

# 用于虚拟现实的360度视频编码技术新进展\*

艾达<sup>1</sup>, 董久军<sup>1</sup>, 林楠<sup>2</sup>, 刘颖<sup>1</sup>

(1. 西安邮电大学 公安部电子信息勘查应用技术重点实验室, 西安 710121; 2. 圣塔克拉拉大学 计算机工程系, 美国 圣塔克拉拉 95053)

**摘要:** 随着虚拟现实技术不断发展,360度视频编码开始成为研究热点。应用于虚拟现实的360度视频与传统的视频相比分辨率更高,编码数据量更大,在实际应用中面临传输带宽有限的瓶颈,编码效率问题有待解决。归纳分析了国际标准组织联合视频研究组(Joint Video Exploration Team, JVET)正在制定的360度视频编码投影变换技术及其编码优化方法,综合对比了各变换技术的编码性能。根据当前最新研究成果,对下一步研究需要解决的问题进行了探讨总结,给出未来的研究方向和思路。

**关键词:** 虚拟现实技术; 360度视频; 投影方式; 运动估计; 视频编码

**中图分类号:** TN911.73      **文献标志码:** A      **文章编号:** 1001-3695(2018)06-1606-07

**doi:**10.3969/j.issn.1001-3695.2018.06.002

## Advance of 360-degree video coding for virtual reality: a survey

Ai Da<sup>1</sup>, Dong Jiujun<sup>1</sup>, Lin Nan<sup>2</sup>, Liu Ying<sup>1</sup>

(1. Key Laboratory of Electronic Information Application Technology of Site-survey of the Ministry of Public Security, Xi'an University of Posts & Telecommunications, Xi'an 710121, China; 2. Dept. of Computer Engineering, Santa Clara University, Santa Clara 95053, USA)

**Abstract:** With the development of virtual reality technology, 360-degree video coding has become a hot topic in the current research. Compared with the traditional video, 360-degree video for virtual reality requires higher resolution and more data volume, which needs much higher coding efficiency on account of the limited bandwidth. JVET is developing future standards for 360-degree video coding. This paper classified and analyzed the latest projection formats and coding schemes of 360-degree video, and compared the coding performance of each method synthetically. On the basis of the latest research, it put forward the issues that need to be addressed and potential directions in the next research.

**Key words:** virtual reality technology; 360-degree video; projection format; motion estimation; video coding

## 0 引言

随着消费级虚拟现实(virtual reality, VR)设备正式推入市场,用户对虚拟现实技术全三维、沉浸式体验的需求不断增长,随之出现了大量的虚拟现实内容。目前虚拟现实内容以视频游戏体验为主,可分为两种类型:一种是通过Unity或Unreal等游戏引擎创建的纯虚拟场景视频;另一种是通过相机阵列拍摄自然场景得到的360度视频<sup>[1]</sup>。其中360度视频内容增长迅速,Google和Facebook不仅推出了360度视频内容分享平台,还发布了各自的360度全景相机<sup>[2]</sup>。可以预见,360度视频将会成为一个新颖的内容载体,在虚拟现实领域中占据重要地位。

360度视频亦称为沉浸式视频,在同一位置改变观察方向,能够浏览四周全部场景<sup>[3]</sup>。在景观展示、数字娱乐、视频会议、远程医疗和会诊、远程监控、机器人自动导航、军事对抗仿真与虚拟战场等领域得到了广泛应用<sup>[4,5]</sup>。

360度视频包含了四周空间的全部视觉信息,需要更高的清晰度(视频分辨率4K及以上)以支持更广阔的视野,数据量是普通视频的数倍,视频传输面临带宽瓶颈<sup>[6,7]</sup>,传统的视频编/解码方案已不能满足存储和传输的需求<sup>[8]</sup>。改进360度视

频编码方案提高压缩效率,是目前主要的研究热点。

360度视频应用主要包括采集、拼接、编码及传输和播放几个部分<sup>[9]</sup>。360度视频采集需要多台摄像机同时完成,完成同步采集后,需要将多相机采集的视频进行拼接。不同角度下拍摄的图像并不在同一投影平面上,直接对重叠的图像无缝拼接,会破坏景物视觉的一致性<sup>[10-12]</sup>,所以需要先对图像进行投影变换,再进行拼接。由于360度视频投影方式不同,投影变换过程中存在着图像失真,分辨率要求较高,传统视频编码技术压缩360度视频效果并不理想<sup>[13-15]</sup>。所以,由ITU-T/VCEG和ISO-IEC/MPEG成立的联合视频研究组(JVET)将360度视频纳入到未来视频编码标准范围内<sup>[16]</sup>,由不同研究团队分别进行了大量的实验测试工作,但相关的综合对比研究工作仍有待开展。

文献[17]对虚拟现实技术的建模方法、虚拟现实表现技术、人机交互及设备、虚拟现实开发平台与支撑环境等进行了全面的论述;文献[18]论述了增强现实与增强虚拟环境的技术特点、设备发展以及移动互联网上的虚拟增强技术与应用;文献[19]是基于视觉沉浸和深度感知的综述。

360度视频是近两年随着VR设备的出现而非常热门的一

**收稿日期:** 2017-04-28; **修回日期:** 2017-06-13      **基金项目:** 国家公安部科技强警项目(2016GABJC51); 国家留学基金资助项目

**作者简介:** 艾达(1973-),男,副教授,硕士,博士,主要研究方向为数字视频图像处理(aida@xupt.edu.cn);董久军(1993-),男,硕士,主要研究方向为刑侦图像处理;林楠,男,教授,主要研究方向为数字视频图像处理;刘颖(1972-),女,高级工程师,硕士,主要研究方向为数字视频图像处理。

种虚拟现实内容,在这方面尚未有综述文章提及。鉴于此,为了对该问题进行系统的研究与阐述,笔者整理、阅读和分析了IEEE、ACM、Elsevier、CNKI以及国际标准组织JVET会议等相关程度较高、质量较好的文献作为参考。

本文总结了360度视频研究的最新进展,包括360度视频的内容和特点,着重分析了360度视频常用的投影方式,分类比较了不同投影方式对编码效率的影响,以及编码优化方案等,探讨其发展面临的问题,为后续的研究提供参考依据。

## 1 投影变换

同步采集后的360度视频图像可以表示在一个球体表面上,但不适于存储、传输和压缩等处理,需要投影变换<sup>[20]</sup>。投影(projection)变换是指将全视域的三维场景展开到二维平面上的映射过程。

作为全景内容的一种重要载体,投影图像不仅应当包含拍摄的全部内容,还要避免过多的扭曲失真产生的回放质量损失<sup>[21,22]</sup>。本文总结了JVET编码标准制定过程中使用的几种投影方式<sup>[23]</sup>:

### a) 球面投影。

球面投影(equirectangular projection, ERP)是把球形表面的内容映射到平面的矩形图中,其原理如图1所示,已广泛应用于计算机图形处理。

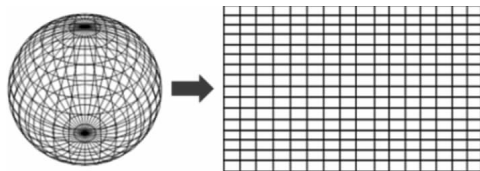


图1 球面投影

这种投影对自然图像有良好的兼容性,但在转换过程中不同纬度有不同的伸展程度,在两极区域失真较大,像素的数量增加,产生大量的冗余信息<sup>[24~27]</sup>。

### b) 立方图投影。

立方图投影(cubemap projection, CMP)原理如图2(a)所示,通过外接立方体,将球体表面内容投影到立方体表面。展开后的布局格式有两种,即 $4 \times 3$ 和 $3 \times 2$ 布局。

$4 \times 3$ 布局有50%的区域空闲,编码效率和内存管理都不理想<sup>[28]</sup>。 $3 \times 2$ 布局是立方图投影中较好的一种,可以通过模块旋转或者互换来改变布局格式,如图2(c)中2和3投影块的互换。柏拉图立体中的立方体模型已应用于VR视频投影<sup>[29]</sup>。立方体投影可从球面投影转换而来,比球面投影节省了25%的比特率,但存在像素密度减少造成的图像模糊问题。

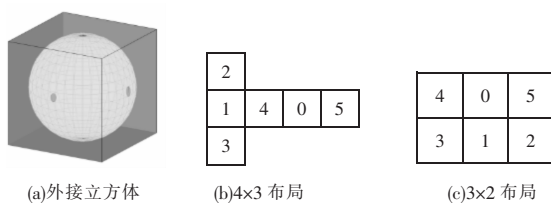


图2 立方图投影

### c) 八面体投影。

八面体投影(octahedron projection, OHP)由八个等边三角形组成,分为常规八面体布局OHP(图3(b))和改进的八面体投影(ROHP)(图3(c))。常规八面体布局含有16条不连续

的边,ROHP含有8条不连续的边,这两种布局所表示的全景内容具有不连续性<sup>[30]</sup>。

相比传统的八面体投影,文献[31]提出了一种布局更为密集,具有最少不连续边的投影方式,如图3(d)所示。

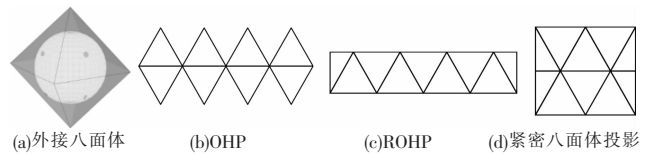


图3 八面体投影

### d) 二十面体投影。

二十面体投影(icosahedron projection, ISP)由12个顶点和20个三角形组成,分为两种布局,其中常规二十面体布局ISP(图4(b))存在大量空闲区域,编码效率不高;改进二十面体布局(RISP)如图4(c)所示,将底部的三角形移到上部,消除了空闲区域<sup>[32]</sup>。二十面体投影也可由球面投影转换而来,比球面投影节省约23%的比特率。

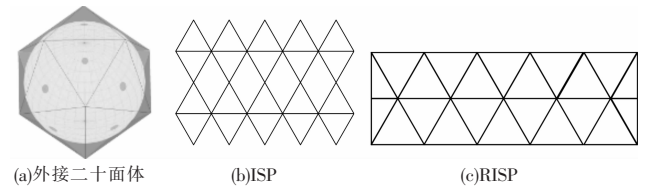


图4 二十面体投影

### e) 球面分隔投影。

球面投影和立方图投影在投影变换过程中存在插值,产生大量非信息承载像素,造成比特率的开销<sup>[33]</sup>。球面分隔投影(segmented sphere projection, SSP)近似等面积映射,有效减少了冗余信息,提高了压缩效率。

球面分隔投影的原理如图5所示。

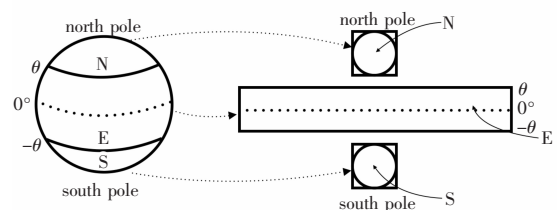


图5 三段球面分隔投影

将球体沿纬度线分成若干球形帽和环,球形帽压平成圆形,半径相当于极点到极圈的弧长;环状部分切段并展开,所有部分变成矩形块。两个压平的球形帽扩展到其外接正方形;放在低纬度区域的上、下两部分,布局成陀螺状。

因低纬度区域与球面投影相同,沿 $45^\circ$ 纬线分隔球面成三部分的投影方案编码性能最好,主流360度视频的宽高比为 $4:3$ <sup>[34,35]</sup>。

### f) 正方棱台投影。

360度视频的高帧速率和高分辨率对传输带宽提出了更高的要求。正方棱台投影(truncated square pyramid, TSP)适用于视区切换方案,对当前视区的内容以高分辨率传输,其余视区内容以低分辨率传输<sup>[36,37]</sup>。

正方棱台投影的映射原理如图6所示,TSP投影的六个面代表整个球面的内容。

在传输时,先发送第一个观看者当前视区的高质量TSP图,再根据观看者头部的方向切换到第二个TSP图。如果第

二个 TSP 图切换或传输被延迟,可向观看者继续呈现第一个 TSP 图,直到第二个 TSP 图到达。

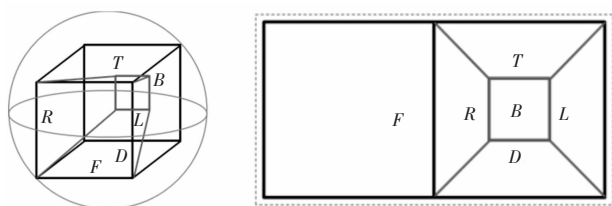


图 6 球面到立方体或 TSP 几何映射

根据以上几种投影方式的介绍,对其投影特点总结如表 1 所示。

表 1 投影方式特点总结

投影方式	布局	转换特点
球面投影 (ERP)		易转换,兼容性好;伸展时出现几何失真,非信息承载像素增加
立方图投影 (CMP)	原生 4×3 布局 紧密 3×2 布局	易转换,几何失真少,像素分布均匀;像素密度减少造成图像模糊
八面体投影 (OHP)	常规八面体布局 改进八面体布局 紧密八面体布局	有效压缩存储,投影布局多样;直接渲染困难,极点区域出现奇异性
二十面体投影 (ISP)	常规二十面体布局 紧密二十面体布局	节省存储空间,变形失真小
球面分隔投影 (SSP)	三段球面分隔 五段球面分隔	像素冗余少,几乎等效映射
正方棱台投影 (TSP)		高效压缩文件大小,减少带宽占用

## 2 投影编码性能分析

不同的投影方式有不同的结构和特点,为研究这些不同对 360 度视频压缩效率的影响,文献[38]给出了测试序列及其参数,包括帧率、帧数、分辨率等。根据 JVET 大量的测试实验结果,对投影编码的性能评估总结如下。

### 2.1 立方图投影编码

文献[39]比较了 4×3 和 3×2 布局的立方图投影编码效率,以参考软件 HM 16.9 为测试平台,测试了四种编码模式,即全帧内编码模式(AI)、随机访问模式(RA)以及两种低延时模式(LD-B/LD-P)。量化参数分别设置为 22、27、32 和 37。

综合测试结果发现,4×3 相比 3×2 布局,在 AI、RA、LD-B 和 LD-P 四种编码模式上平均比特率分别多出 1.4%、1.3%、1.7% 和 1.7%。

图 7 展示了序列 building 在四种编码模式下,4×3 相比 3×2 布局具体的比特率增益。3×2 布局编码时所需要的内存仅为 4×3 布局的 50%,3×2 布局的各种排列方式压缩效率基本接近。文献[40]指出不同投影方式间转换时,分辨率对编码效率有影响。

如图 8 所示,立方体投影视频序列 jump,从球面投影(anchor)转换到立方图投影(CMP)时,不同分辨率下编码效率不同,建议转换投影格式时选择合适的分辨率。

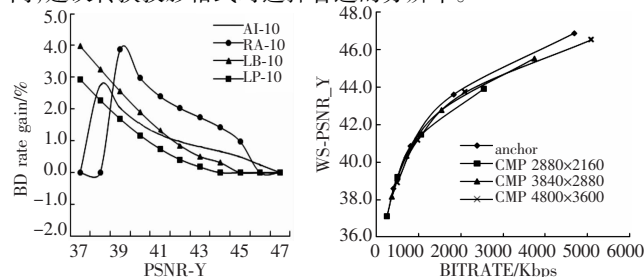


图 7 序列 building 四种编码模式下的比特率增益

图 8 序列 jump 率失真曲线

### 2.2 八面体投影编码

对于 360 度视频投影,布局时尽可能保持面之间的连续性,其压缩效率更高<sup>[41,42]</sup>。文献[31]提出的新八面体投影布局(图 3(d))重新排列了这八个面,不连续边缘最少。以参考软件 HM 16.12 为平台,实验比较了该投影布局与 ROHP、OHP 的编码效率。

序列 building 在此投影布局下与 ROHP 相比的比特率增益效果如图 9 所示。

新八面体布局的比特率显著减少,具体表现为:

a) 与 OHP 相比,在 AI、RA、LD-B 和 LD-P 四种编码模式上平均比特率分别少 14.05%、9.62%、4.52% 和 4.54%。

b) 与 ROHP 相比,平均比特率分别少 5.54%、4.73%、5.59% 和 5.67%。

新八面体布局的编码效率比 OHP 和 ROHP 布局更高,节省一定量的码率,提高了传输效率<sup>[31]</sup>。

### 2.3 二十面体投影编码

文献[43]比较了两种二十面体投影布局 ISP 和 RISP 的编码效率,实验平台为参考软件 HM 16.9。常规二十面体投影布局(ISP)与改进二十面体布局(RISP)相比,在 AI、RA、LD-B 和 LD-P 四种编码模式上平均比特率分别多出 9.5%、4.9%、3.0% 和 3.3%。RISP 编码时占用的内存为 ISP 的 60.6%,RISP 更适合用于 360 度视频编码。

图 10 展示了序列 dancing 的两种二十面体投影布局在不同编码模式下的比特率增益。

从图 10 中可知,在同一 PSNR-Y 下,AI 编码模式下比特率增益最大。

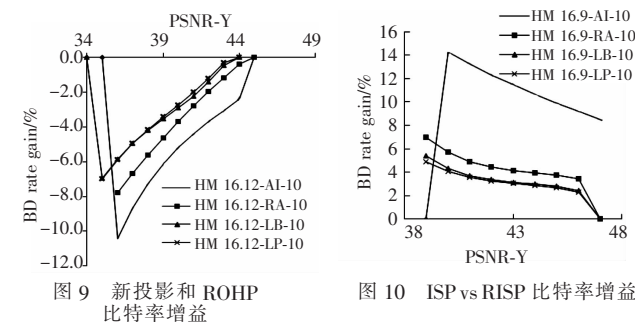


图 9 新投影和 ROHP 比特率增益

图 10 ISP vs RISP 比特率增益

综上所述,将各种投影方式的编码性能汇总如表 2 所示,以供对比分析。

表 2 投影编码性能总结

投影编码	参照	测试	软件	比特率增益/%				特性
				AI	RA	LD-B	LD-P	
立方体投影	3×2 布局	4×3 布局	HM 16.9	1.4	1.3	1.7	1.7	排列紧密,内容连续的布局,编码效率更高
八面体投影	新布局	OHP	HM 16.12	-14.05	-9.62	-4.52	-4.54	
		ROHP		-5.54	-4.73	-5.59	-5.67	
二十面体投影	RISP	ISP	HM 16.9	9.5	4.9	3.0	3.3	

## 3 编码优化

360 度视频数据量较大,投影布局新颖,传统视频编码方法压缩 360 度视频不理想,需针对其特点进行编码优化,有效降低信息数据量,节省存储空间和网络传输带宽。最新提出的优化编码方法如下。

### 3.1 HEVC 局部优化方法

#### 3.1.1 共投影平面的运动补偿预测

立方体投影方式中两个相邻面属于不同投影平面,在面边



界附近的区域存在纹理不连续性,会影响现有运动补偿预测的效率<sup>[44,45]</sup>。在基于共投影面(co-projection-plane, CPP)的运动补偿预测方法中,当运动矢量指向不同投影面中的参考像素时,导出其在当前投影平面中的对应像素,作为新的参考像素。利用布局信息进一步导出参考图片中的新参考像素的位置,最终导出位于当前像素共投影面的新参考像素的位置。

共投影平面的运动补偿预测如图11所示。图11中,面 $ABCD$ 和 $A'B'C'D'$ 是共投影平面, $N$ 的位置就是立方体表面中参考像素的位置,然后根据 $N$ 的位置就可以推出共投影面中新参考像素的位置。

根据相似三角形的特征,由式(1)(2)可以推出 $N$ 的位置。

$$L_{SJ} = \frac{a/2 \times y}{a/2 + x} \quad (1)$$

$$L_{SN} = \frac{a/2 \times x}{a/2 + x} \quad (2)$$

其中: $a$ 是立方体的边长; $L_{JK} = x$ , $L_{KT} = y$ , $L$ 表示线的长度<sup>[46]</sup>。

文献[46]将该方法在HM 16.6的基础上实现,实验测试了七个 $4 \times 3$ 布局的立方体投影视频序列。与原HM 16.6相比,共投影平面的运动补偿预测方法在RA、LD-B和LD-P测试模式中, $Y$ 分量上平均实现了1.2%、1.7%和1.6%的比特率降低,在LD-B配置中,有的甚至高达6.6%。

### 3.1.2 运动补偿预测的几何填充

现有的视频编/解码器仅用于常规2D视频编码。运动补偿预测使用参考图像边界之外的任何样本时,将简单地从图像边界复制样本值来执行填充。然而360度视频包含整个球体的信息,不再具有边界<sup>[47]</sup>。HEVC中传统的复制填充方式不适合于360度视频的编码。几何填充考虑了360度视频的循环属性,在360度视频编码的运动补偿预测中引入不同的填充方法。

a) 球面投影的几何填充相对简单,因为ERP定义在有经度和纬度的球体之上。给定一个点 $(u, v)$ (位于ERP的外部)来填充,用于导出填充样本的点 $(u', v')$ 可以计算如下,其中 $W$ 和 $H$ 是球面投影的宽度和高度。

$$u < 0 \text{ 或 } u \geq W \text{ 且 } 0 \leq v < H, \text{ 则 } u' = u \% W, v' = v$$

$$v < 0, \text{ 则 } v' = -v - 1, u' = (u + W/2) \% W$$

$$v \geq H, \text{ 则 } v' = 2 \times H - v - 1, u' = (u + W/2) \% W$$

b) 对于立方体投影,要使用几何填充方法扩展投影中的每一个面。

立方体投影的几何填充如图12所示。图12中,点 $P$ 在面 $F_1$ 上,但在面 $F_1$ 的边界之外。几何填充不是使用点 $R$ 处的样本值来填充点 $P$ 处的样本值,而是使用点 $Q$ 处的样本值来填充点 $P$ 处的样本值。与HEVC中使用的重复填充方法相比,几何填充可以为CMP表面边界之外的区域提供更有意义的样本。

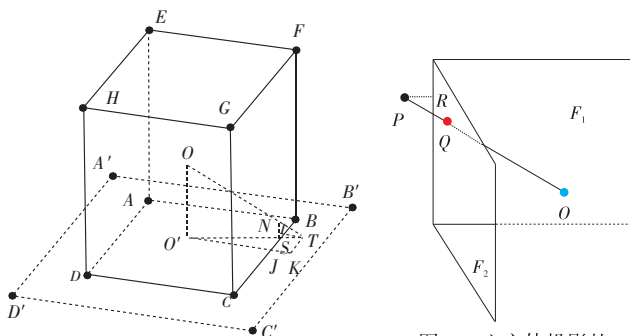


图11 共投影平面的运动补偿预测

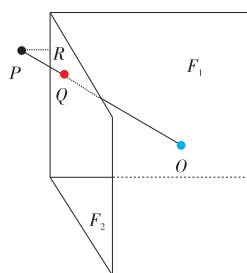


图12 立方体投影的几何填充

几何填充在基于HM 16.12的PCT360投影转换软件<sup>[48]</sup>的基础上实现。文献[49]实验测试了ERP、 $3 \times 2$ 立方体布局两种投影,均使用随机访问模式(RA)。与HEVC中常规2D填充方法相比,几何填充的编码效果如下:a)对于8K序列,在 $Y$ 分量上,球面投影方式下比特率平均降低0.3%,立方体投影方式下平均降低0.8%;b)对于4K序列,在 $Y$ 分量上,球面投影方式下比特率平均降低0.2%,立方体投影方式下平均降低1.0%。几何填充对立方体投影(CMP)的改进更大,因为CMP有六个面,更多的边界外采样受到影响。对于含有快速移动对象的视频序列,几何填充的方法压缩效率更高<sup>[50,51]</sup>。

### 3.1.3 帧内预测的参考样本推导方法

HEVC中帧内编码参考样本通常使用当前块正上方和左侧的样本,新提出的方法在推导该块所属当前面之外的参考样本时,依赖于几何填充方法。

360度视频的几何投影由许多个面组成,每个面位于不同平面,面与面之间的连续性被破坏。当展开为2D图时,不同面之间的不连续性会增加<sup>[52]</sup>。对于360度视频编码的运动补偿预测,文献[49]已提出用几何填充代替重复填充。同样,对于帧内编码,当参考样本位于当前块所属面之外时,应用几何填充方法推导参考样本,在 $Y$ 分量上实现了比特率平均约0.3%的降低,几乎对于所有序列存在一致的亮度和色度性能增益<sup>[53]</sup>。

## 3.2 JEM 整体方案

JEM(joint exploration model)是JVET提出的增强型视频编码技术,以HEVC测试模型为基础<sup>[54~57]</sup>,包括HEVC中基本的编码和解码流程,如块结构、帧内和帧间预测、系数变换、环路滤波器和熵编码等,添加了新的编码工具,修改了一些编码模块。JEM5新的编码模型如表3所示。所列方法已集成到JEM测试软件中<sup>[58]</sup>。

表3 JEM5 编码模块

模块	内容
块结构	二叉树加二叉树(QTBT)块结构 <sup>[59,60]</sup>
帧内预测改进	65个帧内预测方向 <sup>[61~64]</sup> 用于帧内预测的四点插值滤波器 <sup>[61,62]</sup> 边界滤波器应用于除了水平和垂直方向之外的其他方向 <sup>[61,62]</sup>
帧间预测改进	子PU级运动矢量预测 <sup>[61,65,68]</sup> 局部自适应运动矢量方案(AMVR) <sup>[61,65]</sup> 1/16像素运动矢量存储精度 重叠块运动补偿(OBMC) <sup>[61,65]</sup> 局部光照补偿(LIC) <sup>[65,70]</sup>
转换	显式多核变换 <sup>[61,65]</sup> 模式相关非可分二次变换 <sup>[65,71]</sup>
环路滤波器	内容自适应裁剪 自适应环路滤波器(ALF) <sup>[61,65]</sup>
增强CABAC设计 <sup>[42,43]</sup>	变换系数级别的上下文模型选择 多假设概率估计 上下文模型的初始化

结果表明,JEM 3.0与HM 16.9相比,在AI、RA、LD-B和LD-P四种编码模式上,平均比特率的减少分别为:

- 球面投影方式下,16.9%、27.4%、21.9%和24.1%。
- $3 \times 2$ 立方图投影方式下,17.5%、24.8%、22.5%和25.0%。
- 紧密二十面体投影方式下,22.4%、24.5%、21.8%和

25.1%。

JEM 算法在 360 度视频的压缩效率上比 HM 16.9 更为出色,但结果仍低于预期,在这方面要做更多的工作,以提高视频编码的压缩效率。分析归纳以上四种编码优化方法的性能如表 4 所示。其中优化效果以 Y 分量的比特率增益为参考,取四种编码模式下 Y 分量的均值。

表 4 编码优化方案总结

优化编码方法	所对比方法	优化内容	优化效果 (Y 分量)/%
JEM 算法	HEVC	部分编码模块修改,添加新的编码工具	-18.9
共投影平面的运动补偿预测	运动补偿预测(MCP)	立方体投影边缘纹理的不连续对 MCP 造成的影响	-1.5
运动补偿预测的几何填充	复制填充方式	360 度视频内容球形循环属性对填充扩展造成的影响	-0.9
帧内预测参考样本推导方法	原帧内编码参考样本	当前块所属面之外的参考样本推导问题	-0.3

#### 4 结束语

本文总结了 360 度视频的投影方式,综合分析了其对编码效率的影响。通过对比发现:

a) 图像变形失真少,排列布局紧密,内容连续的投影方式,在编码中的压缩效率更高。正方棱台投影兼具以上特性,更适于对转换的场景视区编码。

b) 给定投影方式下的 360 度视频,根据视频序列的不同,适于编码的最佳分辨率亦不同。

360 度视频与传统的视频存在着许多差别,HEVC 达不到理想的压缩效果。本文分析了目前提出的一些新的解决方案,包括下一代视频编码技术 JEM 的提案;针对 360 度视频特点优化的共投影平面的运动补偿技术、几何填充、帧内预测的参考样本推导方法;归纳了各个方案重点优化的内容,为研究者提供参考和启示。

360 度视频在虚拟现实中的研究已经取得了一定的进展,然而为了实现观察舒适、体验真实的虚拟现实效果,在后续的研究中,以下问题和建议需要考虑:

a) 图像失真。由于硬件的限制(360 度视频的捕获相机数量少,如三星 Gear 360 只有两个相机)和广角镜头的使用(如鱼镜头),捕获的视频图像会出现不同程度的扭曲失真。此类镜头的畸变校正技术中心思想是通过对镜头的曲率特征、光学折射线路和透光亮度进行综合分析,采用一定算法把边缘畸变严重的图形拉伸整合为适合观看的正常比例图像。其中所用算法的合理性、编程水平的高低及最终系统资源占用率共同决定了此算法的优劣。

b) 图像拼接。拼接只有几部分图像的 360 度视频具有较大的难度,它们的重叠区域极其有限,因此会出现更多的失真、对象丢失和重影抖动。360 度视频主要是通过多个相机同时拍摄不同角度的视频,后期的拼接融合环节非常重要,需要解决视频同步、亮度与颜色均衡处理、视频稳定、方向控制、特征匹配和像素匹配等问题。在保障拼接精度的前提下,如何提高拼接速度是一个重要的研究方向,未来的趋势就是要在实现 360 度视频拍摄的同时完成实时拼接,降低视频制作成本。

c) 渲染问题。虚拟现实的使命是渲染出一个与现实世界一样逼真的虚拟世界,这本身就是一种挑战。近两年随着硬件设备的发展,人们可以通过虚拟现实头盔体验到沉浸的效果,

但眩晕问题是这一挑战中最难以解决的问题。想要保证不晕,VR 设备输出的画面必须要有足够高的刷新率(120 Hz 及以上)以及足够高的分辨率(4K 及以上),但目前硬件设备的处理能力不足以实现此效果,在渲染视频时普遍存在视觉不同步、声音不同步、景深不同步等问题,这就造成了眩晕感。尽可能地缓解眩晕是一个长久的研究方向,以下几个解决思路可供参考:通过算法优化降低延迟、提升声场的逼真程度、提升景深的逼真程度<sup>[73]</sup>。

d) 高分辨率和高帧速率。在接近眼球的位置播放视频,需要足够高的分辨率和帧率,360 度视频文件就会比较大,传输带宽是传统视频的 4~6 倍。在维持视频高质量的同时,降低带宽占用、减少缓冲时间实现快速播放,成为新的技术需求。其中 360 度视频的投影方式对数据量影响很大,如正方棱台投影格式比原视频减少了近 80% 的文件大小。另外一个研究思路是视区自适应编码(viewport-adaptive encoding),即对人眼聚焦的地方以及周围的图像采用高分辨率渲染,而其他部分以普通分辨率的渲染<sup>[37]</sup>,这项技术的提出可有效减少带宽占用,其眼部跟踪技术需要持续深入研究。

e) 编/解码计算量。对于高分辨率和高帧率的 360 度视频,在编码器和解码器两端存在巨大的数据压缩和计算复杂度,使用 Gear VR 设备 30 min,智能手机就会出现严重发热现象<sup>[74]</sup>。360 度视频编码优化问题,本文第 3 章进行了详细介绍,HEVC 适用于传统的 2D 视频,压缩 360 度视频并不理想。针对 360 度视频投影布局的特点,对其帧内和帧间预测方法作出相应的改进,这方面的研究工作具有重要的意义。

f) 高动态范围(high dynamic range, HDR)影像内容。高动态范围格式要求每像素 10~16 bit 的比特深度。拍摄 HDR 视频是相对简单的,但大多数编/解码器和视频播放器都不支持 HDR 内容,因为它的有用性在 2D 中受到限制。在 VR 中, HDR 将允许观看者根据其方向体验到不同的曝光水平,感受到接近现实世界中的光照变化和明暗对比。利用计算机的性能提升图片显示和动态图片处理能力,拓展亮度的动态表现范围,进一步提升视觉效果是 HDR 研究的重要方向。

构建新一代沉浸式技术就意味着解决新的、更艰巨的工程挑战。对于 360 度视频,需要一种非常高效的视频压缩技术。随着虚拟现实技术的流行和下一代视频编码标准的制定,360 度视频已经越来越受到学术界和工业界的重视,创新视频投影方式,高效压缩编码以满足带宽和质量要求,优化用户体验是研究探索的核心问题。

#### 参考文献:

- [1] Routhier P P. Virtually perfect: factors affecting the quality of a VR experience and the need for a VR content quality standard[C]//Proc of SMPTE Annual Technical Conference and Exhibition. Piscataway, NJ: IEEE Press, 2016: 1-20.
- [2] Kopf J. 360° video stabilization[C]//Proc of ACM SIGGRAPH Asia. 2016.
- [3] 张洋,李庆忠,臧风妮.一种多摄像机全景视频图像快速拼接算法[J].光电子·激光,2012,23(9):1821-1826.
- [4] Budagavi M, Furton J, Jin Guoxin, et al. 360 degrees video coding using region adaptive smoothing[C]//Proc of IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2015: 750-754.
- [5] Järvinen A. Virtual reality as trend contextualising an emerging con-

- sumer technology into trend analysis [C]//Proc of Future Technologies Conference. Piscataway, NJ: IEEE Press, 2016: 1065-1070.
- [6] Dai Feng, Shen Yanfei, Zhang Yongdong, *et al.* Selection of the most efficient tile size in tile-based cylinder panoramic video coding and transmission [J]. *Visual Computer*, 2007, 23(9-11): 892-893.
- [7] Ng K T, Chan S C, Shum H Y. Data compression and transmission aspects of panoramic videos [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2005, 15(1): 82-95.
- [8] 张洪彬, 伏长虹, 苏卫民, 等. 3D-HEVC 深度图像帧内编码单元划分快速算法 [J]. *电子与信息学报*, 2016, 38(10): 2523-2530.
- [9] 田华伟, 肖延辉, 冯文刚, 等. 用于头戴式虚拟现实眼镜的视频水印 [J]. *科学技术与工程*, 2016, 16(5): 193-195.
- [10] Jin Guoxin, Ankur S, Madhukar B. Motion estimation and compensation for fisheye warped video [C]//Proc of IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2015: 2751-2755.
- [11] Schäfer R, Kauff P, Skupin R, *et al.* Interactive streaming of panoramas and VR worlds [J]. *SMPTE Motion Imaging Journal*, 2017, 126(1): 35-42.
- [12] Zheng Jiali, Zhang Yongdong, Shen Yanfei, *et al.* Panoramic video coding using affine motion compensated prediction [C]//Lecture Notes in Computer Science, vol 4577. Berlin: Springer, 2007: 119-121.
- [13] De la Fuente Y S, Skupin R, Schierl T. Video processing for panoramic streaming using HEVC and its scalable extensions [J]. *Multimedia Tools and Applications*, 2016, 76(4): 5631-5659.
- [14] Zhu Zhigang, Xu Guangyou, Edward M R, *et al.* Fast construction of dynamic and multi-resolution 360° panoramas from video sequences [J]. *Image and Vision Computing*, 2006, 24(1): 15-17.
- [15] Zheng Jiali, Shen Yanfei, Zhang Yongdong, *et al.* Adaptive selection of motion models for panoramic video coding [C]//Proc of IEEE International Conference on Multimedia and Expo. Piscataway, NJ: IEEE Press, 2007: 1319-1322.
- [16] Norkin A, Yang H, Ohm J R, *et al.* Call for test materials for future video coding standardization [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [17] 赵沁平. 虚拟现实综述 [J]. *中国科学 F 辑: 信息科学*, 2009, 39(1): 2-46.
- [18] 周忠, 周颐, 肖江剑. 虚拟现实增强技术综述 [J]. *中国科学 F 辑: 信息科学*, 2015, 45(2): 157-180.
- [19] Moreau G. Visual immersion issues in virtual reality: a survey [C]//Proc of the 26th Conference on Graphics, Patterns and Images Tutorials. Washington DC: IEEE Computer Society, 2013: 6-14.
- [20] Hosseini M, Swaminathan V. Adaptive 360 VR video streaming: divide and conquer [C]//Proc of IEEE International Symposium on Multimedia. Washington DC: IEEE Computer Society, 2016: 107-110.
- [21] Ashfaq A, Diane G, Xin Tong, *et al.* Immersion in cardboard VR compared to a traditional head-mounted display [C]//Proc of International Conference on Virtual, Augmented and Mixed Reality. Cham: Springer, 2016: 269-276.
- [22] Choi K P, Vladyslav Z, Choi M, *et al.* Test sequence formats for virtual reality video coding [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [23] Alshina E, Boyce J, Abbas A, *et al.* JVET common test conditions and evaluation procedures for 360° video [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [24] Huang Kaichen, Chien P Y, Chien C A, *et al.* A 360-degree panoramic video system design [C]//Proc of International Symposium on VLSI Design, Automation and Test. Piscataway, NJ: IEEE Press, 2014: 1-4.
- [25] Youvalari R G, Aminlou A, Hannuksela M M, *et al.* Efficient coding of 360-degree pseudo-cylindrical panoramic video for virtual [C]//Proc of IEEE International Symposium on Multimedia. Piscataway, NJ: IEEE Press, 2016: 525-528.
- [26] 王广生, 高新, 孙树文. 全景视频与个性化分发给数字博物馆中的应用 [J]. *北京联合大学学报*, 2015, 29(3): 5-9.
- [27] Lin H C, Lin Jianliang, Chang Shenkai, *et al.* AHG8: an alternative arrangement for cube format (double-cross layout) [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [28] He Yuwen, Vishwanath B, Xiu Xiaoyu, *et al.* AHG8: InterDigital's projection format conversion tool [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [29] Kuzyakov E, Pio D. Next-generation video encoding techniques for 360 video and VR [EB/OL]. (2016-07-15). <https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/>.
- [30] Choi K, Zakharchenko V, Choi M, *et al.* Test sequence formats for virtual reality video coding [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [31] Lin H C, Li C Y, Lin Jianliang, *et al.* AHG8: an efficient compact layout for octahedron format [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [32] Zakharchenko V, Alshina E, Choi K P, *et al.* AHG8: icosahedral projection for 360-degree video content [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [33] Fadlallah F A, Khalifa O O, Abdalla A H. Video streaming based on frames skipping and interpolation techniques [C]//Proc of International Conference on Computer and Communication Engineering. Piscataway, NJ: IEEE Press, 2016: 475-479.
- [34] Zhang Chuanyi, Lu Yao, Li Jisheng, *et al.* AHG8: segmented sphere projection (SSP) for 360-degree video content [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [35] Li Jisheng, Wen Ziyu, Li Sihan, *et al.* Novel tile segmentation scheme for omnidirectional video [C]//Proc of IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2016: 370-374.
- [36] Auwera G V, Coban M, Hendry, *et al.* AHG8: truncated square pyramid projection (TSP) for 360 video [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [37] Sreedhar K K, Aminlou A, Hannuksela M M, *et al.* Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications [C]//Proc of IEEE International Symposium on Multimedia. Washington DC: IEEE Computer Society, 2016: 583-586.
- [38] Jeon S, Kim N, Khanh D Q, *et al.* New GoPro test sequences for virtual reality video coding [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [39] Zhou Minhua. AHG8: a study on compression efficiency of cube projection [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.

- [40] Sun Yule, Lu Ang, Lu Yu. AHG8: a study on the influence of different projection schemes [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [41] Hosseini M, Swaminathan V. Adaptive 360 VR video streaming based on MPEG-DASH SRD [C]//Proc of IEEE International Symposium on Multimedia. Piscataway, NJ: IEEE Press, 2016: 407-408.
- [42] Sreedhar K K, Aminlou A, Hannuksela M M, *et al.* Standard-compliant multiview video coding and streaming for virtual reality applications [C]//Proc of IEEE International Symposium on Multimedia. Washington DC: IEEE Computer Society, 2016: 295-300.
- [43] Zhou Minhua. AHG8: a study on compression efficiency of icosahedral projection [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [44] Bao Yanan, Wu Huasen, Ramli A A, *et al.* Viewing 360 degree videos: motion prediction and bandwidth optimization [C]//Proc of the 24th IEEE International Conference on Network Protocols. Washington DC: IEEE Computer Society, 2016: 1-2.
- [45] Yang Hua, Kenneth R. Optimizing motion compensated prediction for error resilient video coding [J]. *IEEE Trans on Image Processing*, 2010, 19(1): 108-118.
- [46] Ma Xiang, Yang Haitao, Zhao Zhijie, *et al.* Co-projection-plane based motion compensated prediction for cubic format VR content [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [47] Liu Tsuming, Ju Chicheng, Huang Yuhao, *et al.* A 360-degree 4K × 2K panoramic video recording over smart-phones [C]//Proc of IEEE International Conference on Multimedia & Expo Workshops. Piscataway, NJ: IEEE Press, 2016.
- [48] He Yuwen, Vishwanath B, Xiu Xiaoyu, *et al.* AHG8: algorithm description of InterDigital's projection format conversion tool (PCT360) [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [49] He Yuwen, Ye Yan, Hanhart P, *et al.* AHG8: geometry padding for 360 video coding [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [50] Bao Yanan, Wu Huasen, Zhang Tianxiao, *et al.* Shooting a moving target: motion-prediction-based transmission for 360-degree videos [C]//Proc of IEEE International Conference on Big Data. Piscataway, NJ: IEEE Press, 2016: 1161-1170.
- [51] Argyriou L, Economou D, Bouki V, *et al.* Engaging immersive video consumers: challenges regarding 360-degree gamified video applications [C]//Proc of the 15th International Conference on Ubiquitous Computing and Communications and International Symposium on Cyberspace and Security. Washington DC: IEEE Computer Society, 2016: 145-152.
- [52] Hanhart P, He Yuwen, Ye Yan. AHG8: reference samples derivation using geometry padding for intra coding [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2016.
- [53] Yu M, Lakshman H, Girod B. A framework to evaluate omnidirectional video coding schemes, in mixed and augmented reality [C]//Proc of IEEE International Symposium on Mixed and Augmented Reality. Washington DC: IEEE Computer Society, 2015: 31-36.
- [54] Rec. ITU-T H. 265 and ISO/IEC 23008-2, High efficiency video coding (HEVC) [S]. 2013.
- [55] Sullivan G J, Ohm J R, Han W J, *et al.* Overview of the high efficiency video coding (HEVC) standard [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668.
- [56] Misra K, Segall A, Horowitz M, *et al.* An overview of tiles in HEVC [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2013, 7(6): 969-977.
- [57] Sánchez Y, Skupin R, Schierl T. Compressed domain video processing for tile based panoramic streaming using HEVC [C]//Proc of IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2015: 2244-2248.
- [58] Chen J, Alshina E, Sullivan G J, *et al.* Algorithm description of joint exploration test model 5 [C]//Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 2017.
- [59] Karczewicz M, Chen J, Chien W J, *et al.* Study of coding efficiency improvements beyond HEVC [C]//SPIE Optical Engineering + Applications. 2015.
- [60] An J, Chen Y W, Zhang K, *et al.* Block partitioning structure for next generation video coding, MPEG doc. m37524 and ITU-T SG16 doc COM16-C966 [K]. 2015.
- [61] 单娜娜, 周巍, 段哲民, 等. 高性能视频编码帧间预测的单元划分优化算法 [J]. *电子与信息学报*, 2016, 38(5): 1194-1201.
- [62] Chen J, Chien W J, Karczewicz M, *et al.* Further improvements to HMkTA-1.0, ITU-T SG16/Q6 doc. VCEG-AZ07 [K]. 2015.
- [63] Alshina E, Alshin A, Min J H, *et al.* Known tools performance investigation for next generation video coding [C]//SPIE Optical Engineering + Applications. 2015: 95991B.
- [64] Choi K, Alshina E, Alshin A, *et al.* Information on coding efficiency improvements over HEVC for 4K content, MPEG doc. m37043 [K]. 2015.
- [65] Chen Jianle, Chen Ying, Karczewicz M, *et al.* Coding tools investigation for next generation video coding, ITU-T SG16 doc. COM16-C806 [K]. 2015.
- [66] Said A, Zhao Xin, Chen Jianle, *et al.* Position dependent intra prediction combination, MPEG doc. m37502 and ITU-T SG16 doc. COM16-C1016 [K]. 2015.
- [67] Filippov A, Ruffitskiy V. Reference sample adaptive filtering for intra coding, MPEG doc. m37526 and ITU-T SG16 doc. COM16-C983 [K]. 2015.
- [68] Chien W J, Karczewicz M. Extension of advanced temporal motion vector predictor, ITU-T SG16/Q6 doc. VCEG-AZ10 [K]. 2015.
- [69] Lin S, Chen H, Zhang H, *et al.* Affine transform prediction for next generation video coding, MPEG doc. m37525 and ITU-T SG16 doc. COM16-C1016 [K]. 2015.
- [70] Liu H, Chen Y, Chen J, *et al.* Local illumination compensation, ITU-T SG16/Q6 doc. VCEG-AZ06 [K]. 2015.
- [71] Zhao Xin, Chen Jianle, Karczewicz M. Mode-dependent non-separable secondary transform, ITU-T SG16/Q6 doc. COM16-C1044 [K]. 2015.
- [72] Lan C, Xu J, Wu F. Enhancement of HEVC using signal dependent transform (SDT), MPEG doc. m37503 and ITU-T SG16/Q6 doc. VCEG-AZ08 [K]. 2015.
- [73] Atena N. Challenges of 360 degree video [EB/OL]. 2016. <https://atenanguyen.wordpress.com/2016/04/20/challenges-of-360-degree-video/>.
- [74] Atena N. 360-degree video compression: challenges, related work and motivations [EB/OL]. (2016). <https://atenanguyen.wordpress.com/2016/04/21/360-degree-video-compression-challenges-related-work-and-motivations/>.