

虚拟现实视频传输架构和关键技术

Architecture and Key Technologies of Virtual Reality Video Transmission and Streaming

郭宗明/GUO Zongming
班怡璇/BAN Yixuan
谢澜/XIE Lan

(北京大学, 北京 100871)
(Peking University, Beijing 100871, China)

虚拟现实(VR)视频, 又称全景视频或360°视频, 以其充分的沉浸性、真实性、制作简单性等特点正受到人们的广泛关注。它通过全景摄像机记录三自由度或六自由度的自然图像, 降低了VR内容制作的门槛, 提供现有普通视频所不能达到的真实感。近几年, 随着VR设备、网络和终端技术的成熟, VR视频正越来越多地用于影视、游戏、医疗、教育等应用中。据思科的统计报告显示^[1]: VR应用在2016年的平均月流量达到13.3 PB, 预计到2021年达到140 PB, 年增长率达到60%。如此广泛的VR视频给人们的生活和工作带来了新体验, 大量的网络基础设施和智能穿戴设备也推动了VR视频应用的发展。

然而, 高质量的VR视频服务仍面临若干挑战。其中, 网络传输是一个关键性问题。VR视频对网络带宽有较高的消耗, 例如: 一个分辨率为2 K×1 K的虚拟现实视频经过压缩编码, 需要4 Mbit/s的数据量; 若分辨率达到4 K×2 K, 则数据量将会上涨至

收稿日期: 2017-09-26

网络出版日期: 2017-11-09

基金项目: 国家自然科学基金(61471009)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 06-0019-005

摘要: 介绍了虚拟现实(VR)视频传输架构, 提出了两种传输方式: 全景传输和视点传输。全景传输可提供多角度质量均匀的图像, 但存在高码率和带宽浪费问题; 视点传输仅传输视点范围内图像, 可减少码率浪费, 但存在头部运动延迟问题。提出的视点自适应传输技术、视点预测方法, 可以较好地实现低码率、高质量的VR视频传输。

关键词: VR视频; 视点自适应传输; 非均匀映射传输; 分块传输; 视点预测; 用户质量优化

Abstract: In this paper, a main framework of virtual reality (VR) video transmission is introduced, and two schemes are also proposed: panorama transmission and viewport transmission. The panorama transmission can provide 360° symmetric image with high bit-rate and bandwidth wasting; the viewport transmission only transmits a part of images so as to reduce bit-rate wasting, but can cause head movement delay. In this paper, viewport-adaptive streaming and viewport prediction technologies are introduced, which enable high quality service of VR video streaming.

Key words: VR video; viewport adaptation streaming; asymmetric projection streaming; tile-based streaming; viewport prediction; QoE optimization

20 Mbit/s, 至少需要25 Mbit/s的网络带宽才可以承载^[2]。现有互联网带宽难以满足如此高码率虚拟现实视频的流畅播放, 进而严重影响其应用。

为了解决VR视频的传输瓶颈, 工业界和学术界提出了众多解决方案。早期, Google、YouTube采用可兼容现有视频编解码器的全景传输模式, 但存在码率高、视频变形等问题。2015年, Facebook提出了视点自适应传输和金字塔映射模式, 可根据用户视点区域传输视频内容, 减少总体码率。2016年, 三星、高通等公司提出了多种非均匀映射方式, 动态图像专家组(MPEG)也专门成立一个I-AHG沉浸式视频专题组, 制订全景视频文件格式、映射、传输标准。在中

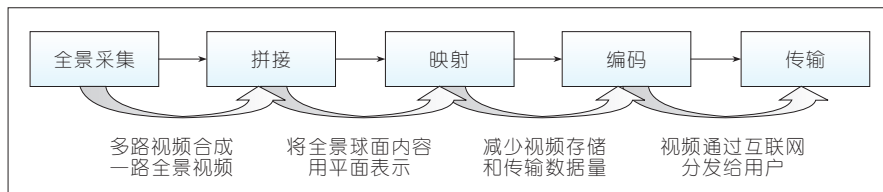
国, 数字音视频编解码技术标准工作组(AVS)标准化组织也于2016年开始制订AVS虚拟视频编码和系统标准, 中兴通讯、北京大学、上海交通大学等单位相继提出多种映射和传输方案。

文中, 我们将详细介绍VR视频传输的发展现状和关键技术, 重点从传输框架、全景传输、自适应传输等方面展开。

1 虚拟现实视频传输框架

完整的虚拟现实视频传输架构包括5个部分: 全景采集、拼接、映射、编码、传输, 如图1所示。

(1) 全景采集。VR视频是由多摄像头组合成的全景相机拍摄而



▲图1 虚拟现实视频传输架构

成。利用全景摄像机采集自然图像,可以极大简化虚拟现实内容的制作。近几年,全景相机硬件技术得到飞速发展。从早期由多个普通摄像头合成的全景相机,发展到各种专用相机,例如:Facebook Surround 360、三星 Gear 360等。可以提供从视频采集到拼接等一系列功能,方便虚拟现实视频的创作。

(2)视频拼接。视频拼接是全景视频采集的后处理过程,它将若干个摄像头的视频合成一路全景视频,形成一个完整的虚拟现实视频提供给用户。目前已经有多种商业化或开源的视频拼接软件,例如:Video Stitch、Facebook Surround 360 SDK、Nuke等。

(3)映射过程。为了便于存储和压缩编码,球面视频内容需要被几何映射到平面。映射影响了VR视频编码前的像素量,在一定程度上也决定了视频内容所包含的信息量。映射过程与传输紧密相关,分为两类:均匀映射和非均匀映射。均匀映射保留全部全景视频信息;非均匀映射仅在部分区域保留完整信息,其他区域保留部分信息,并用较少像素表示。

(4)视频编码。VR视频使用压缩编码减少视频中的冗余信息。为了兼容现有编码器和播放设备,目前虚拟现实视频主要采用H.264,高效视频编码(HEVC)等编码标准。

(5)网络传输。VR视频需要通过互联网分发给用户,网络传输是目前最大的挑战。为了提供清晰、流畅的虚拟现实视频体验,映射和传输方法起到了关键性作用。传输方法主要分为两类:

- 全景传输方法。将360°全景

视频以同等质量、完整的发送给用户。可以保证映射内容完整保留了原始球面的所有内容,保留信息量最大;但由于播放设备和视角的限制,在某一时刻,用户只关心视角范围内的部分内容,全景传输方法势必会造成巨大的资源浪费。

- 视点自适应传输方法。顾名思义,指的是客户端可根据用户视点,动态传输视角范围内的视频,避免了资源浪费。然而,若用户当前视点与下载视频的视点不符,会出现黑屏或视频质量降低等问题。

在虚拟现实视频传输框架中,映射与传输方式紧密相关。其中,全景传输对应均匀映射,完整的映射和传输360°视频;视点自适应传输对应非均匀映射和分块传输,按用户视角映射和传输局部视频。

下面将详细介绍这两种虚拟视频传输方式,以及各自优缺点。

2 全景传输

全景传输方法是将球面视频均

匀的映射到平面上,按照矩形布局编码和传输完整的360°视频内容。全景传输可直接使用现有编码器和传输方式,因为其简单和通用而成为当下虚拟现实视频使用最广泛的传输方法。

均匀映射是全景传输的主要映射方式,决定了编码效率和带宽消耗。代表性的均匀映射方法有矩形映射(ERP)、立方体映射(CMP)、八面体映射等,如图2所示。

ERP方法是使用最多的虚拟现实视频映射方法,这种方法在球的纬度上按照赤道的采样频率对内容进行采样并用平面进行表示,生成的虚拟现实视频的宽高比为2:1。但ERP方法在南北两极存在着严重的过采样现象,使得整体的冗余像素增加。

立方体映射是将球面内容投影到外接的一个立方体的6个面上,并将6个面拼接重组成一个完整的虚拟现实视频。

除此之外,更多的多面体被应用于VR视频的映射方法中,例如:正八面体(OHP)、正二十面体(ISP)等。理论上,使用面数越多的正多面体,可以使得映射后的数据量越小。但是,这些正多面体都需要通过“再拼接”形成矩形平面。拼接过程导致内容上的不连续现象,在一定程度上也会影响编码效率。

图2 ▶
矩形、六面体等均匀映射方法示例

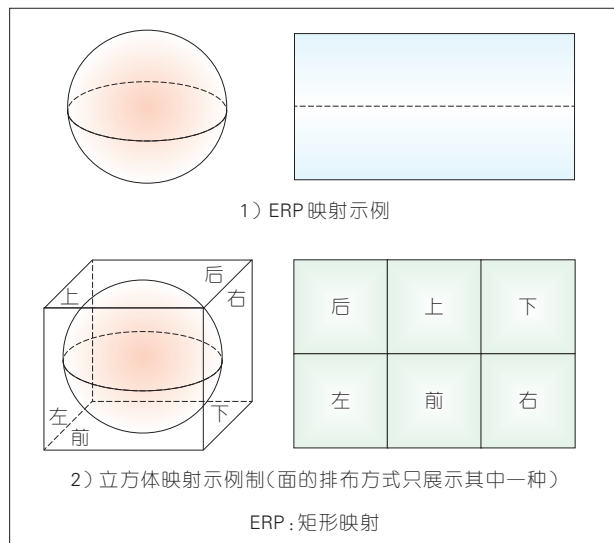


表1总结了目前主要的均匀映射方式的优缺点。其中,像素比例代表映射后面积和原始球面的比例关系,

▼表1 均匀映射方法比较

映射方法	像素比例	优点	缺点
矩形映射	1.57	映射方法简单	南北两极存在严重过采样
立方体映射	1.91	每个面失真小	6个面拼接存在不连续现象,影响编码效率
八面体映射	1.65	映射面积小	多面拼接存在不连续现象,影响编码效率
二十面体映射	1.21	映射面积小	多面拼接存在不连续现象,影响编码效率
柱状映射	1.20	映射面积小	存在空白区域

是影响视频编码码率的一个因素。

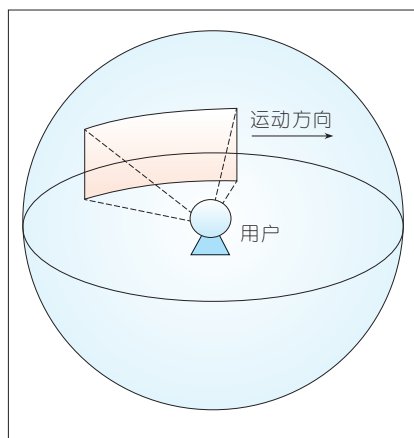
3 视点自适应传输

视点自适应传输技术是一种根据用户视点,按需下载视频内容的方法。如图3所示,对于用户视角内的内容可以选择下载高质量的版本,对于视角外的部分可下载低质量的版本,或者不进行下载。

视点自适应传输可有效减少虚拟现实视频的带宽浪费问题。在相同带宽条件下,它可分配更多带宽给视角区域,从而提高用户视窗内视频的质量,增强用户的观看体验。

按照传输内容的构成形式,视点自适应方法可以分为两类:非均匀映射和分块传输。

(1)非均匀映射^[3]。传输一个质量不均匀的360°全景视频,用户视点范围内是高分辨率,其他区域是低分辨率,从而减少整体码率。



▲图3 视点自适应技术原理

(2)分块传输^[4]。将全景视频从空间上分为不同视频块,根据视点范围传输部分视频块。

两种的主要差别为:非均匀映射需传输360°的全景视频,可防止头部快速运动导致的黑场;分块方法可传输局部或全部视频内容,灵活性高。

视点预测是视点自适应传输的另一个重要技术。VR视频要求显示延迟低于20ms^[5]。同时,由于存在网络传输延迟,在头部快速运动时,要避免视野范围内出现黑场,必须对视点区域进行预测和预取。因此视点预测的准确度极大影响虚拟现实视频的传输质量。

下面将详细介绍视点自适应的非均匀映射、分块传输和视点预测等关键技术。

3.1 非均匀映射传输技术

非均匀映射传输技术是一种通过将原始球面视频进行非均匀映射处理后进行传输的技术。其在对球面内容进行采样时,令球面上的像素点有不同的权重,使得关键视频内容得到保留,而不重要的区域被下采,仅保留少部分关键信息,如图4所

示。其优点是在保证关键区域质量的同时,从整体上降低视频的大小,极大地减轻了网络传输和客户端解码的压力。

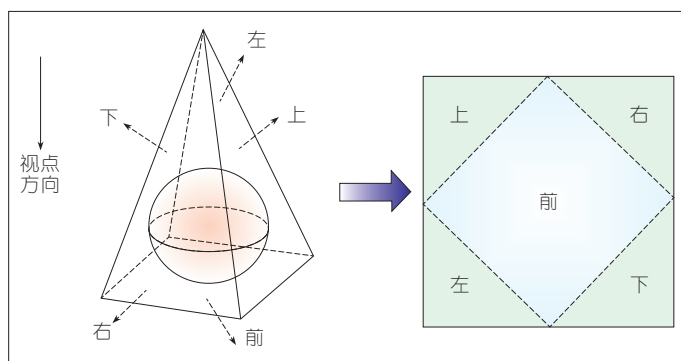
目前已有多种非均匀映射传输方式,例如:Facebook提出的多分辨率立方体、金字塔映射、偏移立方体等,高通提出的阶段金字塔等。它们共同目标是在保持视点区域质量的前提下,降低整体分辨率,但不同映射方式存在变形、失真或边界等问题,对编码效率有不同影响。

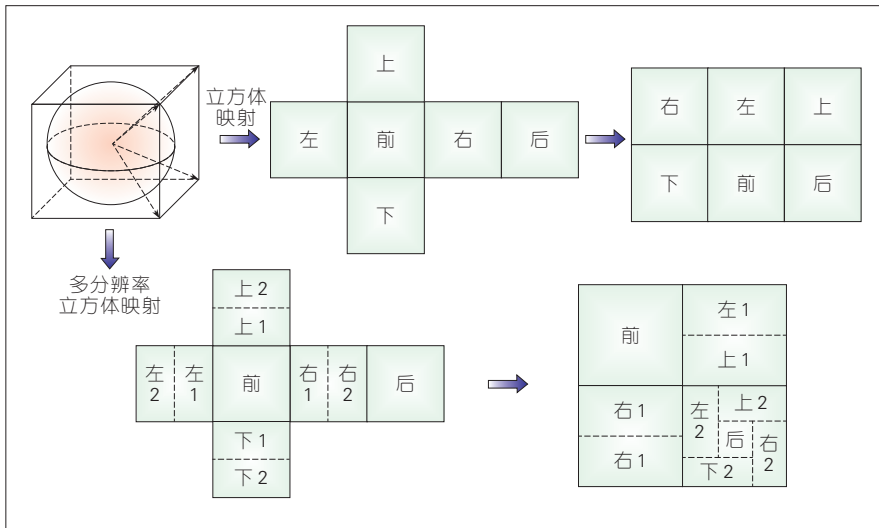
Facebook最早提出了一种多分辨率立方体映射方法,它是通过将立方体映射方式进行变换后得到的。保持正前面和与其相邻的上、下、左、右4个面的一半分辨率不变,对这4个面剩下的一半和后面的视频内容进行下采样,最终重新组合成一个二维平面视频帧,如图5所示。这种方法的缺陷是在上、下、左、右面的中间部位由于采样率突变,存在明显的边界,影响用户的观看体验。

除Facebook外,高通曾提出一种截断的金字塔映射方法^[6],并被MPEG OMAF标准采纳。这种方法与Facebook的金字塔映射方法类似,只不过非关注区域投影至四棱台的顶部,而不是投影为一个像素点,如图6所示。这种方法相较于金字塔方法,在一定程度上消除了由于映射造成的斜边效应,在内容表示上更加连续,有利于视频编码。

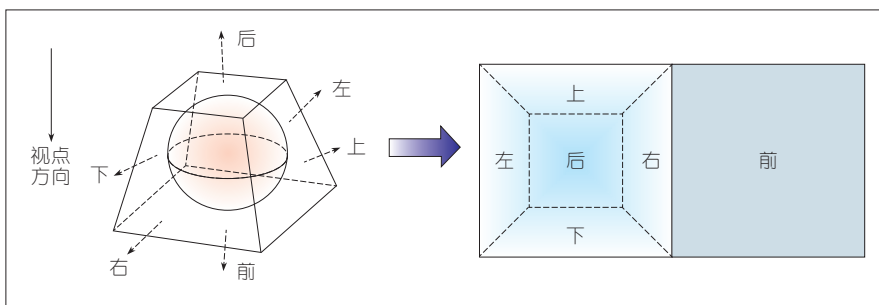
为了支持非均匀映射传输,需要针对每一个主视点都预先生成不同的视频版本。一旦用户发起请求,就

►图4 非均匀映射的原理





▲图5 多分辨率立方体映射



▲图6 截断金字塔映射

立即将对应视点的版本提取出来进行传输,在保证视频质量的同时减少了网络流量。

3.2 分块传输技术

分块传输技术是另一种视点自适应传输方式。它将全景视频按照空间划分为若干个子视频块,客户端可以根据网络状况和用户头部运动有针对性的向服务器端请求视频片段,具体过程如图7所示。

与非均匀映射传输方式不同,分

块传输仅传一部分内容,进一步减少了传输数据量。它可以自由地选择各个分块的质量,增加了传输的灵活性。同时,最新的HEVC编码标准支持tile方式编码,可用一个编解码器对整个视频流进行编解码,大大减少了客户端的解码复杂度。所以,分块传输是目前最实用的全景视频传输技术之一。

分块传输可有多种切分方式,例如:六面几何式^[7]、18格矩形分割方式^[8]、12格矩形分割方式^[8]等。不同

的切分方式将会影响编码效率、传输块数量和边界失真等。

图8以18格和12格矩形分割方式为例。18格划分是将水平方向按照60°间隔分成6列,垂直方向按照45°-90°-45°方式分3行,共计18个视频块。12格划分是水平方向按照90°间隔分成四列,垂直方向同上。

值得注意的是:在全景视频传输过程中,切分粒度越小,视频块的组合就越贴近用户视角,浪费的视频内容也就越少,也即自适应传输技术就越灵活。但是,一味减小视频块大小并不会使视频数据量有持续显著性下降。这是由于在切分后像素之间的相关性被破坏,使得切分后编码数据量之和略大于切分前的数据量,即便是在高压缩率的HEVC标准下,18格划分方式都会招致5.82%的压缩损失^[9],在极端条件下,浪费的视频内容大小甚至不足以抵消分块带来的损失。所以,如何决定分块大小也是一个值得关注的问题。

3.3 视点预测方法

视点预测是视点自适应传输的关键技术之一。由于网络传输存在延迟,为保证视点自适应播放的连续性,减少头部运动带来的卡顿甚至黑屏,如何准确预测视角位置是一个较大的挑战。

例如:线性预测方法中^[9],预测用户未来1s的头部位置时,准确率可达到90%以上,而将预测时间延长至3s时,准确率会大幅降至70%。一旦预测错误,由于本地缓存中缺少相应视角的视频片段,会导致用户视窗中出现黑窗,极大地影响用户的观看

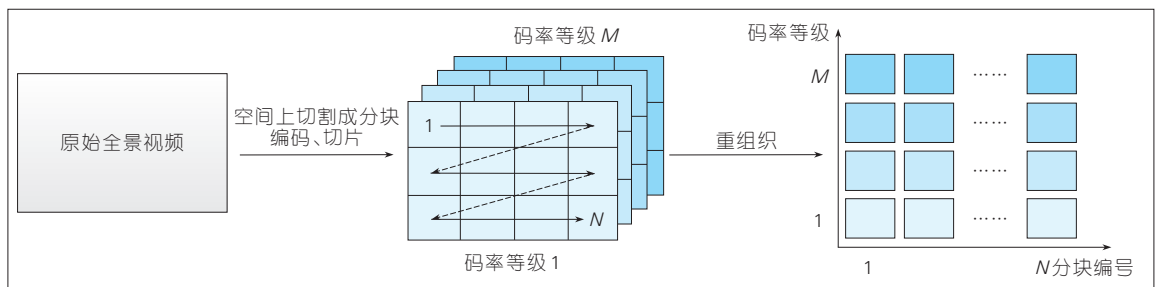


图7 分块传输示意

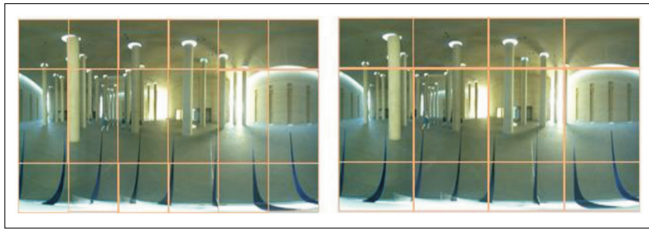


图8
18格/12格划分示意

体验。

目前视点预测方式主要分为两类:运动预测和内容分析方法。运动预测是根据用户的历史浏览行为,预测未来视角位置,包括均值算法^[9]、线性回归^[9]、概率统计^[10]、运动估计^[11]、行为学习^[12]等方法。

其中,线性回归算法尤为重要,它是一种利用数理统计中的回归分析来预测下一时刻用户位置的算法。具体来说,它先将用户在滑动窗口内不同时刻的运动状况进行记录,再通过记录下的头部数据来训练回归模型,得到回归曲线,预测下一时刻观看概率最大位置,如图9所示。

另外一类是内容分析方法。它是一种是基于视频中各个对象的显著性特征来进行预测。显著性特征代表着物体吸引用户的程度,显著性越强,表示用户越关注,同时也代表用户观看概率越高,显著性越弱,则代表观看概率越低。目前,学术界主要是通过机器学习的方法得到视频的显著性特征^[13],再进行预测。另外一种内容分析方法^[13]是同时基于内容以及用户位置进行预测的神经网络。其中,视频内容包括显著性特征

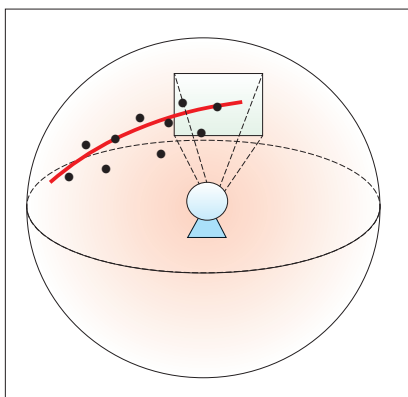


图9 线性回归预测视点

和运动特征。它需要在视频上架前对视频内容和测试头部数据集进行训练,学习出用户最可能出现的行为模式,进而在用户实际观看时根据学习结果预测下一时刻用户头部位置。

4 结束语

VR视频正得到越来越广泛的应用,但由于带宽消耗巨大,传输瓶颈已成为它面临的重大挑战之一。文章中,我们系统地介绍了VR视频传输框架,以及关键技术的发展现状,详细比较各种技术的优缺点,总结VR视频传输的发展方向。随着VR视频传输技术的发展,在有限带宽上传输高质量VR视频将成为现实。

参考文献

- [1] Cisco Visual Networking Index: Globe Mobile Data Traffic Forecast Update[R]. Technical report, 2017
- [2] Internet Connection Speed Recommendations[EB/OL]. [2017-10-12]. <https://help.netflix.com/en/node/306>
- [3] XU Z M, XIE L, ZHANG X G, et al. Optimal Viewport Adaptive Streaming for 360-Degree Videos[C]// IEEE International Conference on Multimedia&Expo(ICME '2017). USA:IEEE, 2017. DOI:10.1109/ICC.2017.7996611
- [4] BAN Y X, XIE L, ZHANG X G, et al. An Optimal Spatial-Temporal Smoothness Approach for Tile-Based 360-Degree Video Streaming[C]//IEEE Conference on Visual Communications and Image Processing (VCIP). USA:IEEE, 2017
- [5] YAO R, HEATH T, DAVIES A, et al. Oculus VR Best Practices Guide[EB/OL]. (2014-03-17)[2017-10-11]. <http://brianschrank.com/vrgames/resources/OculusBestPractices.pdf>
- [6] VR/360 Video Truncated Square Pyramid Geometry for OMAF: ISO/IEC JTC1/SC29/WG11/M[S]. US: MPEG, 2016
- [7] HOSSEINI, MOHAMMAD, and SWAMINATHAN V. Adaptive 360 VR Video Streaming: Divide and Conquer[C]//2016 IEEE International Symposium on Multimedia (ISM). USA:IEEE, 2016
- [8] ZARE A, AMINLOU A, HANNUKSELA M, and GABBOUJ M. HEVC-Compliant Tile-Based Streaming of Panoramic Video for Virtual

Reality Applications[C]// the ACM International Conference on Multimedia(ACM MM). USA:ACM,2016

- [9] QIAN F, JI L, HAN B, et al. Optimizing 360 Video Delivery over Cellular Networks[C]// 5th ACM Workshop on All Things Cellular, 2016. USA:ACM, 2016:1-6. DOI: 10.1145/2980055.2980056
- [10] XIE L, XU Z, ZHANG X, et al. 360ProbDASH: Improving QoE of 360 Video Streaming using Tile-based HTTP Adaptive Streaming[C]//ACM International Conference on Multimedia(ACM MM). USA: ACM, 2017
- [11] BATTLE L, CHANG R, and STONEBRAKER M. Dynamic Prefetching of Data Tiles for Interactive Visualization[C]//the ACM International Conference on Management of Data, 2016
- [12] LO W-C, FAN C L, LEE J, HUANG C Y, et al. 360° Video Viewing Dataset in Head-Mounted Virtual Reality[C]// 8th ACM on Multimedia Systems Conference. USA: ACM, 2017
- [13] FAN C L, LEE J, LO W-C, et al. Fixation Prediction for 360° Video Streaming in Head-Mounted Virtual Reality[C]//the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video. USA:ACM, 2017:67-72. DOI: 10.1145/3083165.3083180

作者简介



郭宗明, 北京大学博导, 北京大学计算机科学技术研究所所长; 主要研究方向为数字视频处理及传输、彩色图像处理; 曾主持、领导研发团队攻克多项数字视频处理技术与工程难关, 研制成功多项工程应用系统并获得广泛应用, 促进了中国电视行业从模拟播出到数字播出的技术变革, 促进了中国网络视频行业的发展和普及; 获中国高等学校十大科技进展、国家科技进步二等奖、教育部科技进步一等奖和二等奖、教育部技术发明奖二等奖、广电总局科技创新二等奖等; 发表论文150余篇, 申请专利120余项。



班怡璇, 北京大学计算机科学技术研究所在读本科生; 研究方向为网络视频传输、HTTP动态自适应流媒体; 发表论文1篇。



谢澜, 北京大学计算机科学技术研究所在读硕士生; 研究方向为HTTP动态流传输、全景视频视点自适应传输; 已发表学术论文4篇, 拥有专利2项、标准提案3项。