

面向沉浸式视频编码和传输的高效投影方法

Efficient Projections for Immersive Video Coding and Transmission

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 06-0050-06

摘要: 全景视频投影技术是全景视频应用中的一项关键技术,为此成立了 IEEE 1857.9 沉浸视频内容编码工作组,致力于制订高效的沉浸视频投影和编码方法。IEEE 1857.9 沉浸视频内容编码工作组设计了一系列的映射格式,其中双极方形的投影格式相较于球面投影技术(ERP)取得了 11.56% 的编码性能增益。此外,IEEE 1857.9 工作组还针对流切换的场景设计了基于主视点的投影模型,在保证主视点区域视频质量不变的情况下,可获得 70% 的编码性能增益。

关键词: 虚拟现实;沉浸视频;投影;编码

Abstract: Since the panoramic video projection technology is a key technology in panoramic video applications, the IEEE 1857.9 immersive video content coding working group was established so as to develop efficient immersive video projection and coding methods. A series of projection schemes are designed by the IEEE 1857.9 immersion video content coding working group, in which the bipolar square projection format achieves 11.56% coding performance gain compared to the equirectangular projection (ERP). In addition, for the flow switching scenario, a projection model based on the main view point is designed by the IEEE 1857.9 working group, which can achieve 70% coding performance gain while guaranteeing the video quality of the main view area.

Key words: virtual reality; immersive video; projection; encoding

王悦名/WANG Yueming
王荣刚/WANG Ronggang
高文/GAO Wen

(北京大学深圳研究生院,广东深圳
518055)
(Shenzhen Graduate School, Peking
University, Shenzhen 518055, China)

条带,然后根据图像内容以及条带的纬度来进行自适应的下采样)^[1]、条带分割投影(将球面视频根据纬度分割成多个环带,两极投影为平面,中间的环带投影为矩形)^[2]等投影方案被纷纷提出。这些投影方法可以减少全景视频的传输带宽,但是由于全景视频需要提供 360° 的视频内容,因此全景视频的传输带宽和播放复杂度仍然比传统视频大得多。此外,随着虚拟现实技术的视频采集设备的发展,全景视频的分辨率也提高到 8K 甚至更高。为了解决上述问题,IEEE 1857.9 沉浸视频内容编码工作组于 2015 年 12 月 5 日成立,该工作组致力于制订高效的沉浸视频投影和编码方法。

IEEE1857.9 工作组致力于推动用于压缩、解压缩和重建沉浸式视觉内容的高效编码工具标准化。该标准针对的应用场景和服务对象包括但不限于 VR,例如:基于无人机的 VR、增强现实、全景视频和其他视频/音频驱动的服务,以及诸如沉浸式视频流、广播、存储和通信之类的

虚拟现实(VR)是一种计算机仿真技术,它使用头戴式设备(有时与物理空间或多投影环境相结合)产生逼真的图像、声音和其他感觉,给用户带来沉浸式的体验。沉浸视频(或球面视频,全景视频)是通过全景摄像机在同一时刻捕获 360° 范围内的视频来获得,在观看过程中,观看者能看到全方位的视频,感受到身临其境的感觉。然而现有的编码和存储技术不支持球形视频的处理,因

此球形全景视频需要投影到二维平面上进行存储和编码。

将球面全景视频映射到二维平面上方法多种多样^[1],其中最常见的是球面投影技术(ERP)^[2],但是 ERP 技术在高纬度区域存在很严重的过采样,因此会浪费很多传输带宽。随后,等面积圆柱投影(与 ERP 类似,但其通过降低两极在纬度方向的采样密度是球面总体的采样密度一致)^[3]、立方体投影(通过透视投影的方法,将球面投影到立方体平面上)^[4]、自适应条带投影(将 ERP 格式的全景图像根据纬度分割成多个

收稿日期: 2017-10-20
网络出版日期: 2017-11-08

应用。最近,IEEE 1857.9 沉浸视频内容编码工作组拟定的标准草案即将定稿。本文对 IEEE1857.9 工作组采纳或研究的投影技术进行了综述。

1 全景投影技术

在众多的投影方法中,ERP 投影是最常用的投影格式,但是其编码效率却非常低。为了提高 VR 视频的编码效率,IEEE1857.9 工作组研发了一系列针对 VR 视频的高效投影格式,并且在标准中采纳了其中的部分投影格式。

1.1 ERP 投影

最常见的全景视频的投影格式是 ERP 投影^[2]。ERP 投影根据等间隔的经度和纬度将球面投影到二维平面上,如图 1 所示。

ERP 在 VR 视频中使用很广泛,但是 ERP 存在很多问题,比如这种投影方式会造成两极区域的过采样和失真,导致使用 ERP 投影得到的平面视频的编码效率很低。

1.2 多面体投影

多面体投影^[6]通常通过透视投影将球面投影到外切多面体上。在 IEEE1857.9 工作组会议中,多种多面体投影模型被提出,包括:立方体投影、八面体投影、二十面体投影。

多面体的面越多,越接近球面,采样密度也就越均匀,多面体投影的多种投影模型如表 1 所示。

1.3 双极方形投影

分析发现 ERP 投影格式在两极区域存在较严重的过采样,而在中间区域采样密度比较均匀(但不完全均匀),针对以上特点,设计出了一种双极方形的投影格式^[7]。如图 2 所示,双极方形投影将球面根据纬度(南北纬 45°)划分为 3 个区域,根据等间距的经度和纬度将球的区域 II 投影到二维平面上(投影方法与 ERP 相同),并将区域 I 和区域 III 投影到由多个

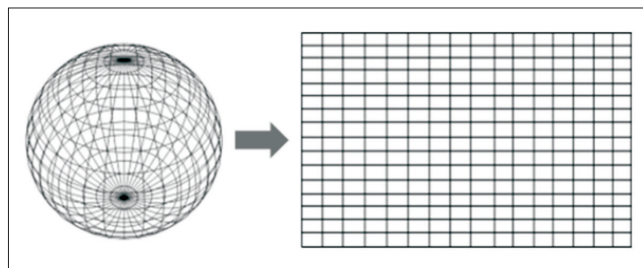


图 1 ERP 投影示意

表 1 多种多面体投影示意

投影方式	3D 模型	投影得到的 2D 平面
立方体投影 (6 面)		
八面体投影 (8 面)		
二十面体投影 (20 面)		

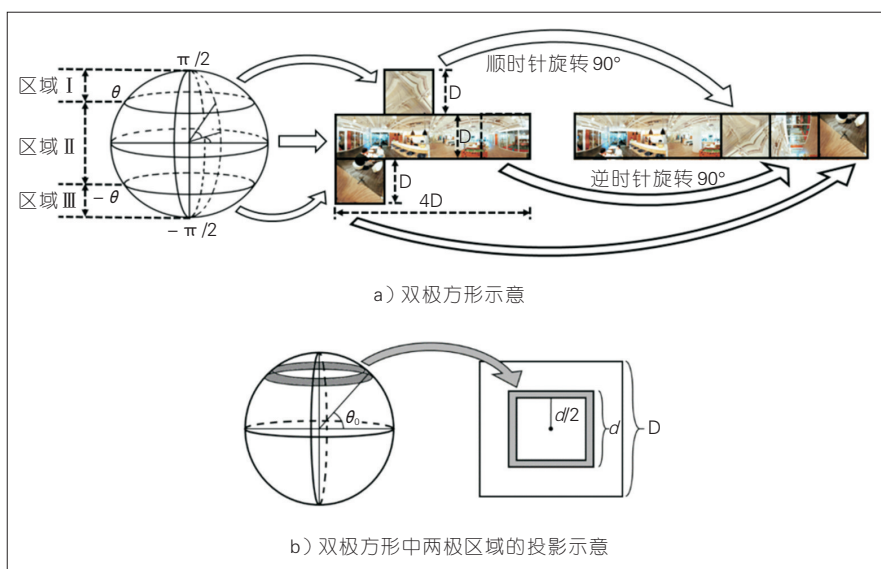


图 2 双极方形投影

同心方形环组成的矩形平面中。

1.4 双环带投影

双环带投影^[8]通过北纬 30° 和南纬 30° 的两条纬线将球面分成 3 部分。中间区域是一个环形区域,称为“环区域”;另外的两部分分别是顶区域和底区域。然后,双环带投影进一步将环区域分成 6 个均匀的子区域,将顶和底区域分别分成 4 个子区域,球面的区域划分如图 3 所示。双环

带投影进一步将这 14 个子区域投影为平面上的 14 个方向平面,并进一步将 14 个方向平面拼接成一个矩形平面。

2 非对称投影技术

在具有一对多信道和反馈信道(如直播广播系统、视频点播系统等)的应用场景中,动态流切换技术能在很大程度上减少传输带宽和播放的复杂度。在动态流切换技术中,全景

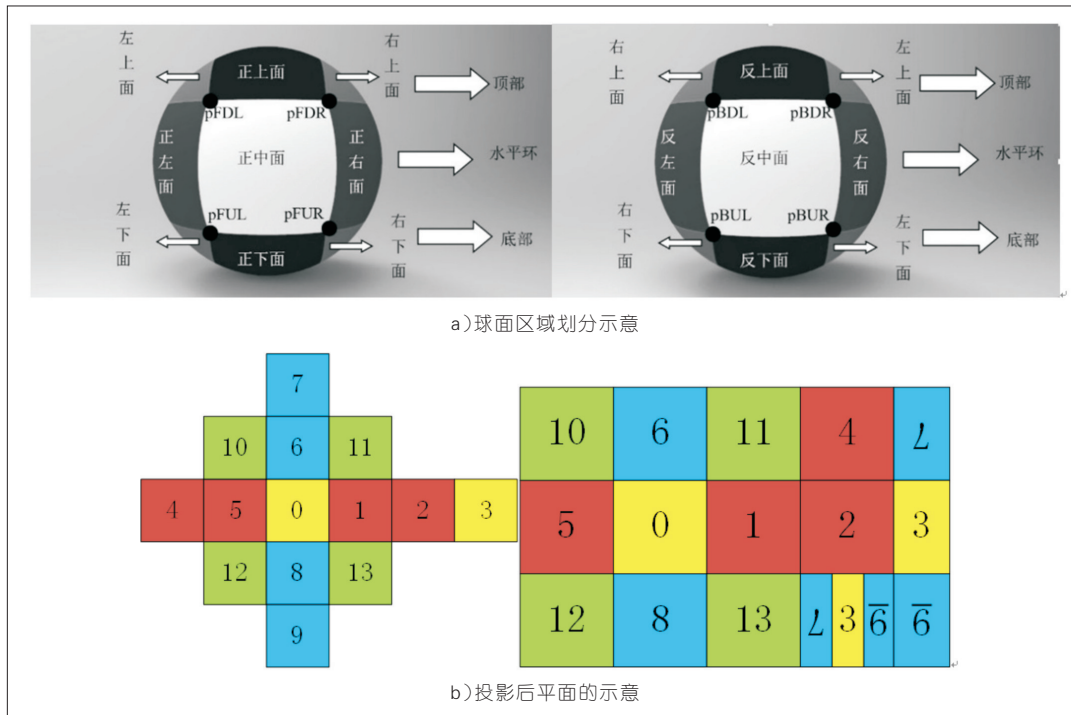


图3 双环带投影示意

视频由覆盖整个全景视频的多个主视点不同的非对称投影序列表示。根据人观看VR视频时头部转动的方向,每一时刻只传输一个主视点与当前头部朝向最接近的非对称投影序列的码流。下一时刻,如果观看者头部的转动幅度超过现有的主视点范围,服务端将传输对应主视点区域的非对称投影序列的码流到客户端。通常,全景视频的非对称投影格式的主视点区域的采样密度较高,而非主视点区域的采样密度很低,并且整个非对称投影格式的全景视频的分辨率低于全分辨率的全景视频。IEEE1857.9工作组设计了几种非对称投影格式,有效地降低了传输带宽和播放复杂度。

2.1 非对称圆形投影

非对称圆形投影(ASC)^[9]根据每个区域与主视点中心的角度将VR视频的球面划分成3个区域。如下图所示,C是主视点的中心,区域I是主视点区域,区域I与主视点中心之间的角度小于 Z_1 ,区域II是非主视点区域,区域和主视点中心之间的角度

从 Z_1 到 Z_2 ,区域III也是非主视点区域,区域和主视点中心之间的角度大于 Z_2 。ASC将图4 a)球面上的区域I、区域II和区域III分别投影到图4 b)中的二维平面上的区域I(半径为 $0\sim\rho_1$)、区域II(半径为 $\rho_1\sim\rho_2$)和区域III(半径为 $\rho_3\sim 0$)。

ASC在主视点区域使用等积投影,以确保主视点的采样密度高且均匀;在非主要视点区域,采样密度随着区域与主视点中心角度的增加而减小。另外,主视点区域的大小是可变的, Z_1 、 Z_2 和 ρ_1 可以根据应用场景、网络状况或其他因素自定义,因

此ASC投影具有很高的灵活性。

2.2 等角金字塔投影

等角金字塔投影(EAP)^[10]首先通过透视投影将球面投影到金字塔,然后将金字塔投影到二维平面。图5 a)是主视点中心为D时,通过透视投影将球面投影到金字塔的示意图。金字塔的底面对应于主视点区域(主视点区域的角度为 θ),而金字塔的侧面对应于非主视点区域。图5 b)是展开后金字塔的示意图。图5 c)是通过将金字塔投影到二维平面的示意图。

金字塔底面对应主视点区域,

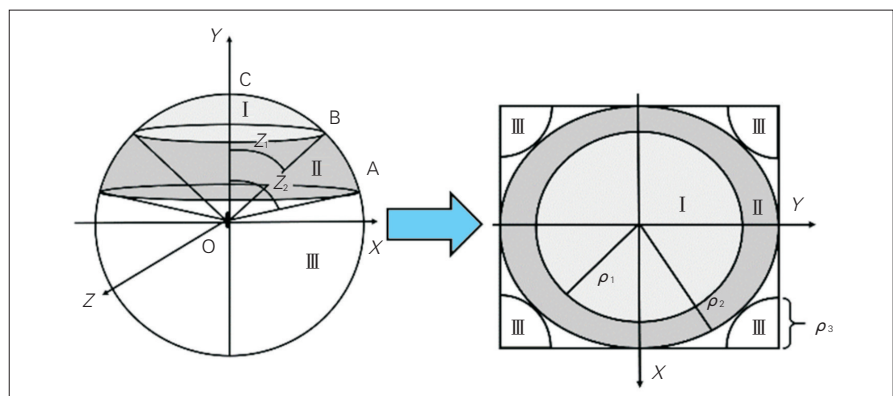


图4 非对称圆形投影示意

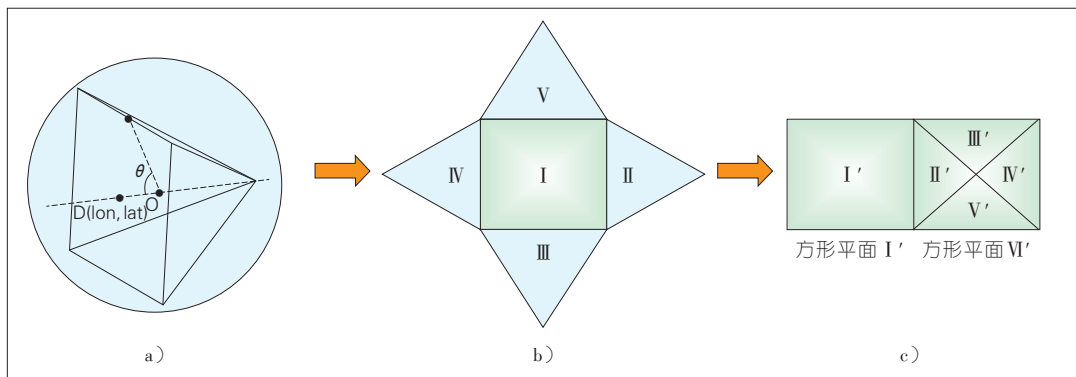


图5 等角金字塔投影的示意

EAP对金字塔的主视点区域使用等角投影;金字塔侧面对应非主视点区域,使用采样密度按梯度下降的方式投影,离主视点越远,采样密度越低。

2.3 非对称投影的自适应滤波方法

在非对称投影中,由于非主视点区域的采样密度较低,该区域产生锯齿现象。为了消除非主视点区域的锯齿现象,提出了一种用于非对称投影的自适应滤波方法^[11]。该方法将非主视点区域划分为多个子区域,并对不同的子区域使用不同强度的滤波器进行滤波。

如图6所示,对于EAP,自适应滤波方法将非主视点区域划分为4个子区域(A、B、C、D),分别对子区域A、B、C和D分别进行强滤波、中等强度滤波、弱滤波和无滤波。自适应滤波可以消除锯齿效应,降低码率。

3 不同映射格式编码效率的测试

为了评估不同投影格式的编码

效率,IEEE1857.9工作组进行了一系列的测试。

IEEE1857.9工作组使用LETIN VR^[12]提供的测试序列集,其中包含8个分辨率为4 096×2 048的ERP格式的VR视频序列,4个分辨率为8 192×4 096的ERP格式的VR视频序列。测试序列的长度为300帧,其帧率为30 f/s。

实验流程如图7所述。将分辨率更高的ERP格式的原始序列作为参考基准,然后使用不同的投影格式将原始序列投影到不同的平面格式。

本实验使用RD-VR16.1编码和解码不同投影格式生成的序列。最后,计算解码后的序列和原始序列之间的球面峰值信噪比(S-PSNR)值。

在实验中,将ERP格式作为实验基准。实验结果如表2所示,立方体投影、八面体投影、二十面体投影、双极方形投影和双环带投影的增益分别为:3.76%、2.93%、8.39%、11.56%和8.61%。

对非对称投影格式的测试流程如图8所示,使用RD-VR16.1(RA的默认配置)对ERP格式的序列进行编

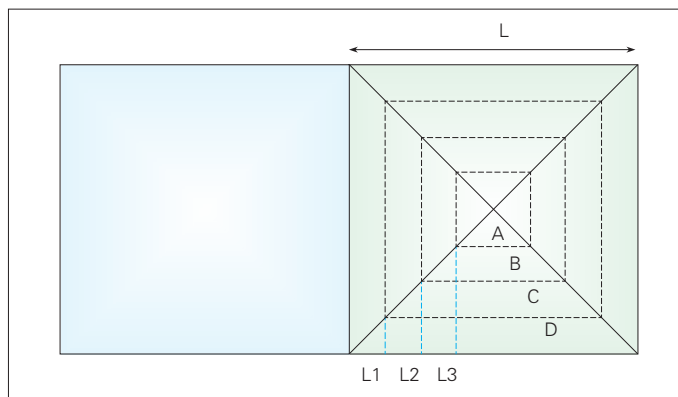


图6 自适应滤波示意

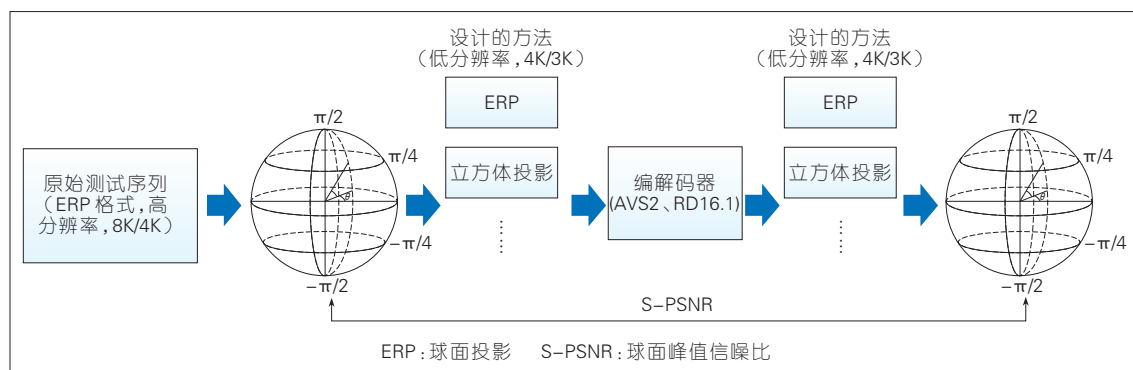


图7 实验流程

表2 不同格式的编码效率的实验结果

原始分辨率	序列	立方体投影	八面体投影	二十面体投影	双极方形投影	双环带投影
4 096x2 048	Fengjing_1	-19.81%	-22.16%	-25.40%	-29.27%	-25.17%
	Fengjing_3	-3.35%	-5.55%	-14.34%	-23.18%	-17.35%
	Hangpai_1	-8.09%	-3.52%	-7.16%	-10.89%	-9.28%
	Hangpai_2	-14.73%	-10.33%	-12.45%	-13.85%	-14.88%
	Hangpai_3	-7.08%	-3.46%	-7.87%	-9.04%	-9.30%
	Xinwen_1	-0.83%	6.38%	3.07%	-1.88%	0.25%
	Xinwen_2	-0.37%	10.74%	3.88%	-0.24%	3.62%
	Yanchanghui_2	3.73%	8.85%	0.64%	-7.31%	-1.59%
8 192x4 096	THEPLACE	16.14%	11.87%	0.29%	1.9%	5.74%
	Natatorium	-1.94%	-0.23%	-9.86%	-11.43%	-10.26%
	Highway	-13.75%	-19.43%	-26.87%	-27.25%	-25.00%
	Canolafield	6.02%	-1.83%	-4.58%	-6.24%	-0.15%
4K 均值	-6.31%	-2.38%	-7.45%	-11.96%	-9.21%	
8K 均值	1.62%	-2.40%	-10.25%	-10.75%	-7.42%	
所有序列的均值	-3.67%	-2.39%	-8.39%	-11.56%	-8.61%	

码和解码,然后计算解码序列与原始序列之间的加窗 S-PSNR。

非对称投影的实验过程如图9所示:首先将ERP格式的序列分别投影为ASC和EAP格式;然后,使用RD-VR16.1(RA的默认配置)对ASC和EAP格式的序列进行编码和解码;最

后,计算解码序列与原始序列之间的加窗 S-PSNR。实验参数的设置如下:ASC的分辨率为1 448x1 448, Z_1 、 Z_2 和 ρ_1 分别设置为45°、90°和512。EAP分辨率为2 048x1 024, θ 为45°。

实验结果如表3所示。由于ASC的主视点区域是圆形的,因此添加了

使用循环窗口计算加窗 S-PSNR 的一组数据。实验结果表明:ASC和EAP能在主视点区域实现超过50%的增益,并且EAP的BD-rate增益比ASC高出4%~10%,这主要是由于ASC存在被浪费的区域。但是ASC中的所有参数(Z_1 、 Z_2 和 θ)都是可调的,因此ASC比EAP更灵活。例如:当主视点区域设置为120°时,ASC可以实现比EAP更好的性能。另外,自适应滤波可以在EAP的基础上进一步将码率降低约11%,并且消除非主视点区域的锯齿现象。

4 结束语

IEEE1857.9工作组致力于标准化用于压缩、解压缩和重建沉浸式视觉内容的高效编码工具。该工作组已经设计了一系列针对VR视频的投影格式。被采纳的全景视频投影格式最多能节省30%的码率。此外,工作组为动态流切换应用设计了各种投影格式和自适应滤波方法,实现了70%的编码性能的增益。未来IEEE 1857.9工作组还将继续致力于沉浸

图8 基准实验流程

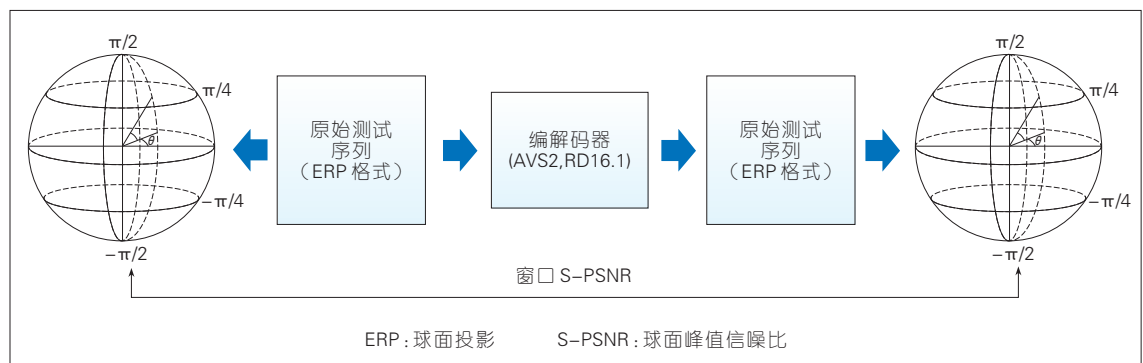
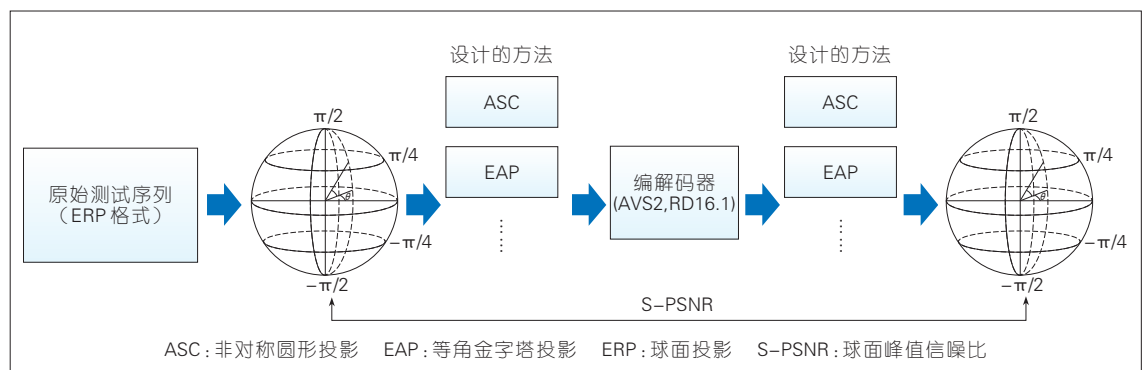


图9 非对称投影实验流程



▼表3 实验结果

窗口大小	方形窗口计算的S-PSNR			圆形窗口计算的S-PSNR
	ASC	EAP	EAP + 自适应滤波	ASC
70°	-54.71%	-59.28%	-70.19%	-55.08%
80°	-52.98%	-59.45%	-70.31%	-54.87%
90°	-49.07%	-58.45%	-69.61%	-54.36%
100°	-42.47%	-54.63%	-66.96%	-46.82%
110°	-36.56%	-50.70%	-63.89%	-40.94%
120°	-30.91%	-46.30%	-58.83%	-35.56%
130°	-24.84%	-41.20%	-50.71%	-30.25%
140°	-18.59%	-35.40%	-33.39%	-24.28%
150°	-11.32%	-28.21%	-11.65%	-18.02%

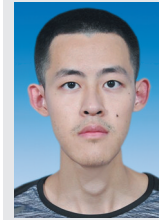
ASC: 非对称圆形投影 EAP: 等角金字塔投影 S-PSNR: 球面峰值信噪比

式视觉内容的高效编码工具的研究, 寻求新的高效的映射格式, 此外还将在3D沉浸式视觉内容的映射和编码工具上投入更多的研究。

参考文献

- [1] ZORIN D, BARR A H. Correction of Geometric Perceptual Distortions in Pictures [C]// Conference on Computer Graphics and Interactive Techniques. USA: ACM, 1995: 257-264. DOI:10.1145/218380.218449
- [2] SMOLIC A, MCCUTCHEN D. 3DAV Exploration of Video-Based Rendering Technology in MPEG[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2004, 14(3):348-356. DOI: 10.1109/TCSVT.2004.823395
- [3] NG K T, CHAN S C, SHUM H Y. Data Compression and Transmission Aspects of Panoramic Videos[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2005, 15(1):82-95. DOI: 10.1109/TCSVT.2004.839989
- [4] YU M, LAKSHMAN H, GIROD B. Content Adaptive Representations of Omnidirectional Videos for Cinematic Virtual Reality[C]// The 3rd International Workshop on Immersive Media Experiences. USA:ACM, 2015:1-6. DOI: 10.1145/2814347.2814348
- [5] LI J, WEN Z, LI S, et al. Novel Tile Segmentation Scheme for Omnidirectional Video[C]// IEEE International Conference on Image Processing. USA: IEEE, 2016:370-374.
- [6] CHOIK P, VLADYSLAVZ, CHOIM, et al. On 2D Representation Format of Panoramic Video: IEEE1857.9-04-M1027[S]. Guiyang: IEEE, 2016:6
- [7] WANG Y M, WANG R G, WANG Z Y, et al. A New Panoramic Video Projection Scheme: IEEE1857.9-04-M1028[S]. Guiyang: IEEE, 2016:6
- [8] LIN C, GU X, WU C J, et al. A New Projection Method of Omnidirectional Video: IEEE1857.9-04-M1025[S]. Guiyang: IEEE, 2016:6
- [9] WANG Y M, WANG R G, WANG Z Y, et al. A New Asymmetric Projection Scheme Based on ROI: IEEE1857.9_M1060/AVSVR_M1060 [S]. Haikou: IEEE, 2016:12
- [10] WANG Y M, WANG R G, WANG P, et al. Improved Pyramid Projection: IEEE1857.9_M1107/AVSVR_M1107[S]. Dalian: IEEE, 2017:8
- [11] SHENG X J, WU Y X, YIN H B, et al. An Improved Pyramid Projection Based on Adaptive Filtering: IEEE1857.9_M1111/AVSVR_M1111[S]. Dalian: IEEE, 2017:8
- [12] CHEN J, ZHANG J. Projection CE: Summary Report: IEEE1857.9_M1072/AVSVR_M1072 [S]. Haikou: IEEE, 2016:12

作者简介



王悦名, 北京大学信息工程学院数字媒体研究中心在读硕士研究生; 主要研究方向为视频编解码技术和虚拟现实视频编码技术; 已发表论文2篇, 申请发明专利3项。



王荣刚, 北京大学信息工程学院教授、院长助理, 国际MPEG互联网视频压缩标准(IVC)专题联合组长; IEEE1857.9专题组组长; 主要研究方向为新一代视频编解码技术、视频增强和虚拟现实视频技术等; 已发表论文100余篇, 申请国际/国内发明专利70余项。



高文, 北京大学教授、国家自然科学基金委员会副主任、中国工程院院士、ACM/IEEE Fellow; 主要研究方向为智能人机交互与多媒体技术、计算机视觉、多媒体数据压缩等; 曾获国家技术发明二等奖、科技进步二等奖和自然科学二等奖; 已出版著作6部, 在IEEE T-CSVT、IEEE T-MM、IEEE T-LP等顶级国际期刊发表论文近百篇。

《中兴通讯技术》2018年专题计划

期次	专题名	专题名
1	5G承载网技术和优化组网	中国联通研究院 张云勇院长
2	大数据智能化无线网络技术	重庆邮电大学 陈前斌教授
3	毫米波与太赫兹通信技术	东南大学 洪伟教授 东南大学 王海明教授
4	5G回传网络光电子器件技术	中兴光电子技术有限公司 孙笑晨技术总监 中兴光电子技术有限公司 徐积勇总经理
5	可再生能源供电的无线通信	清华大学 牛志升教授
6	区块链技术及其物联网应用	北京航空航天大学 刘建伟教授