

摘要

近年来,虚拟现实和全景视频技术有了巨大的发展,在娱乐表演、体育赛事、医疗事业、教育产业等虚拟现实视频播放时给用户以身临其境的感受,带来了越来越多的关注度的同时也实现了跨越式的大发展。然而全景视频需要 4K 以上分辨率才能获得可接受的观看质量,巨大的数据量使得其需要更大的网络传输带宽,这极大地限制了其应用推广。特别是在当前用户群体主要使用移动网络来观看视频的情况下,实际网络传输带宽相对更不稳定,由此带来的漫长的缓存时间延迟成为高质量全景视频服务的重重阻碍,难以满足正常全景视频的流畅播放,极大地降低了用户的观看体验。

本文在视点编码,视点预测和视点传输这三个方面提出了如下的创新贡献:

1. 提出了一种基于用户视点运动状况的自适应块级码率分配传输方案:本文方法将全景视频切分为多个不同视点区域的切片并组织成多层级结构,针对用户主要视点区域的块分配更多的比特数,同时在播放视频时根据当前用户视点运动状况来选择相应的视点区域切片。该视点编码方法嵌入常规编码器中,通过调整块级码率分配的方式来编码,在正常的视频编码量化过程中只需要修改块的量化参数,能够自适应的切换视点区域的范围,更好地适应当前用户视点的运动状态、更准确的匹配当前用户视点。在基于分块的系统的视点质量与本文方法一致的情况下,所提方法相比基于分块的系统能普遍节省 30%~40%的比特率。同时为了减小视点切片长度、让播放器能够更灵活的切换视点,本文系统使用了基于分组的超短切片编码方法,对于超短切片在 QP 较小时使用 1x2 矩形重组方法能平均节省至少 15%的网络传输带宽,而且在 QP 逐渐升高后重组方法的效果越来越好,对于超短切片在 QP 较大时使用 2x2 正方形重组方法能平均节省最多 30%的网络传输带宽。
2. 提出了基于深度学习等多种方法的视点预测策略:本文尝试了基于统计规律的视点预测方法、基于线性回归的视点预测方法和基于深度学习的视点预测方法这三类全景视频视点预测方法。基于统计规律的视点预测方法主要利用实验数据集里的规律,实验数据表明用户视点移动保持着一定程度的连续性,当前位置的视点与之前的位置有相当大的关系;基于线性回归的视点预测方法根据之前用户视点的运动状况做线性回归,目前尝试了使用前 1 秒、前 2 秒、前 4 秒内的运动来预测用户视点在接下来的运动状况;基于深度学习的视点预测方法主要使用基于端到端的卷积神经网络的视点预测,类似 Alex 的卷积神经网络结构,主要由 5 层卷积层以及 3 层全连接层组成。实验比较了这三种方法的视点预测偏差率和优缺点,基于深度学习的视点预测方法在视点预测偏差率上是最优的,高达 80%的概率预测角度偏差在 10 度

以内，但是由于深度学习方法的计算代价太大，智能手机这一类终端设备还不可能在较短时间内承载如此之高的计算量，综合权衡后，实际系统中使用的还是基于统计规律的视点预测方法。

3. 基于 DASH 标准传输协议，提出了扩展视点区域和视点范围的切换传输方法：本文系统放弃传统的流媒体传输方法转而采用和拓展基于 DASH 标准的传输方法，在 DASH 原有的码率和分辨率切换的功能上传输已经预测的对应视点。为了从整体上评价本文所描述系统在降低了网络传输带宽后的画面质量，本文设计了一项基于双盲原则的主观实验，由用户分别观看一组相同视频内容、分别由传统传输方法和本文所描述系统进行传输的视频，然后进行打分并最后对比传统传输方法和本文所描述系统的打分数据来评判本文所描述系统的画面质量，在平均节省 32% 的网络传输带宽的情况下，平均主观最终得分为 0.91，可见在牺牲掉小部分主观观看质量的同时节省了大量的网络传输带宽，总体效果良好。

关键词：全景视频，视点，码率分配，多尺度，自适应传输