式视觉感受,基于沉浸式视频的质量评价研究成为视频处理领域的研究重点并逐渐引起研究者的重视。如何量化沉浸式视频的质量损耗并进行质量评价,设计符合人眼视觉感知的算法,用以准确高效地评估视频质量,对于沉浸式视频处理研究领域的发展具有学术研究意义和实际应用价值。本文首先回顾了沉浸式视频主观质量评价数据库和客观质量评价方法的研究,然后介绍了近期涉及沉浸式视频质量评价的主要技术,即深度信息和虚拟视图合成,最后对沉浸式视频质量评价研究工作进行总结和展望。

关键词: 沉浸式视频; 质量评价; 数据库; 深度信息; 虚拟视图合成

中图分类号: TN911.73 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2022.09.008

引用格式: 曾焕强,丁瑞,黄海靓,等. 沉浸式视频质量评价综述[J]. 信号处理,2022,38(9): 1853-1862. DOI: 10.16798/j.issn.1003-0530.2022.09.008.

Reference format: ZENG Huanqiang, DING Rui, HUANG Hailiang, et al. Immersive video quality assessment: An overview[J]. Journal of Signal Processing, 2022, 38(9): 1853-1862. DOI: 10.16798/j.issn.1003-0530.2022.09.008.

Immersive Video Quality Assessment: An Overview

ZENG Huanqiang^{1,2} DING Rui² HUANG Hailiang² CHEN Jing² ZHU Jianqing¹

(1. School of Engineering, Huaqiao University, Quanzhou, Fujian 362021, China;

2. School of Information Science and Engineering, Huaqiao University, Xiamen, Fujian 361021, China)

Abstract: In recent years, immersive media technology has developed rapidly and gotten initial implementation in many fields with the computer vision research and the continuous emergence of virtual reality devices. Considering that the quality loss generated by distorted immersive video is reflected in the user's intuitive perception, quality assessment research based on immersive video has become the study focus in the field of immersive video processing and has gradually attracted the attention of researchers. How to quantify the quality loss of immersive video and to evaluate the quality, and design algorithms that match the visual perception of human eyes for accurate and efficient assessment of video quality are of academic research significance and practical application value for the development of the research field of immersive video processing. This paper reviews recent research on immersive video subjective quality assessment databases and objective quality assessment methods. In addition, this paper analyzes the latest techniques related to immersive video quality assessment, such as depth information and virtual view synthesis. Finally, the research work on immersive video quality assessment is summarized and prospected.

Key words: immersive video; quality assessment; database; depth information; virtual view synthesis

收稿日期: 2022-06-02; 修回日期: 2022-09-06

基金项目: 国家重点研发计划项目(2021YFE0205400); 国家自然科学基金项目(61871434); 福建省自然科学基金(2022J06023, 2022J01294); 福厦泉国家自主创新示范区协同创新平台项目(2021FX03); 福建省教改项目(FBJG20180038)

1 引言

随着视觉媒体技术的快速发展,人们在学习、工作和休闲等各个方面都渴望更加真实的视觉体验,沉浸式视频(Immersive video)获得了越来越多的关注。沉浸式的概念早期出现于电影领域,通过覆盖人眼至少 120° (水平) $\times 70^\circ$ (垂直)视场角的巨幕,与音频及特效系统相辅助,能让观众完全沉浸在视听环境中并有身临其境的感觉^[1]。目前,沉浸式视频通过人机交互、虚拟现实(Virtual Reality, VR)等技术,可将现实环境营造成兼具画面包围感和沉浸式音响的场景,并通过头戴式显示器(Head-Mounted Display, HMD)、传感器及手柄等设备实现沉浸式体验。

伴随着数字时代技术的更新,沉浸式视频也在不断更新迭代,高分辨率、高帧率、高色深、广色域、高动态的视频能提供更为丰富的画面层次,保持更为精致的画面细节,进一步强化沉浸式视频的空间表现力和临场感,更好地提升沉浸式视听效果。与此同时,通过HMD观察的沉浸式视频格式也不断在变化,MPEG会议上针对沉浸式媒体的标准化工作制定了MPEG-I标准^[2],其根据视频格式将沉浸式视频分为了三个阶段:3个自由度(3DoF)、3DoF+(3 Degrees of Freedom Plus)和6DoF^[3-5]。

与传统视觉信号相同,沉浸式视频在采集、处理、传输和展示等过程中,不可避免地受到外界干扰而导致视频质量下降。所以如何量化沉浸式视频的质量损耗并进行质量评价,对于沉浸式视频处理研究领域的发展具有学术研究意义和现实应用价值。近年来,基于3DoF的沉浸式视频质量评价研究已引起学术界的广泛关注,并取得了初步进展^[6-10],关于6DoF沉浸式视频的相关研究也在逐步探索中。沉浸式视频质量评价的研究目标是设计符合人眼视觉主观感知的算法模型,用以准确高效地评估视频质量。研究内容主要包括:主观质量评价数据库和客观质量评价方法。

本文其余部分的安排如下:第2节介绍基于3DoF的沉浸式视频质量评价数据库和质量评价算法,第3节介绍涉及6DoF沉浸式视频质量评价的主要技术,第4节给出总结和未来的展望。

2 3DoF 沉浸式视频的质量评价

如图1所示,3DoF 沉浸式视频指观察者在固定位置上于三个旋转度(偏航角(Yaw)、俯仰角(Pitch)、滚转角(Roll))上自由活动,即在一个固定的观看位置上环顾四周,如 360° 视频^[11],即具有 360° 宽广视野感知范围的视频,用户可视作球体的中心点,与球体表面的不同视频信息具有相同距离,并可通过改变方向来选择观看内容。3DoF+沉浸式视频是对3DoF的有限修改,即增加了三个位置维度(X, Y, Z)上头部的有限范围活动(脚步固定,不允许走动),其更加符合用户感知事物的真实过程,并能兼容混合现实(Mixed Reality, MR)和增强现实(Augmented Reality, AR)。6DoF 沉浸式视频是3DoF+的进一步改进,在3DoF视频的基础上额外提供3个平动自由度,即视频内容可根据观众位置的移动而改变,支持用户更大范围的观察行动。

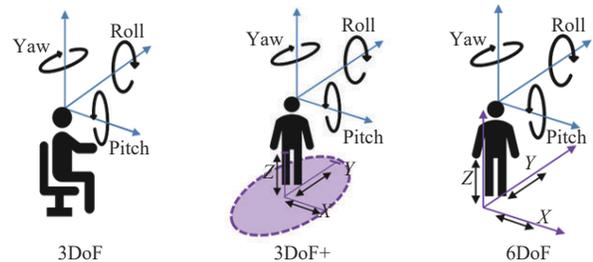


图1 关于3DoF(左)、3DoF+(中)、6DoF(右)的划分^[3]

Fig. 1 About the classification of 3DoF(left), 3DoF+(middle), 6DoF(right)^[3]

在3DoF的概念提出之前,通过HMD实现的沉浸式体验通常基于自然/屏幕视频和 360° 视频,而现有的3DoF沉浸式视频一般都指代 360° 视频(或称为全向视频、全景视频)。用户可通过改变方向来选择观看内容。

2.1 3DoF 沉浸式视频的主观质量评价数据库

沉浸式视频主观评价方法一般遵照国际标准ITU-R BT. 500-13^[12]中所规定的实验设置搭建主观质量评价平台,在特定环境内观看测试序列并基于各项规定方法进行打分,再对评分进行处理分析得到主观平均得分(Mean Opinion Score, MOS)或主观差异平均得分(Differential Mean Opinion Score, DMOS),以此反映失真沉浸式视频的感知质量退化。现有的主观评价方法主要分为以下几种:

(1)双刺激损伤尺度法(Double Stimulus Impairment Scale, DSIS):将参考视频和失真视频以“视频对”的形式展示给受试者,观看顺序为参考视频在前,失真视频在后,即对参考视频有主观印象后再进行失真视频的评分,通常采用表 1 所示的 5 分制评分表。

表 1 DSIS 的 5 分制失真测度

Tab. 1 Distortion measures on the 5-point scale for DSIS

5	察觉不到的
4	能察觉但不令人讨厌
3	稍微令人讨厌的
2	令人讨厌的
1	非常令人讨厌的

(2)双刺激连续质量尺度法(Double Stimulus Continuous Quality Scale, DSCQS):与 DSIS 观看方式大致相同,区别是参考和失真视频的播放顺序随机,且二者都需要评分,通常采用图 2 所示的 5 分制失真测度。

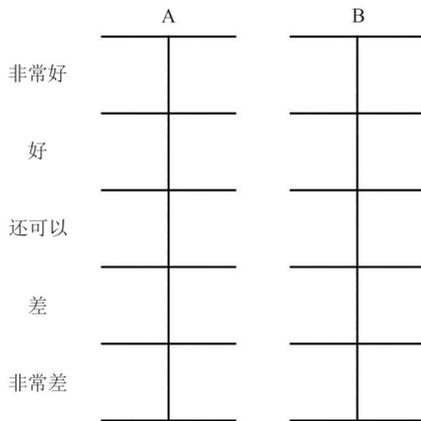


图 2 DSCQS 的评分测度

Fig. 2 Scoring measures of DSCQS

(3)单刺激法(Single Stimulus Methods, SSM):在不观看参考视频的前提下,以随机方式显示测试视频并进行主观评分,具体实现时根据播放次数分为两种方式:不重复播放视频序列(Single Stimulus, SS)和重复播放测试序列(Single Stimulus with Multiple Repetition, SSMR)。除了常用的 5 分制评分,还可以通过采用图 3 所示 9 分制或 11 分制提高精度^[12]。

(4)单刺激连续质量评估方法(Single Stimulus

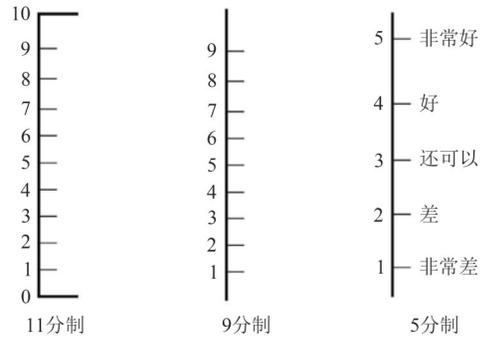


图 3 SSM 评分测度

Fig. 3 Scoring measures of SSM

Continuous Quality Evaluation, SSCQE):在不观看参考视频的前提下,选择序列的时间较长,分值取平均,对序列评分时不仅考虑分值大小,还需关注评分时长。

(5)同时双激励连续质量评估方法(Simultaneously Double Stimulus Continuous Evaluation, SDSCE):该方法以 SSCQE 为基准条件而制定,用以衡量较长序列的视频质量。

(6)多媒体视频质量主观评估方法(Subjective Assessment Method for Video Quality evaluation, SAMVIQ)^[13]:1)训练测试人员,播放一组参考视频及相应的测试序列(质量从高到低排列),给定测试人员以基本的预期得分;2)进行预测试,播放一组包含参考视频和三个失真视频的视频集合,当测试人员的评分结果与预期得分相匹配时可正式测试;3)测试人员观看多类视频序列,每组包含一个原始视频和三个随机的失真序列,并基于连续评分测度进行打分,即在 0 到 100 之间的连续尺度(分 5 个等级分别表示很好、好、一般、差、很差)上对视频序列进行打分。

沉浸式视频质量评价数据库反映了人类视觉系统对于沉浸式视频的主观质量感知,是设计和评判沉浸式视频客观质量评价方法的重要依据。目前,基于 3DoF 沉浸式视频的主观质量评价数据库已经较为成熟,考虑到数据库的多样性和实用性,本文具体介绍几种常用的数据库,如表 2 所示。

1)NAMA3DS1-COSPAD1 数据库^[14]:由 10 个原始参考视频、100 个失真视频及对应主观评分组成,失真类型包括 H.264/AVC 压缩失真、JPEG2000 压缩失真、图像锐化和下采样锐化等。

表2 几种常用沉浸式视频数据库
Tab. 2 Several common immersive video databases

常用沉浸式视频质量评价数据库				
年份	数据库	参考视频	失真视频	失真类型
2012	NAMA3DS1-COSPADI ^[14]	10	100	H.264/AVC、JPEG 2000、下采样、图像锐化、下采样锐化
2016	QI-SVQA ^[15]	9	450	高斯模糊、H.264压缩失真
2017	IVQAD 2017 ^[16]	10	150	帧率、分辨率、码率
2017	VR-VQA48 ^[17]	12	36	QP值
2018	VQA-ODV ^[18]	60	540	H.265量化参数压缩、投影失真
2018	IIP-IVQD ^[19]	10	50	HEVC压缩失真
2018	VRQ-TJU ^[20]	13	364	H.264、JPEG2000压缩失真
2022	VOD-VQA ^[21]	18	774	帧大小、帧率、QP值

2) QI-SVQA 数据库^[15]: 包括9个原始视频和450个失真视频, 失真类型包括高斯模糊和H.264压缩失真。

3) IVQAD 2017 数据库^[16]: 由 Insta 360 4K 球形VR摄像机拍摄的10个不同场景构成, 涵盖了人物、草坪、建筑物等室内外场景, 通过设置不同的帧率、分辨率和码率模拟失真情况。

4) VR-VQA48 数据集^[17]: 采集了12个参考视频, 并通过设置4种QP值模拟压缩失真。

5) VQA-ODV^[18] 数据库: 参考视频涉及60个不同的场景, 包括真实场景与CG合成场景。该数据库的失真视频构建考虑了沉浸式视频在传输、重建以及显示过程中可能引入的失真效应, 即3种不同量化参数(QP=27、37、42)下的H.265压缩失真与3种不同投影格式(ERP、RCMP、TSP)的组合失真, 每个视频序列对应9种质量损失程度。

6) IIP-IVQD 数据集^[19]: 在10个参考场景下通过HEVC压缩失真模拟获取50个失真视频。

7) VRQ-TJU 数据集^[20]: 包含13个VR原始视频, 并基于H.264和JPEG2000压缩失真构建失真视频数据集, 其中包含104个对称视频和260个非对称视频。

8) VOD-VQA 数据库^[21]: 选择18个360°视频序列作为参考视频, 通过设置不同的帧大小、帧率和QP值得到774个失真视频。

主观实验的观看设备通常为HMD, 后续再采用不同的主观评分标准进行打分。值得注意的是, 现有的主观实验仅能实现HMD观看者的口述评分与手动记录数据, 主观质量评价实验所需的大量数据

处理工作仍然是亟待解决的一个问题。

2.2 3DoF 沉浸式视频客观质量评价方法

沉浸式视频的客观质量评价方法是通过数学统计方法预测失真视频的质量损耗, 可以有效弥补主观质量评价方法适用范围的局限, 主观实验物力及人力耗费过高, 做不到实时性, 同时结果易受观察者和观测环境的影响。客观质量评价方法计算简便, 实用性强。根据对参考视频的依赖程度可分为全参考、半参考和无参考方法^[22]。现有的沉浸式视频客观评价方法主要是针对3DoF沉浸式视频, 包括全参考质量评价和无参考质量评价两种类型。

关于全参考质量评价方法, Sun等^[22]提出了一种基于球形均匀加权的方法来准确评估失真360°视频的客观质量, 通过对投影平面上每个像素点的误差引入对应的权重值, 避免了由重采样表示空间转换到观察空间时产生的误差传播。Azevedo等^[23]提出了一种基于视口的多度量融合方法, 通过组合从360°视频的视口中提取的多种时空客观质量指标(特征)来实现失真360°视频的质量评估。Gao等^[24]考虑到失真空间特征和对应时间变化对眼视觉的影响, 提出了一种时空建模方法, 并将三种现有的VQA指标(S-PSNR、CPP-PSNR、WS-PSNR)集成进该方法, 提高了失真360°视频的质量评估准确性。

关于无参考沉浸式视频质量评价方法, Zhang等^[25]通过整合代表不同失真因素的质量因子, 构建了一种VR环境下的视听质量评估框架。Li等^[26]同时考虑于视口方案(viewport proposal)和视口显著性预测, 提出了基于视口方案的CNN方法

来预测失真 360° 视频的质量评分。Zhang 等^[27]基于球域提取 360° 视频的空间和时间特征,并通过多核学习(Multiple Kernel Learning, MKL)回归计算失真视频质量分数。Yang 等^[28]设计了一种端到端的神经网络模型来实现 360° 视频质量评价,该算法通过结合球形 CNN 和非局部神经网络,有效地提取 360° 视频的复杂时空特征。

然而,360° 视频不支持人眼的运动视差,观看 360° 视频时,物体的相对位置信息无法根据观看者与物体的相对位置变化而更新,这与用户在现实世界中的体验背道而驰,会导致感知的不适应。3DoF+沉浸式视频作为 3DoF 至 6DoF 阶段的过渡产物,除了提供三个旋转度上的体验之外,还允许一定范围的头部活动,而 6DoF 视频较 3DoF+视频进一步增加了用户自由度,即视频内容可以根据用户观看位置的移动而改变。因此,6DoF 沉浸式视频较 3DoF 视频可提供正确的运动视差,带来与真实世界更为相似的观看体验,并可进一步满足沉浸式视频的相关需求。

因此,观测自由度及视差信息的增加使得基于 3DoF 沉浸式视频所设计的客观质量评价方法不适用于 6DoF 视频,如何探索出针对人类视觉系统和 6DoF 沉浸式视频特征的高效客观质量评价算法模型是当前视频处理领域的研究热点。

3 6DoF 沉浸式视频概念及相关技术

3.1 6DoF 视频概念

MPEG-I 制定了沉浸式视频标准,具体涉及 3DoF、3DoF+视频及 6DoF 视频等等^[29]。其中提及 6DoF 视频具有图 4 所示的三种类型:窗口式 6DoF^[30]、全向 6DoF^[31]以及 6DoF。其中,窗口式 6DoF 需要借助显示器的帮助,在窗口受限的情况下观看视频场景,窗口式 6DoF 内容是由普通 2D 相机拍摄。全向 6DoF 指提供不受限制的自由旋转角度(Yaw、Pitch、Roll)和有限体积内的小幅度身体平移活动,是一种受限制的 6DoF 形式,与 3DoF+相似,但允许更大的视角范围,全向 6DoF 内容是由全向(360° 视频)摄像机拍摄。

MPEG-I 定义了 6DoF 沉浸式视频的源视频格式,主要分为多视图视频加深度(Multiview Video Plus Depth, MVD)、多平面图像(Multiplane Image,

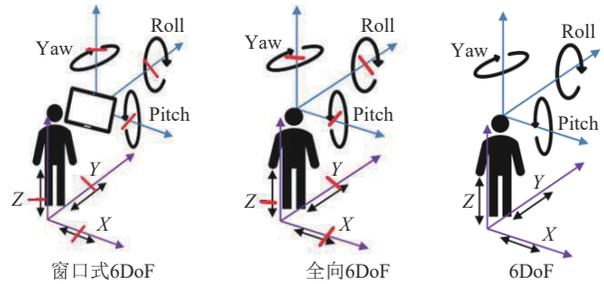


图 4 窗口式 6DoF(左)、全向 6DoF(中)和 6DoF(右)^[3]
Fig. 4 Windowed 6DoF(left)、Omnidirectional 6DoF(middle)和 6DoF(right)^[3]

MPI) 和多球形图像(Multi Sphere Image, MSI)。6DoF 视频主要采用 MVD 格式,如图 5 所示,具有每帧相关联的多视图纹理视频和深度视频,视频子集为具有多个视图的立体视频格式,通常由多相机系统采集^[32-35]。图 5(b)显示的深度视频序列场景与图 5(a)中的纹理视频相对应,并可通过深度估计算法从后者提取。MVD 格式源自单视图视频加深度(Single-View Video Plus Depth, VPD),可利用深度图来渲染虚拟视图。由于 MVD 格式表示的视图合成基于多个纹理视图和深度视图,其合成视图的质量通常优于 VPD 格式的质量。MPI 格式或 MSI 格式主要应用于沉浸式光场视频。将视频的每一层像素固定为某一深度,MPI 视频可近似为一叠排列在不同的深度的半透明彩色层,如图 6(a)所示,其是



(a) MVD 视频中的纹理视频



(b) MVD 视频中的深度视频

图 5 6DoF 沉浸式视频的 MVD 视频源格式

Fig. 5 MVD video source format for 6DoF immersive video



(a) MPI视频

(b) MSI视频

图6 沉浸式光场视频的MPI视频和MSI视频源格式

Fig. 6 MPI video and MSI video source formats for immersive light field video

由二维规则网格图像组成的单一矩形视频。这些纹理视图的分辨率相同,是由多个相机(如微透镜阵列光场相机)拍摄得到的^[36-37]。将来自相机阵列的视频流编码为一组具有RGBA纹理的同心球体,得到如图6(b)所示的MSI视频。但由于MSI格式所占空间过大,存储数据过多不易压缩,不利于传输与展示,通常将其转换为MPI格式以作后续处理。

MPEG-I标准除了对沉浸式视频做出的规划和技术支持,其还提供部分可下载测试参考序列(Common Test Conditions, CTC)^[38],其中包含了纹理信息和深度信息。这些序列涵盖了18种不同场景,分别以普通视口的自然视图或ERP投影格式下的球面虚拟视图呈现。目前,MPEG-I标准架构下已提出了许多关于6DoF视频的算法模型,如视频的编码和解码技术,用于提取纹理视频中深度信息的深度估计技术,以及对针对场景中某一目标视点的虚拟视图合成技术等。考虑到深度估计技术和虚拟视图合成技术对沉浸式视觉感知的影响,二者已受到计算机图形学领域的广泛关注,并有望应用于沉浸式质量评价研究工作中。接下来我们分别介绍影响人眼视觉感知的6DoF关键技术:深度估计技术和虚拟视图合成技术。

3.2 深度估计技术

针对6DoF沉浸式视频的特征提取方案无法仅依靠纹理图像来表征场景的全部信息,而深度信息表征了拍摄对象与相机之间的空间距离,可通过计算各点之间的相对距离对场景信息进行额外补充。具体来说,相机与拍摄场景中某个像素点的相对距离可由深度信息推断,而深度信息一般以归一化视差的形式来适应人眼视觉系统,即目标像素点的灰度值反映了该点与摄像机的距离,因此更亮的区域对应更高的视差(更小的相机拍摄距离)。然而,自然场景信息采集

时可能无法直接获取深度信息,深度估计技术应运而生,即利用纹理信息来计算出场景深度信息^[39-40]。

目前,获取深度信息的方式主要分为主动方式和被动方式两种。其中,主动方式需要采集设备自身发射能量,常用设备包括结构光相机^[41]、TOF相机和激光雷达等。以TOF相机^[42]为例,其通过比较发射脉冲和反射脉冲的相位差来推算传输过程中的传输延迟,进而计算场景中拍摄对象与相机的距离。然而TOF相机仍存在一些缺陷,比如相机的造价较高、采集过程易受噪声的干扰、相机获取的深度图像分辨率不及纹理图像的分辨率大,以及物体的边缘深度信息不准确等。此外,在被动获取深度信息的方式中,双目立体视觉^[43-44]是比较通用的方法,即对同时获得的两张同一场景图像使用匹配算法,找寻二者间相适配的像素点,再应用三角原理得出视差,针对多张图像进行相同运算得到场景深度信息。类似的,该算法同样适用于同一场景下不同角度的图像深度信息获取。但基于双目立体视觉获取视差信息的过程易受像素点匹配精度的影响,其生成的视图仍然存在一定误差。除此之外,深度信息也能通过间接估算图像的光度特征^[45]、明暗特征^[46]等特征获取。以上方法主要都针对3DoF视频设计,对于6DoF视频具有局限性,不能完全满足其深度估计的相关需求。

为此,MPEG-I标准提出一种高质量的深度估计软件(Depth Estimation Reference Software, DERS),目前已经更新到DERS 8.0^[47-49]。DERS主要由表3所示的三个部分组成,即匹配成本、时间增强和图像切割。首先,基于所有可能的标签(如深度信息)为参考图像的每个像素计算一个误差成本;其次,将每个像素和标签的最终误差设定为图像对的最小误差;最后,将其进行图形切割优化,以找到每个像素的最佳深度估计。

3.3 虚拟视图合成技术

6DoF沉浸式视频支持双目立体视差、运动视差,提供了与现实世界更相似的观看体验。但是由于采集和存储条件的限制,任意视点间的视觉信息不能完全获取,即在位移、遮挡等情况下导致采集时的信息丢失。现有情况下,如何有效利用已知视点预测目标视点是6DoF视频重建及展示阶段亟须解决的一项重要问题。

表 3 DERS算法框架^[39]

Tab. 3 Basic framework of DERS^[39]

算法 1
概述
需求: $R =$ 参考图像, $I =$ 搜寻图像, $N =$ 标签数量
for all $l \in [0, N]$ do
for all $i \in I$ do
for all $p \in R$ do
基于参考图像的运动信息 $R[p]$ 计算误差成本 $[p][l][i]$
end for
最终误差 $E[p][l] = \min_i(\text{误差成本}[p][l][i])$
end for
end for
graphCut(E) //图形切割优化

基于图像的视图合成算法通过参考已知视点的信息,利用 3D-Warping 投影、插值等手段直接合成虚拟视图,高效生成符合需求的虚拟视图,因此基于图像的虚拟视图合成算法得到了广泛应用。此外,目前基于深度图像探索算法模型对已知场景进行虚拟视图合成的研究领域具有一定的热点。基于深度图像目标视点的虚拟视图合成方法 (Depth Image Based Rendering, DIBR)^[50-52]通过一组视频序列 (纹理视频和深度视频) 和相应的信息来生成非获取视点的图像,其算法框架如图 7 所示,主要分为三步: 3D-Warping 投影、融合和图像修复。首先,对输入纹理视图及深度视图进行变换,各自生成一个虚拟的目标视图;其次,融合两幅生成的虚拟视图;最后,对融合后产生的空洞进行填充修复。然而, DIBR 方法存在缺陷,其主要针对传统 2D 场景。由于高自由度沉浸式视频的映射方式区别于 2D 视频,满足高自由度沉浸式视频的虚拟视图合成模型亟待探索。

针对 6DoF 沉浸式视频, MPEG-I 标准提出两种参考视图合成工具 RVS (Reference View Synthesizer) 及 VSRS (View Synthesis Reference Software)^[53-55], 克服了低数量参考视图合成虚拟视图的限制, 支持任意数量的输入视图进行视图合成, 采用自然、逼真的渲染且无需事先进行 3D 建模。值得注意的是, 与 RVS 相比, VSRS 在非平面相机对准和步进方面的结果较差, 然而这两种软件工具都显示出较好的发展趋势, 即有助于在未来工作中推断最佳输入的相机位置及数量, 满足高自由度沉浸式视频的采集和渲染需求。

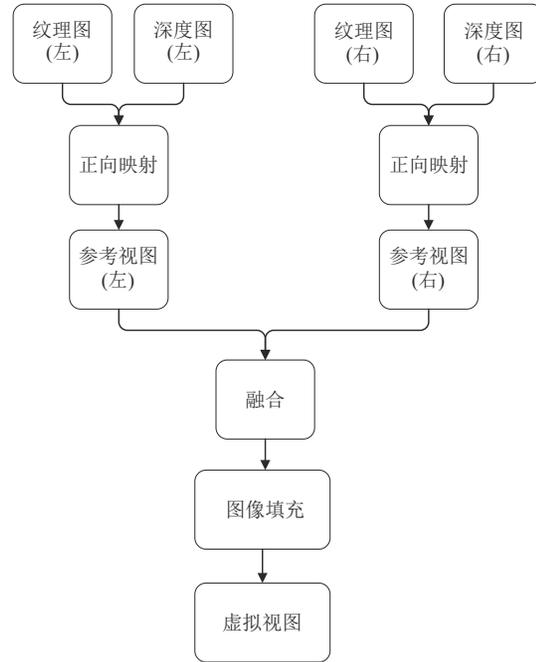


图 7 DIBR 算法基本框架

Fig. 7 Basic framework of DIBR

3.4 6DoF 沉浸式视频质量评价

深度估计技术和虚拟视图合成技术对沉浸式视觉感知尤为重要, 6DoF 沉浸式视频相比于 3DoF 沉浸式视频, 表征的信息更多, 因此更能带来沉浸式的视觉感知体验。对于沉浸式视频的质量评价而言, 除了常用的纹理信息、边缘信息和色度信息等, 也需要考虑深度信息, 用于模拟主观感知场景过程中对景深信息的感知, 因此深度估计技术对 6DoF 视频主观评价过程具有重要意义。而虚拟视图合成技术能够对任意已知视图间的虚拟视图进行合成, 视图能够进行任意切换满足了 6DoF 视频观看时更高的沉浸式视觉感知体验。在 6DoF 沉浸式视频质量评价研究工作中, 除了利用多数经典算法所考虑的纹理信息、色度信息等信息, 如何利用这两种技术设计出性能优越、鲁棒性好的算法模拟视觉感知, 以此满足 6DoF 沉浸式视频身临其境的体验需求, 仍有待探索。

4 总结和展望

本文首先概述沉浸式视频的概念和研究背景, 然后就 3DoF 沉浸式视频的主观质量评价数据库和客观质量评价算法展开介绍, 最后分析影响 6DoF 沉浸式视频质量评价的相关技术, 并简要介绍了最

新成果。随着人们在学习、工作和休闲等各个方面对视觉媒体不断增强的视觉体验需求,不同应用场景下的视频质量评估和质量感知优化工作需求日益递增。未来的研究工作中,端到端的沉浸式视频处理系统在拍摄、制作、编转码、传输分发和呈现设备中仍存在很大的发展空间,高自由度沉浸式视频也将在医疗、教育、虚拟游戏等领域得到广泛应用。因此,针对沉浸式视频的质量评价研究同时具有学术研究意义和现实应用意义,其主观和客观研究方向有望深入探索并应用于更多具体场景。

参考文献

- [1] 周屹,周耀平,欧阳玥.《5G高新视频-沉浸式视频技术白皮书(2020)》解读[J].广播与电视技术,2020,47(11):21-25.
ZHOU Yi, ZHOU Yaoping, OUYANG Yue. Interpretation of "white paper on 5G advanced-format and new-concept video-immersive video technology (2020)" [J]. Radio & TV Broadcast Engineering, 2020, 47(11): 21-25. (in Chinese)
- [2] HSU C F, HUNG T H, HSU C H. Optimizing immersive video coding configurations using deep learning: A case study on TMIV [J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2022, 18(1): 1-25.
- [3] ALAIN M, ZERMAN E, OZCINAR C. Immersive imaging technologies: From capture to display [C]//MM '20: Proceedings of the 28th ACM International Conference on Multimedia, 2020: 4787-4788.
- [4] SEO B S, HWANG S, LEE Y H, et al. Viewpoint-switching delay of multi-viewpoint 360 VR in a bandwidth-limited environment [C]//2021 IEEE Region 10 Symposium. Jeju, Korea, Republic of. IEEE, 2021: 1-5.
- [5] KIM H H, LIM S G, LEE G, et al. Efficient patch merging for atlas construction in 3DoF+ video coding [J]. IEICE Transactions on Information and Systems, 2021, 104(3): 477-480.
- [6] BARREDA-ÁNGELES M, ALEIX-GUILLAUME S, PEREDA-BAÑOS A. Virtual reality storytelling as a double-edged sword: Immersive presentation of nonfiction 360°-video is associated with impaired cognitive information processing [J]. Communication Monographs, 2021, 88(2): 154-173.
- [7] SINGLA A, FREMEREY S, ROBITZA W, et al. Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution [C]//Thematic Workshops '17: Proceedings of the on Thematic Workshops of ACM Multimedia 2017, 2017: 511-519.
- [8] ZHANG Bo, ZHAO Junzhe, YANG Shu, et al. Subjective and objective quality assessment of panoramic videos in virtual reality environments [C]//2017 IEEE International Conference on Multimedia & Expo Workshops. Hong Kong, China. IEEE, 2017: 163-168.
- [9] XU Mai, SONG Yuhang, WANG Jianyi, et al. Predicting head movement in panoramic video: A deep reinforcement learning approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(11): 2693-2708.
- [10] JIN Y, GOODALL T, PATNEY A, et al. A foveated video quality assessment model using space-variant natural scene statistics [C]//2021 IEEE International Conference on Image Processing. Anchorage, AK, USA. IEEE, 2021: 1419-1423.
- [11] FAN C L, LO W C, PAI Y T, et al. A survey on 360° video streaming [J]. ACM Computing Surveys, 2020, 52(4): 1-36.
- [12] NARITA N. Method for the subjective assessment of the quality of television pictures recommended by CCIR rec. 500-5 [J]. The Journal of the Institute of Image Information and Television Engineers, 1993, 47: 979-982.
- [13] KOZAMERNIK F, STEINMANN V, SUNNA P, et al. SAMVIQ—A new EBU methodology for video quality evaluations in multimedia [J]. SMPTE Motion Imaging Journal, 2005, 114(4): 152-160.
- [14] URVOY M, BARKOWSKY M, COUSSEAU R, et al. NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences [C]//2012 Fourth International Workshop on Quality of Multimedia Experience. Melbourne, VIC, Australia. IEEE, 2012: 109-114.
- [15] QI Feng, ZHAO Debin, FAN Xiaopeng, et al. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models [J]. Signal, Image and Video Processing, 2016, 10(4): 737-744.
- [16] DUAN Huiyu, ZHAI Guangtao, YANG Xiaokang, et al. IVQAD 2017: An immersive video quality assessment database [C]//2017 International Conference on Systems, Signals and Image Processing (IWSSIP). Poznan, Poland. IEEE, 2017: 1-5.
- [17] XU Mai, LI Chen, LIU Yufan, et al. A subjective visual quality assessment method of panoramic videos [C]//2017 IEEE International Conference on Multimedia and Expo. Hong Kong, China. IEEE, 2017: 517-522.
- [18] LI Chen, XU Mai, DU Xinzhe, et al. Bridge the gap be-

- tween VQA and human behavior on omnidirectional video: A large-scale dataset and a deep learning model [C]//MM '18: Proceedings of the 26th ACM International Conference on Multimedia, 2018: 932-940.
- [19] ZHANG Yingxue, WANG Yingbin, LIU Feiyang, et al. Subjective panoramic video quality assessment database for coding applications[J]. IEEE Transactions on Broadcasting, 2018, 64(2): 461-473.
- [20] YANG Jiachen, LIU Tianlin, JIANG Bin, et al. 3D panoramic virtual reality video quality assessment based on 3D convolutional neural networks [J]. IEEE Access, 2018, 6: 38669-38682.
- [21] MENG Yu, MA Zhan. Viewport-based omnidirectional video quality assessment: Database, modeling and inference[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(1): 120-134.
- [22] SUN Yule, LU Ang, YU Lu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video [J]. IEEE Signal Processing Letters, 2017, 24(9): 1408-1412.
- [23] DE A AZEVEDO R G, BIRKBECK N, JANATRA I, et al. A viewport-driven multi-metric fusion approach for 360-degree video quality assessment[C]//2020 IEEE International Conference on Multimedia and Expo. London, UK. IEEE, 2020: 1-6.
- [24] GAO Pan, ZHANG Pengwei, SMOLIC A. Quality assessment for omnidirectional video: A spatio-temporal distortion modeling approach[J]. IEEE Transactions on Multimedia, 2022, 24: 1-16.
- [25] ZHANG Bo, YAN Zhaoyu, WANG Jing, et al. An audio-visual quality assessment methodology in virtual reality environment[C]//2018 IEEE International Conference on Multimedia & Expo Workshops. San Diego, CA, USA. IEEE, 2018: 1-6.
- [26] LI Chen, XU Mai, JIANG Lai, et al. Viewport proposal CNN for 360° video quality assessment[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 10169-10178.
- [27] ZHANG Yingxue, LIU Zizheng, CHEN Zhenzhong, et al. No-reference quality assessment of panoramic video based on spherical-domain features [C]//2021 Picture Coding Symposium (PCS). Bristol, United Kingdom. IEEE, 2021: 1-5.
- [28] YANG Jiachen, LIU Tianlin, JIANG Bin, et al. Panoramic video quality assessment based on non-local spherical CNN [J]. IEEE Transactions on Multimedia, 2021, 23: 797-809.
- [29] BOYCE J M, DORÉ R, DZIEMBOWSKI A, et al. MPEG immersive video coding standard[J]. Proceedings of the IEEE, 2021, 109(9): 1521-1536.
- [30] 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.
- [31] ATTAL B, LING S, GOKASLAN A, et al. MatryODShka: Real-time 6DoF video view synthesis using multi-sphere images [C]//European Conference on Computer Vision. Springer, Cham, 2020: 441-459.
- [32] TANIMOTO M. FTV (free-viewpoint TV) [C]//2010 IEEE International Conference on Image Processing. Hong Kong, China. IEEE, 2010: 2393-2396.
- [33] SCHENKEL A, BONATTO D, FACHADA S, et al. Natural scenes datasets for exploration in 6DOF navigation[C]//2018 International Conference on 3D Immersion (IC3D). Brussels, Belgium. IEEE, 2018: 1-8.
- [34] STANKIEWICZ O, DOMAŃSKI M, DZIEMBOWSKI A, et al. A free-viewpoint television system for horizontal virtual navigation [J]. IEEE Transactions on Multimedia, 2018, 20(8): 2182-2195.
- [35] GOORTS P, DUMONT M, ROGMANS S, et al. An end-to-end system for free viewpoint video for smooth camera transitions[C]//2012 International Conference on 3D Imaging (IC3D). Liege, Belgium. IEEE, 2012: 1-7.
- [36] LI Jisheng, HE Yuze, JIAO Jinghui, et al. Extending 6-DoF VR experience via multi-sphere images interpolation [C]//MM '21: Proceedings of the 29th ACM International Conference on Multimedia, 2021: 4632-4640.
- [37] 马晓辉, 曾焕强, 陈婧, 等. 基于多视点伪序列的光场图像压缩[J]. 信号处理, 2019, 35(3): 378-385.
- MA Xiaohui, ZENG Huanqiang, CHEN Jing, et al. Light field image compression based on multi-view pseudo sequence [J]. Journal of Signal Processing, 2019, 35(3): 378-385. (in Chinese)
- [38] PARK M W, PARK J T, PARK G H, et al. Chrominance compensation for multi-view video coding [J]. IEICE Transactions on Information and Systems, 2009, E92.D(2): 353-356.
- [39] ROGGE S, BONATTO D, SANCHO J, et al. MPEG-I depth estimation reference software [C]//2019 International Conference on 3D Immersion (IC3D). Brussels, Belgium. IEEE, 2019: 1-6.
- [40] MIELOCH D, STANKIEWICZ O, DOMAŃSKI M. Depth map estimation for free-viewpoint television and virtual navigation[J]. IEEE Access, 2020, 8: 5760-5776.
- [41] CAMPLANI M, MANTECON T, SALGADO L. Depth-color fusion strategy for 3-D scene modeling with Kinect [J]. IEEE Transactions on Cybernetics, 2013, 43(6): 1560-1571.

- [42] FOIX S, ALENYA G, TORRAS C. Lock-in time-of-flight (ToF) cameras: A survey [J]. IEEE Sensors Journal, 2011, 11(9): 1917-1926.
- [43] SCHARSTEIN D, SZELISKI R, ZABIH R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), 2001: 131-140.
- [44] HARTLEY R, ZISSERMAN A. Multiple View Geometry in Computer Vision [M]. Cambridge: Cambridge University Press, 2004.
- [45] IKEUCHI K. Determining surface orientations of specular surfaces by using the photometric stereo method [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1981, PAMI-3(6): 661-669.
- [46] ZHANG Ruo, TSAI P S, CRYER J E, et al. Shape-from-shading: a survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690-706.
- [47] MIELOCH D, GARUS P, MILOVANOVIĆ M, et al. Overview and efficiency of decoder-side depth estimation in MPEG immersive video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022.
- [48] SENOH T, TETSUTANI N, YASUDA H. Depth estimation and view synthesis for immersive media [C]//2018 International Conference on 3D Immersion (IC3D). Brussels, Belgium. IEEE, 2018: 1-8.
- [49] SENOH T, WAKUNAMI K, SASAKI H, et al. Fast depth estimation using non-iterative local optimization for super multi-view images [C]//2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Orlando, FL, USA. IEEE, 2015: 1042-1046.
- [50] NDJIKI-NYA P, KOPPEL M, DOSHKOV D, et al. Depth image-based rendering with advanced texture synthesis for 3-D video [J]. IEEE Transactions on Multimedia, 2011, 13(3): 453-465.
- [51] FEHN C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV [C]//Proc SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI, 2004, 5291: 93-104.
- [52] SUN W, XU L, AU O C, et al. An overview of free viewpoint depth-image-based rendering (DIBR) [C]//APSIPA Annual Summit and Conference, 2010: 1023-1030.
- [53] FACHADA S, BONATTO D, SCHENKEL A, et al. Free navigation in natural scenery with DIBR: RVS and VSRS in MPEG-I standardization [C]//2018 International Conference on 3D Immersion (IC3D). Brussels, Belgium. IEEE, 2018: 1-6.
- [54] GU Song, ZHANG Wenhong, WANG Ronggang. Enhanced DIBR framework for free viewpoint video [C]//2021 7th International Conference on Computer and Communications (ICCC). Chengdu, China. IEEE, 2021: 911-916.
- [55] FACHADA S, BONATTO D, SCHENKEL A, et al. Depth image based view synthesis with multiple reference views for virtual reality [C]//2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON). Helsinki, Finland. IEEE, 2018: 1-4.

作者简介



曾焕强 男,1984年生,福建惠安人。华侨大学工学院、信息科学与工程学院教授,工学院院长,博士,IEEE高级会员,中国电子学会信号处理分会委员。主要研究方向为图像处理、视频编码、计算机视觉。

E-mail: zeng0043@hqu.edu.cn



丁瑞 男,1999年生,江西吉安人。华侨大学信息科学与工程学院硕士研究生,主要研究方向为视频处理。

E-mail: 21014082004@stu.hqu.edu.cn



黄海靓 女,1997年生,浙江温州人。华侨大学信息科学与工程学院硕士研究生,主要研究方向为屏幕视频质量评价和视频处理。

E-mail: 19014082026@stu.hqu.edu.cn



陈婧 女,1980年生,福建厦门人。华侨大学信息科学与工程学院副教授,博士,主要研究方向为图像处理和视频编码。

E-mail: chenjing8005@hqu.edu.cn



朱建清 男,1987年生,福建莆田人。华侨大学工学院教授,博士,主要研究方向为模式识别和机器视觉。

E-mail: jqzhu@hqu.edu.cn