

doi:10.13756/j.gtxyj.2020.03.006

光通信系统与网络技术

面向教育的多视点视频结合 CDN 解决方案

郭湘南, 王功乾, 田 路

(烽火通信科技股份有限公司, 武汉 430074)

摘要:多视点视频是由多个相机同时对同一场景拍摄形成视频的三维化技术。针对教育行业要求数字化转型的问题,文章提出将多角度呈现教学过程的多视点视频技术和内容分发网络(CDN)流媒体传输技术结合应用于远程教学。实验结果表明,面对远程教育的应用场景,多视点视频技术和 CDN 结合能够较好地解决教学中的需求和挑战,是未来的发展方向。

关键词:多视点视频;内容分发网络;平滑切换

中图分类号:TN919.8

文献标志码:A

文章编号:1005-8788(2020)03-0030-03

Multi-View Video Combined with CDN Solution for Education

GUO Xiang-nan, WANG Gong-qian, TIAN Lu

(FiberHome Telecommunication Technologies Co., Ltd., Wuhan 430074, China)

Abstract: Multi-view video is a three-dimensional technology that uses multiple cameras to shoot the same scene at the same time. Aiming at the problem of digital transformation required by the education industry, the article proposes a solution that combines multi-view video technology and Content Delivery Network (CDN) streaming media transmission technology. This solution is applied to distance learning and suitable for presenting the teaching process from multiple angles. The results show that in the face of distance education application scenarios, the combination of multi-view video technology and CDN can effectively solve the needs and challenges in teaching, which is the future development direction.

Key words: multi-view video; CDN; smooth switching

0 引 言

近年来,教育行业正持续进行数字化转型,视频媒体技术作为数字化转型的重要部分在教育领域得到广泛的应用,诸如远程教育和泛在教育等新的教育形式都依赖于视频媒体技术的发展。随着教育形式的多样化,教师和学生对视频的呈现形式提出了越来越高的要求,相较于传统视频的单视角二维表现,视频教学更加需要有三维的表现方式。

多视点视频^[1]是一种满足人们三维观看视频需求的方法,其在录制视频时,由多个布置在不同视点的相机从不同视角对同一场景进行拍摄,从而得到一组视频信号。用户在观看时,可以自由切换多视点视频中的视点来看到不同视角的内容,从而得到“环视”和“运动视差”的效果。

本文将从多视点视频在教育场景下应用的需求和挑战出发,提出结合内容分发网络(Content Delivery Network, CDN)的解决方案,并对其中的关键技术进行探讨。

1 教育场景下多视点视频应用的需求与挑战

多视点视频技术适用于实操型的远程教学场景,如:各种类型的实验、手工制作和体育教学等。为达到更好的教学效果,学生需能够对视频呈现的视角进行变换和控制,同时还需能从不同角度观看教学过程。

远程观看的多视点教学视频需要支持多路视频并发传输的实时直播、时移回看和在线点播等功能,并支持播放过程中学生通过不同的交互方式(如遥控器方向键/触摸板、鼠标拖动和手机/平板电脑点触等)对视点进行自由切换。

多视点视频的教学应用主要面临以下 3 个方面的挑战:

(1) 带宽的挑战:多视点视频使用的相机越多,相机阵列覆盖的范围就越大,用户体验也越好,而带宽需求随视点个数的增加线性增加。以 4K 视频为例,单路相机录制的视频采用高效视频编码

收稿日期:2019-12-11

作者简介:郭湘南(1974—),男,湖南宁乡人。高级工程师,硕士,主要研究方向为视讯业务及 CDN 技术。

通信作者:田路,工程师。E-mail: tianlu@fiberhome.com

(High Efficiency Video Coding, HEVC) 压缩后所需带宽约为 30 Mbit/s, 20 路相机同时传输需 600 Mbit/s 带宽, 更高的带宽将给用户带来更佳的服务体验, 但高带宽也带来传输瓶颈的问题。

(2) 时延和丢包的挑战: 用户在进行视点切换时, 通过终端发出视点切换命令, 终端需要与服务器进行信令交互, 并由服务器切换视点码流。切换命令发出到切换后新视点码流显示的时延若超过 500 ms 将带来明显的卡顿, 降低用户的使用体验。

(3) 平滑视点切换的挑战: 由于相机个数限制以及排列间距影响, 视点切换播放时会使用户有不连续的感觉。为保证切换的平滑性, 需要依据两视点的视频码流计算出其间的虚拟视点视频。但此种计算会消耗大量计算资源。

CDN 是媒体内容分发和传输的工具, 将其与多视点视频结合, 可最大限度降低内容传输对带宽的需求, 降低用户交互响应时延, 提高用户的体验。同时, 利用 CDN 的边缘计算能力, 可实现平滑视点切换的计算, 减少终端的计算压力。

2 教育场景下多视点视频端到端应用解决方案与关键技术

图 1 所示为一种教育场景下多视点视频端到端应用解决方案, 该方案包括: 视频采集、编码、传输和解码/播放等。

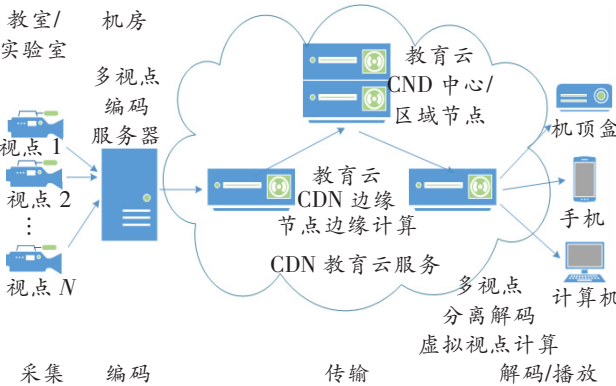


图 1 多视点视频端到端应用

2.1 采集

教学场景下多视点视频的采集需要多个相机组成相机阵列, 面对教师从多个角度进行拍摄。相机阵列的配置主要有一维平行、环形汇聚、交叉、面阵和球型等方式, 目前使用较多的是一维平行和环形汇聚阵列, 如图 2 所示。

采集模块设置要点:

(1) 相机的个数越多, 覆盖角度越大。但受限

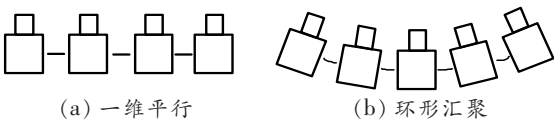


图 2 多视点相机一维平行和汇聚阵列

于传输带宽, 相机个数的选择应结合分辨率要求、压缩能力以及带宽消耗综合考虑。

(2) 相机的参数应该一致, 如分辨率、水平/垂直角度和镜头焦距/光圈/畸变等。

(3) 相机的间距一般要小于人的双眼瞳孔中心间距。间距越小, 显示的深度范围越大。

(4) 所有相机采集同步是保证显示效果及虚拟视点合成的基础。

2.2 编码

多视点相机阵列采集的原始视频占用带宽巨大。以 4K 画质、20 路相机为例, 原始视频需求带宽高达 120 Gbit/s, 不适合直接传输, 在采集的同时需要在本地进行压缩编码以降低带宽的占用。

原始视频的压缩编码可选分路与合路两种方式:

(1) 视点分路编码

对相机阵列生成的每一路原始视频各自独立进行压缩编码, 多路并行发送^[2], 当前主流的编码标准包括 HEVC 和音视频编码标准 (Audio Video coding Standard, AVS)2 等。

分路编码的优点是只需在采集视频的本地进行一次编码, 终端接收单路视频可直接解码播放; 缺点在于未消除多视点间存在的冗余信息, 存在部分带宽浪费。

需要注意的是, 如果后端存在虚拟视点计算的需求, 则要求多路编码在时间上同步。编码时应配置参数保证多路视频流在相同时间点生成 I 帧。

(2) 视点合路编码

针对同一对象进行多视点拍摄采集的视频, 多个视点间图像存在相关性, 编码时可利用其他视点已解码图片作为参考图像进行包括帧间预测以及位移估值在内的视点间预测, 从而降低编码后文件的大小。通过利用视点间的相关性, 可以将所有原始视频经多视点合路压缩编码到一个文件中。

现有主流的多视点合路编码标准包括 3D 视频编码扩展开发联合合作团队 (Joint Collaborative Team on 3D Video Coding Extension Development, JCT-3V) 工作组开发的 3D-HEVC 标准^[3] 与国内 AVS 工作组开发的 3D-AVS2 标准^[4], 二者均

利用视差技术和纹理图/深度图相关性^[5]实现多视点视频的压缩合路。

合路压缩编码的优点在于能够提高多视点编码的压缩率,缺点是消耗算力巨大。经测试,普通中央处理器编码一帧图像以分钟计^[6],且在解码端需要占用同样的计算资源对合路视频进行解码分离。

2.3 传输

编码后的视频需要使用流媒体形式封装传输,主流的传输协议包括实时流传输协议(Real Time Streaming Protocol, RTSP)、实时消息传输协议(Real Time Messaging Protocol, RTMP)、超文本传输协议(Hyper Text Transfer Protocol, HTTP)、实时流(HTTP Live Streaming, HLS)和基于 HTTP 的动态自适应流(Dynamic Adaptive Streaming over HTTP, DASH HTTP)等。表 1 所示为这几种流媒体协议的特点对比。RTSP 多用于传统交互式网络电视(Internet Protocol Television, IPTV),RTMP 和 HLS 多用于互联网视频,DASH 则是新一代的流媒体承载标准,在加载速度、扩展性和兼容性等多方面有更好的表现。

表 1 主流流媒体协议对比

协议	RTSP	RTMP	HLS	DASH
传输方式	TCP/UDP	TCP	HTTP	HTTP
视频封装格式	TS	FLV	TS	MP4/TS
实时性	高	较高	低	中
数据分段	连续流	连续流	切片文件	切片文件
倍速播放	支持	不支持	不支持	不支持
自适应码率播放	不支持	支持	支持	支持

注:TCP 为传输控制协议;UDP 为用户数据报协议;TS 为传输流;FLV 为 Flash 视频;MP4 为 MPEG-4 第 14 部分;MPEG 为动态图像专家组。

流媒体协议的选择应基于场景需求,例如直播需要较高的实时性,传输线路可靠性低时应支持自适应码率播放和快速加载需要视频切片等。同时,在多视点视频分路编码条件下,需要在流媒体协议的保留字段中定义视频帧的时间同步信息,以便在切换视点以及虚拟视点计算时对齐前后视点同一时刻的视频帧。

在教育多视点视频应用中,视频内容的传输由 CDN 教育云服务承载。采集编码的多视点视频内容首先注入到内容制作侧边缘 CDN 节点,由其发布到对应区域节点,再发布到中心节点,通过中心节点向其他区域节点推送。而用户侧边缘 CDN 节点接收到用户的直播/点播请求会向相应 CDN 区域节点请求内容。面向用户提供服务、缓存视频内容

和信令交互都由用户侧边缘 CDN 节点完成。CDN 教育云服务通过缩短服务端与用户间的网络距离和分散不同地区用户的服务请求,降低了多视点视频的传输时延。

2.4 解码/播放

在远程教育场景中,学生通过不同终端(电视机顶盒、手机、平板和计算机等)向教育云的 CDN 边缘服务节点请求多视点内容的流媒体播放,CDN 拉取多视点直播码流(直播)或从存储中读取相应的多视点视频(点播/时移回看),利用其边缘计算能力,分离出某个视点码流提供给用户。针对分路编码的码流只需将单路视点的视频码流取出,而合路编码的码流则需要 CDN 对合路视频拆分解码,再取出单路视点视频码流,会消耗较多 CDN 计算资源。

学生在观看多视点教学视频的过程中通过终端控制视点的切换,终端在信令中携带视点的偏移量和时间点发送至边缘 CDN 节点,由 CDN 进行下一个视点码流的发送。为保证切换过程的平滑性,可利用 CDN 的边缘计算能力,将两个相邻视点间的视频帧拼接裁剪出多个虚拟视点输出。虚拟视点的计算会消耗大量计算资源,可以考虑 3 种渐进式的实现:

(1) 切换视点时,由 CDN 直接提供下个视点的码流,不进行虚拟视点合成,CDN 在不同视点码流间做好时间同步即可,此方式适合对平滑切换无要求的场景。

(2) 静态画面合成虚拟视点,即在用户发出切换视点命令时,CDN 通过算法拼接截取出虚拟视点显示的画面发送到终端显示。这种方式由于仅对静态画面进行处理,是当前可行的多视点平滑切换方案。

(3) 动态画面合成虚拟视点,即在视频播放过程中根据用户命令计算动态的视点切换视频,此种方式需要将多个视点的每一帧图像取出进行虚拟视点合成计算,对 CDN 边缘计算能力要求极高,是今后改进的方向。

3 结束语

本文提出了一种多视点视频结合 CDN 应用的端到端解决方案,分析了其关键技术,并论证了解决方案的可行性。随着视频编码、网络和 CDN 技术的发展,多视点视频正逐步由实验研究转向实际应用,以其为代表的视频三维化与实际教学体验相结合,将成为未来视频演进的一个发展趋势。

(下转第 78 页)