

HEVC 兼容的全景视频运动补偿预测算法

Motion Compensation and Prediction for Panoramic Video Coding on HEVC

胡强/HU Qiang¹
马文广/MA Wenguang²
石志儒/SHI Zhiru²

(1. 上海交通大学, 上海 200240;
2. 上海交通大学, 上海 201210)
(1. Shanghai Jiao Tong University,
Shanghai 200240, China;
2. ShanghaiTech University, Shanghai
201210, China)

1 三维球面投影及所提 编码框架

随着消费级虚拟现实 (VR) 技术的不断完善, HTC VIVE、Oculus Rift、三星 Gear VR、Deepon 等 VR 设备得到井喷式发展。作为能够快速制作推广的全景视频也随之进入公众的视野, 在景观展示、视频会议、远程医疗和会诊、远程监测控制、机器人自动驾驶等领域得到广泛应用。目前 VR 视频主要有两种生产途径: 一种是通过渲染引擎创作的纯虚拟场景; 另一种是通过相机阵列拍摄拼接的 360° 视频^[1]。

使用者在观看 360° 全景视频, 在同一位置通过改变观察方向, 能够浏览四周全部场景^[2]。360° 视频包含了四周空间的全部运动信息, 利用虚拟现实头盔显示观看时, 只选取其中一部分视野进行渲染显示, 所以全景视频需要更清晰 (至少空间分辨率 4 K 以上), 才能匹配现在人们对高清

收稿日期: 2017-10-20

网络出版日期: 2017-11-13

基金项目: 上海市科技人才计划项目 (15XD1502900); 上海市科委项目 (17JC1403800)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 06-0028-004

摘要: 利用图像渲染技术, 全景视频能够为终端用户提供沉浸的虚拟现实体验。不同于传统二维视频, 360° 全景视频不仅包含平移运动还包括仿射运动, 后者主要由于球面映射所产生的畸变。特别是对于广泛使用的经纬球面投影, 其两极区域的扭曲失真特别严重。提出了一种全新的运动补偿预测算法, 其将当前像素投影回球面并寻找其匹配的参考像素块。在此基础上利用球面上运动相对一致, 提出了一种新的运动向量预测 (MVP) 模式。实验结果表明: 该算法与传统高效视频编码器 (HEVC) 相比, 最高提升 3.65% 平均码流差 (BDBR) 的编码性能。

关键词: 全景视频; 虚拟现实; 运动补偿; 视频编码

Abstract: Virtual reality can provide immersive panoramic video for the end user experience by image rendering technology. Different from the traditional 2D video, 360° panoramic video includes the translational motion and the affine motion. The affine motion is mainly due to the distortion generated by the spherical mapping. For the latitude and longitude spherical projection which is widely used, the polar area distortion is especially serious. In this paper, a novel motion compensated prediction scheme is proposed to deal with non-translational motions for panoramic video. Furthermore, a compensation of motion vector predictor (MVP) is employed on the sphere plane to further improve the coding performance. Experimental results show that the proposed method reaches a maximum 3.65% Bjontegaard Delta Bitrate (BDBR) reductions compared to high efficiency video coding (HEVC).

Keywords: panoramic video; virtual reality; motion compensation; video coding

分辨率的体验。全景视频的数据量相比于传统视频是成倍的增长趋势, VR 全景视频的点播和直播的传输面临带宽瓶颈^[3]。传统视频编解码的方案不太适应全新的视频格式的需求, 需要根据全景视频的特性提出有针对性的编解码方案, 满足存储和传输的需求^[4]。

360° 全景视频应用主要包括前端相机阵列的视频同步采集、拼接、编码及传输和播放。其中多相机阵列同步采集的各视角视频, 进行校正配准之后, 对重叠的图像无缝拼接, 需要将图像进行投影变换到三维球面模上。投影到球体表面模型进

行拼接处理, 然后再以一定的投影模式变换成二维图像, 进行压缩传输和储存。所谓的投影变换是指将全景视域的三维场景和二维平面相互映射过程, 作为全景视频内容的最广泛的转播载体, 投影后的二维视频, 最终通过渲染在虚拟现实终端设备进行观看。在投影变化和反变换过程中, 存在着图像失真, 运动一致被破坏等问题, 传统的视频编码技术在应对 360° 全景视频的效果并不理想^[5-7]。

针对以上存在的问题, 由国际电信联盟电信标准分局 (ITU-T)/视频编码专家组 (VCEG) 和国际标准化组织-国际电工委员会 (ISO-IEC)/动态

图像专家组(MPEG)成立的联合视频研究组(JVET)将 360° 全景视频纳入到未来视频编码标准范围内^[8]。与传统的高效视频编码器(HEVC)/H.265编码标准^[9]相比,正在制订的联合视频探测组(JVET)标准,将通过优化投影变换避免过多的扭曲失真,作为一个重要的实验测试和研究方向,从而避免过多的扭曲失真产生的回放质量。目前使用最为广泛的是球面投影(ERP)和立方图投影(CMP)。

球面投影是把球型表面的内容映射到平面的矩形图中,如图1所示,其对自然图像有良好的兼容性,在转换过程中不同维度区域有不同的延展程度,赤道区域的失真比两级区域小很多,像素的数量在两级被放大,运动方向扭曲,产生大量的时间和空间的冗余信息^[10]。而立方图投影通过外接立方体,将球体表面内容投影到立方体表面。展开后按照一定的布局格式进行布局,通常的布局方法有 3×2 布局和 4×3 布局,如图2所示。不用的布局方案对编码效率和内存管理都有不同程度的影响。相比球面投影,立方图投影节省了25%的像素面积,对比特率有较大的提升,但存在像素密度减少照成的模糊问题^[11]。文章中,我们针对全景视频投影过程中产生畸变和运动扭曲,所带来的编码问题,提出了一种全新的运动补偿算法。

在传统的视频编码算法中,基于块匹配的运动估计和运动补偿得到广泛的应用。其建立在相邻时域帧内的二维物体所做的都是平移运动,大小形状不改变的基础上,对动态帧之间的时域冗余有非常高压缩效果。但是对于全景视频,由于从三维球面坐标投影到二维平面坐标转换的过程中,产生了一系列相机旋转、缩放、错切运动的形变,影响了运动的连贯性和一致性,从而降低了帧间预测的效率。

针对全景视频出现的大量非平

移运动,我们提出了一种全新的运动补偿和预测的模式,用于提高 360° 全景视频的时域编码效率。为避免几何失真对运动预测带来的影响,首先将当前编码块的像素投影到三维球面坐标,利用球面运动一致性的特点,在球面坐标系中找相对应的参考像素,并最终将三维球面坐标投影回二维坐标。更进一步,我们提出了基于球面坐标的运动向量预测方法,避免了投影变换带来非平移运动的影响,视频编码框架如图3所示。我们提出的一整套运动补偿预测算法都在块划分的基础上提出,可以作为HEVC编码的附加模式,对 360° 全景视频有着更高效的编码压缩性能。

2 基于三维球面投影的运动补偿算法

为了提升 360° 全景视频的编码压缩性能,我们针对全景视频中大量产生的非平移运动,进行了深入的分

析,其中绝大部分非平移运动,是由投影变换产生的畸变导致。针对这种情况,我们在HEVC编码框架下提出了一种全新的运动补偿模式;基于三维球面投影的运动补偿(SphMC)。整体算法流程如图3所示。本算法作为 360° 视频的一种补充,同时存在于传统的帧间预测的框架内。HEVC编码过程中,最优的预测模式仍然是遍历所有块划分和预测模式后,选择最小率失真值的模式,作为最优的编码模式。其中SphMC模式的加入,提供了一种针对全景视频的运动补偿模式,对整体编码性能有明显的增益。

基于前面章节分析,纹理的不连续性会严重影响运动预测的性能,针对全景投影视频的有效参考像素,需要从运动连续的三维球面坐标系中推导而来。文中涉及到的投影变换都以立方体投影为例,SphMC模式的参考块像素点的展示在图4中。假

图1 球面三维坐标与球面投影变换

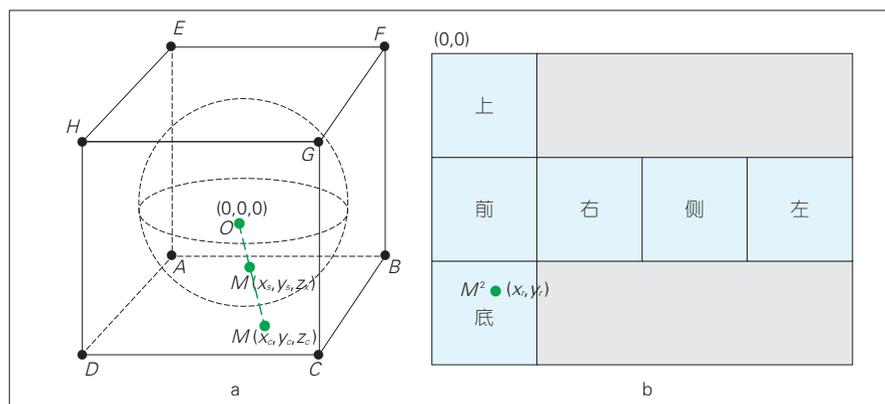
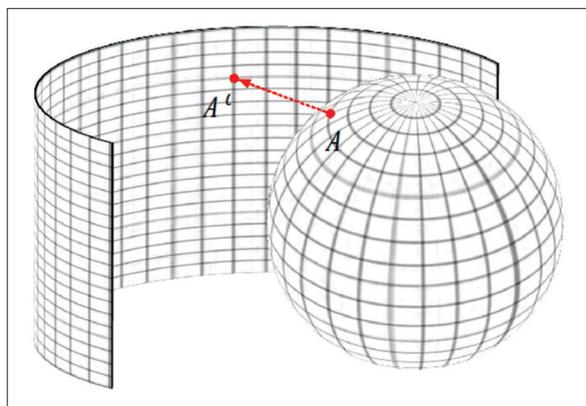
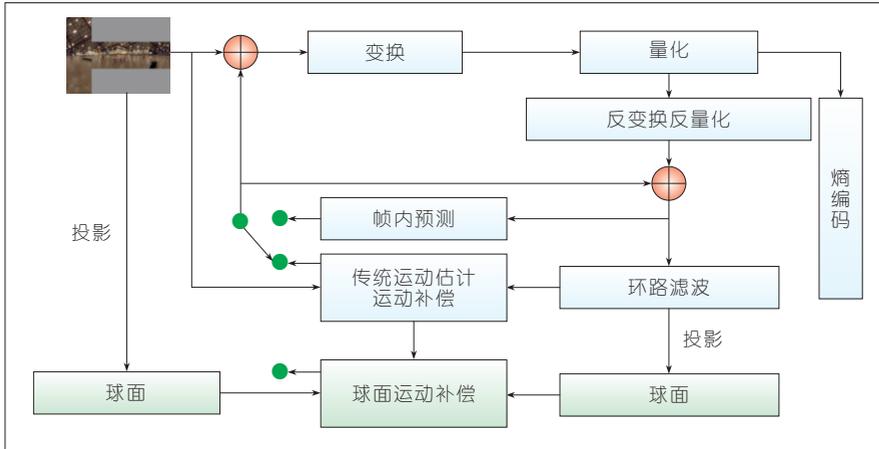


图2 球面三维坐标与立方图投影变换



▲图3 基于三维球面投影的运动补偿和预测模式

设当前编码块所在帧 t_i ，该块中心位置 n_0 移动到帧 t_i 帧的 n_2 位置，基于三维球面的投影问题就简化为：在参考帧中利用球面投影变换，寻找当前编码块 n_1 的对应参考点 n_3 。

为了在参考帧得到出对应的点 n_3 的位置，其在三维球面上的坐标位置首先会被推出来，然后再投影回立方图。在这个投影关系中，立方图和三维球面坐标是一一对应。 $s = P(n)$ 代表了从立方图平面 N 到三维全面 S 的映射，反变换由 $n = P^{-1}(s)$ 表示。所以点 n_0 、 n_1 和 n_2 的坐标投影到三维球面分别为： $s_0 = P(n_0)$ 、 $s_1 = P(n_1)$ 和 $s_2 = P(n_2)$ 。在三维球面的表达中，绝大多数的像素点和运动具有一致性，所以 s_3 的可通过以下式(1)得到：

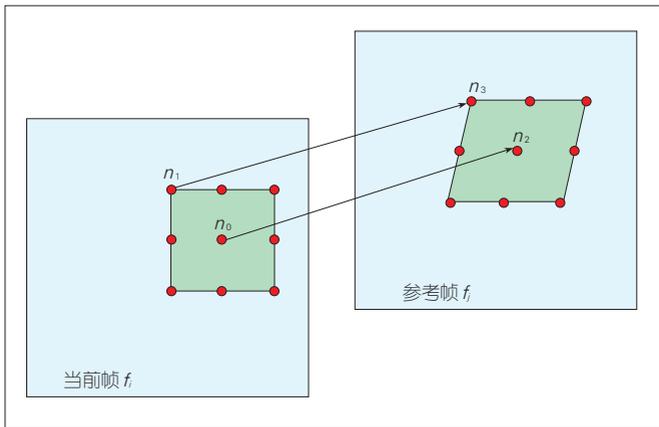
$$s_3 = P(n_2) + P(n_1) - P(n_0) \quad (1)$$

得到在三维球面坐标点 s_3 后，其

在立方图坐标系中的相对应的二维坐标 n_3 可以通过反投影 $n_3 = P^{-1}(s_3)$ 得到。其余参考点可以通过相似的方法推演得到。对于非整像素的位置，可通过双线性差值得到。最后，基于三维球面投影得到的参考区域与编码块进行运动补偿的计算。

3 基于三维球面投影的运动向量预测算法

在 HEVC 帧间预测过程中，会针对空域和时域上已经确定的运动向量，来预测当前块运动估计的初始向量方位，该过程通过相邻预测单元 (PU) 的运动向量来进行预测。但对于 360° 全景视频，由于投影变换带来的空间不连续，传统的基于临块的运动预测已经不太适用。本章中，我们提供了一种基于三维球面投影的预测向量，其针对不同 PU 的相对距离，



▲图4 基于三维球面投影的运动补偿算法

如图5所示。

针对 360° 全景视频的帧间预测，假设 n_4 和 n_5 分别为分块 C 和分块 B_1 的中心点， MV_1 为分块 B_1 的运动向量，而 MV_0 代表了当前编码块 C 的预测向量。类似于我们提出的运动补偿方法， n_4 和 n_5 都利用二维立体图平面投影回三维球面坐标，从而寻找正确的运动预测向量。我们可以得到以下关系：

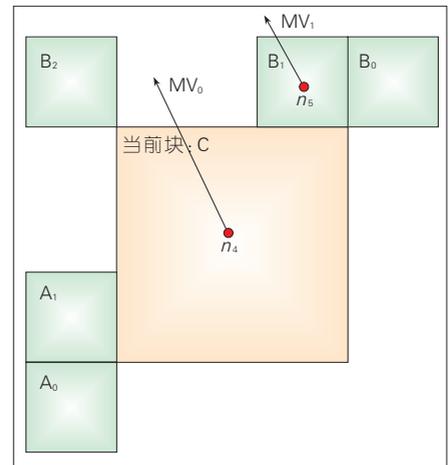
$$\begin{aligned} P(n_2 + MV_0) - P(n_4) = \\ P(n_5 + MV_1) - P(n_5) \end{aligned} \quad (2)$$

所以，基于球面投影的预测向量 MV_0 可以由此得到：

$$MV_0 = P^{-1}(P(n_5 + MV_1) + P(n_4) - P(n_5)) - n_4 \quad (3)$$

4 实验结果

为了验证所提算法的有效性，基于投影变换的运动补偿算法已集成到了 HEVC 标准 HM16.15^[12] 测试软件中，通过对比 HM16.15 与该算法的性能，证明了改算法的有效性。基于 HEVC 常用测试条件^[12]，我们主要测试 3 种配置：randomaccess_main (RA_main)、lowdelay_main (LD_main)、lowdelay_P_main (LP_main)。所采用的 QP 分别是 22、27、32、37，包括高比特率和低比特率的情况。鉴于该算法的目标在于更好地表征 360° 视频立



▲图5 针对球面投影模式的运动向量预测

体图投影中的无规则运动,我们选择了一些相对于立方体 4×3 格式有更大运动的序列来表明该算法的优势。测试序列信息如表1所示。另外,由于这些序列都具有极高的空间分辨率,因此我们只测试了1s所对应的帧数来证明算法的有效性。

RA_main、LD_main、LP_main的实验结果分别如表2、表3所示。从这些表格中,我们可以看到:通过该算法,在RA_main、LD_main两种配置下,视频Y通道分别可达到1.83%,1.21%,1.19%平均码流差(BDBR)^[13]的减少。鉴于编码顺序相邻帧之间

▼表1 实验所用测试序列

测试序列	分辨率	帧数
SkateboardInLot	4 736 × 3 552	32
DrivingInCountry	3 840 × 2 880	24
Bicyclist	3 840 × 2 880	24
Glacier	3 840 × 2 880	24
ParamotorTraining	3 840 × 2 880	24
TimelapseBuilding	3 840 × 2 880	24

▼表2 RA-Main配置下所提算法实验结果

测试序列	BDBR/(%)	BDBR/(%)	BDBR/(%)
SkateboardInLot	-1.33	-0.98	-0.85
DrivingInCountry	-2.72	-0.73	-1.31
Bicyclist	-1.15	-1.02	-1.44
Glacier	-3.65	-2.72	-3.04
ParamotorTraining	-0.42	-0.33	-0.56
TimelapseBuilding	-1.72	-1.24	-1.44
平均	-1.83	-1.17	-1.44

BDBR: 平均码流差

▼表3 LD-Main配置下所提算法实验结果

测试序列	BDBR/(%)	BDBR/(%)	BDBR/(%)
SkateboardInLot	-1.24	-0.23	-0.65
DrivingInCountry	-1.83	-0.68	-1.12
Bicyclist	-1.12	-0.31	-1.08
Glacier	-1.53	-0.62	-0.93
ParamotorTraining	-0.34	-0.15	-0.21
TimelapseBuilding	-1.23	-0.75	-0.71
平均	-1.21	-0.45	-0.78

BDBR: 平均码流差

的平均距离要比LD_main和LP_main中的距离大一些,RA中的不规则运动会更加明显。这就是为什么我们能实现更好的R-D性能。特别是对于具有较大运动的序列Glacier,算法可将RA_main中Y通道的平均码流差减少3.65%。因此,对于具有较大运动的序列,就可实现更显著的平均码流差的减少。

5 结束语

360°全景视频与传统的视频存在着许多差别,传统的HEVC编码不能达到最理想的压缩效果,我们分析了由于投影变换带来的运动失真的问题,并相应地提出了一种基于球面投影的运动补偿和预测算法,其主要利用三维球面的运动具有较强的一致性,通过立方体二维坐标与球面三维坐标的相互转换,寻找出新的参考像素和运动向量预测。经过大规模的实验,结果表明:该算法与传统HEVC相比,最高提升3.65%BDBR的编码性能。在后续的研究中,以下问题将会着重进行考虑:提出针对360°全景视频的运动估计算法,并且修改相应的率失真准则,以更好地适应360°全景视频的编码需求。

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]// Processing of SMPTE Annual Technical Conference and Exhibition. USA:IEEE, 2016: 1-20. DOI: 10.5594/M001733
- [2] 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
- [3] NG K T, CHAN S C, SHUM H Y. Data Compression and Transmission Aspects of Panoramic Videos [J]. IEEE Transaction on Circuits and Systems for Video Technology, 2005, 15 (1): 82-95. DOI: 10.1109/TCSVT.2004.839989
- [4] 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
- [5] DE LA, FUENTE Y S, SKUPIN R, et al. Video Processing for Panoramic Streaming Using HEVC and its Scalable Extensions [J]. Multimedia Tools and Applications, 2016: 1-29
- [6] ZHU Z G, XU G Y, 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
- [7] ZHENG J L, SHEN Y F, ZHANG Y D, et al. Adaptive Selection of Motion Models for Panoramic Video Coding [C]// Process of IEEE International Conference on Multimedia and Expo.USA:IEEE, 2007: 1319-1320
- [8] 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, ITU, 2016
- [9] SULLIVAN G J, OHM J, WOO J H, et al. Overview of the High Efficiency Video Coding (HEVC) Standard [C]//IEEE Transaction on Circuits System Video Technology. USA:IEEE, 2012,(22):1649-1668
- [10] LIN H, LIN J L, CHANG S K, 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. ITU, 2016
- [11] KUZUYAKOV E, PIO D. Next-Generation Video Encoding Techniques for 360 Video and VR [EB/OL]. (2016-07-15)[2017-11-07].https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/. 2016. 7. 15
- [12] JCT-VC Subversion Respository for the HEVC Test Model Version HM16.15[EB/OL]. [2017-11-07]. https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.15
- [13] BOSSEN F. Common HM Test Conditions and Software Reference Configurations, Document JCTVC-K1100[R]. China: Shanghai, 2012

作者简介



胡强,上海交通大学图像通信与网络工程研究所在读博士;研究方向主要包括图像处理和视频编码;已发表论文4篇,获得专利4项。



马文风,上海科技大学信息与技术学院研究助理;研究方向主要包括光场摄影技术、计算成像技术及相关设备制作以及虚拟现实技术。



石志儒,2012年在英国萨里大学工程系获得博士学位,之后于上海交通大学电子信息与工程学院从事博士后研究,现任上海科技大学副研究员;从事计算机视觉、视频编解码、虚拟现实等方面的研究;已发表学术论文10余篇。