中图法分类号:TN919.8　　**文献标识码**:A　　**文章编号**:1006-8961(2023)06-1863-28

论文引用格式:Wang X, Liu Q, Peng Z J, Hou J H, Yuan H, Zhao T S, Qin Y, Wu K J, Liu W Y and Yang Y. 2023. Research progress of six degree of freedom (6DoF) video technology. Journal of Image and Graphics, 28(06):1863-1890(王旭, 刘琼, 彭宗举, 侯军辉, 元辉, 赵铁松, 秦熠, 吴科君, 刘文予, 杨铀. 2023. 6DoF视频技术研究进展. 中国图象图形学报, 28(06):1863-1890)[DOI:10.11834/jig.230025]

6DoF视频技术研究进展

王旭[1],刘琼[2],彭宗举[3],侯军辉[4],元辉[5],赵铁松[6],秦熠[7],吴科君[8],
刘文予[2],杨铀[2*]

1. 深圳大学计算机与软件学院,深圳　518060; 2. 华中科技大学电子信息与通信学院,武汉　430074;
3. 重庆理工大学电气与电子工程学院,重庆　400054; 4. 香港城市大学计算机科学系,香港;
5. 山东大学控制科学与工程学院,济南　250061; 6. 福州大学物理与信息工程学院,福州　350300;
7. 华为技术有限公司,上海　201206; 8. 南洋理工大学电气与电子工程学院信息科学与系统研究中心,新加坡　639798,　新加坡

摘　要:随着元宇宙概念的兴起,以6自由度(six degree of freedom, 6DoF)视频为代表的新一代交互式媒体技术得到产业界和学术界的广泛关注。6DoF视频隶属于多媒体通信领域,通过计算重构的方式向用户提供包括视角、光照、焦距和视场范围等多个维度的媒体交互与内容变化,能使千里之外的用户有身临其境、千人千面之感,与元宇宙具有的感知、计算、重构、协同和交互等技术特征具有高度重合性。因此,6DoF视频涵盖的技术体系可作为实现元宇宙的替代技术框架。本文提出了6DoF视频10个方面的40个问题,并将6DoF视频端到端技术链条归纳为生成、分发和呈现3个宏观阶段,随后围绕这3个技术阶段分别从内容采集与预处理、编码压缩与传输优化以及交互与呈现等方面阐述国内外研究进展。其中,在内容采集与预处理阶段,阐述了多视点联合采集、多视点与深度联合采集、深度图与点云预处理;在视频压缩与传输阶段,阐述了多视点视频编码、多视点+深度视频编码、光场图像压缩、焦栈图像压缩、点云编码压缩、6DoF视频传输优化;在交互与显示阶段,阐述了解码后滤波增强和虚拟视点合成。最后,本文围绕该领域当下的挑战,对未来趋势进行了讨论。

关键词:元宇宙;6DoF视频;内容采集;编码压缩;视点合成

Research progress of six degree of freedom (6DoF) video technology

Wang Xu[1], Liu Qiong[2], Peng Zongju[3], Hou Junhui[4], Yuan Hui[5], Zhao Tiesong[6],
Qin Yi[7], Wu Kejun[8], Liu Wenyu[2], Yang You[2*]

1. College of Computer Science and Software Engineering, Shenzhen University, Shenzhen　518060, China;
2. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan　430074, China;
3. School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing　400054, China;
4. Department of Computer Science, City University of Hong Kong, Hong Kong, China; 5. School of Control Science and Engineering, Shandong University, Jinan　250061, China; 6. College of Physics and Information Engineering, Fuzhou University, Fuzhou　350300, China; 7. Huawei Technologies Co., Ltd., Shanghai　201206, China;
8. School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore　639798, Singapore

收稿日期:2023-01-14;**修回日期**:2023-04-07;**预印本日期**:2023-04-14

*通信作者:杨铀　yangyou@hust.edu.cn

基金项目:国家自然科学基金项目(61871270,61971203);江苏省前沿引领技术基础研究重大项目(BK20222006)

Supported by: National Natural Science Foundation of China (61871270, 61971203); Major Project of Fundamental Research on Frontier Leading Technology of Jiangsu Province (BK20222006)

Abstract: The six degree of freedom based (6DoF-based) video technique is featured by interaction between video content and users, and it is focused on its 1) linear-derived multiple capacities, 2) horizontal straightness, 3) vertical straightness, 4) pitch, 5) yaw, and 6) roll motions of users. In this manner, users can change multiple audio-visual dimensions, including: viewing perspective, lighting condition or directions, focal length or spot, field of view through ground truth-compared computational or synthesized content reconstruction. The 6DoF video can be used to change conventional behavior of video watching, in which the user-video interaction is limited to different span of channels and the relations between video contents is restricted as well. The 6DoF-based technique can offer immersive experience for users because the homogeneity of video-watching receptive content can be in consistency per their motion. In this way, the 6DoF video can be recognized as an epoch-making type of video for academia and industries. At the same time, metaverse-driven 6DoF video has also been recognized as a new generation of interactive media technology, which is recognized as one of the key technologies for Metaversein related domains. All these features make users experience feel depth-immersive and diversified. This mutual-benefited status is in relevance to the metaverse-based perception, computing, reconstruction, collaboration, interaction, and other related technical features. Basically, 6DoF video is originated from the framework of typical multimedia communication system, where it can be suitable to meet the basic procedure requirement of video-contextual multimedia communication like its capturing, content process, video compression, transmission, decode and display. To realize intelligent human-terminal interaction, it brings a new look beyond traditional 3D video communication system, and the requirements for interaction range and intelligence are still acomplicated. Therefore, such newly techniques are in support of new type of video to a certain extent. Our proposed technical framework of 6DoF-relevant multimedia communication system is demonstrated on the three aspects of generation, distribution, and visualization. Forty scientific and technical challenges of this domain are illustrated and it can be categorized them into 10 different directions. We carry out literature review of its growth of per one of these 10 directions on the aspects of content acquisition and pre-processing, coding compression and transmission optimization, interaction, and presentation. For techniques analysis, it is focused on such aspects of 1) content generation-derived multiview video-captured content, 2) multiview video plus depth, and 3) point cloud. The data-acquired systems can be categorized by 2 types of multiview and multiview plus depth system, and different types of contents can be thus obtained via these systems. To describe the 3D structure of the spot scene initially, multiview color videos can be captured without any affiliated information, but it is a challenging issue for subsequent data processing techniques. After that, multiview plus depth system is proposed to handle this problem, while data can be classified into two types of i) color plus depth and ii) point cloud. Data-heteogenous volume is a big challenge for these kinds of data representation to some extent. The video compression techniques-after can be focused on in terms of the video contents. Popular compression techniques for multiview video, multiview video plus depth, light fields, and point clouds are discussed further, including their origination, mechanism, performance, and credible application standards. Subsequently, transmission techniques for 6DoF video are illustrated as well after the video bitstream is obtained. Such techniques like bit allocation, interaction oriented transmission, standards and protocols are all mentioned and discussed. Its quality evaluation and synthesized-view for user-terminal interaction are analyzed as well. It can be reached to user-friendly in terms of a "capture to display" based 6DoF video system. Pixel-based methods are still discussed and optimized but computational cost is challenged there. Recent learning based methods are more concerned about terminal-oriented applications, especially for its synthesized view. To meet the requirements from practical applications, 40 scientific and technical challenges mentioned above are still to be resolved further.

Key words: metaverse; six degree of freedom (6DoF) video; content capturing; coding compression; view synthesis

0 引 言

6自由度(six degrees of freedom,6DoF)视频具体表现为在观看视频过程中,用户站在原地时头部与视频内容之间的x、y、z 3个自由度的交互和用户位姿发生移动时与内容之间的另外x、y、z 3个自由度的交互(Boyce等,2021)。6DoF视频有多视点视频、

多视点+深度视频、光场视频、焦栈图像和点云序列等多种数据表示方式(Wien等,2019)。用户可以通过体感、视线、手势、触控和按键等交互方式来选取任意方向和位置的观看视角。视频系统在获得用户交互参数后,通过虚拟视点绘制技术完成视角平滑切换,在沉浸式体验上更加出色。6DoF视频体现了用户与视频内容的高度交互性,全面打破了人们被动接受视频内容的传统模式,能够实现千人千面的视觉体验,是当前多媒体通信、计算机视觉、人机交互和计算显示等多个学科领域的交叉与前沿。一方面,6DoF视频通过计算重构的方式向用户提供包括视角、光照、焦距和视场范围等多个视听维度的交互与变化,使千里之外的用户有身临其境之感,这与元宇宙所具有的感知、计算、重构、协同和交互等技术特征高度重合。因此,6DoF视频所涵盖的技术体系可用做实现元宇宙的替代技术框架。另一方面,6DoF视频从采集、处理、编码、传输、显示、交互和计算等方面改变了数字媒体端到端全链条的生产制作模式,给内容提供商、运营商、设备商和用户带来巨大的改变,因此也受到国防训练、数字媒体和数字教育的高度关注。

本文将围绕6DoF视频内容的生产、分发与呈现中存在的关键问题(如图1所示),从内容采集与预处理、编码压缩与传输优化以及交互与呈现等方面阐述国内外研究进展,并围绕该领域当下挑战及未来趋势开展讨论。

图1　6DoF视频系统中的关键问题

Fig. 1　Key problems in 6DoF video systems

1　6DoF内容采集与预处理

6DoF视频以3维场景为观察对象,以3维时空分布的点云、图像等为数据表达,可用模型$f(x, y, z, \theta, \varphi, \lambda, t)$刻画,包含空间$(x, y, z)$、角度$(\theta, \varphi)$、光谱$(\lambda)$和时间$(t)$等。如何获取3维场景的视觉信息是6DoF视频采集与生成需要实现的任务与目标。相机一直以来作为获取视觉信息的主要工具,将分布在3维时空(x, y, z, t)中的光降维到2维时空(x, y, t)上形成图像或视频。基于相机的视觉获取无法得到深度z,因此如何通过相机来实现3维场景的视觉信息获取,长期以来是一个挑战性的难题。从技术演进的角度,3维场景的视觉信息获取可分为多视点联

合采集、多视点与深度联合采集这两个方向和阶段。

1.1 多视点联合采集

虽然单相机的视觉获取只能得到平面图像，但是仿照人眼的双目视觉系统，只要能够利用2个及以上的相机进行多视点同步采集，就能够在得到的多视点图像基础上进行立体匹配，从而得到深度z的信息（Marr和Poggio，1976）。为此，科研人员以6DoF视频为目标，研制出了不同类型的多视点视频采集系统。如图2所示，以影视内容制作为目标，工程技术人员于1999年首次搭建了由上百台相机共同构成的多视点联合采集系统。该系统在几何排布上具有线性环绕的特点，并形成了著名的“子弹时间”影视效果（Stankiewicz等，2018）。观众可通过这种方式在屏幕上直接得到立体的观感。通过该多视点联合采集系统所形成的交互式媒体内容具有非常震撼的视觉效果，但同时也有明显的缺陷，如不能拍动态的视频、几何排布复杂不利于后期视觉计算以及成本高昂难以商业推广等。因此，降低相机数量，简化几何排布方式，研发多相机标定方法成为多视点联合采集面临的关键需求。

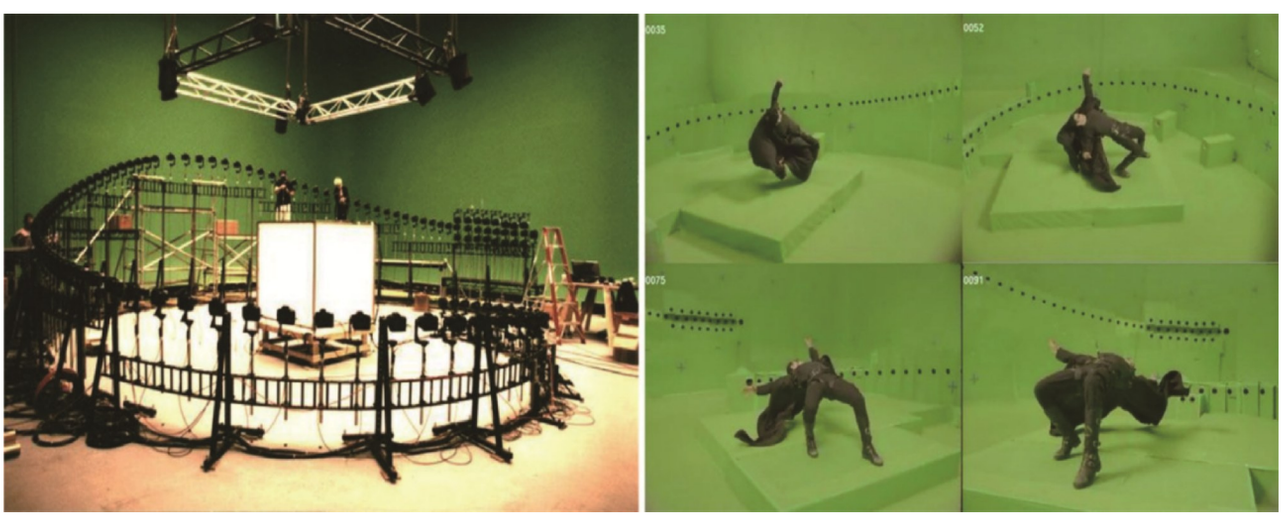
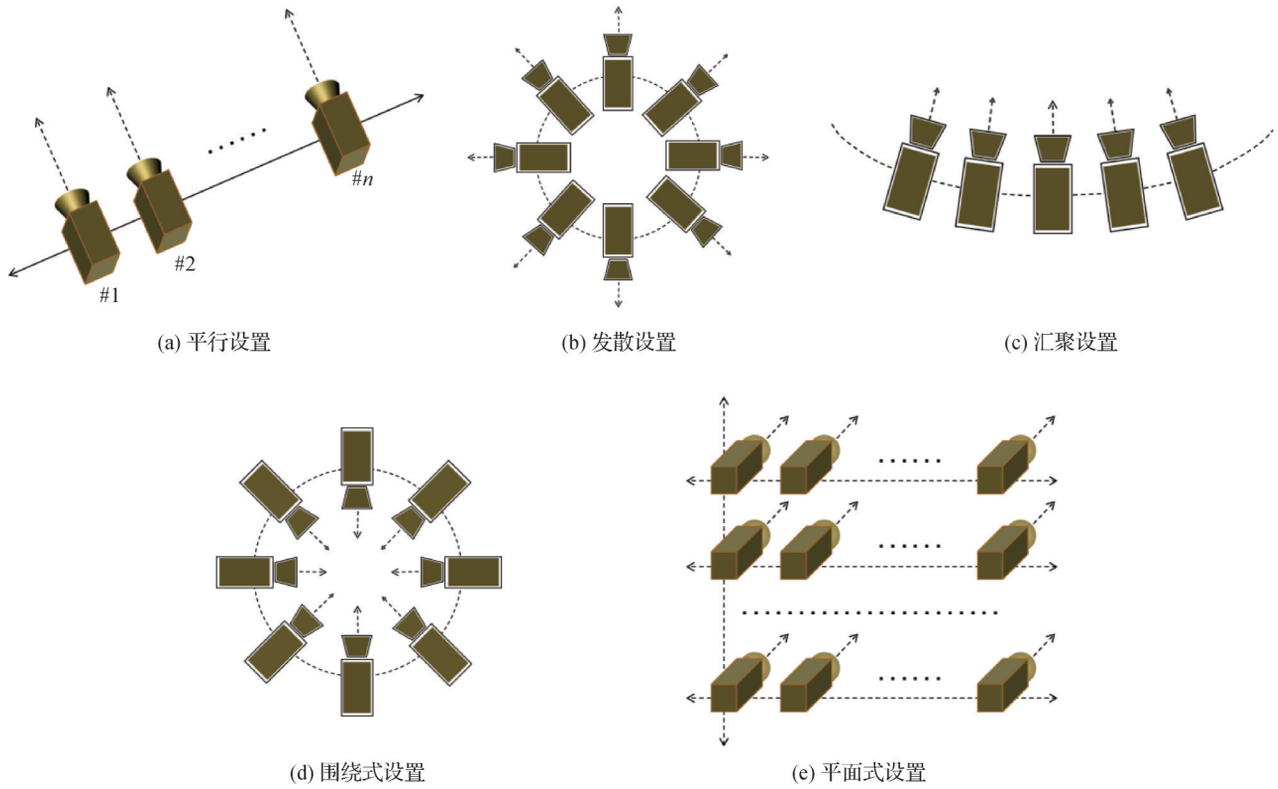


图2 影视制作中的多视点视频采集系统

Fig. 2 Multi-viewpoint video capturing system in film and television production

为了解决上述问题，研究者提出了几种典型的几何排布模式，如图3所示。图3(a)所示的平行模式以直线分布、光轴平行的方式进行排布，视点之间的图像原则上不存在垂直偏移，在交互过程中体现为水平移动。稀疏的（间距20 cm及以上）平行模式是MPEG（motion picture expert group）中典型的多视点视频数据表达形式（Merkle等，2007），而稠密的平行模式则可较为方便地构成光线空间（ray space）（Tanimoto，2012），从而实现平移之外的纵向交互。图3(b)所示的发散模式是所有相机的光轴后延线共圆心，从形式上不局限于水平共心，也可以是球面发散的共心方式。这种模式可较方便地形成全景视频用于3自由度交互，并在许多商业应用中取得了成功。图3(c)所示的汇聚模式在排布模式上是平行模式的简单变化，在直线分布的基础上将光轴汇聚到一个点上，视点之间的图像原则上不存在垂直偏移，在交互过程中体现为具有弧度的水平移动。然而，在实际操作中汇聚模式有许多问题，如汇聚点的确定、相机间的几何标定问题等，导致大部分的汇聚模式最后退化到图2的模式，即交互只在真实相机之间做切换，较少通过视觉计算的方式去绘制虚拟视点。图3(d)所示的围绕模式不局限于平面，也可以进一步拓展成半球体、圆球体的布置形式。与汇聚模式类似，同样面临着汇聚点确定、相机间几何标定的难题，而且难度更大，因为每一个相机一定会有另外一个相机与之完全相对，无法通过构建两个视点之间公共特征点的匹配关系以完成几何标定所需的有关参数。华中科技大学团队突破了这一限制，通过视点传递的方式克服了环绕相机阵列（Abedi等，2018）以及球面相机阵列（An等，2020）的几何标定问题，为后续720°交互奠定了基础。图3(e)所示的平面模式在几何分布上是平行模式的简单扩充，但是在实际应用中产生了许多变型，并逐步演化成光场采集系统，催生了许多交互式媒体之外



(a) 平行设置　(b) 发散设置　(c) 汇聚设置

(d) 围绕式设置　(e) 平面式设置

图 3　几种典型的多视点视频采集系统的几何排布方式

Fig. 3　Geometric arrangement of typical multi-viewpoint video capturing systems((a)parallel setting;(b)diverging setting;(c)converging setting;(d)surrounding setting;(e)planar setting)

的新型应用(Levoy 和 Hanrahan,1996)和亿像素采集系统(Brady 等,2012)。

1.2　多视点与深度联合采集

典型的多视点联合采集需通过后期计算的方式得到深度,如果能够直接得到深度信息,则可以大幅提升采集效率。然而,直接获得场景的深度信息并不是一件容易的事情,进而在获取深度信息的基础之上是否能够多视点获取,又是另外一个难题。

直接获取场景深度信息的方式大体分为被动式和主动式两个技术方向。被动式探测以双目立体匹配为代表(Zhang,2012)。主动式探测方法以结构光技术为代表,并根据光源的不同又分为点扫描(Franca 等,2005)、线扫描(Scharstein 和 Szeliski,2002)和面结构光(Van der Jeught 和 Dirckx,2016)。点扫描和面扫描中激光器发出点状或条状光束,进而通过旋转或平移,实现完整的 3 维测量。面结构光方法投射 2 维编码图案,无需移动投影设备即可重建目标表面,具有更高的效率(苏显渝 等,2014)。此外,面结构光中投影图案通常与编码技术进行结合,提取块级/像素级/亚像素级的码字用于视差匹配,以获得更高的精度和效率。面结构光的编码通常包括空域编码、时域编码和相位编码,通过多个编码对场景进行多次扫描来获得目标场景的深度。上述模式都是通过扫描的方式才能得到场景的深度信息,因此不适宜动态场景的深度获取。采用点—面结合技术的 Kinect 深度传感器克服了这个难题(Lilienblum 和 Al-Hamadi,2015),虽然深度图的质量、图像分辨率、时间分辨率和探测距离等基本参数还有很大的提升空间,但是该设备的出现首次将场景的深度感知从静态提升至了动态,给产业界和学术界同时带来一轮新的研究热潮。后来出现了基于光调制的 ToF(time of flight)技术及相关设备,包括 ToF 相机和激光雷达(laser radar,LiDAR)等,大幅度提升了探测距离,但是在深度图质量、图像分辨率和时间分辨率等参数上也都与 Kinect 一样面临相同的问题。

将多个深度传感器与多个彩色相机相互配合对场景进行视觉采集,则形成了多视点与深度联合采集方案。在这些方案中,几何排布上可以借鉴多视点联合采集方案。多视点与深度联合采集的关键难点在于多深度采集中所出现的视点间干扰、彩色视

频与深度视频时间分辨率不匹配以及空间分辨率差距过大等问题。多深度相机之间的干扰来自其成像原理本身，如不同视角的Kinect会使用相似甚至相同的点—面结构光，不同视角的ToF相机对同一波长的光进行相同的调制，这些都会导致解码失败。为了解决这个问题，华中科技大学团队从机理层面进行了探索，针对多种原理的深度传感器分别设计了包括M-序列等方法在内的多深度相机联合采集方案，较好地解决了上述难题（Yan等，2014；Li等，2015；Xiang等，2015）。此外，还进一步针对深度视频与彩色视频时间分辨率不匹配的问题，以及由此导致的深度图运动模糊问题，提出了时域上采样法（Yang等，2012）和时域校正法（Yang等，2015c；Gao等，2015）等多种方法，为运动场景的立体感知提供了丰富的工具集。

1.3 深度图与点云预处理

如前所述，动态场景的深度图或点云数据往往具有空间分辨率低、时间分辨率低、画面噪声多等问题。为了保证下游任务的精度，需要进行预处理。从处理技术上来分，主要包括深度图预处理和点云数据预处理两个类型。

1.3.1 深度图预处理

深度信息不直接用于人眼观测，而是作为辅助信息帮助参考视点图像映射到正确的虚拟视点上。深度图像上的失真会传播至虚拟视点图像，造成主客观质量的下降。因此，在虚拟视点内容生成前，需通过深度预处理技术尽可能获得最接近场景实际距离的深度图像。Ibrahim等人（2020a）较详细地对深度图预处理技术工作进行了系统性的梳理。总体而言，深度图、点云的去噪与图像去噪技术是同步发展的，但同时也有自身的一些特点。典型的图像滤波器，如多边滤波器（Choudhury和Tumblin，2005）、流型滤波器（Gastal和Oliveiray，2012）和非区域均值（Buades等，2005）等都可以直接作用于深度图的去噪，但这些滤波器都只能解决以像素为单位的深度图噪声。一旦噪声区域过大，如Kinect深度传感器的噪声多以成片区域深度值缺失为特点，则传统的滤波器都会失效（Xie等，2015）。为了解决这个问题，Kopf等人（2007）提出了联合双边滤波方法。该方法是对双边滤波的改进，引入了参考图像为指导，能够较好地处理大面积深度值缺失的难题，但同时也引入了彩色图中的边缘和纹理信息，给去噪后的深度图带来了伪纹理。Liu等人（2017）利用对齐彩色图像特征来引导深度图像修复，通过彩色信息引导权重并结合双边插值方法来进行深度图空洞修复。Wang等人（2015）提出一种面向Kinect深度图像恢复的三边约束稀疏表示方法，在惩罚项上考虑了参考块与目标块间的强度相似度和空间距离的约束，在数据保真度项下考虑了目标块质心像素的位置约束，通过对纹理图像的特征学习，预测出深度图像空洞恢复的最优解。为了有效克服伪纹理的问题，Ibrahim等人（2020b）引入条件随机场方法以抑制在彩色图引导过程中的纹理干扰问题。随着深度学习技术的发展，人们也开始探索单一深度图（张洪彬 等，2016）、彩色与深度图联合（Zhu等，2017）的滤波方案，总体上遵循了彩色图滤波的基本架构，包括特征提取、图像重建等模块。基于深度学习框架的深度图滤波虽然能够取得较好的去噪效果，但是目前仍面临物体边缘滤波模糊的难题。

多视点联合滤波也是一个值得关注的课题。如果将每一个视点的深度图单独处理，势必会导致视点间深度不稳定的问题，为此需要将多个视点联合在一起考虑。华中科技大学团队He等人（2020b）提出了跨视点跨模态的联合滤波框架，建立了视点之间的映射模型与关联方式，能够较好地克服多种类型的噪声在不同视点间的蔓延。针对平面相机阵列，Mieloch等人（2021）考虑到纹理信息的使用会在深度修正中引入误差，仅用多个视点的深度信息对所选视点的信息进行交叉验证，通过多次迭代，增强了多个深度图像的视点间一致性，且可以自由设置需要修正的视点位置和数目。

1.3.2 点云预处理

深度相机和激光雷达传感器产生的原始点云通常是稀疏、不均匀和充满噪声的，需要进行去噪或补全。现有的点云补全的方法大致分为基于几何或对齐的方法和基于表示学习的方法两类。

基于几何或对齐的方法包括基于几何的方法和基于对齐的方法。基于几何的方法通过先前的几何假设，直接从观察到的形状部分预测不可见的形状部分（Hu等，2019）。更具体地，一些方法通过生成平滑插值来局部填充表面孔。例如拉普拉斯平滑（Nealen等，2006）和泊松表面重建（Kazhdan和Hoppe，2013），这些方法直接从观察区域推断缺失数据并显示出令人印象深刻的结果，但是需要为特

定类型的模型预定义几何规则，并且仅适用于不完整程度较小的模型。基于对齐的方法在形状数据库中检索与目标对象相似的相同模型，然后将输入与模型对齐，随后对缺失区域进行补全。目标对象包括整个模型（Pauly 等，2005）或其中的一部分（Kim 等，2013）。除此以外，还有一些方法使用变形后的合成模型（Rock 等，2015）或非 3D 几何图元，例如平面（Yin 等，2014）和二次曲面（Chauve 等，2010）代替数据库中的 3D 形状。这些方法在 3D 模型的类型上具有较强的泛化性，但在推理优化和数据库构建过程中成本高，且对噪声敏感。

基于表示学习的方法是一种点云补全的方法。Dai 等人（2017）提出了基于 3D 体素的编码器—解码器架构 3D-EPN（3D-encoder-predictor）。尽管基于 3D 体素化的表示学习方法可以直接扩展使用定义在 2D 规则网格上的神经层或算子，但精细对象的重建需要消耗大量显存和算力。随着基于点表示学习的 PointNet（Qi 等，2017a）和 PointNet++（Qi 等，2017b）等模型的出现，人们提出了 TopNet（Tchapmi 等，2019）、PCN（point cloud net）（Yuan 等，2018）和 SA-Net（shuffle attention net）（Wen 等，2020a）等基于点编码器—解码器框架的点云修复模型。该类模型首先通过编码器从不完整的点云中提取全局特征，再利用解码器根据提取的特征推断完整的点云。

现有基于表示学习的点云补全任务的相关研究主要分为两类。1）基于先进的深度学习框架。为了提高点云生成的完整形状的真实性和一致性，人们提出了基于对抗生成网络的 RL-GAN-Net（reinforcement learning generative adversarial network）（Sarmad 等，2019）、基于变分自动编码器的 VRCNet（variational relational point completion network）（Pan 等，2021）和基于注意力机制的 PoinTr（Yu 等，2021）、SnowflakeNet（Xiang 等，2021）、PCTMA-Net（point cloud transformer with morphing atlas-based point generation network）（Lin 等，2021）、MSTr（Liu 等，2022）等模型，这些模型能更好地挖掘 3D 形状的全局和局部几何结构，从而更有利于补全点云中的不完整部分。2）基于任务特性的算子。为了保留更多的精细特征信息，SoftPool++（Wang 等，2022a）设计了 softpool 算子替代 PointNet 中的最大池化算子。Wu 等人（2021）提出基于密度感知的倒角距离，以改善原有损失函数对点云局部密度不敏感或精细结构保护不足等缺陷。

考虑实际应用需求，渐进式点云补全任务也开始得到关注，人们提出了 CRN（cascaded refinement network）（Wang 等，2022b）、PF-Net（point fractal network）（Huang 等，2020b）、PMP-Net++（point cloud completion by transformer-enhanced multi-step point moving paths）（Wen 等，2023）等模型，以实现 3D 点云的渐进细化。总体而言，基于学习的点云补全方法在性能提升上效果显著，但在模型泛化上仍有很大的提升空间。如何结合几何先验以提升模型的泛化性是一个潜在的研究方向。

2 6DoF视频压缩与传输

6DoF 视频有多视点视频、多视点+深度视频、光场图像、焦栈图像和点云序列等多种数据表示方式，本节根据各种数据表示方式的特点，对 6DoF 视频压缩与传输的研究进展展开介绍。

2.1 多视点视频编码

自从 1988 年 CCITT（Consultative Committee International for Telegraph and Telephone）制定了视频编码标准 H.261 后，视频编码技术的应用越来越广泛，并涌现出大量的视频编码标准，包括 H.264/AVC（Wiegand 等，2003）、H.265/HEVC（high efficiency video coding）（Ohm 等，2012）和 H.266/VVC（versatile video coding）（Bross 等，2021）。最简单的多视点视频编码 MVC（multi-view video coding）方案是独立地对各个视点进行编码，但是这样不能充分去除视点间冗余，于是产生了时域—视点域结合的编码压缩方案研究。

1）多视点视频扩展国际编码标准。MPEG-2 标准中已采用了多视点视频配置来编码立体或者多视点视频信号。由于压缩标准的局限性、显示技术和硬件处理能力的限制，MPEG-2 的多视点扩展没有得到实际应用。2005 年，MPEG 组织在 H.264/AVC 的基础上提出了 MVC 扩展标准（Vetro 等，2011），并形成了联合多媒体模型（joint multiview model，JMVM）。该模型集成了视点间亮度补偿、自适应参考帧滤波、MotionSkip 模式以及视点合成预测等基于宏块的编码工具。类似于 H.264/AVC 的 MVC，JCT-3V 在 H.265/HEVC 的基础上提出了扩展编码标准 MV-HEVC（multi-view HEVC）（Tech 等，2016）。

我国从1996年开始参加MPEG专家组的工作，不断有提案被接受，在视频压缩的技术成果逐渐具备了国际竞争力。2002年6月，我国成立了数字音视频编解码技术标准工作组AVS（audio-video standard），目标是制定一个拥有自主知识产权的音视频编码标准。至今，其版本已经发展到AVS3。基于国际编码标准，国内学者在MVC快速算法、率失真控制和基于深度学习的多视点编码等方面进行了深入研究，取得了极大的进展。

除了高效的压缩编码标准之外，精心设计的预测编码结构能充分利用多视点视频信号中的时空相关性和视点间的相关性。目前，MVC中广泛采用的分层B帧编码结构（hierarchical B pictures，HBP）结合运动估计和视差估计，获得了较高的压缩效率和优秀的率失真性能。

2）面向编码的多视点视频预处理。利用多视点视频扩展编码标准压缩多视点视频信号时，能在编码标准框架下同时消除时空冗余和视点间冗余。然而，多视点视频信号往往存在几何偏差和颜色偏差，影响了编码压缩效率。因此，多视点视频信号的预处理也能提升压缩性能。Doutre和Nasiopoulos（2009）对多视点视频信号进行颜色校正，提升了视点之间颜色一致性和MVC的视点间预测性能。Fezza等人（2014）提出了基于视点间对应区域直方图匹配方法的多视点颜色校正算法，以提升压缩性能。福州大学团队Niu等人（2020）针对多视点视频信号中存在的全局、局部和时间颜色差异，提出了由粗到细的多阶段颜色校正算法。

3）多视点视频快速编码。由于各种编码标准集成了多种复杂技术，且多视点视频巨大的数据量也会带来巨大的时间开销。因此，多视点彩色视频编码的计算复杂度问题长期以来都是难题。针对各种编码标准和多视点扩展编码标准，学者们广泛地开展了快速编码算法研究。典型的手段包括减少搜索点数（Cernigliaro等，2009）、利用MVC的编码模式的时空相关性和视点相关性减少当前编码宏块的搜索数量（Zeng等，2011）以及基于像素级与图像组级的并行搜索算法（Jiang和Nooshabadi，2016）等。

国内学者也提出了若干快速编码算法。Li等人（2008）通过减小搜索范围和参考帧数目来提高MVC速度。在MVC快速宏块模式选择方面，Shen等人（2010）利用相邻视点的宏块模式辅助当前视点的宏块模式选择，提高编码速度。Ding等人（2008）通过共享视点间编码信息（例如率失真代价、编码模式和运动矢量）来降低MVC的运动估计的计算复杂度。MVC中，大量宏块的最优模式为DIRECT/SKIP模式。根据此特性，Zhang等人（2013b）提出了Direct模式的提前判断方法，从而避免所有宏块模式的搜索过程。Yeh等人（2014）利用已编码视点的最大和最小率失真代价形成阈值条件，用于提前终止当前编码视点的每个宏块编码模式选择过程。Pan等人（2015）提出了一种Direct模式的快速模式决策算法，并利用MVC特性，设计了运动和视差估计的提前终止算法。Li等人（2016b）利用宏块模式的一致性和率失真代价的相关性，提出了Direct模式的判定方法。

4）MVC的码率控制。码率控制旨在提高网络带宽利用率和视频重建质量。与单视点视频编码的码率控制不同，MVC的码率控制需要考虑视点级的码率分配。Vizzotto等人（2013）在帧级和宏块级实现了一种分层MVC比特控制方法，该方法充分利用了当前帧和以编码相邻帧比特分布的相关性。Yuan等人（2015）提出了视点间编码依赖关系模型，认为视点间的依赖关系主要由编码器的跳跃（SKIP）模式导致，并据此提出了理论上最优的多视点视频码率分配与控制算法。

5）基于深度学习的MVC。Lei等人（2022）提出了基于视差感知参考帧生成网络（disparity-aware reference frame generation network，DAG-Net）生成深度虚拟参考帧。该网络包含多级感受野模块、视差感知对齐模块和融合重建模块，能转换不同视点之间的视差关系，生成更可靠的参考帧。这些参考帧插入到3D-HEVC的参考帧列表中，能提升MVC的编码效率。Peng等人（2022）提出了基于多域相关学习和划分约束网络的深度环路滤波方法。其中，多域相关学习模块充分利用多视点的时间和视点相关性来恢复失真视频的高频信息，分割约束重建模块通过设计分割损失减少压缩伪影。

2.2 多视点+深度视频编码

多视点彩色加深度（multiview video plus depth，MVD）是一种典型的场景表示方式，MVD信号包括多视点视频信号和对应的深度视频信号。多视点视频信号是利用相机阵列对在同一场景从不同位置采集得到，而对应深度视频可采用深度相机获取或者

利用软件估计得到。与传统的视频信号相比，MVD的数据量随着相机数目的增加而成倍增加。

1）多视点+深度视频国际编码标准。为了编码MVD信号，JCT-3V基于HEVC提出了3D-HEVC的扩展编码标准（Tech等，2016），该标准能充分利用深度视频的特性和视点之间的相关性，提升MVD信号的编码性能。针对沉浸式视频的最新编码压缩标准为ISO/IEC MIV（MPEG immersive video），该标准定义了比特流格式和解码过程。沉浸式视频参考软件TMIV（test model for immersive video）包括编码器、解码器和渲染器等，并提供了测试用例、测试条件、质量评估方法和实验性能结果等。在TMIV中，多个纹理和几何视图使用传统的2D视频编解码器编码为补丁的图集，同时优化比特率、像素率和质量。

2）多视点+深度视频快速编码。在基于H. 265/HEVC及多视点视频扩展标准方面，学者们提出了基于MV-HEVC和3D-HEVC标准的多视点深度视频快速编码算法（张洪彬 等，2016）。由于深度视频编码深度视频信息反映3D场景的几何信息，最简单的方法是对深度视频下采样，降低编码复杂度和降低码率，代价为丢失场景信息，导致绘制失真。Tohidypour等人（2016）利用已编码信息，结合在线学习的方法，调节3D-HEVC编码中非基础视点彩色视频的运动搜索范围和降低模式搜索的复杂度。Chung等人（2016）提出了新的帧内/帧间预测和快速四叉树划分方案，既提高了3D-HEVC的深度视频的压缩率，又提高了压缩速度。Zhang等人（2018）针对3D-HEVC中深度视频编码模式引入的额外编码复杂度，提出了两种深度视频的帧内模式决策方法。Xu等人（2021）基于MV-HEVC编码平台，提出了复杂度分配和调节，实现了MVC的编码复杂度优化，已适应于不同的视频应用系统。在多视点深度视频方面，Lei等人（2015）利用MVD视频信号中的视点相关性、彩色和深度视频的相关性，提出了多视点深度视频快速编码算法。Peng等人（2016）和黄超等人（2018）基于3D-HEVC提出了联合预处理和快速编码系列算法，增强了MVD信号中深度视频的时间不一致性，提高了压缩效率和编码速度。

3）多视点+深度视频编码码率控制。与MVC的码率控制仅需要考虑视点级的码率分配不同，MVD编码进一步需要考虑彩色与深度视频之间的码率分配。Yuan等人（2011，2014）最早确定了虚拟视点失真和多视点纹理和深度视频的编码失真之间的解析关系，进而将多视点+深度视频编码码率控制问题建模为拉格朗日优化问题，并求得理论上的最优解。Chung等人（2014）提出一种基于新型视点综合失真模型的比特分配算法，在纹理和深度数据之间优化分配有限的比特预算，以最大化合成的虚拟视图和编码的真实视图的质量。Klimaszewski等人（2014）提出一种新的多视点深度视频压缩质量控制方法，建立了深度和纹理量化参数计算的数学模型。De Abreu等人（2015）提出一种在相关约束条件下有效选择预测结构及其相关纹理和深度量化参数的算法，具有较优的压缩效率和较低的计算复杂度，为交互式媒体应用提供了一种有效的编码解决方案。Fiengo等人（2016）利用最新的对凸优化工具，提出了帧级比特最优速率分配的算法，其码率控制性能超越标准MV-HEVC。Domański等人（2021）提出一种可用于比特率控制的视频编码器模型，该模型适用于MVD编码，从AVC的模型中，可以快速推导出HEVC和VVC的模型。Paul（2018）提出一种基于3维帧参考结构来提高交互和降低计算时间，增加一个参考帧来提高遮挡区域的率失真性能，采用视觉注意的比特分配以提供更好的视频感知质量。Liu等人（2011）提出一种MVD的视点、彩色/深度级和帧级的联合码率控制算法，利用预编码及数理统计分析方法实现视点级、彩色/深度级的比特分配。Zhang等人（2013a）提出了基于视点合成失真模型的多视点深度视频编码的区域位分配和率失真优化算法，测试序列的编码效率得到显著提高。Li等人（2021b）提出了一种基于视图间依赖性和时空相关性新的多视图纹理视频编码位分配方法，建立了一个基于视图间依赖关系的联合多视图率失真模型。该方法在率失真性能方面优于其他最先进的算法。

4）基于深度学习的深度视频编码。相比于彩色视频，深度视频具有更加平滑的内容和更大的空域冗余，可以以更小的分辨率进行编码，以提高编码效率。针对深度视频编码，Li等人（2022）提出了基于深度上采样的多分辨率预测框架，该框架对于不同复杂度的深度块，使用最优的分辨率进行编码，以提高深度视频编码效率。

2.3 光场图像压缩

光场图像压缩的目的在于去除子视点图像内部冗余以及子视点图像间冗余。传统2D图像编码中

成熟的帧内压缩技术可以直接应用于光场图像压缩去除子视点图像内部冗余。因此，光场图像压缩的相关研究主要致力于去除视点间冗余（Liu等，2019）。光场图像的各子视点图像由于视差变化具有不规则变化的显著特点，根据建模方法，现有的光场图像压缩研究大体可分为基于伪视频序列的方法、基于优化的方法和基于视点重建的方法3类。

1）基于伪视频序列的方法。光场图像压缩的关键在于如何充分利用子视点图像间的相关性。相邻的子视点图像之间存在着极大比例的重复场景，且由于视差引起的场景变化平缓，与传统视频中前后帧中的场景变化较为相似。自然而然地，早期的光场图像压缩引入了传统2D视频编码的框架，将光场图像中的子视点图按照一定的扫描顺序重组为伪视频序列，将视点间冗余转化为伪视频序列的帧间冗余，直接利用成熟的视频压缩标准中的帧间预测技术去除视点间冗余。因此，此类研究方案的重点在于如何构建合理的子视点排列顺序以及预测结构，从而在伪视频序列的帧间编码过程中尽量减少编码视点与参考视点间的残差信息，增加压缩效率。针对扫描顺序，国内一些早期的工作（Dai等，2015）中提出了横向、纵向、之字形和环形的扫描方案，且均取得了一定的性能提升。而在此类工作中，影响力较大的是由中国科学技术大学Li等人（2017）提出的2维层级编码框架。在此框架中，首先将所有视点图划分为4个象限，再在每个象限中按固定位置划分为4个编码层次，沿用传统视频编码中多层次编码的框架，即在编码过程中首先使用高保真编码方案压缩低层次视点图，并且在高层次视点图压缩时作为参考视点。此外，在选取参考视点图的过程中，通过衡量与不同参考视点间的距离确定最佳的参考视点，进一步提升压缩效率。此工作为较早提出的完整的光场编码框架，经常被后续研究引用作为评价标准。此外，Liu等人（2016）将传统视频编码中的可伸缩编码思想应用到光场压缩中，提出了一个包括3层分辨率和质量可伸缩的光场编码框架。

基于伪视频序列的压缩方法致力于将视点间的相关性转换为时域相关性，从而得以利用视频编码技术中的帧间预测技术去除伪视频序列的时域冗余。然而，传统视频编码的帧间预测技术中，只考虑了前后帧场景间的平移运动，用表征上下、左右位移的2维的运动向量表示。而光场图像中各个子视点图像场景间更多的是由于视角变化引起的不规则运动，这与传统视频存在本质上的差异。所以，由于缺少针对光场图像特性的适应性优化，基于伪视频序列的光场编码方案难以取得最优的压缩性能。

2）基于优化的方法。在基于伪视频序列压缩方案的基础上，一部分研究者致力于研究子视点间场景不规则运动的模型，优化原有光场编码框架中的部分模块，以期进一步提升编码效率。这些研究包括基于单应性变化矩阵、图变换等优化方案。Chang等人（2006）针对视点间物体的不规则变化，首先利用传统的图像分割方法获取物体形状，继而提出了一种视差补偿算法来估计相邻子视点图中该物体的形状变化，据此提升预测效率。此外，此工作也在光场编码基础框架上提出了改进方案，即使用聚类算法对子视点图像进行排序，根据聚类结果调整伪视频帧的排序。Jiang等人（2017）提出了基于单应性变化矩阵的光场图像编码框架优化方案。具体的，该方法利用单应性或者多应性变化矩阵将所有子视点图统一映射到一个或者多个深度面上，继而在此基础上求取光场图像的低秩表示。最后，通过单应性矩阵参数与低秩矩阵的联合优化，以实现光场低秩表示数据的压缩。Dib等人（2020）基于超射线表示的视差模型提出了一个局部低秩逼近方法。超射线由与所有子视点图像都相关的超像素点构建，通过施加形状与大小的约束，使得超射线得以表达复杂的场景变换，继而通过参数化的视差模型描述每条超射线表示帧内的视差局部变化。此模型的最佳参数将通过交替搜索估计的方法确定。

由于图信号也能较好地描述图像中物体的不规则运动，部分研究者进而将图变换应用于光场压缩的视点间预测模块。基于图变换的优化框架最早由Su等人（2017）提出，该方法依据深度信息将所有像素分类并构建图表示，并在此基础上对子视点图间场景的不规则变化进行预测。然而此方案依赖于深度信息，并且基于图变换的运动预测大幅增加了整体模型的复杂度。针对于此，Rizkallah等人（2021）提出了一个局部图变换的方法，通过图规约技术以及谱聚类来减少图的维度，从而控制算法的整体复杂度，并提出了不同规约方案下重建子视点图的率失真准则模型，以实现在特定复杂度限制下寻找最优图构建的目的。

在光场图像压缩乃至传统视频压缩领域中，如

何描述邻近视点或帧间场景间的不规则运动是一个长久以来悬而未决的难题。类似于图变化或者单应性变化矩阵等基于人工设计函数的优化方案受限于其预测的准确率，对整体编码性能提升较为有限，且极大地增加了整体编码框架的复杂度，给实际应用带来了挑战。

3）基于视角重建的方法。相比于传统使用手工设计函数描述复杂运动的优化方案，直接使用智能图像生成技术以重建邻近视点图的方案更为简洁、高效。深度神经网络中的先验知识显著减少了重建光场图像所需要传递的信息，大幅提升了光场图像压缩框架的效率，因此成为当前光场压缩研究的重要方向。

该类方法首先在所有待压缩的子视角图中选取数幅作为关键视角（Chen 等，2018），压缩并传送至解码端。然后，在编码非关键子视角图时，将重建后的关键视角图作为输入，利用图像生成网络合成非关键视角图。最后，合成的非关键视角图与原图之间的残差将被压缩并传送至解码端。如香港城市大学 Hou 等人（2019）使用基于深度学习的角度超分辨率模型用于预测非关键视角图。北京大学 Jia 等人（2019）使用对抗生成模型来学习子视角图像结构中的角度以及空间变化，从而得以实现更准确的预测非关键视角帧的预测。针对低码率条件下的光场压缩，Ahmad 等人（2020）提出了基于剪切小波变换的非关键视角预测方法。Bakir 等人（2021）提出了一种自适应的非关键视点丢弃的策略，并在解码端对生成的非关键帧进行图像增强后处理，以进一步提升整体压缩效率。

2.4 焦栈图像压缩

焦栈图像是光场图像的降维，其压缩是一个全新的课题。相比于传统 2D 图像的固定视点、固定对焦的采样模式，焦栈图像需要在某一时刻对不同深度的场景进行稠密采集，以获取完整的场景图像数据。焦栈图像序列与普通视频具有不同的成像特性和冗余模型，普通视频帧之间的冗余模型通过运动矢量来刻画，而焦栈图像序列则通过焦深来刻画，因此现有编码框架不适用于焦栈图像压缩的目标。

焦栈图像编码方法可分为两类，即基于静态图像的编码和基于视频的编码。在基于静态图像的编码方法中，Sakamoto 等人（2012a）将焦栈图像序列划分为尺寸为 8 的 3D 像素块，然后对每个 3D 像素块进行 3D-DCT（3D discrete cosine transform）变换和线性量化，并按照频率从低到高的顺序排列为 1D（one dimension）信号，最后利用霍夫曼编码方法将信号写入码流完成编码。为了抑制图像退化噪声，Sakamoto 等人（2012b）进一步利用 3D 离散小波变换对焦栈图像进行处理，相比于基于 3D 离散余弦变换的方法，有效抑制了编码产生的块效应失真。Khire 等人（2012）提出的方法采用差分脉冲编码调制和相邻图像的信息来估计冗余度，获得了比 JPEG 和 JPEG2000 更高的压缩效率。

基于视频的编码方法考虑了序列各帧之间的相关性，通过运动搜索进行帧间预测，相比于基于静态图像的编码方法可获得更高的压缩性能。如 van Duong 等人（2019）面向光场重聚焦应用，将焦栈图像排列为视频序列，直接使用 HEVC 编码器进行压缩。然而，这显然不能挖掘图像间的焦深冗余。为此，Wu 等人（2020b，2022）分别提出了基于高斯 1D 维纳滤波的块模式单向/双向焦深预测，以及分层焦深预测的方法，较早地开展了焦深冗余模型的构建。该类型相比于直接利用视频编码的方案，压缩性能上有了极大提升。然而，需要强调的是，焦栈图像压缩的研究刚刚起步，尚有许多未知的问题需要探索和研究。

2.5 点云编码压缩

3D 点云是具有法线、颜色和强度等属性的无序 3D 点集。大规模 3D 点云数据的高效编码压缩技术具有广泛的市场应用前景。现有研究主要可分为传统压缩方法和智能压缩方法两类。

1）传统压缩方法。为了实现点云数据的高效压缩，工业界和学术界提出了多种解决方案（Mekuria 等，2017）。点云压缩方法是通过八叉树等表示方法将点云进行预处理，主要思路有 3 种。第 1 种是通过映射，将 3 维点云转换成 2 维图像后，采用传统的图像或者视频编码工具进行编码操作；第 2 种是首先直接将数据矢量线性变换为合适的连续值表示，独立地量化其元素，然后再使用多种无损的熵编码对得到的离散表示进行熵编码操作；第 3 种是将八叉树空间索引信息直接进行编码。根据组织机构不同，主要可分为运动图像专家组（MPEG）提出的点云压缩（point cloud compression，PCC）标准、音视频标准组（audio video coding standards workgroup，AVS）提出的点云压缩参考模型（point cloud refer-

ence model, PCRM)和谷歌公司研发的“Draco”编码软件3类。

2017年MPEG启动了关于点云压缩的技术征集提案,此后一直在评估和提升点云压缩技术的性能。根据点云压缩的不同应用场景,MPEG划分了3类点云数据,并针对3类点云开发了3种不同的编码模型,分别是用于自动驾驶的动态获取点云的模型(LiDAR point cloud compression, L-PCC)、针对用于表示静止对象和固定场景的静态点云模型(surface point cloud compression, S-PCC)和针对用于沉浸式多媒体通信的动态点云的模型(video-based point cloud compression, V-PCC)。其中,动态获取点云指点云获取设备一直处于运动状态,获取的点云场景也处在实时变化之中;静态点云指被扫描物体与点云获取设备均处于静止状态;动态点云指被扫描物体是运动的,但是点云获取设备处于静止状态。由于L-PCC和S-PCC的编码框架相似,2018年1月MPEG对现有的L-PCC和S-PCC进行整合,推出了全新的测试模型(geometry-based point cloud compression, G-PCC)。2022年MPEG公布了第1代点云压缩国际标准V-PCC (ISO/IEC 23090-5)和G-PCC (ISO/IEC 23090-9)(Schwarz等,2019)。其中,V-PCC适用于点分布相对均匀且稠密的点云,G-PCC适用于点分布相对稀疏的点云。G-PCC的几何信息编码部分主要是通过坐标变换和体素化(Schnabel和Klein, 2006)的方法进行位置量化与重复点移除,然后通过八叉树构建将3维空间划分为层次化结构,将每个点编码为它所属的子结构的索引,最后通过熵编码生成几何比特流信息。属性信息部分则是通过预测变换、提升变换(Liu等,2020)和区域自适应分层变换(region-adaptive hierarchical transform, RAHT)(de Queiroz和Chou, 2016)等进行冗余消除。V-PCC则通过将输入点云分解为块集合,这些块可以通过简单的正交投影独立地映射到常规的2D网格,再通过诸如HEVC和VVC等传统2维视频编码器来处理纹理信息及附加元数据。

为了保障我国数字媒体相关产业的安全发展,AVS也成立了点云工作组,并在2019年12月发布了国内第1个点云压缩编码参考模型PCRM(point cloud reference model)。PCRM的核心编码思想与G-PCC类似,同样是依据点云的几何结构直接编码。PCRM的几何编码主要是通过多叉树结构对点云划分,利用节点之间的关系和占位信息对点云编码。PCRM的属性编码有两种方案,一种是直接预测编码;另一种是基于变换的编码,即对点云的属性信息进行离散余弦变换。

Draco架构是谷歌媒体团队提出的开源3D数据压缩解决方案,使用k-维树等多种空间数据索引方法对属性和几何信息进行量化、预测压缩以及熵编码以达到高效压缩目的。

2)智能压缩方法。随着深度学习的发展及其在数据编码领域的应用,研究人员提出了基于深度学习的端到端点云编码方法。2021年MPEG也开展了基于深度学习的点云编码(artificial intelligence-point cloud compression, AI-PCC)技术探索,并提出标准测试流程。基于深度学习的端到端点云编码方法主要涉及基于体素表示、基于点表示和深度熵模型3种方式。

基于体素表示的方法是将点云转换为体素化的网格表示,再对体素进行编码与压缩。Quach等人(2019, 2020)和Wang等人(2021b)受基于学习的图像压缩方法的启发,使用基于3D卷积的自编码器,在体素上提取潜在表示作为点云的几何编码并在体素上执行二分类任务以重建点云几何信息。由于点云的稀疏性,点云占据的体素只占全部空间的小部分,体素网格中的大部分空间保持空白,导致存储和计算的浪费。为了克服这一缺陷,南京大学Wang等人(2021a)利用稀疏体素代替稠密体素,并通过Minkowski稀疏卷积来降低内存要求以提升编码性能。

基于点表示的方法直接使用神经网络处理点云,而不需要额外的体素化。浙江大学Huang等人(2019)直接使用自编码器用于点云几何压缩。深圳大学Wen等人(2020b)提出了一种用于大规模点云的自适应八叉树划分模块,并使用动态图卷积神经网络作为点云自编码器的核心骨干网络。为了获得更好的率失真性能,Wiesmann等人(2021)使用核点卷积,南京大学Gao等人(2021)使用神经图采样来充分利用点的局部相关性。

深度熵模型将点云构建成八叉树形式,并在八叉树上应用神经网络估计概率熵模型。Huang等人(2020a)使用简单多层感知机,根据在八叉树上收集到的上下文信息来进行熵估计。Biswas等人(2020)考虑点云序列间的上下文,并将该上下文信息引入

到神经网络估计的熵模型中，以提升点云序列编码与压缩的性能。北京大学Fu等人(2022)基于注意力机制，充分利用长距离的上下文信息，以进一步提升编码与压缩性能。为了避免过多的上下文信息所引入的额外编解码复杂度，南京大学Wang等人(2022a)提出了轻量级SparsePCGC(sparse point cloud grid compression)模型，该模型已参与了最新的MPEG AI-PCC的基线评测。目前，使用深度学习技术进行点云属性压缩的工作较少，是一个有待于进一步探索的领域。目前代表性的方法是由中山大学Fang等人(2022)提出的3DAC(three dimensional attribute coding)算法，该方法首先将带有属性的点云构建为RAHT树，并使用神经网络为RAHT树构建上下文熵模型，以消除统计冗余。此外，Tang等人(2018，2020)提出基于隐函数表示的自编码器结构，以实现3D/4D点云数据的高效压缩。

2.6 6DoF视频传输优化

6DoF视频的典型应用是扩展现实(extended reality，XR)(Hu等，2020)。XR业务的典型特征是高数据速率和严格的时延预算，因此被归类在5.5G愿景中的eMBB(enhanced mobile broadband)和URLLC(ultra reliable low latency communication)业务之间。早在2016年，3GPP(3rd generation partnership project)已开展支撑XR业务的标准化工作，其中服务和系统工作组定义了高速率和低延迟XR应用程序。2018年，多媒体编解码器、系统和服务工作组继续开展这项工作，报告了相关流量特征。与此同时，系统架构和服务工作组标准化了新的5G服务质量标识符，以支持包括XR在内的交互式服务。各种XR应用程序和服务都有其用户设置、流量和服务质量指标，3GPP SA4为XR业务确定了20多个XR用例，对无线解决方案的性能评估提出了挑战。在此基础上，3GPP建议将XR用例分为3个基本类别，即虚拟现实(virtual reality，VR)、增强现实(augmented reality，AR)和云游戏(cloud game，CG)。对于无线传输来说，XR业务的两个关键性能指标是容量和功耗。在方案对比之前，所有参会组织为容量和延迟约束定义了以用户为中心的联合度量方式，即满足用户数。由于XR业务对时延敏感，因此延迟接收到的数据包与丢失的数据包是等同的，这些超时接收到的数据包将被统计到误包率中。

目前较为主流的VR服务模式是基于视场角的数据流(viewport-dependent streaming，VDS)。VDS是一种自适应流方案，使用网络状态和用户姿势信息来调整3D视频的比特率(Yaqoob等，2020)。具体而言，就是基于用户的位置和方向将全景视频在3D空间上划分为独立的子图像，流服务器通过存储不同质量(即视频分辨率、压缩和帧率)的子图像提供多种表示，由用户动作来触发新视频内容的传输。下载视场(field of view，FOV)中的所有子图后，用户的XR终端设备将进行渲染，然后进行显示。VDS的使用意味着VR服务伴随着上行频繁更新的动作、控制信号，会带来高速的下行传输速率。对于XR CG，控制信号包括手持控制器输入和3DoF/6DoF运动样本，即旋转数据(“滚动”、“俯仰”和“偏航”)以及用户设备的3D空间位移数据。相关研究工作主要包括基于用户视口轨迹的预测方案和基于混合方法的预测方案两类。

1)基于用户视口轨迹的预测方案。Nasrabadi等人(2020)提出了一种基于聚类的视口预测方法，该方法结合当前用户的视口变化轨迹和以前观看者的视口轨迹。算法每隔一定的时间将以前的用户基于他们的视口模式进行聚类，并决定当前用户所属类别，从而利用该类中的视口变化模式预测当前用户的未来视口。Feng等人(2020)提出的LiveDeep方法采用了一种混合方法来解决VR直播流媒体的训练数据不足的问题，并基于卷积神经网络(convolutional neural network，CNN)模型分析视频内容，通过长短时记忆循环神经网络对用户感知轨迹进行预测，以消除单一模型造成的不准确性。类似地，Xu等人(2018)为了避免头部运动预测错误，提出了一种概率视口预测模型，该模型利用了用户方向的概率分布。Yuan等人(2020)采用高斯模型估计用户未来运动视角，并采用Zipf模型估计不同视角的优先级，进而保障用户观看视角的时间—空间质量一致性。Hou等人(2021)提出了基于长短时记忆循环神经网络的视口预测模型。该模型使用过去的头部运动来预测用户注视点的位置，实现了最优段预取方法。

Fan等人(2020)提出利用传感器和内容特性来预测未来帧中每个Tile的观看概率。为了提高预测性能，提出了几种新的增强方法，包括生成虚拟视口、考虑未来内容、降低特征采样率以及使用更大的数据集进行训练。Chen等人(2021)提出了一种用

户感知的视口预测算法Sparkle。该方法首先进行测量研究,分析真实的用户行为,观察到视图方向存在急剧波动,用户姿势对用户的视口移动有显著影响。此外,跨用户的相似性在不同的视频类型中是不同的。基于此,该方法进一步设计了基于用户感知的视口预测算法,通过模拟用户在分片地图上的视口运动,并根据用户的轨迹和其他类似用户在过去时间窗口的行为来确定用户将如何改变视口角度。

2)基于混合方法的预测方案。该类方法在视口预测时除了考虑用户的头部跟踪历史数据,还结合了其他能反映视频内容特性的数据。Nguyen等人(2018)将全景显著性检测模型与头部跟踪历史数据相结合,以实现头部运动预测的精细化预测。Ban等人(2018)利用360°视频自适应流媒体中的跨用户行为信息进行视口预测,试图同时考虑用户的个性化信息和跨用户行为信息来预测未来的视口。与以往基于图像像素级信息的视口预测方法不同,Wu等人(2020a)提出了基于语义内容和偏好的视口预测方法,从嵌入的观看历史中提取用户的语义偏好作为空间注意,以此帮助网络找到未来视频中感兴趣的区域。类似地,Feng等人(2021b)提出的LiveROI(live region of interest)视口预测方案采用实时动作识别方案来理解视频内容,并根据用户轨迹动态更新用户偏好模型,在不需要历史用户或视频数据的情况下有效预测视口。实时视口预测机制LiveObj(live object)通过对视频中的对象进行语义检测并跟踪,再通过强化学习算法实时推断,从而实现用户的视口预测。Zhang等人(2021b)将头部运动预测任务建模为稀疏有向图学习问题。在最新的研究中,Maniotis和Thomos(2022)将VR视频在边缘缓存网络中的内容放置看做马尔可夫决策过程,然后利用深度强化学习算法确定最优缓存放置。Kan等人(2022)提出了一种名为RAPT360(rate adaptive with prediction and trilling 360)的策略,通过拟合不同预测长度下基于拉普拉斯分布的预测误差概率密度函数,以提高视口预测方法的准确性。提出的视口感知自适应平铺方案可根据视口的2维投影的形状和位置分配3种类型的平铺粒度。

当前,6DoF视频传输优化的研究重心已逐渐从全景视频码流转向点云码流。随着数据量的显著增大,6DoF视频传输优化不仅需要考虑视口的自适应预测,还要在编码压缩时考虑到码流容错和纠错能力。此外,为了应对移动终端算力不足的限制,还需要考虑边缘服务器的动态配置与卸载。

3 6DoF视频交互与显示

6DoF视频允许用户自由选择观看视角,这就需要给用户提供大量可供自由选择的视点内容。然而,对任意视角进行视觉内容的采集需要记录的数据量非常大,给采集、存储和传输过程造成很大的负担。因此,在实际的场景环境中,通常采集场景中有限的视点信息,并借助已有视点信息,依靠虚拟视点绘制技术绘制出未采集的视点(即虚拟视点)画面,以供用户自由切换。

现有的虚拟视点图像绘制技术研究正向6DoF方向发展(Jin等,2022)。虚拟视点技术的相关研究与应用大部分还停留在水平基线绘制阶段。考虑到平移自由度是沉浸式视频系统中向用户提供运动视差的关键,MPEG开展了关于平移自由度的探索实验。其中,基于4参考视点的虚拟视点视觉内容绘制算法可以在用户切换观看视点时提供更多的平移自由度,成为近年来的研究热点。绘制算法存在影响用户感知的伪影、背景渗透等绘制失真,且3维映射环节存在计算冗余导致绘制速度较慢,同时参考视点的数量增长进一步增加了3维映射环节的时间消耗,所以绘制技术还存在改进的空间。以下将从解码后滤波增强和虚拟视点合成两个角度展开讨论。

3.1 解码后滤波增强

3.1.1 深度图滤波

由于深度图纹理较少,通常会在编码端以高压缩比进行编码,从而使得解码端的深度图质量较低,这给虚拟视点绘制带来挑战。Yang等人(2015a)提出了直接利用编码参数(如运动矢量、块模式等)来进行深度图滤波的方法。Yuan等人(2012)证明3D视频编码误差服从平稳白噪声的分布规律,并据此首次提出了基于维纳滤波的深度图滤波和虚拟视图滤波方法。Yang和Zheng(2019)提出了一种新型局部双边滤波器,为不太可能受到噪声影响的像素赋予了更高的权重,但没有彻底解决边缘轮廓中的不连续性问题。Yang等人(2019)和He等人(2020a)提出了一种跨视点的多边滤波方法,最终提升了虚拟

视点绘制质量。He等人(2020b)针对有损编码造成的深度失真提出了一种跨视点优化滤波方法，该方法设计了一个互信息度量来模拟跨视点质量一致性的约束，其中包括数据精度和空间平滑性，可以恰当地处理对象边缘上的振铃和错位伪影。

3.1.2　点云上采样

点云上采样任务的目标是对低分辨率稀疏点云进行上采样，生成一个密集、完整且均匀的点云，并需要保持目标物体的形状。现有的点云上采样的方法大致可以分为基于优化和基于深度学习两大类。

1)基于优化方法的模型。该类型方法一般依赖于几何先验知识或者一些额外的场景属性。为了上采样稀疏点集，Alexa等人(2003)提出在局部切线空间的Voronoi图顶点处插入点。Lipman等人(2007)引入了局部最优投影算子来重新采样点并基于L_1范数重建曲面。Huang等人(2009)设计了一种带迭代正态估计的加权策略，以整合具有噪声、异常值和非均匀性的点集。Huang等人(2013)提出边缘感知的点集重采样方法，以实现渐进式点集上采样。Wu等人(2015)通过引入新的点集表示方法，以改善孔洞和缺失区域的填充质量。由于上述方法在建模时依赖于目标点云的先验假设，仅适用于光滑平面，对含有大量噪声的稀疏点云上采样效果有限。

2)基于数据驱动的模型。Yu等人(2018b)首次提出了基于数据驱动的点云上采样模型PU-Net(point cloud upsampling network)。相比基于优化方法的模型，PU-Net显著提升了点云上采样的性能。为了充分利用点云中的全局与局部几何结构，EC-Net(edge-aware point set consolidation network)(Yu等，2018a)实现了边缘感知点云上采样，进一步提高了表面重建质量。为了处理大规模点集，Wang等人(2019)提出的MPU模型在训练集生成时，将上采样目标物体分割成小尺度的片元。

根据模型改进的手段不同，现有的研究工作主要可分为4类。1)基于先进的神经网络架构。如PU-GAN(point cloud upsampling adversarial network)(Li等，2019a)通过利用生成对抗网络学习合成潜伏空间中均匀分布的点。PU-GCN(Qian等，2021)基于图卷积网络来高效提取点云局部结构信息。PU-Transformer(Qiu等，2022)借助多头自注意力机制和位置编码，以增强模型的表示学习能力。PUFA-GAN(Liu等，2022)通过分析点云的频域信息，进一步增强模型的表达和学习能力。2)基于几何先验的模型设计。如PUGeo-Net(geometry-centric network for 3D point cloud upsampling)(Qian等，2020)不仅利用点云的坐标信息，还使用了点云的法向量信息来显式学习目标物体的局部几何表示。深圳大学Zhang等人(2021a)提出了基于可微渲染的点云上采样网络，通过最小化含有重建损失和渲染损失的复合损失函数来生成高质量的稠密点云。Dis-PU(point cloud upsampling via disentangled refinement)(Li等，2021a)首先生成一个能覆盖物体表面的稠密点云，然后再通过微调点的位置来保证点云的分布均匀性。3)任意倍数上采样策略。Meta-PU(meta point cloud upsampling)(Ye等，2022)采用元学习的方式动态调节上采样模块的权重，从而使得模型训练一次就可以支持不同倍率上采样需求。在线性近似理论的基础上，Qian等人(2021)自适应地学习插值权重以及高阶近似误差。Mao等人(2022)在归一化流约束下的特征空间中构建可学习的插值过程。Zhao等人(2022)选择多个靠近物体隐式表面的体素化的点云中心作为种子点，再将种子点密集且均匀地投射到物体的隐式表面，最后通过最远点采样，实现任意倍率的点云上采样任务。4)自监督学习策略。为了提升模型的泛化性。SPU-Net(self-supervised point cloud upsampling)(Liu等，2022)将自监督学习应用在点云上采样任务中。总体而言，现有基于学习的方法依赖于数据集特性，在实际应用时的泛化性能仍有很大提升空间。未来结合优化和数据驱动方法，提升点云上采样任务的性能是一个很有潜力的研究方向。

3.2　虚拟视点合成

按照绘制机理不同，虚拟视点合成方法可根据6DoF视频内容划分为基于模型的绘制(model based rendering，MBR)和基于图像的绘制(image based rendering，IBR)两类。MBR是利用3维网格或者点云数据建立3维立体模型，从而重建出趋于真实的场景(Chen等，2019)。其中，在基于网格的表示方式中，通过基于三角形的方式来表示场景中的对象，对于静态场景可以较好地通过数十、数百或者数千幅输入图像的匹配特征进行划分，获得明确的3D模型。然而，由于网格的不规则性和低细节，从重建的场景中生成动态的新对象是一项困难的任务。MBR方法适用于简单场景，复杂场景中数据量会随着场景

复杂度的增加而急剧增长，不适用于追求强烈交互感的沉浸式场景。IBR方法是使用获取的图像的颜色值来恢复场景的外观，目前有两种方式，即基于光场图像的绘制方法和基于深度图像的绘制方法（depth image-based rendering，DIBR）（Bonatto 等，2021）。与DIBR技术相比，基于光场图像的绘制由于光场数据中含有大量不易压缩的高频信息，实际采集、存储、传输以及终端内容生成的任务都更重，而且产生重影、伪影等失真的概率也更大。DIBR使用的数据更简单，易于处理，技术复杂度低，对设备要求不高，可以生成更具真实感的视觉内容。随着深度估计算法和多视点视觉内容获取技术的长足进步，DIBR技术已成为实现6DoF视频的基础技术。基于神经辐射场的视点合成方法得到了广泛关注（Xu等，2021）。本部分将重点介绍基于深度图像的虚拟视点绘制技术和基于神经辐射场的视点合成技术。

3.2.1　基于深度图像的虚拟视点绘制

DIBR技术包括3维映射（3D-Warping）、视点融合和空洞填补3个环节，考虑到深度图的质量对绘制虚拟视点质量也具有重要意义，因此围绕DIBR技术的研究可划分为3D-Warping优化与加速、视点融合优化和空洞填补优化。

1）3D-Warping优化与加速。3D-Warping是DIBR的核心环节，这一环节对虚拟视点生成的质量和速度有重要影响。Nonaka等人（2018）提出了利用图形处理器并行编程的实时虚拟视点视觉内容绘制方法，大幅降低了绘制一帧图像所需的时间。但这类方法对用户使用的硬件配置提出了较高的要求，另一方面，在算法层面上不去除冗余，仍会占用一定的开销。

针对由3D-Warping环节所引起的绘制质量不佳问题，Ni等人（2009）提出了一种针对汇聚相机阵列的启发式融合插值算法，融合插值过程中考虑了深度、映射像素位置和视点位置，然而难以自适应地确定合适尺寸的窗口。Fachada等人（2018）提出一种支持宽基线场景的视点绘制方法，参考视点图像被划分为以像素中心为顶点的三角形，在映射图像中重新形成的三角形中的像素通过三线性插值进行填充，提高了切向曲面的绘制质量。

针对由3D-Warping环节所引起的绘制速度过慢问题，国内研究者提出利用专用的现场可编程逻辑门阵列设备（Li等，2008）和超大规模集成电路设备（黄超 等，2018）来解决。为了从算法层面提升绘制速度，Jin等人（2016）提出了区域级的映射方法，根据区域的不同特征将区域分类，仅对包含重要信息的区域进行映射操作，避免计算中的冗余信息，大幅减少了映射时间，但由于不同区域利用的是来自不同视点的信息，生成的图像中存在明显的区域边界。在提升绘制质量方面，Fu等人（2017）提出一种基于变换域的用于多视点混合分辨率图像的超分辨率方法，并基于目标低分辨率视点和辅助高分辨率视点之间相关性的最优权重分配算法，可以为低分辨率帧的视点图像提供更多细节信息。Nie等人（2017）针对宽基线街道图像提出了一种新颖的单应性限制映射公式，该公式通过利用映射网格的一阶连续性来增强相邻超像素间单应性传播的平滑度，可以消除重叠、拉伸等小伪影。

2）视点融合优化。不同的融合策略会影响虚拟视点绘制图像绝大部分区域的内容。Vijayanagar等人（2013）根据1维邻域中非空洞像素的数量来优化左右参考视点映射图像的融合权重，但该方法仅能改善空洞附近的失真。Lee等人（2016）利用边缘信息提取出深度图的不可靠区域，根据颜色相似性、深度可靠性和深度值进行视点融合，减少了伪影和模糊。Wegner等人（2016）采用Z-Buffer技术对深度差区域进行视点融合，但该方法需要准确的深度图。Ceulemans等人（2018）提出了一种针对宽基线相机阵列的多视点绘制框架，首先对深度图进行预处理以避免不可靠的信息在整个帧中传播，并且利用加权颜色混合结合直方图匹配确保了参考摄像机的颜色直方图之间的平滑过渡。Sharma和Ragavan（2019）利用几何信息得到纹理匹配概率，自适应地融合参考视点的纹理和深度信息。de Oliveira等人（2021）采用快速分层超像素算法来计算视差和颜色相似性，增强了图像中结构的一致性。

针对平面相机阵列，Chang和Hang等人（2017）提出了一种改进的多参考视点融合算法，选择距离最接近的参考视点作为主导参考视点，并根据其他辅助参考视点的深度和颜色信息修复深度边缘区域中的错误像素。但由于视点切换过程中主导参考视点会发生变化，用户自由巡航时易产生不连续感和出画感。Kim等人（2021）通过直方图匹配去除了由于图像对比度不一致而导致的误差，解决了图像之

间差异较大时出现的失真。Qiao等人(2019)采用多项式拟合方法进行视点亮度校正,提升了虚拟视点融合准确度。

3)空洞填补优化。由于遮挡、采样精度不够高、计算中的舍入误差以及视野的局限性等原因,融合后的虚拟视点图像中存在部分缺失信息的区域需要填补以协调图像的整体视觉效果。空洞填补是利用DIBR过程进行虚拟视点绘制的困难挑战之一,根据参考信息来源可以分为基于图像修复的方法、基于时域的方法和基于空域的方法。

Criminisi等人(2004)提出的修复方法可以在不引入模糊伪影的情况下填充较大的空洞。该方法通过复制来自虚拟视点图像非空洞区域的最佳匹配块来填充空洞,但是有时会错误地采用前景纹理来填充孔洞。因此,基于邻域信息传播的算法会在空洞附近产生模糊伪影。Kim和Ro(2017)提出了一种具有时空一致性和双目对称性的可靠标签传播方法,将相邻视图和前一帧中使用的可靠标签传播到要填充的目标图像,可以避免前景用于空洞填充的发生。Kanchana等人(2022)基于深度学习的方法进行空洞填补,结合时间先验和归一化深度图来预测填充向量,可以提高绘制视点的时空一致性。

实际上,当视点切换时,捕捉时域上的信息更难,所以一些研究者提出了基于空域信息的空洞填补方法。Yao等人(2014)利用时域信息来辅助空洞填补。首先利用纹理和深度信息的时间相关性来生成背景参考图像,然后将其用于填充与场景的动态部分关联的孔洞;而对于静态部分,则使用传统的修补方法。该方法可以避免部分区域的闪烁效应,但是会产生时延现象。Luo等人(2018)提出一种基于快速马尔可夫随机场的空洞填补方法,将图像修复作为能量优化问题并通过循环置信传播来解决,而且利用深度信息来阻止前景纹理错误填充。Lie等人(2018)提出一种建立背景子画面模型填充空洞的方法,通过将视频的空间和时间信息逐步整合到统一的背景子模型中,从而利用真实的背景信息来恢复空洞,但其需要每一帧模型的更新维护和额外的过程,会导致时间复杂度增加。Rahaman和Paul等人(2018)采用高斯混合模型(Gaussian mixed model,GMM)方法来分离背景和前景像素,并通过对相应的GMM模型和映射图像像素亮度的自适应加权平均来恢复映射过程中引入的缺失像素,但其学习率需预先训练得到且无法改变,鲁棒性较差。Thatte和Girod(2019)通过挖掘空洞区域的特性,设计出一种统计模型来预测视点切换而导致虚拟视点图像中丢失数据的可能性,但只能用于单自由度视点切换的情况。Zhu和Gao(2019)针对GMM对于往复运动的局限性,提出了一种改进方法,使用深度信息来调整GMM的学习率,提高了辨别前景像素和背景像素的准确性。Luo等人(2020)提出了一种包括前景提取、运动补偿、背景重构和空洞填补4个模块的空洞填充框架,可使用或扩展现有的大部分背景重建方法和图像修复方法作为该框架的模块。

现有的空洞填补算法存在一定的局限,且不可避免地会引入边缘模糊,无法完全恢复出空洞中的真实信息。基于四参考视点的DIBR算法通过引入更多参考视点的方式显著减少了空洞区域,尤其是消除了位于视野边界的空洞,仅剩余部分公共小块空洞,提升了虚拟视点图像的主客观质量。

3.2.2 基于神经辐射场的视点合成

Mildenhall等人(2020)提出了基于神经辐射场的视点合成方法NeRF(nueral radiance field),该算法使用全连接(非卷积)深度网络表示场景,其输入是单个连续5D坐标(3维空间位置和观察方向),输出是可支持任意视角查看的3维体素场景。算法通过沿相机光线查询5D坐标来合成视图,并使用经典的体渲染技术将输出颜色和密度投影到图像中。因为体积渲染是自然可微的,所以优化表示所需的唯一输入是一组具有已知相机姿势的图像。该算法描述了如何有效地优化神经辐射场以渲染具有复杂几何和外观的场景的逼真的新颖视图,并展示了优于先前神经渲染和视点合成工作的结果。在此基础上,Barron等人(2021)提出了Mip-NeRF的解决方案,扩展了NeRF以连续值的比例表示场景。通过有效地渲染抗锯齿圆锥截头体而不是射线,Mip-NeRF减少了锯齿伪影并显著提高了NeRF表示精细细节的能力。针对全景视频输入,Barron等人(2022)提出了解决采样和混叠问题的NeRF变体Mip-NeRF360,使用非线性场景参数化、在线蒸馏和基于失真的正则化器来克服无界场景带来的模糊或低分辨率的渲染问题。Wang等人(2021c)提出了一种双向阴影渲染方法来实时渲染全景视频中真实和虚拟对象之间的阴影。Hong等人(2022)将神经辐射场与人体头部的参数表示相结合,提出了基于

NeRF的参数化头部模型HeadNeRF，可以在GPU（graphics processing unit）上实时渲染高保真头部图像，并支持直接控制生成图像的渲染姿势和各种语义属性。总体而言，基于神经辐射场的视点合成方法已得到产业界和学界的广泛关注，随着模型训练速度的大幅提升和渐进式渲染技术的广泛研究，将具有非常大的应用潜力。

4 发展趋势与展望

6DoF视频技术的发展将为未来元宇宙时代的到来奠定基础，并且将呈现多维度的发展，包括感官丰富程度的提升、分辨率和码率的提升、时延和可靠性需求的提升以及与现实的交互程度的提升。从这些维度出发，对6DoF视频技术的内容采集与预处理、压缩与传输以及交互与显示提出了更高的要求与挑战。

1）6DoF内容采集与预处理。内容采集的难度以及后期制作技术的复杂程度直接影响了6DoF视频内容制作的难度，因此长期以来是限制6DoF视频发展的主要原因。从发展需求来看，未来的研发方向包括两个方面：（1）轻量化和低成本的视频采集系统。例如，手持彩色3维扫描仪、手持多视点采集系统等装备已经开始具有这些特点，但是距离实际应用还有较长的演进路线；（2）高效、智能的视频内容处理技术。当前技术在几何标定、深度图去噪等方面已经有较好的积累，但适用范围还比较有限，亟需适应面更广、处理流程更智能的技术。

2）6DoF视频压缩与传输。该方向的研究热点主要集中于高效点云压缩和数据传输策略。一方面，现有的点云压缩算法仍存在数据分布刻画难、场景先验利用少和计算复杂度高等挑战。基于3维场景智能分析的大规模3D点云压缩研究，可以实现非结构化点云数据的场景—目标—要素多目标层次化表示，然后根据应用场景类型和目标特性做针对性压缩，以改善重建点云中存在的细节丢失和全局形变等问题，进而实现高效的点云数据编码压缩，是潜在的发展趋势。另一方面，相对于传统视频流式传输场景，点云视频特有的传输方式对资源调度优化引入了新的挑战。例如，在码流传输过程中需要考虑预测视口大小与点云质量等指标之间的平衡。将强化学习在传统视频流式传输场景中的应用迁移到点云视频流式传输场景中，并针对新场景进行适应性的改进与优化，是一个有潜力的研发方向。

3）6DoF视频交互与显示。未来云渲染架构下，大量的视点合成和渲染计算工作都位于云端服务器上完成，可以有效降低终端的计算负载和功耗，同时也使终端的佩戴重量尽可能降低。同时，借助终端的异步时间扭曲技术，实时视频的端到端时延要求可放松至70 ms，实现无眩晕感的沉浸式视频体验。如何对端、管、云三者高效协同，将是未来6DoF视频交互与显示的重要技术方向。

致　谢　本文由中国图象图形学学会图像视频通信专业委员会组织撰写，该专委会链接为http://www.csig.org.cn/detail/2383。

参考文献（References）

Abedi F, Yang Y and Liu Q. 2018. Group geometric calibration and rectification for circular multi-camera imaging system. Optics Express, 26(23): 30596-30613 [DOI: 10.10364/OE.26.030596]

Ahmad W, Vagharshakyan S, Sjöström M, Gotchev A, Bregovic R and Olsson R. 2020. Shearlet transform-based light field compression under low bitrates. IEEE Transactions on Image Processing, 29: 4269-4280 [DOI: 10.1109/TIP.2020.2969087]

Alexa M, Behr J, Cohen-Or D, Fleishman S, Levin D and Silva C T. 2003. Computing and rendering point set surfaces. IEEE Transactions on Visualization and Computer Graphics, 9(1): 3-15 [DOI: 10.1109/tvcg.2003.1175093]

An P, Liu Q, Abedi F and Yang Y. 2020. Novel calibration method for camera array in spherical arrangement. Signal Processing: Image Communication, 80: #115682 [DOI: 10.1016/j. image. 2019. 115682]

Bakir N, Hamidouche W, Fezza S A, Samrouth K and Déforges O. 2021. Light field image coding using VVC standard and view synthesis based on dual discriminator GAN. IEEE Transactions on Multimedia, 23: 2972-2985 [DOI: 10.1109/TMM.2021.3068563]

Ban Y X, Xie L, Xu Z M, Zhang X G, Guo Z M and Wang Y. 2018. CUB360: exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming//Proceedings of 2018 IEEE International Conference on Multimedia and Expo. San Diego, USA: IEEE: 1-6 [DOI: 10.1109/ICME.2018.8486606]

Barron J T, Mildenhall B, Tancik M, Hedman P, Martin-Brualla R and Srinivasan P P. 2021. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Nashville, USA: IEEE: 5835-5844 [DOI: 10.1109/ICCV48922.2021.00580]

Barron J T, Mildenhall B, Verbin D, Srinivasan P P and Hedman P.

2022. Mip-NeRF 360: unbounded anti-aliased neural radiance fields//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5460-5469 [DOI: 10.1109/CVPR52688.2022.00539]

Biswas S, Liu J, Wong K, Wang S L and Urtasun R. 2020. MuSCLE: multi sweep compression of LiDAR using deep entropy models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1859 [DOI: 10.48550/arXiv.2011.07590]

Bonatto D, Hirt G, Kvasov A, Fachada S and Lafruit G. 2021. MPEG immersive video tools for light field head mounted displays//Proceedings of 2021 International Conference on Visual Communications and Image Processing. Munich, Germany: IEEE: #9675317 [DOI: 10.1109/VCIP53242.2021.9675317]

Boyce J M, Doré R, Dziembowski A, Fleureau J, Jung J, Kroon B, Salahieh B, Vadakital V K M and Yu L. 2021. MPEG immersive video coding standard. Proceedings of the IEEE, 109(9): 1521-1536 [DOI: 10.1109/JPROC.2021.3062590]

Brady D J, Gehm M E, Stack R A, Marks D L, Kittle D S, Golish D R, Vera E M and Feller S D. 2012. Multiscale gigapixel photography. Nature, 486(7403): 386-389 [DOI: 10.1038/nature11150]

Bross B, Wang Y K, Yan Y, Liu S, Chen J L, Sullivan G J and Ohm J R. 2021. Overview of the versatile video coding (VVC) standard and its applications. IEEE Transactions on Circuits and Systems for Video Technology, 31(10): 3736-3764 [DOI: 10.1109/TCSVT.2021.3101953]

Buades A, Coll B and Morel J M. 2005. A non-local algorithm for image denoising//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE: 60-65 [DOI: 10.1109/CVPR.2005.38]

Cernigliaro G, Jaureguizar F, Ortega A, Cabrera J and García N. 2009. Fast mode decision for multiview video coding based on depth maps//Proceedings of SPIE 7257, Visual Communications and Image Processing 2009. San Jose, USA: SPIE: #72570N [DOI: 10.1117/12.806861]

Ceulemans B, Lu S P, Lafruit G and Munteanu A. 2018. Robust multiview synthesis for wide-baseline camera arrays. IEEE Transactions on Multimedia, 20(9): 2235-2248 [DOI: 10.1109/TMM.2018.2802646]

Chang C L, Zhu X Q, Ramanathan P and Girod B. 2006. Light field compression using disparity-compensated lifting and shape adaptation. IEEE Transactions on Image Processing, 15(4): 793-806 [DOI: 10.1109/TIP.2005.863954]

Chang H R and Hang H M. 2017. Wide angle virtual view synthesis using two-by-two Kinect V2//Proceedings of 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Kuala Lumpur, Malaysia: IEEE: 1083-1091 [DOI: 10.1109/APSIPA.2017.8282189]

Chauve A L, Labatut P and Pons J P. 2010. Robust piecewise-planar 3D reconstruction and completion from large-scale unstructured point data//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 1261-1268 [DOI: 10.1109/CVPR.2010.5539824]

Chen J, Hou J H and Chau L P. 2018. Light field compression with disparity-guided sparse coding based on structural key views. IEEE Transactions on Image Processing, 27(1): 314-324 [DOI: 10.1109/TIP.2017.2750413]

Chen J, Watanabe R, Nonaka K, Konno T, Sankoh H and Naito S. 2019. Fast free-viewpoint video synthesis algorithm for sports scenes//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE: 3209-3215 [DOI: 10.1109/IROS40897.2019.8967584]

Chen J Y, Luo X Z, Hu M, Wu D and Zhou Y P. 2021. Sparkle: user-aware viewport prediction in 360-degree video streaming. IEEE Transactions on Multimedia, 23: 3853-3866 [DOI: 10.1109/TMM.2020.3033127]

Choudhury P and Tumblin J. 2005. The trilateral filter for high contrast images and meshes//Proceedings of the ACM SIGGRAPH 2005 Courses. Los Angeles, United States: ACM: #1198565 [DOI: 10.1145/1198555.1198565]

Chung K L, Huang Y H, Lin C H and Fang J P. 2016. Novel bitrate saving and fast coding for depth videos in 3D-HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 26(10): 1859-1869 [DOI: 10.1109/TCSVT.2015.2473296]

Chung T Y, Sim J Y and Kim C S. 2014. Bit allocation algorithm with novel view synthesis distortion model for multiview video plus depth coding. IEEE Transactions on Image Processing, 23(8): 3254-3267 [DOI: 10.1109/TIP.2014.2327801]

Criminisi A, Perez P and Toyama K. 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 13(9): 1200-1212 [DOI: 10.1109/TIP.2004.833105]

Dai A, Qi C R and Nießner M. 2017. Shape completion using 3D-encoder-predictor CNNs and shape synthesis//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6545-6554 [DOI: 10.1109/CVPR.2017.693]

Dai F, Zhang J, Ma Y K and Zhang Y D. 2015. Lenselet image compression scheme based on subaperture images streaming//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE: 4733-4737 [DOI: 10.1109/ICIP.2015.7351705]

De Abreu A, Frossard P and Pereira F. 2015. Optimizing multiview video plus depth prediction structures for interactive multiview video streaming. IEEE Journal of Selected Topics in Signal Processing, 9(3): 487-500 [DOI: 10.1109/JSTSP.2015.2407320]

de Oliveira A Q, da Silveira T L T, Walter M and Jung C R. 2021. A hierarchical superpixel-based approach for DIBR view synthesis.

IEEE Transactions on Image Processing, 30: 6408-6419 [DOI: 10.1109/TIP.2021.3092817]

de Queiroz R L and Chou P A. 2016. Compression of 3D point clouds using a region-adaptive hierarchical transform. IEEE Transactions on Image Processing, 25(8): 3947-3956 [DOI: 10.1109/TIP.2016.2575005]

Dib E, Le Pendu M, Jiang X R and Guillemot C. 2020. Local low rank approximation with a parametric disparity model for light field compression. IEEE Transactions on Image Processing, 29: 9641-9653 [DOI: 10.1109/TIP.2020.3029655]

Ding L F, Tsung P K, Chien S Y, Chen W Y and Chen L G. 2008. Content-aware prediction algorithm with inter-view mode decision for multiview video coding. IEEE Transactions on Multimedia, 10(8): 1553-1564 [DOI: 10.1109/TMM.2008.2007314]

Domański M, Al-Obaidi Y and Grajek T. 2021. Universal modeling of monoscopic and multiview video codecs with applications to encoder control//Proceedings of 2021 IEEE International Conference on Image Processing. Anchorage, USA: IEEE: 2144-2148 [DOI: 10.1109/ICIP42928.2021.9506735]

Doutre C and Nasiopoulos P. 2009. Color correction preprocessing for multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology, 19(9): 1400-1406 [DOI: 10.1109/TCSVT.2009.2022780]

Fachada S, Bonatto D, Schenkel A and Lafruit G. 2018. Depth image based view synthesis with multiple reference views for virtual reality//Proceedings of 2018 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video. Helsinki, Finland: IEEE: 1-4 [DOI: 10.1109/3DTV.2018.8478484]

Fan C L, Yen S C, Huang C Y and Hsu C H. 2020. Optimizing fixation prediction using recurrent neural networks for 360° video streaming in head-mounted virtual reality. IEEE Transactions on Multimedia, 22(3): 744-759 [DOI: 10.1109/TMM.2019.2931807]

Fang G C, Hu Q Y, Wang H Y, Xu Y L and Guo Y L. 2022. 3DAC: learning attribute compression for point clouds//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, United States: IEEE: 14799-14808 [DOI: 10.1109/CVPR52688.2022.01440]

Feng X L, Li W T and Wei S. 2021b. LiveROI: region of interest analysis for viewport prediction in live mobile virtual reality streaming// Proceedings of the 12th ACM Multimedia Systems Conference. Istanbul, Turkey: Association for Computing Machinery: 132-145 [DOI: 10.1145/3458305.3463378]

Feng X L, Liu Y and Wei S. 2020. LiveDeep: online viewport prediction for live virtual reality streaming using lifelong deep learning// Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces. Atlanta, USA: IEEE: 800-808 [DOI: 10.1109/VR46266.2020.00104]

Fezza S A, Larabi M C and Faraoun K M. 2014. Feature-based color correction of multiview video for coding and rendering enhancement. IEEE Transactions on Circuits and Systems for Video Technology, 24(9): 1486-1498 [DOI: 10.1109/TCSVT.2014.2309776]

Fiengo A, Chierchia G, Cagnazzo M and Pesquet-Popescu B. 2016. Convex optimization for frame-level rate allocation in MV-HEVC//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA: IEEE: 2157-2161 [DOI: 10.1109/ICIP.2016.7532740]

Franca J G D M, Gazziro M A, Ide A N and Saito J H. 2005. A 3D scanning system based on laser triangulation and variable field of view// Proceedings of 2005 IEEE International Conference on Image Processing. Genova, Italy: IEEE: 425-428 [DOI: 10.1109/ICIP.2005.1529778]

Fu C Y, Li G, Song R, Gao W and Liu S. 2022. OctAttention: octree-based large-scale contexts model for point cloud compression//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press: 625-633 [DOI: 10.1609/aaai.v36i1.19942]

Fu Z Z, Li Y, Xu J, Wu H G and Lai Y W. 2017. Super resolution for multiview mixed resolution images in transform-domain with optimal weight. Multimedia Tools and Applications, 76(2): 3031-3045 [DOI: 10.1007/s11042-016-3258-9]

Gao L Y, Fan T Y, Wang J Q, Xu Y L, Sun J and Ma Z. 2021. Point cloud geometry compression via neural graph sampling//Proceedings of 2021 IEEE International Conference on Image Processing. Anchorage, USA: IEEE: 3373-3377 [DOI: 10.1109/ICIP42928.2021.9506631]

Gao Y, Yang Y, Zhen Y and Dai Q H. 2015. Depth error elimination for RGB-D cameras. ACM Transactions on Intelligent Systems and Technology, 6(2): #13 [DOI: 10.1145/2735959]

Gastal E S L and Oliveiray M M. 2012. Adaptive manifolds for real-time high-dimensional filtering. ACM Transactions on Graphics, 31(4): #33 [DOI: 10.1145/2185520.2185529]

He X, Liu Q and Yang Y. 2020a. MV-GNN: multi-view graph neural network for compression artifacts reduction. IEEE Transactions on Image Processing, 29: 6829-6840 [DOI: 10.1109/TIP.2020.2994412]

He X, Liu Q and Yang Y. 2020b. Make full use of priors: cross-view optimized filter for multi-view depth enhancement. ACM Transactions on Multimedia Computing, Communications, and Applications, 16(4): #127 [DOI: 10.1145/3408293]

Hong Y, Peng B, Xiao H Y, Liu L G and Zhang J Y. 2022. HeadNeRF: a realtime NeRF-based parametric head model//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20342-20352 [DOI: 10.1109/CVPR52688.2022.01973]

Hou J H, Chen J and Chau L P. 2019. Light field image compression based on Bi-level view compensation with rate-distortion optimization. IEEE Transactions on Circuits and Systems for Video Technology, 29(2): 517-530 [DOI: 10.1109/TCSVT.2018.2802943]

Hou X S, Dey S, Zhang J Z and Budagavi M. 2021. Predictive adaptive streaming to enable mobile 360-degree and VR experiences. IEEE Transactions on Multimedia, 23: 716-731 [DOI: 10.1109/TMM.2020.2987693]

Hu F H, Deng Y S, Saad W, Bennis M and Aghvami A H. 2020. Cellular-connected wireless virtual reality: requirements, challenges, and solutions. IEEE Communications Magazine, 58(5): 105-111 [DOI: 10.1109/MCOM.001.1900511]

Hu W, Fu Z Q and Guo Z M. 2019. Local frequency interpretation and non-local self-similarity on graph for point cloud inpainting. IEEE Transactions on Image Processing, 28(8): 4087-4100 [DOI: 10.1109/TIP.2019.2906554]

Huang C, Peng Z J, Miao J C and Chen F. 2018. Joint depth video enhancement and fast intra encoding algorithm in 3D-HEVC. Journal of Image and Graphics, 23(4): 500-509 (黄超, 彭宗举, 苗瑾超, 陈芬. 2018. 联合深度视频增强的3D-HEVC帧内编码快速算法. 中国图象图形学报, 23(4): 500-509) [DOI: 10.11834/jig.170452]

Huang H, Li D, Zhang H, Ascher U and Cohen-Or D. 2009. Consolidation of unorganized point clouds for surface reconstruction. ACM Transactions on Graphics, 28(5): 1-7 [DOI: 10.1145/1618452.1618522]

Huang H, Wu S H, Gong M L, Cohen-Or D, Ascher U and Zhang H. 2013. Edge-aware point set resampling. ACM Transactions on Graphics, 32(1): #9 [DOI: 10.1145/2421636.2421645]

Huang H C, Wang Y C, Chen W C, Lin P Y and Huang C T. 2019. System and VLSI implementation of phase-based view synthesis//Proceedings of the ICASSP 2019 —— 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Brighton, UK: IEEE: 1428-1432 [DOI: 10.1109/ICASSP.2019.8682399]

Huang L L, Wang S L, Wong K, Liu J and Urtasun R. 2020a. OctSqueeze: octree-structured entropy model for LiDAR compression//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1310-1320 [DOI: 10.1109/CVPR42600.2020.00139]

Huang Z T, Yu Y K, Xu J W, Ni F and Le X Y. 2020b. PF-Net: point fractal network for 3D point cloud completion//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 7662-7670 [DOI: 10.1109/CVPR42600.2020.00768]

Ibrahim M M, Liu Q, Khan R, Yang J Y, Adeli E and Yang Y. 2020a. Depth map artefacts reduction: a review. IET Image Processing, 14(12): 2630-2644 [DOI: 10.1049/iet-ipr.2019.1622]

Ibrahim M M, Liu Q and Yang Y. 2020b. Adaptive colour-guided non-local means algorithm for compound noise reduction of depth maps. IET Image Processing, 14(12): 2768-2779 [DOI: 10.1049/iet-ipr.2019.0074]

Jia C M, Zhang X F, Wang S S, Wang S Q and Ma S W. 2019. Light field image compression using generative adversarial network-based view synthesis. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9(1): 177-189 [DOI: 10.1109/JETCAS.2018.2886642]

Jiang C Y and Nooshabadi S. 2016. A scalable massively parallel motion and disparity estimation scheme for multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology, 26(2): 346-359 [DOI: 10.1109/TCSVT.2015.2402853]

Jiang X R, Le Pendu M, Farrugia R A and Guillemot C. 2017. Light field compression with homography-based low-rank approximation. IEEE Journal of Selected Topics in Signal Processing, 11(7): 1132-1145 [DOI: 10.1109/JSTSP.2017.2747078]

Jin C C, Peng Z J, Chen F and Jiang G Y. 2022. Subjective and objective video quality assessment for windowed-6DoF synthesized videos. IEEE Transactions on Broadcasting, 68(3): 594-608 [DOI: 10.1109/TBC.2022.3165473]

Jin J, Wang A H, Zhao Y, Lin C Y and Zeng B. 2016. Region-aware 3-D warping for DIBR. IEEE Transactions on Multimedia, 18(6): 953-966 [DOI: 10.1109/TMM.2016.2539825]

Kan N W, Zou J N, Li C L, Dai W R and Xiong H K. 2022. RAPT360: reinforcement learning-based rate adaptation for 360-degree video streaming with adaptive prediction and tiling. IEEE Transactions on Circuits and Systems for Video Technology, 32(3): 1607-1623 [DOI: 10.1109/TCSVT.2021.3076585]

Kanchana V, Somraj N, Yadwad S and Soundararajan R. 2022. Revealing disocclusions in temporal view synthesis through infilling vector prediction//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 3093-3102 [DOI: 10.1109/WACV51458.2022.00315]

Kazhdan M and Hoppe H. 2013. Screened poisson surface reconstruction. ACM Transactions on Graphics, 32(3): #29 [DOI: 10.1145/2487228.2487237]

Khire S, Cooper L, Park Y, Carter A, Jayant N and Saltz J. 2012. ZPEG: a hybrid DPCM-DCT based approach for compression of Z-stack images//Proceedings of 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego, USA: IEEE: 5424-5427 [DOI: 10.1109/EMBC.2012.6347221]

Kim D M, Yoon Y S, Shin S Y and Suh J W. 2021. Rendering-based free-view image synthesis using peripheral view and depth images//Proceedings of the 36th International Technical Conference on Circuits/Systems, Computers and Communications. Jeju, Korea (South): IEEE: #9501463 [DOI: 10.1109/ITC-CSCC52171.2021.9501463]

Kim H G and Ro Y M. 2017. Multiview stereoscopic video hole filling considering spatiotemporal consistency and binocular symmetry for synthesized 3D video. IEEE Transactions on Circuits and Systems for Video Technology, 27(7): 1435-1449 [DOI: 10.1109/TCSVT.2016.2515360]

Kim V G, Li W, Mitra N J, Chaudhuri S, DiVerdi S and Funkhouser T.

2013. Learning part-based templates from large collections of 3D shapes. ACM Transactions on Graphics, 32(4): #70 [DOI: 10.1145/2461912.2461933]

Klimaszewski K, Stankiewicz O, Wegner K and Domański M. 2014. Quantization optimization in multiview plus depth video coding// Proceedings of 2014 IEEE International Conference on Image Processing. Paris, France: IEEE: 3223-3227 [DOI: 10.1109/ICIP.2014.7025652]

Kopf J, Cohen M F, Lischinski D and Uyttendaele M. 2007. Joint bilateral upsampling. ACM Transactions on Graphics, 26(3): #1276497 [DOI: 10.1145/1276377.1276497]

Lee T C, Chien C L and Hang H M. 2016. Virtual view synthesis quality refinement//Proceedings of 2016 3DTV-Conference: the True Vision —— Capture, Transmission and Display of 3D Video. Hamburg, Germany: IEEE: 1-4 [DOI: 10.1109/3DTV.2016.7548958]

Lei J J, Sun J, Pan Z Q, Kwong S, Duan J H and Hou C P. 2015. Fast mode decision using inter-view and inter-component correlations for multiview depth video coding. IEEE Transactions on Industrial Informatics, 11(4): 978-986 [DOI: 10.1109/TII.2015.2446769]

Lei J J, Zhang Z Q, Pan Z Q, Liu D, Liu X R, Chen Y and Ling N. 2022. Disparity-aware reference frame generation network for multiview video coding. IEEE Transactions on Image Processing, 31: 4515-4526 [DOI: 10.1109/TIP.2022.3183436]

Levoy M and Hanrahan P. 1996. Light field rendering//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA: ACM: 31-42 [DOI: 10.1145/237170.237199]

Li G, Lei J J, Pan Z Q, Peng B and Ling N. 2022. Multiple resolution prediction with deep up-sampling for depth video coding. IEEE Transactions on Circuits and Systems for Video Technology, 32(9): 6337-6346 [DOI: 10.1109/TCSVT.2022.3157074]

Li L, Li Z, Li B, Liu D and Li H Q. 2017. Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression. IEEE Journal of Selected Topics in Signal Processing, 11(7): 1107-1119 [DOI: 10.1109/JSTSP.2017.2725198]

Li L H, Xiang S, Yang Y and Yu L. 2015. Multi-camera interference cancellation of time-of-flight (TOF) cameras//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE: 556-560 [DOI: 10.1109/ICIP.2015.7350860]

Li R H, Li X Z, Fu C W, Cohen-Or D and Heng P A. 2019a. PU-GAN: a point cloud upsampling adversarial network//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 7202-7211 [DOI: 10.1109/ICCV.2019.00730]

Li R H, Li X Z, Heng P A and Fu C W. 2021a. Point cloud upsampling via disentangled refinement//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 344-353 [DOI: 10.1109/CVPR46437.2021.00041]

Li T S, Yu L, Wang H K and Kuang Z. 2021b. A bit allocation method based on inter-view dependency and spatio-temporal correlation for multi-view texture video coding. IEEE Transactions on Broadcasting, 67(1): 159-173 [DOI: 10.1109/TBC.2020.3028340]

Li X M, Zhao D B, Ma S W and Gao W. 2008. Fast disparity and motion estimation based on correlations for multiview video coding. IEEE Transactions on Consumer Electronics, 54(4): 2037-2044 [DOI: 10.1109/TCE.2008.4711270]

Li Y, Yang G B, Chen N, Zhu Y P and Ding X L. 2016b. Early DIRECT mode decision for MVC using MB mode homogeneity and RD Cost correlation. IEEE Transactions on Broadcasting, 62(3): 700-708 [DOI: 10.1109/TBC.2016.2570018]

Lie W N, Hsieh C Y and Lin G S. 2018. Key-frame-based background sprite generation for hole filling in depth image-based rendering. IEEE Transactions on Multimedia, 20(5): 1075-1087 [DOI: 10.1109/TMM.2017.2763319]

Lilienblum E and Al-Hamadi A. 2015. A structured light approach for 3-D surface reconstruction with a stereo line-scan system. IEEE Transactions on Instrumentation and Measurement, 64(5): 1258-1266 [DOI: 10.1109/TIM.2014.2364105]

Lin J J, Rickert M, Perzylo A and Knoll A. 2021. PCTMA-Net: point cloud transformer with morphing atlas-based point generation network for dense point cloud completion//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague, Czech Republic: IEEE: 5657-5663 [DOI: 10.1109/IROS51168.2021.9636483]

Lipman Y, Cohen-Or D, Levin D and Tal-Ezer H. 2007. Parameterization-free projection for geometry reconstruction. ACM Transactions on Graphics, 26(3): #1276405 [DOI: 10.1145/1276377.1276405]

Liu D, Wang L, Li L, Xiong Z, Wu F and Zeng W. 2016. Pseudo-sequence-based light field image compression//Proceedings of 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, USA: IEEE: #7574674 [DOI: 10.1109/ICMEW.2016.7574674]

Liu H, Yuan H, Liu Q, Hou J and Liu J. 2020. A comprehensive study and comparison of core technologies for MPEG 3D point cloud compression. IEEE Transactions on Broadcasting, 66(3): 701-717 [DOI: 10.1109/TBC.2019.2957652]

Liu H, Yuan H, Hou J, Hamzaoui R and Gao W. 2022. PUFA-GAN: a frequency-aware generative adversarial network for 3D point cloud upsampling. IEEE Transactions on Image Processing, 31: 7389-7402 [DOI: 10.1109/TIP.2022.3222918]

Liu W, Chen X G, Yang J and Wu Q. 2017. Robust color guided depth map restoration. IEEE Transactions on Image Processing, 26(1): 315-327 [DOI: 10.1109/TIP.2016.2612826]

Liu X H, Liu X C, Liu Y S and Han Z Z. 2022. SPU-Net: self-supervised point cloud upsampling by coarse-to-fine reconstruction with self-projection optimization. IEEE Transactions on Image Processing, 31: 4213-4226 [DOI: 10.1109/TIP.2022.3182266]

Liu Y W, Huang Q M, Ma S W, Zhao D B, Gao W, Ci S and Tang H. 2011. A novel rate control technique for multiview video plus depth based 3D video coding. IEEE Transactions on Broadcasting, 57(2): 562-571 [DOI: 10.1109/TBC.2011.2105652]

Liu Y Y, Zhu C and Guo H W. 2019. Survey of light field data compression. Journal of Image and Graphics, 24(11): 1842-1859 (刘宇洋, 朱策, 郭红伟. 2019. 光场数据压缩研究综述. 中国图象图形学报, 24(11): 1842-1859) [DOI: 10.11834/jig.190035]

Luo G B, Zhu Y S and Guo B. 2018. Fast MRF-based hole filling for view synthesis. IEEE Signal Processing Letters, 25(1): 75-79 [DOI: 10.1109/LSP.2017.2720182]

Luo G B, Zhu Y S, Weng Z Y and Li Z T. 2020. A disocclusion inpainting framework for depth-based view synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(6): 1289-1302 [DOI: 10.1109/TPAMI.2019.2899837]

Maniotis P and Thomos N. 2022. Viewport-aware deep reinforcement learning approach for 360° video caching. IEEE Transactions on Multimedia, 24: 386-399 [DOI: 10.1109/TMM.2021.3052339]

Mao A, Du Z, Hou J, Duan Y, Liu Y and He Y.2022. Pu-flow: a point cloud upsampling network with normalizing flows. IEEE Transactions on Visualization and Computer Graphics: #05893 [10.48550/arXiv.2107.05893]

Marr D and Poggio T. 1976. Cooperative computation of stereo disparity: a cooperative algorithm is derived for extracting disparity information from stereo image pairs. Science, 194(4262): 283-287 [DOI: 10.1126/science.968482]

Mekuria R, Blom K and Cesar P. 2017. Design, implementation, and evaluation of a point cloud codec for tele-immersive video. IEEE Transactions on Circuits and Systems for Video Technology, 27(4): 828-842 [DOI: 10.1109/TCSVT.2016.2543039]

Merkle P, Smolic A, Muller K and Wiegand T. 2007. Efficient prediction structures for multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology, 17(11): 1461-1473 [DOI: 10.1109/TCSVT.2007.903665]

Mieloch D, Dziembowski A and Domański M. 2021. Depth map refinement for immersive video. IEEE Access, 9: 10778-10788 [DOI: 10.1109/ACCESS.2021.3050554]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]

Nasrabadi A T, Samiei A and Prakash R. 2020. Viewport prediction for 360° videos: a clustering approach//Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video. Istanbul, Turkey: Association for Computing Machinery: 34-39 [DOI: 10.1145/3386290.3396934]

Nealen A, Igarashi T, Sorkine O and Alexa M. 2006. Laplacian mesh optimization//Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia. Kuala Lumpur, Malaysia: ACM: 381-389 [DOI: 10.1145/1174429.1174494]

Nguyen A, Yan Z S and Nahrstedt K. 2018. Your attention is unique: detecting 360-degree video saliency in head-mounted display for head movement prediction//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): Association for Computing Machinery: 1190-1198 [DOI: 10.1145/3240508.3240669]

Ni Z F, Tian D, Bhagavathy S, Llach J and Manjunath B S. 2009. Improving the quality of depth image based rendering for 3D video systems//Proceedings of the 16th IEEE International Conference on Image Processing. Cairo, Egypt: IEEE: 513-516 [DOI: 10.1109/ICIP.2009.5413941]

Nie Y W, Zhang Z S, Sun H Q, Su T and Li G Q. 2017. Homography propagation and optimization for wide-baseline street image interpolation. IEEE Transactions on Visualization and Computer Graphics, 23(10): 2328-2341 [DOI: 10.1109/TVCG.2016.2618878]

Niu Y Z, Zheng X H, Zhao T S and Chen J H. 2020. Visually consistent color correction for stereoscopic images and videos. IEEE Transactions on Circuits and Systems for Video Technology, 30(3): 697-710 [DOI: 10.1109/TCSVT.2019.2897123]

Nonaka K, Watanabe R, Chen J, Sabirin H and Naito S. 2018. Fast plane-based free-viewpoint synthesis for real-time live streaming//Proceedings of 2018 IEEE Visual Communications and Image Processing. Taichung, China: IEEE: 1-4 [DOI: 10.1109/VCIP.2018.8698648]

Ohm J R, Sullivan G J, Schwarz H, Tan T K and Wiegand T. 2012. Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC). IEEE Transactions on Circuits and Systems for Video Technology, 22(12): 1669-1684 [DOI: 10.1109/TCSVT.2012.2221192]

Pan L, Chen X Y, Cai Z G, Zhang J Z, Zhao H Y, Yi S and Liu Z W. 2021. Variational relational point completion network//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8520-8529 [DOI: 10.1109/CVPR46437.2021.00842]

Pan Z Q, Zhang Y and Kwong S. 2015. Efficient motion and disparity estimation optimization for low complexity multiview video coding. IEEE Transactions on Broadcasting, 61(2): 166-176 [DOI: 10.1109/TBC.2015.2419824]

Paul M. 2018. Efficient multiview video coding using 3-D coding and saliency-based bit allocation. IEEE Transactions on Broadcasting, 64(2): 235-246 [DOI: 10.1109/TBC.2017.2781118]

Pauly M, Mitra N J, Giesen J, Gross M and Guibas L J. 2005. Example-based 3D scan completion//Proceedings of the 3rd Eurographics Symposium on Geometry Processing. Vienna, Austria: Eurographics Association: #23

Peng B, Chang R J, Pan Z Q, Li G, Ling N and Lei J J. 2022. Deep

in-loop filtering via multi-domain correlation learning and partition constraint for multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology: #3213515 [DOI: 10.1109/TCSVT.2022.3213515]

Peng Z J, Han H M, Chen F, Jiang G Y and Yu M. 2016. Joint processing and fast encoding algorithm for multi-view depth video. Eurasip Journal on Image and Video Processing, 2016(1): #24 [DOI: 10.1186/s13640-016-0128-3]

Qi Charles R, Su H, Kaichun M and Guibas L J. 2017a. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 77-85 [DOI: 10.1109/CVPR.2017.16]

Qi C R, Yi L, Su H and Guibas L J. 2017b. PointNet++: deep hierarchical feature learning on point sets in a metric space//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5105-5114 [DOI: 10.48550/arXiv.1706.02413]

Qian G C, Abualshour A, Li G H, Thabet A and Ghanem B. 2021. PU-GCN: point cloud upsampling using graph convolutional networks//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 11678-11687 [DOI: 10.1109/CVPR46437.2021.01151]

Qian Y, Hou J H, Kwong S and He Y. 2020. PUGeo-Net: a geometry-centric network for 3D point cloud upsampling//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 752-769 [DOI: 10.1007/978-3-030-58529-7_44]

Qian Y, Hou J, Kwong S and He Y. 2021. Deep magnification-flexible upsampling over 3D point clouds. IEEE Transactions on Image Processing, 30: 8354-8367 [10.1109/TIP.2021.3115385]

Qiao Y G, Jiao L C, Yang S Y, Hou B and Feng J. 2019. Color correction and depth-based hierarchical hole filling in free viewpoint generation. IEEE Transactions on Broadcasting, 65(2): 294-307 [DOI: 10.1109/TBC.2019.2901391]

Qiu S, Anwar S and Barnes N. 2022. PU-transformer: point cloud upsampling transformer//Proceedings of the 16th Asian Conference on Computer Vision. Macau, China: Springer: 2475-2493 [DOI: 10.48550/arXiv.2111.12242]

Quach M, Valenzise G and Dufaux F. 2019. Learning convolutional transforms for lossy point cloud geometry compression//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 4320-4324 [DOI: 10.1109/ICIP.2019.8803413]

Quach M, Valenzise G and Dufaux F. 2020. Improved deep point cloud geometry compression//Proceedings of the 22nd IEEE International Workshop on Multimedia Signal Processing. Tampere, Finland: IEEE: #928707 [DOI: 10.1109/MMSP48831.2020.9287077]

Rahaman M D and Paul M. 2018. Virtual view synthesis for free viewpoint video and multiview video compression using Gaussian mixture modelling. IEEE Transactions on Image Processing, 27(3): 1190-1201 [DOI: 10.1109/TIP.2017.2772858]

Rizkallah M, Maugey T and Guillemot C. 2021. Rate-distortion optimized graph coarsening and partitioning for light field coding. IEEE Transactions on Image Processing, 30: 5518-5532 [DOI: 10.1109/TIP.2021.3085203]

Rock J, Gupta T, Thorsen J, Gwak J, Shin D and Hoiem D. 2015. Completing 3D object shape from one depth image//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 2484-2493 [DOI: 10.1109/CVPR.2015.7298863]

Sakamoto T, Kodama K and Hamamoto T. 2012a. A novel scheme for 4-D light-field compression based on 3-D representation by multi-focus images//Proceedings of the 19th IEEE International Conference on Image Processing. Orlando, USA: IEEE: 2901-2904 [DOI: 10.1109/ICIP.2012.6467506]

Sakamoto T, Kodama K and Hamamoto T. 2012b. A study on efficient compression of multi-focus images for dense light-field reconstruction//Proceedings of 2012 Visual Communications and Image Processing. San Diego, USA: IEEE: #6410759 [DOI: 10.1109/VCIP.2012.6410759]

Sarmad M, Lee H J and Kim Y M. 2019. RL-GAN-Net: a reinforcement learning agent controlled GAN network for real-time point cloud shape completion//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5898-5907 [DOI: 10.1109/CVPR.2019.00605]

Scharstein D and Szeliski R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47(1/3): 7-42 [DOI: 10.1023/A:1014573219977]

Schnabel R and Klein R. 2006. Octree-based point-cloud compression//Botsch M, Chen B Q, Pauly M and Zwicker M, eds. Symposium on Point-Based Graphics. [s. l.]: The Eurographics Association: 1811-7813 [DOI: 10.2312/SPBG/SPBG06/111-120]

Schwarz S, Preda M, Baroncini V, Budagavi M, Cesar P, Chou P A, Cohen R A, Krivokuća M, Lasserre S, Li Z, Llach J, Mammou K, Mekuria R, Nakagami O, Siahaan E, Tabatabai A, Tourapis A M and Zakharchenko V. 2019. Emerging MPEG standards for point cloud compression. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9(1): 133-148 [DOI: 10.1109/JETCAS.2018.2885981]

Sharma M and Ragavan G. 2019. A novel image fusion scheme for FTV view synthesis based on layered depth scene representation and scale periodic transform//Proceedings of 2019 International Conference on 3D Immersion. Brussels, Belgium: IEEE: 1-8 [DOI: 10.1109/IC3D48390.2019.8975902]

Shen L Q, Liu Z, Yan T, Zhang Z Y and An P. 2010. Early SKIP mode decision for MVC using inter-view correlation. Signal Processing: Image Communication, 25(2): 88-93 [DOI: 10.1016/j.image.

2009.11.003］

Stankiewicz O, Lafruit G and Domański M. 2018. Multiview video: acquisition, processing, compression, and virtual view rendering// Chellappa R and Theodoridis S, eds. Academic Press Library in Signal Processing, Volume 6. Amsterdam, the Netherlands: Elsevier: 3-74［DOI: 10.1016/B978-0-12-811889-4.00001-4］

Su X, Rizkallah M, Maugey T and Guillemot C. 2017. Graph-based light fields representation and coding using geometry information// Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE: 4023-4027［DOI: 10.1109/ICIP.2017.8297038］

Su X Y, Zhang Q C and Chen W J. 2014. Three-dimensional imaging based on structured illumination. Chinese Journal of Lasers, 41(2): #0209001 (苏显渝, 张启灿, 陈文静. 2014. 结构光3维成像技术. 中国激光, 41(2): #0209001)［DOI: 10.3788/CJL201441.0209001］

Tang D H, Dou M S, Lincoln P, Davidson P, Guo K W, Taylor J, Fanello S, Keskin C, Kowdle A, Bouaziz S, Izadi S and Tagliasacchi A. 2018. Real-time compression and streaming of 4D performances. ACM Transactions on Graphics, 37(6): #256［DOI: 10.1145/3272127.3275096］

Tang D H, Singh S, Chou P A, Häne C, Dou M S, Fanello S, Taylor J, Davidson P, Guleryuz O G, Zhang Y D, Izadi S, Tagliasacchi A, Bouaziz S and Keskin C. 2020. Deep implicit volume compression//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1290-1300［DOI: 10.1109/CVPR42600.2020.00137］

Tanimoto M. 2012. FTV: free-viewpoint television. Signal Processing: Image Communication, 27(6): 555-570［DOI: 10.1016/j.image.2012.02.016］

Tchapmi L P, Kosaraju V, Rezatofighi H, Reid I and Savarese S. 2019. TopNet: structural point cloud decoder//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 383-392［DOI: 10.1109/CVPR.2019.00047］

Tech G, Chen Y, Müller K, Ohm J R, Vetro A and Wang Y K. 2016. Overview of the multiview and 3D extensions of high efficiency video coding. IEEE Transactions on Circuits and Systems for Video Technology, 26(1): 35-49［DOI: 10.1109/TCSVT.2015.2477935］

Thatte J and Girod B. 2019. A statistical model for disocclusions in depth-based novel view synthesis//Proceedings of 2019 IEEE Visual Communications and Image Processing. Sydney, Australia: IEEE: 1-4［DOI: 10.1109/VCIP47243.2019.8966071］

Tohidypour H R, Pourazad M T and Nasiopoulos P. 2016. Online-learning-based complexity reduction scheme for 3D-HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 26(10): 1870-1883［DOI: 10.1109/TCSVT.2015.2477955］

van der Jeught S and Dirckx J J J. 2016. Real-time structured light profilometry: a review. Optics and Lasers in Engineering, 87: 18-31［DOI: 10.1016/j.optlaseng.2016.01.011］

van Duong V, Canh T N, Huu T N and Jeon B. 2019. Focal stack based light field coding for refocusing applications//Proceedings of 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Jeju, Korea (South): IEEE: 1-4［DOI: 10.1109/BMSB47279.2019.8971928］

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］

Vijayanagar K R, Kim J, Lee Y and Kim J B. 2013. Efficient view synthesis for multi-view video plus depth//Proceedings of 2013 IEEE International Conference on Image Processing. Melbourne, Australia: IEEE: 2197-2201［DOI: 10.1109/ICIP.2013.6738453］

Vizzotto B B, Zatt B, Shafique M, Bampi S and Henkel J. 2013. Model predictive hierarchical rate control with markov decision process for multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology, 23(12): 2090-2104［DOI: 10.1109/TCSVT.2013.2270400］

Wang J Q, Ding D D, Li Z, Feng X X, Cao C T and Ma Z. 2022a. Sparse tensor-based multiscale representation for point cloud geometry compression. IEEE Transactions on Pattern Analysis and Machine Intelligence: #3225816［DOI: 10.1109/TPAMI.2022.3225816］

Wang J Q, Ding D D, Li Z and Ma Z. 2021a. Multiscale point cloud geometry compression//Proceedings of 2021 Data Compression Conference. Snowbird, United States: IEEE: 73-82［DOI: 10.1109/DCC50243.2021.00015］

Wang J Q, Zhu H, Liu H J and Ma Z. 2021b. Lossy point cloud geometry compression via end-to-end learning. IEEE Transactions on Circuits and Systems for Video Technology, 31(12): 4909-4923［DOI: 10.1109/TCSVT.2021.3051377］

Wang L L, Wang H, Dai D Q, Leng J Y and Han X G. 2021c. Bidirectional shadow rendering for interactive mixed 360° videos//Proceedings of 2021 IEEE Virtual Reality and 3D User Interfaces. Lisboa, Portugal: IEEE: 170-178［DOI: 10.1109/VR50410.2021.00038］

Wang X G, Ang M H and Lee G H. 2022b. Cascaded refinement network for point cloud completion with self-supervision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11): 8139-8150［DOI: 10.1109/TPAMI.2021.3108410］

Wang Y F, Wu S H, Huang H, Cohen-Or D and Sorkine-Hornung O. 2019. Patch-based progressive 3D point set upsampling//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5951-5960［DOI: 10.1109/CVPR.2019.00611］

Wang Z Y, Hu J H, Wang S Z and Lu T. 2015. Trilateral constrained sparse representation for Kinect depth hole filling. Pattern Recognition Letters, 65: 95-102［DOI: 10.1016/j.patrec.2015.07.025］

Wegner K, Stankiewicz O and Domański M. 2016. Novel depth-based

blending technique for improved virtual view synthesis//Proceedings of 2016 International Conference on Signals and Electronic Systems. Krakow, Poland: IEEE: 93-98 [DOI: 10.1109/ICSES.2016.7593828]

Wen X, Li T Y, Han Z Z and Liu Y S. 2020a. Point cloud completion by skip-attention network with hierarchical folding//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1939-1948 [DOI: 10.1109/CVPR42600.2020.00201]

Wen X, Xiang P, Han Z Z, Cao Y P, Wan P F, Zheng W and Liu Y S. 2023. PMP-Net++: point cloud completion by transformer-enhanced multi-step point moving paths. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1): 852-867 [DOI: 10.1109/TPAMI.2022.3159003]

Wen X Z, Wang X, Hou J H, Ma L, Zhou Y and Jiang J M. 2020b. Lossy geometry compression of 3D point cloud data via an adaptive octree-guided network//Proceedings of 2020 IEEE International Conference on Multimedia and Expo. London, UK: IEEE: 1-6 [DOI: 10.1109/ICME46284.2020.9102866]

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]

Wien M, Boyce J M, Stockhammer T and Peng W H. 2019. Standardization status of immersive video coding. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9(1): 5-17 [DOI: 10.1109/JETCAS.2019.2898948]

Wiesmann L, Milioto A, Chen X Y L, Stachniss C and Behley J. 2021. Deep compression for dense point cloud maps. IEEE Robotics and Automation Letters, 6(2): 2060-2067 [DOI: 10.1109/LRA.2021.3059633]

Wu C L, Zhang R X, Wang Z and Sun L F. 2020a. A spherical convolution approach for learning long term viewport prediction in 360 immersive video//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, United States: AAAI: 14003-14010 [DOI: 10.1609/aaai.v34i01.7377]

Wu K, 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]

Wu K, Yang Y, Liu Q and Zhang X. 2022. Focal stack image compression based on basis-quadtree representation. IEEE Transactions on Multimedia: #3169055 [DOI: 10.1109/TMM.2022.3169055]

Wu S H, Huang H, Gong M L, Zwicker M and Cohen-Or D. 2015. Deep points consolidation. ACM Transactions on Graphics, 34(6): #176 [DOI: 10.1145/2816795.2818073]

Wu T, Pan L, Zhang J Z, Wang T, Liu Z W and Lin D H. 2021. Density-aware chamfer distance as a comprehensive metric for point cloud completion. [EB/OL]. [2023-01-14]. https://arxiv.org/pdf/2111.12702.pdf

Xiang P, Wen X, Liu Y S, Cao Y P, Wan P F, Zheng W and Han Z Z. 2021. SnowflakeNet: point cloud completion by snowflake point deconvolution with skip-transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5479-5489 [DOI: 10.1109/ICCV48922.2021.00545]

Xiang S, Yu L, Yang Y, Liu Q and Zhou J L. 2015. Interfered depth map recovery with texture guidance for multiple structured light depth cameras. Signal Processing: Image Communication, 31: 34-46 [DOI: 10.1016/j.image.2014.11.004]

Xie J, Feris R S, Yu S S and Sun M T. 2015. Joint super resolution and denoising from a single depth image. IEEE Transactions on Multimedia, 17(9): 1525-1537 [DOI: 10.1109/TMM.2015.2457678]

Xu Y W, Xing K Y, Liu H, Zhao T S and Kwong S. 2021. Flexible complexity optimization in multiview video coding. IEEE Transactions on Circuits and Systems for Video Technology, 31(10): 4096-4106 [DOI: 10.1109/TCSVT.2020.3043005]

Xu Z M, Zhang X G, Zhang K and Guo Z M. 2018. Probabilistic viewport adaptive streaming for 360-degree videos//Proceedings of 2018 IEEE International Symposium on Circuits and Systems. Florence, Italy: IEEE: 1-5 [DOI: 10.1109/ISCAS.2018.8351404]

Yan Z Q, Yu L, Yang Y and Liu Q. 2014. Beyond the interference problem: hierarchical patterns for multiple-projector structured light system. Applied Optics, 53(17): 3621-3632 [DOI: 10.1364/AO.53.003621]

Yang M and Zheng N N. 2019. SynBF: a new bilateral filter for postremoval of noise from synthesis views in 3-D video. IEEE Transactions on Multimedia, 21(1): 15-28 [DOI: 10.1109/tmm.2018.2849605]

Yang Y, Deng H P, Wu J and Yu L. 2015a. Depth map reconstruction and rectification through coding parameters for mobile 3D video system. Neurocomputing, 151: 663-673 [DOI: 10.1016/j.neucom.2014.04.088]

Yang Y, Liu Q, He X and Liu Z. 2019. Cross-view multi-lateral filter for compressed multi-view depth video. IEEE Transactions on Image Processing, 28(1): 302-315 [DOI: 10.1109/TIP.2018.2867740]

Yang Y, Liu Q, Ji R R and Gao Y. 2012. Dynamic 3D scene depth reconstruction via optical flow field rectification. PLoS One, 7(11): #47041 [DOI: 10.1371/journal.pone.0047041]

Yang Y, Wang X, Liu Q, Xu M L and Yu L. 2015c. A bundled-optimization model of multiview dense depth map synthesis for dynamic scene reconstruction. Information Sciences, 320: 306-319 [DOI: 10.1016/j.ins.2014.11.014]

Yao C, Tillo T, Zhao Y, Xiao J M, Bai H H and Lin C Y. 2014. Depth map driven hole filling algorithm exploiting temporal correlation information. IEEE Transactions on Broadcasting, 60(2): 394-404 [DOI: 10.1109/TBC.2014.2321671]

Yaqoob A, Bi T and Muntean G M. 2020. A survey on adaptive 360° video streaming: solutions, challenges and opportunities. IEEE

Communications Surveys and Tutorials, 22(4): 2801-2838 [DOI: 10.1109/COMST.2020.3006999]

Ye S Q, Chen D D, Han S F, Wan Z Y and Liao J. 2022. Meta-PU: an arbitrary-scale upsampling network for point cloud. IEEE Transactions on Visualization and Computer Graphics, 28(9): 3206-3218 [DOI: 10.1109/TVCG.2021.3058311]

Yeh C H, Li M F, Chen M J, Chi M C, Huang X X and Chi H W. 2014. Fast mode decision algorithm through inter-view rate-distortion prediction for multiview video coding system. IEEE Transactions on Industrial Informatics, 10(1): 594-603 [DOI: 10.1109/TII.2013.2273308]

Yin K X, Huang H, Zhang H, Gong M L, Cohen-Or D and Chen B Q. 2014. Morfit: interactive surface reconstruction from incomplete point clouds with curve-driven topology and geometry control. ACM Transactions on Graphics, 33(6): #202 [DOI: 10.1145/2661229.2661241]

Yu L Q, Li X Z, Fu C W, Cohen-Or D and Heng P A. 2018a. EC-Net: an edge-aware point set consolidation network//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 398-414 [DOI: 10.1007/978-3-030-01234-2_24]

Yu L Q, Li X Z, Fu C W, Cohen-Or D and Heng P A. 2018b. PU-Net: point cloud upsampling network//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2790-2799 [DOI: 10.1109/CVPR.2018.00295]

Yu X M, Rao Y M, Wang Z Y, Liu Z Y, Lu J W and Zhou J. 2021. PoinTr: diverse point cloud completion with geometry-aware transformers//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 12478-12487 [DOI: 10.1109/ICCV48922.2021.01227]

Yuan H, Kwong S, Ge C, Wang X and Zhang Y. 2014. Interview rate distortion analysis-based coarse to fine bit allocation algorithm for 3-D video coding. IEEE Transactions on Broadcasting, 60(4): 614-625[DOI: 10.1109/TBC.2014.2361964]

Yuan H, Kwong S, Wang X, Gao W and Zhang Y. 2015. Rate distortion optimized inter-view frame level bit allocation method for MV-HEVC. IEEE Transactions on Multimedia, 17(12): 2134-2146 [DOI: 10.1109/TMM.2015.2477682]

Yuan H, Liu J, Xu H, Li Z and Liu W. 2012. Coding distortion elimination of virtual view synthesis for 3D video system: theoretical analyses and implementation. IEEE Transactions on Broadcasting, 58(4): 558-567 [DOI: 10.1109/TBC.2012.2187612]

Yuan H, Zhao S, Hou J, Wei X and Kwong S. 2020. Spatial and temporal consistency-aware dynamic adaptive streaming for 360-degree videos. IEEE Journal of Selected Topics in Signal Processing, 14(1): 177-193 [DOI: 10.1109/JSTSP.2019.2957981]

Yuan W T, Khot T, Held D, Mertz C and Hebert M. 2018. PCN: point completion network//Proceedings of 2018 International Conference on 3D Vision. Verona, Italy: IEEE: 728-737 [DOI: 10.1109/3DV.2018.00088]

Zeng H Q, Ma K K and Cai C H. 2011. Fast mode decision for multiview video coding using mode correlation. IEEE Transactions on Circuits and Systems for Video Technology, 21(11): 1659-1666 [DOI: 10.1109/TCSVT.2011.2133350]

Zhang H B, Fu C H, Chen R L, Xiao Y Z and Su W M. 2016. Fast intra coding for depth map in 3D-HEVC. Journal of Image and Graphics, 21(7): 845-853 (张洪彬, 伏长虹, 陈锐霖, 萧允治, 苏卫民. 2016. 3D-HEVC深度图像快速帧内编码方法. 中国图象图形学报, 21(7): 845-853) [DOI: 10.11834/jig.20160702]

Zhang H B, Fu C H, Chan Y L, Tsang S H and Siu W C. 2018. Probability-based depth intra-mode skipping strategy and novel VSO metric for DMM decision in 3D-HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 28(2): 513-527 [DOI: 10.1109/TCSVT.2016.2612693]

Zhang P P, Wang X, Ma L, Wang S Q, Kwong S and Jiang J M. 2021a. Progressive point cloud upsampling via differentiable rendering. IEEE Transactions on Circuits and Systems for Video Technology, 31(12): 4673-4685 [DOI: 10.1109/TCSVT.2021.3100134]

Zhang X, Cheung G, Zhao Y, Le Callet P, Lin C Y and Tan J Z G. 2021b. Graph learning based head movement prediction for interactive 360 video streaming. IEEE Transactions on Image Processing, 30: 4622-4636 [DOI: 10.1109/TIP.2021.3073283]

Zhang Y, Kwong S, Xu L, Hu S D, Jiang G Y and Kuo C C J. 2013a. Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model. IEEE Transactions on Image Processing, 22(9): 3497-3512 [DOI: 10.1109/TIP.2013.2265883]

Zhang Y, Kwong S, Xu L and Jiang G Y. 2013b. DIRECT mode early decision optimization based on rate distortion cost property and inter-view correlation. IEEE Transactions on Broadcasting, 59(2): 390-398 [DOI: 10.1109/TBC.2013.2253033]

Zhang Z Y. 2012. Microsoft Kinect sensor and its effect. IEEE Multimedia, 19(2): 4-10 [DOI: 10.1109/MMUL.2012.24]

Zhao W B, Liu X M, Zhong Z W, Jiang J J, Gao W, Li G and Ji X Y. 2022. Self-supervised arbitrary-scale point clouds upsampling via implicit neural representation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 1989-1997 [DOI: 10.1109/CVPR52688.2022.00204]

Zhu J, Zhang J, Cao Y and Wang Z F. 2017. Image guided depth enhancement via deep fusion and local linear regularizaron//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE: 4068-4072 [DOI: 10.1109/ICIP.2017.8297047]

Zhu T T and Gao P. 2019. An improved Gaussian mixture model based hole-filling algorithm exploiting depth information//Proceedings of 2019 IEEE Visual Communications and Image Processing. Sydney,

Australia: IEEE: #8965964 [DOI: 10.1109/VCIP47243.2019.8965964]

作者简介

王旭,男,副教授,博士生导师,主要研究方向为多媒体通信、沉浸式媒体智能编码与处理、3D视觉。
E-mail: wangxu@szu.edu.cn
杨铀,通信作者,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
刘琼,女,教授,主要研究方向为多媒体通信、3D视频处理和理解、3D计算机视觉。E-mail: q.liu@hust.edu.cn
彭宗举,男,教授,博士生导师,主要研究方向为多媒体通信、视频压缩与传输、3D视频系统、视觉计算与应用。
E-mail: pengzongju1@cqut.edu.cn
侯军辉,男,助理教授,博士生导师,主要研究方向为多媒体通信、图像/视频表示学习、3D几何表示学习、视频数据压缩与传输。E-mail: jh.hou@cityu.edu.hk
元辉,男,教授,博士生导师,主要研究方向为沉浸式媒体智能编码与处理、流媒体传输控制、视觉智能。
E-mail: huiyuan@sdu.edu.cn
赵铁松,男,教授,博士生导师,主要研究方向为图像处理和计算机视觉、视觉编码和通信、触觉信息处理。
E-mail: t.zhao@fzu.edu.cn
秦熠,男,博士,主要研究方向为多媒体无线通信、多媒体体验评估、3D视频处理和理解、3D计算机视觉。
E-mail: qinyi4@huawei.com
吴科君,男,博士后,主要研究方向为光场成像、多视点多焦点图像编码、图像复原。E-mail: kejun.wu@ntu.edu.sg
刘文予,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: liuwy@hust.edu.cn