

## 基于球域失真空-时依赖的全景视频编码

杨栩<sup>1,2</sup>, 朱策<sup>3</sup>, 郭红伟<sup>3,4</sup>, 罗雷<sup>1,3</sup>

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 成都师范学院物理与工程技术学院, 四川 成都 611130;  
3. 电子科技大学信息与通信工程学院, 四川 成都 611731; 4. 红河学院工学院, 云南 蒙自 661100)

**摘要:** 全景视频平面编码失真和球域感知失真不同域使主客观质量评价不一致, 进而损失编码性能。此外, 独立率失真优化技术没有考虑球域失真的时域依赖性对编码的影响, 编码性能还有提升空间。针对上述问题, 提出一种空-时域依赖的球域失真模型以优化全景视频编码。首先, 提出一种球域失真到编码失真的空域映射模型, 使主客观质量评价趋近一致; 然后, 提出一种球域失真时域传播模型, 以提升传播链上所有编码单元的整体编码性能; 最后, 计算球域失真空域映射权重和时域传播权重来调整编码参数。实验结果表明, 在低延时编码配置下, 相较于通用视频编码基准 VTM14.0, 所提算法有平均 7.4% (最高达 22.1%) 的码率节省和 9% 的编码时间节省。  
**关键词:** 全景视频编码; 时域依赖率失真优化; 投影; 球域失真; 编码失真

中图分类号: TN919.81

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023197

## Panoramic video coding based on spherical distortion with spatio-temporal dependency

YANG Xu<sup>1,2</sup>, ZHU Ce<sup>3</sup>, GUO Hongwei<sup>3,4</sup>, LUO Lei<sup>1,3</sup>

1. School of Information and Communication Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China  
2. School of Physics and Engineering Technology, Chengdu Normal University, Chengdu 611130, China  
3. School of Communication and Information Engineering, University of Electronic Science and Technology, Chengdu 611731, China  
4. School of Engineering, Honghe University, Mengzi 661100, China

**Abstract:** The planar coding distortion, which affects objective quality, and spherical distortion, which affects subjective quality, as well as existing independent rate-distortion optimization techniques that fail to consider the temporal propagation of spherical distortion and its impact on coding performance, result in coding performance degradation. To address these issues, a spatio-temporal dependent spherical distortion model was proposed for optimizing panoramic video coding. Firstly, a spatial mapping model was proposed to map spherical distortions to coding distortions, aiming to align subjective and objective quality assessments. Secondly, a temporal propagation model for spherical distortions was introduced to enhance the overall coding performance of all coding units in the propagation chain. Finally, the coding parameters were adjusted by computing the weights for spatial mapping and temporal propagation of spherical distortions. Experimental results demonstrate that, under low-delay encoding configurations, compared to the VTM14.0, a state-of-the-art video coding benchmark, the proposed algorithm achieves an average bitrate savings of 7.4% (up to 22.1%) and reduction in coding time of 9%.

**Keywords:** panoramic video coding, temporal dependent rate-distortion optimization, projection, spherical distortion, coding distortion

收稿日期: 2023-07-11; 修回日期: 2023-09-28

通信作者: 朱策, eczhu@uestc.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62020106011, No.U19A2052, No.62061015); 四川省科技厅基金资助项目 (No.2022ZHCG0116); 重庆市自然科学基金资助项目 (No.2023NSCQ-MSX2930); 重庆邮电大学青创基金资助项目 (No.SCIE-QN-2022-05)

**Foundation Items:** The National Natural Science Foundation of China (No.62020106011, No.U19A2052, No.62061015), The Project of the Science and Technology Department in Sichuan Province (No.2022ZHCG0116), The Chongqing Natural Science Foundation (No.2023NSCQ-MSX2930), Youth Innovation Group Support Program of ICE Discipline of CQUPT (No.SCIE-QN-2022-05)

## 0 引言

随着 Web3.0 的到来和元宇宙的兴起, 虚拟现实 (VR, virtual reality) 和增强现实 (AR, augmented reality) [1] 技术为用户提供了沉浸式交互视听体验, 广泛应用于游戏、监控、视频会议等领域。相较于传统 2D 视频, VR 全景视频<sup>[2-3]</sup>包含  $360^{\circ} \times 180^{\circ}$  的全方位球域视觉信息, 具有高分辨率、高帧率以及大视场等特征<sup>[4]</sup>, 数据量极大, 这给传输和存储全景视频带来巨大挑战。因此, 迫切需要高效的全景视频压缩技术来应对这些挑战。

全景视频投影到编码平面后再使用传统视频编码器 (如高效视频编码 (HEVC, high efficiency video coding) 标准 H.265/HEVC<sup>[5]</sup>、通用视频编码 (VVC, versatile video coding) 标准 H.266/VVC<sup>[6]</sup> 对应的编码器) 进行压缩。然而, 全景视频在投影过程中会产生边界和拉伸形变, 影响跨边界的运动搜索和形变区域的编码性能。因此, 在全景视频编码方面, 一些研究聚焦于投影格式 (如等距柱状投影 (ERP, equi-rectangular projection) <sup>[7]</sup>、正六面体投影 (CMP, cubemap projection) <sup>[8]</sup>、正八面体投影 (OHP, octahedron projection) <sup>[3]</sup> 等), 以期找到既具有较少的投影边界又能产生较小拉伸形变的投影格式。然而, 很难找到能同时满足这 2 个条件的投影格式, 因为投影平面越少, 投影边界越少, 但拉伸形变越严重; 反之, 投影平面越多, 拉伸形变越小, 但投影边界也越多, 导致组合后的矩形编码平面的图像不连续。由于 CMP 失真相对较小, 正六面体的展开平面能成为一个独立的矩形编码平面, 符合编码器基于矩形块划分的编码特性, 因此 CMP 成为传统编码和智能编码<sup>[9]</sup> 的主要研究对象之一, 本文将在 CMP 上建立优化模型。

还有一些研究集中在改进编码算法, 以弥补投影带来的编码性能损失。一类研究关注投影边界, 例如文献[8, 10]通过填充 CMP 边界区域来提升跨边界运动搜索效果。然而, 由于投影边界占比较小, 性能提升非常有限, 特别是对运动变化不大的全景视频几乎没有效果。在此基础上, 文献[11]不仅进行了边界填充拓展, 还建立了符合全景视频特征的率失真优化模型, 从而提升了编码性能。另一类研究关注投影产生的拉伸形变, 文献[12]针对 CMP 提出基于熵平衡的比特分配算法以匹配拉伸区域的形变程度, 进而提升率失真性能。文献[13]根据

CMP 在平面中心区域与外围区域的形变程度来调整量化参数 (QP, quantization parameter), 取得和文献[12]差不多的编码性能。文献[14]根据 CMP 的形变程度, 提出一种球域率失真优化算法。虽然这些算法对全景视频编码比较有效, 但是其仅考虑投影区域的拉伸形变对编码性能的影响, 没有考虑编码帧间的时域依赖关系, 因此率失真性能还有较大的提升空间。

视频编码的帧间预测技术使当前编码单元的失真不仅影响其自身的编码性能, 还会影响后续编码帧中参考该单元的编码块的率失真性能, 文献[15-20]对 2D 平面的时域依赖率失真优化模型进行了研究。但该模型无法直接应用于全景视频编码, 并且该模型的复杂度较高, 这对超高分辨率的全景视频编码来说是不可行的。

本文针对正六面体投影的全景视频编码, 提出空-时依赖的球域失真模型, 主要贡献如下。

1) 在 CMP 上建立平面编码失真与球域感知失真的空域映射模型。像素均匀分布在球面, 投影后拉伸平面区域的冗余像素与其面积成正比, 以此来划分平面编码区域及球面环带区域并计算相应面积, 并将面积比作为球域失真空域映射权重。

2) 根据运动补偿预测的时域参考关系建立球域失真时域传播模型。为了降低编码复杂度和减少缓存空间, 建立失真后向传播链。通过运动补偿预测误差和重建误差的关系来计算失真时域影响权重。

3) 根据帧级时域相关性自适应调整帧级量化参数, 根据球域失真空域映射权重和时域影响权重来调整编码树单元 (CTU, coding tree unit) 级拉格朗日乘子和 QP。实验结果显示, 在相同编码质量下, 本文算法在所有全景视频测试序列上有平均超过 7% 的码率节省。

## 1 球域失真真空-时域依赖优化模型

全景视频需要投影到 2D 平面进行编码, 解码后反投影回球面供人观看, 这种框架使主观视觉感受受到球域失真而非编码失真的影响。然而, 客观质量评价受基于均方误差 (MSE, mean square error) 的编码失真影响。不一致的主客观质量评价会损失编码性能。此外, 运动补偿预测误差使球域失真在时域传播, 当前编码单元的失真不仅影响自身的编码性能, 还会影响后续与它具有参考关系的编码单

元的编码性能。针对以上两点，本文提出一种空-时域依赖的球域失真优化模型。

为了使全景视频编码的客观质量评价与主观质量评价保持一致，全景视频在平面编码的率失真优化目标是在给定的目标比特内尽可能地减少球域感知失真而非编码失真，也即

$$\begin{aligned} \min \sum_{i=1}^N D_i^s \\ \text{s.t. } \sum_i R_i \leq R_T \end{aligned} \quad (1)$$

其中， $D_i^s$  及  $R_i$  分别为第  $i$  块编码单元所对应的球域感知失真和平面编码比特， $R_T$  为编码比特预算， $N$  为编码单元总数。

如图 1 所示，球面图像会投影到正六面体的 6 个面以适应编码器的矩形图像编码规则。在实际编码过程中，编码器会将编码平面划分为有限个编码单元，并按照行和列排列。根据球面图像到编码平面再到编码单元划分之间的一一对应关系可以看出，每一个编码单元都对应着唯一的一块球面区域。虽然像素点均匀分布在球面，但是不同位置的球面图像到编码平面的投影距离不同，导致与每个编码单元对应的球面图像在投影后会出现不同程度的拉伸形变，拉伸区域的空缺像素通过插值填充，但这些插值像素并不真实存在，而是成为冗余像素并对编码域的主客观质量评价产生影响。

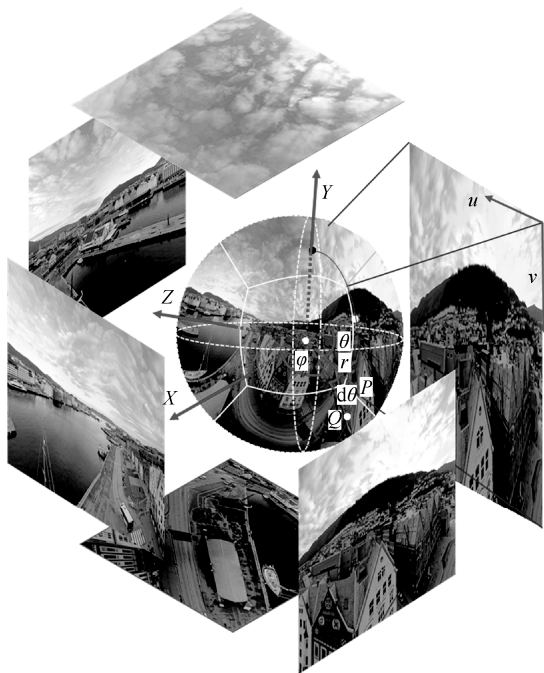


图 1 全景视频正六面体投影

为了使影响球域主观质量的评价指标球域失真  $D_i^s$  和编码域的质量评价指标编码失真  $D_i$  趋于一致，引入编码单元的空域映射权重来抵消冗余像素对编码域编码失真的不利影响。所以，CMP 视频序列的平面编码失真与球域感知失真之间存在以下映射关系

$$\sum_{i=1}^N D_i^s = \sum_{i=1}^N \omega_i^s D_i \quad (2)$$

其中， $\omega_i^s$  和  $D_i$  分别为第  $i$  个编码单元所对应的球域失真映射权重以及编码失真。

根据拉格朗日乘子法<sup>[18]</sup>，合并式(1)和式(2)，将约束性优化问题转换为无约束优化问题，则球域率失真优化模型变为

$$\min_{o_1, \dots, o_N} \left( \sum_{i=1}^N \omega_i^s D_i(o_1, \dots, o_i) + \lambda \sum_{i=1}^N R_i(o_1, \dots, o_i) \right) \quad (3)$$

其中， $\lambda$  为全局拉格朗日乘子，用于平衡失真和码率，较小的  $\lambda$  使失真较小但消耗的码率较大，较大的  $\lambda$  使失真较大但消耗的码率较小； $o_1, \dots, o_N$  为编码单元的编码参数，比如量化参数、编码模式、参考帧索引等。由于编码参数之间存在耦合关系，因此式(3)的全局率失真优化问题在编码中难以实现。

视频编码采用独立率失真优化技术，其易于实现且复杂度较低。编码时，假设编码单元间相互独立，对每个编码单元进行独立率失真优化。当前编码单元的率失真性能仅取决于当前编码单元的编码参数，与其他编码单元的编码参数无关，即

$$\min_{o_i} (D_i(o_i) + \lambda_i R_i(o_i)) \quad (4)$$

其中， $\lambda_i$  及  $o_i$  分别为当前编码单元的拉格朗日乘子及编码参数。然而，独立率失真优化使编码性能还有很大的提升空间。

根据文献[18]，编码单元的码率仅由其自身的编码参数决定，而失真与其参考单元的失真具有很强的相关性，失真会通过运动补偿预测在时域传播。所以，有必要建立失真传播链并量化失真的影响大小。

除非场景切换，视频图像具有时域连续性。帧间预测技术充分利用这种特性来减少视频信号在时域上的冗余，进一步提高了压缩率。已编码单元的失真会影响后续以其为参考的编码单元的失真，从而形成一条失真时域传播链，如图 2 所示。图 2(a)为球域失真传播链；图 2(b)为球面图像投影到六面体后的失真传播链；图 2(c)为编码平面上的失真传播链。

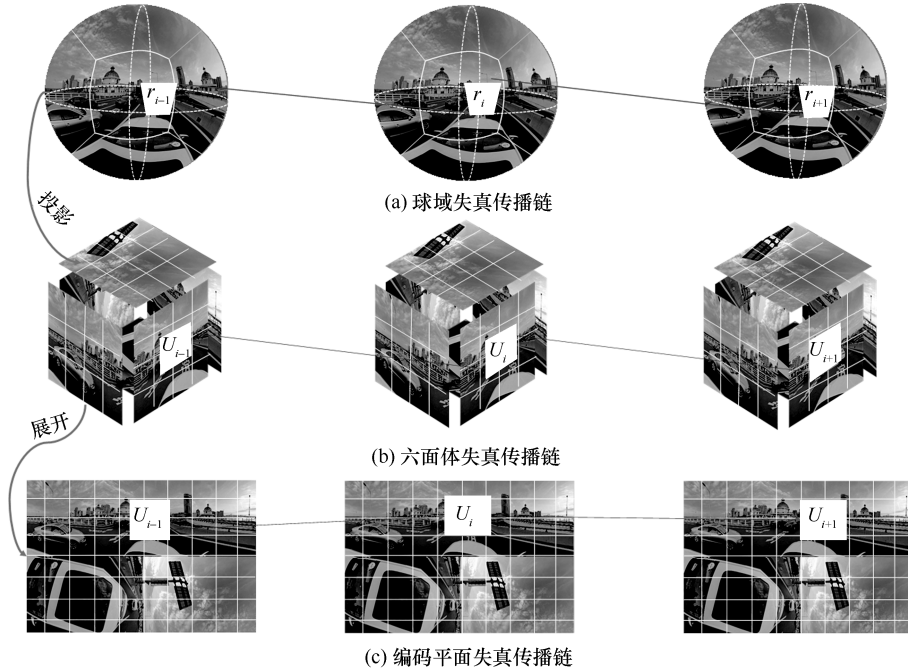


图 2 失真时域传播链

根据文献[18], 失真时域传播链上第  $i+1$  个编码单元  $U_{i+1}$  的重建失真  $D_{i+1}$  和运动补偿预测 (MCP, motion compensation prediction) 失真  $D_{i+1}^{\text{mcp}}$  存在如下关系

$$D_{i+1} = e^{-bR_{i+1}} D_{i+1}^{\text{mcp}} \quad (5)$$

其中,  $b$  为与视频信号源相关的常量;  $R_{i+1}$  为编码码率, 主要由自身的编码参数决定。  $D_{i+1}^{\text{mcp}}$  可以表示为

$$\begin{aligned} D_{i+1}^{\text{mcp}} &= \|U_{i+1} - \bar{U}_i\|^2 = \\ &= \|(U_{i+1} - U_i) + (U_i - \bar{U}_i)\|^2 \approx \\ &= \alpha(\|U_{i+1} - U_i\|^2 + \|U_i - \bar{U}_i\|^2) = \\ &= \alpha(D_{i+1}^{\text{omcp}} + D_i) \end{aligned} \quad (6)$$

其中,  $\alpha$  为接近 1 的常量;  $\|\cdot\|$  为 2 范数运算符;  $\bar{U}_i$  为重建编码单元;  $D_{i+1}^{\text{omcp}}$  为第  $i+1$  个编码块对源参考块的运动补偿预测误差 (OMCP, original MCP), 与编码参数无关。结合式(5)及式(6)可得

$$\begin{aligned} D_{i+1} &= \alpha e^{-bR_{i+1}} (D_{i+1}^{\text{omcp}} + D_i) = \\ &= \beta_{i+1} (D_{i+1}^{\text{omcp}} + D_i) = \\ &= \beta_{i+1} D_i + \gamma_{i+1} \end{aligned} \quad (7)$$

其中,  $\beta_{i+1} = e^{-bR_{i+1}} \alpha$  为失真影响因子;  $\gamma_{i+1}$  为与源运

动补偿预测误差相关项, 与编码参数无关。

依次类推, 传播链上第  $N$  个编码单元的失真  $D_N$  与  $D_i$  的关系可以表示为

$$D_N = \beta_N \beta_{N-1} \cdots \beta_{i+1} D_i + \gamma_N \quad (8)$$

结合式(5)~式(8), 式(3)中的球域率失真优化问题变为

$$\min_{o_i} \left( \omega_i^s \left( 1 + \sum_{p=i+1}^N \prod_{l=i+1}^p \beta_l \right) D_i(o_i) + \lambda R_i(o_i) \right) \quad (9)$$

其中,  $\omega_i^s = \sum_{p=i+1}^N \prod_{l=i+1}^p \beta_l$  代表当前编码单元的失真影响权重, 也用于量化时域依赖大小。

式(9)的球域率失真优化模型进一步表示为

$$\min_{o_i} \left( D_i(o_i) + \frac{\lambda}{\omega_i^s (1 + \omega_i^s)} R_i(o_i) \right) \quad (10)$$

## 2 球域失真真空域映射权重

不同经纬度的球面区域到正六面体的投影距离不同, 导致球面区域投影到平面后出现不同程度的拉伸形变。具体而言, 具有相同弧度  $\phi$  的不同位置球面区域在平面上的投影高度会有所差异, 如图 3 所示, CMP 的切面上高纬度球面区域在投影平面的高度  $d_1$  大于低纬度区域投影高度

$d_2$ 。投影后，通过插值来填补平面拉伸区域的空缺像素，而插值的冗余像素会影响全景视频的编码性能。为解决这个问题，本文需要在球域失真与编码失真之间建立一个符合主客观质量评价一致性的空域映射模型。

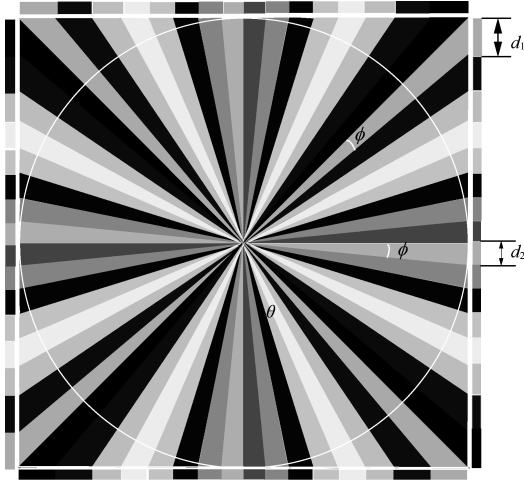


图3 同弧度球面投影高度变化

从六面体的一个面来观察，可以看到球面图像的一部分被投影到该面上。如图4所示，黑色正方形块为投影平面，球面区域被划分为5个环带区域，每个区域宽度相同，分别标记为1~5。假设球面上同一环带区域到投影平面的距离相等，相应区域各编码单元的失真权重相同。由于像素点在整个球面上均匀分布，球面图像投影到编码平面前后的像素点数量与区域面积成正比。因此，本文用面积比来量化球域失真映射权重  $\omega_i^s$ 。

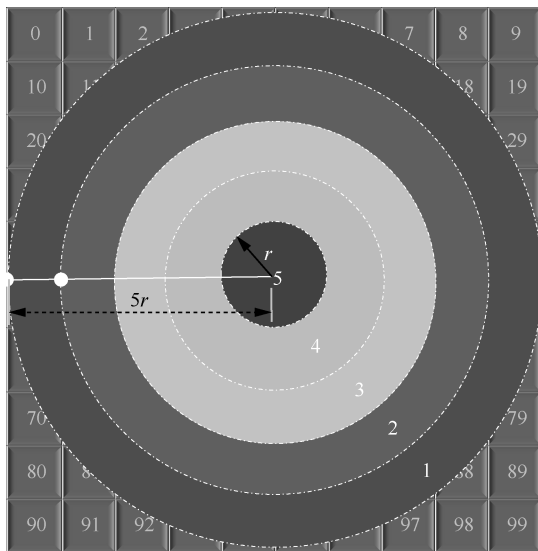


图4 球面区域划分

由于 CMP 的 6 个投影平面的分辨率相同，均为  $1280 \times 1280$ ，将其划分为 100 个分辨率为  $128 \times 128$  的 CTU。为了与编码平面对应，将球面环带宽度  $r$  设定为 CTU 的行宽。则第  $i$  个球面环带区域面积  $S_{\text{sphere}}(i)$  为

$$S_{\text{sphere}}(i) = \pi r^2 [(5-i+1)^2 - (5-i)^2], i \in \{1, 2, 3, 4, 5\} \quad (11)$$

在编码平面，各编码单元对应着唯一的球面区域，将球面环带区域对应的 5 个编码平面区域，分别标记为数字 1~5，如图5所示，同色的编码单元组成同一编码平面区域，同一编码平面区域的球域失真映射权重相同。

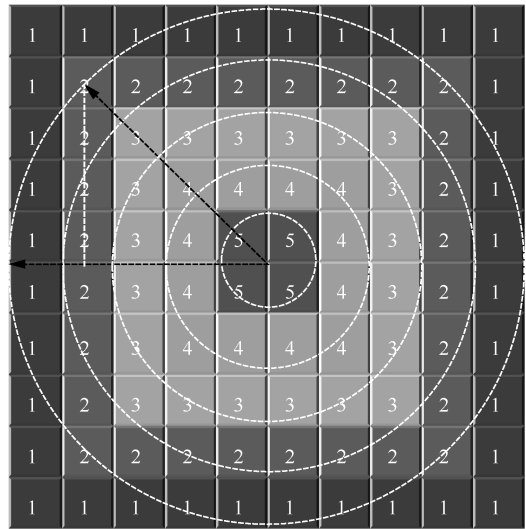


图5 编码平面区域划分

第  $i$  个编码平面区域面积  $S_{\text{cmp}}(i)$  为

$$S_{\text{cmp}}(i) = 4[2(5-i)+1]r^2, i \in \{1, 2, 3, 4, 5\} \quad (12)$$

因此，根据式(11)和式(12)，第  $i$  个编码平面区域内的编码单元的球域失真映射权重  $\omega_i^s$  为

$$\omega_i^s = \frac{S_{\text{sphere}}(i)}{S_{\text{cmp}}(i)} = \frac{\pi r^2 \{(5-i+1)^2 - (5-i)^2\}}{4[2(5-i)+1]r^2}, i \in \{1, 2, 3, 4, 5\} \quad (13)$$

至此，对于正六面体的一个编码平面的球域失真映射权重模型已经建立，而实际编码平面包含 6 个映射平面，将它们按照如图6所示的顺序排列，从而组成一个矩形编码平面。图6(a)展示了投影平面的组合顺序，编码平面的第一行分别为左、前、右投影平面，第二行分别为底、后、顶投影平面。图6(b)展示了编码平面的区域划分，而该区域划分方式决定了编码单元的空域映射权重。

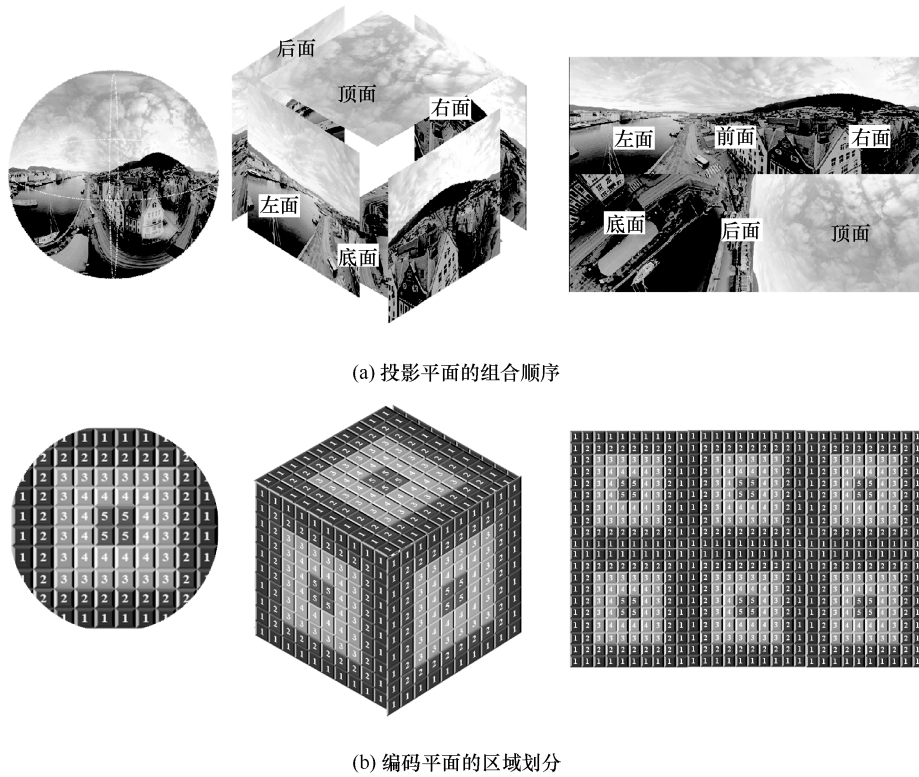


图 6 投影平面的组合顺序和编码平面的区域划分

### 3 球域失真时域影响权重

球域失真映射到编码平面后，由于采用了帧间预测编码技术，编码平面上的压缩失真不仅影响当前编码单元的编码性能，还会对后续参考该编码单元的其他编码单元的编码性能产生影响。因此，当考虑时域依赖编码时，需要根据参考关系来建立全局率失真优化模型。最新的通用视频编码标准 VVC 采用层次编码结构，如图 7 所示，其中  $8n+i$  表示第  $n$  个图像组 (GOP, group of picture) 中每个编码帧的图像序号 (POC),  $i=1,2,\dots,8$ 。在低时延编码配置中，一个 GOP 包含 8 帧，分为 3 层。编码帧参考时域上前一帧以及最近邻的 3 个关键帧，当前编码帧的失真会在时域上进行传播。因此，有必要量化编码单元的失真影响因子，并根据该失真影响因子来调整编码参数，从而提升传播链上所有编码单元的整体编码性能。

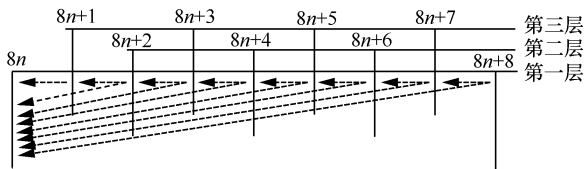


图 7 层次编码结构

根据式(5)~式(8)，可以递推出当前编码单元的失真影响因子，然而，该失真影响因子的求解需要缓存未编码帧，并通过前向运动搜索在源像素上来完成，复杂度极高。为了不增加编码复杂度，根据视频信号连续性特性，可以在已编码帧上建立后向失真传播链，如图 8 所示，根据已编码信息来量化失真影响权重。



图 8 后向失真传播链

根据式(5)~式(7)，在后向失真传播链中，当前编码单元  $U_i$  的失真  $D_i$  与前一编码单元  $U_{i-1}$  的失真  $D_{i-1}$  之间的关系变为

$$D_{i-1} = e^{-bR_{i-1}} D_i^{\text{mcp}} = \beta_{i-1} D_i + \gamma_{i-1} \quad (14)$$

同理，当前编码单元  $U_i$  的失真  $D_i$  对后向失真传播链中编码单元  $U_{i-N}$  的失真影响变为

$$D_{i-N} = \beta_{i-N} \beta_{i-N+1} \cdots \beta_{i-1} D_i + \gamma_{i-N} \quad (15)$$

根据式(14)可得，失真影响因子  $\beta_{i-1}$  为

$$\beta_{i-1} \approx e^{-bR_{i-1}} = \frac{D_{i-1}}{D_{i-1}^{\text{mcp}}} \quad (16)$$

当编码图 8 所示的传播链上的编码单元  $U_i$  时, 后向传播链上的其余编码单元已经完成编码, 所以式(16)中  $D_{i-1}$  及  $D_{i-1}^{\text{mcp}}$  可以获取。进而式(9)变为

$$\min_{\omega_i^s} \left( \omega_i^s \left( 1 + \sum_{p=i-1}^{i-N} \prod_{l=i-1}^p \beta_l \right) D_i(o_i) + \lambda R_i(o_i) \right) \quad (17)$$

综合式(15)~式(17), 编码单元  $U_i$  的失真时域影响权重  $\omega_i^t$  为

$$\begin{aligned} \omega_i^t &= \sum_{p=i-1}^{i-N} \prod_{l=i-1}^p \beta_l = \\ &= \frac{D_{i-1}}{D_{i-1}^{\text{mcp}}} + \frac{D_{i-1}}{D_{i-1}^{\text{mcp}}} \frac{D_{i-2}}{D_{i-2}^{\text{mcp}}} + \dots + \\ &= \frac{D_{i-1}}{D_{i-1}^{\text{mcp}}} \frac{D_{i-2}}{D_{i-2}^{\text{mcp}}} \dots \frac{D_{i-N}}{D_{i-N}^{\text{mcp}}} \end{aligned} \quad (18)$$

## 4 算法实现

第 2 节和第 3 节根据第 1 节建立的球域失真真空-时域依赖模型确立了球域失真空域映射权重  $\omega_i^s$  和时域影响权重  $\omega_i^t$ , 它们在算法实现过程中会调整拉格朗日乘子和量化参数, 从而影响编码模式的选择以及编码失真。因此, 算法实现过程中需要调整量化参数。

### 4.1 帧级量化参数调整

当前主流编码器均采用图 7 所示的层次编码结构, 将一个 GOP 中的各帧分成不同的编码层次, 不同层次的编码帧采用不同的编码策略, QP 级联编码策略<sup>[21]</sup>作为一种重要的层次编码技术, 极大地提升了编码性能。然而, 当前 QP 级联技术采用固定的 QP 偏移策略, 即

$$\text{QP}_{\text{slice}} = \text{QP}_0 + \text{QP}_i, i = 1, 2, 3 \quad (19)$$

其中,  $\text{QP}_{\text{slice}}$  为帧级量化参数;  $\text{QP}_0$  为编码器默认的输入量化参数;  $\text{QP}_i$  为第  $i$  层编码帧的量化参数偏移量, 层次越高, 该偏移值越大。

虽然该技术能体现不同层次编码帧的重要性, 却不能根据图像特征自适应调整量化参数。为此, 本文提出一种自适应帧级 QP 调整策略。

自适应量化参数编码策略是根据帧级时域依赖性的大小来调整量化参数。直观上讲, 对于时域相关性较强的编码帧, 有大部分编码单元采用帧间预测编码, 极限为当前编码帧的失真完全由参考帧决定; 反之, 则可能有较多的编码单元采

用帧内预测编码。所以, 可以通过帧级重建失真误差与运动补偿预测误差的比值来量化帧级时域相关性。然而, 在编码某帧结束之前不能获取重建失真误差与运动补偿预测误差信息, 也就不能量化该编码帧与参考帧的时域相关性大小。为此, 根据视频图像的连续性特点, 只要没有发生场景切换 (在实际编码过程中, 使用帧像素值的方差的变化量来判断视频场景是否发生切换), 就可用时域邻近已编码帧的编码信息来近似量化当前编码帧的时域相关性大小, 即

$$\omega_{\text{slice}} = \frac{D_{\text{frame}_{i-1}}}{D_{\text{frame}_{i-1}}^{\text{mcp}}} = \frac{\sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (O_{m,n} - C_{m,n})^2}{\sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (O_{m,n} - P_{m,n})^2} \quad (20)$$

其中,  $\omega_{\text{slice}}$  为帧级时域依赖权重,  $D_{\text{frame}_{i-1}}$  和  $D_{\text{frame}_{i-1}}^{\text{mcp}}$  分别为当前编码帧的上一帧的重建失真误差和运动补偿预测误差,  $O_{m,n}$ 、 $C_{m,n}$  和  $P_{m,n}$  分别为源像素值、重建像素值和预测像素值,  $H$  和  $W$  为图像分辨率的高和宽。

为了保证关键帧的参考质量, 帧级量化参数调整策略仅用于时域相关性较强的非关键帧, 即

$$\text{QP}_{\text{slice}} = \text{QP}_{\text{slice}} + \lfloor 10|1 - \omega_{\text{slice}}| \rfloor, \omega_{\text{slice}} > 0.7 \quad (21)$$

其中,  $\lfloor \cdot \rfloor$  为向下取整运算。

### 4.2 编码树单元级量化参数调整

将式(10)中的带权重的拉格朗日乘子表示为当前编码单元  $U_i$  的拉格朗日乘子  $\lambda_i$ , 也即其率失真曲线斜率为

$$\lambda_i = \frac{\lambda}{\omega_i^s (1 + \omega_i^t)} \quad (22)$$

其中,  $\lambda$  为参考软件本身的拉格朗日乘子, 由帧级量化参数计算而来, 即

$$\lambda = \varepsilon \omega_i 2^{\frac{\text{QP}_{\text{slice}} - 12}{3}} \quad (23)$$

其中,  $\text{QP}_{\text{slice}}$  为 4.1 节优化后的帧级量化参数,  $\omega_i$  为与编码帧所属层级相关的权重系数,  $\varepsilon$  为与当前编码帧是否作为参考图像相关的系数。

由式(22)计算出每个 CTU 的拉格朗日乘子后, 再使用文献[22]拟合的拉格朗日乘子与 QP 的关系式计算 QP 并用于编码, 即

$$\text{QP}_i = 4.2005 \ln \lambda_i + 13.7122 \quad (24)$$

### 4.3 算法流程

本节给出球域失真真空-时域依赖的球域率失真优化 (STD-SRDO, spatial and temporal dependent based spherical rate-distortion optimization) 算法的详细流程, 如算法 1 所示。该算法在空-时域依赖的球域失真模型上分别进行帧级和编码树单元级量化参数调整, 通过球域失真映射模型获取球域失真空域映射权重, 再通过球域失真时域传播模型获取失真时域影响权重, 最后计算拉格朗日乘子和量化参数以优化编码。假设编码帧数为  $N$ , 每帧包含  $M$  个编码单元。

**算法 1** 球域失真真空-时域依赖的球域率失真优化

输入 当前待编码帧

- 1) for  $j=1$  to  $N$  do
- 2) if  $\omega_{\text{slice}} > 0.7$  and  $N \% 8 \neq 0$  then
- 3)  $QP_{\text{slice}} \leftarrow QP_{\text{slice}} + 10|1 - QP_{\text{slice}}|$
- 4)  $\lambda \leftarrow \varepsilon \omega_j 2^{\frac{QP_{\text{slice}} - 12}{3}}$
- 5) end if
- 6) for  $j=1$  to  $M$  do
- 7)  $S_{\text{sphere}}(j) \leftarrow \pi r^2 \{[5 - i + 1]^2 - [5 - i]^2\}$
- 8)  $S_{\text{cmp}}(j) \leftarrow 4r^2 [2(5 - i) + 1]$
- 9)  $\omega_j^s \leftarrow \frac{S_{\text{sphere}}(j)}{S_{\text{cmp}}(j)}$
- 10)  $\beta_j \leftarrow \frac{D_j}{D_j^{\text{mcp}}}$
- 11)  $\omega_j^t \leftarrow \sum_{p=j-1}^{j-N} \prod_{l=j-1}^p \beta_l$
- 12)  $\lambda_j \leftarrow \frac{\lambda}{\omega_j^s (1 + \omega_j^t)}$
- 13)  $QP_j \leftarrow 4.2005 \ln \lambda_j + 13.7122$
- 14) end for
- 15) end for

## 5 实验

全景视频编码的目标是在最小化码率和编码时间的同时, 保证编码质量。为验证本文算法的有效性, 将基于球域失真真空-时域依赖的全景视频编码算法集成到 VVC 参考软件 VTM14.0 中, 编码器配置为低时延 P 帧 (LDP, low-delay P frame) 和低时延 B 帧 (LDB, low-delay B frame), 实验条件遵守国际编码标

准组织 (JCT-VC, joint collaborative team on video coding) 建议的通用测试条件 (CTC, common test conditions) [23], 全景视频编码的标准测试序列包括 8K、6K 和 4K 三类共计 16 个视频序列, 其帧率、比特深度以及分辨率等特征信息如表 1 所示。实验计算机配置为 Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz  $\times$  2, 内存为 256 GB, 基于 X64 的处理器, 64 位 Windows10 专业版操作系统。

表 1 全景视频测试序列特征信息

类别	序列	帧宽/ 像素	帧高/ 像素	帧数/ 帧	帧率/ (frame·s <sup>-1</sup> )	比特 深度/bit
8K	Trolley	8 192	4 096	300	30	10
	GasLamp	8 192	4 096	300	30	10
	Skateboardinginlot	8 192	4 096	300	30	8
	ChairliftRide	8 192	4 096	300	30	8
	KiteFlite	8 192	4 096	300	30	8
	Harbor	8 192	4 096	300	30	8
	SkateboardTrick	8 192	4 096	600	60	8
	Train	8 192	4 096	600	60	8
6K	Balboa	6 144	3 072	600	60	8
	Broadway	6 144	3 072	600	60	8
	Landing	6 144	3 072	300	60	8
	BranCastle	6 144	3 072	300	60	8
4K	PoleVault	3 840	1 920	300	30	8
	AerialCity	3 840	1 920	300	30	8
	DrivingInCity	3 840	1 920	300	30	8
	DrivingInCountry	3 840	1 920	300	30	8

全景视频大都具有较多的仿射运动, 运动目标通常出现在赤道区域附近 [24], 而赤道附近区域也是人眼视觉关注度较高的区域。图 9 给出了全景视频测试序列。从图 9 可以看出, 全景视频图像具有类型多样、分辨率高、视域广等特征, 有些视频序列的背景图像是静止的, 比如 PoleVault、Harbor 等; 有些视频序列的背景和前景图像都呈现运动变化, 比如 DrivingInCity、Landing 等。

### 5.1 率失真性能分析

本文所提算法以 VTM14.0 为基准, 分别用 QP 值为 22、27、32、37 去编码每个测试序列。编码率失真性能评价指标为 BD-rate。BD-rate 是评价视频编码算法性能的主要参数之一, 表示在相同编码质量下新算

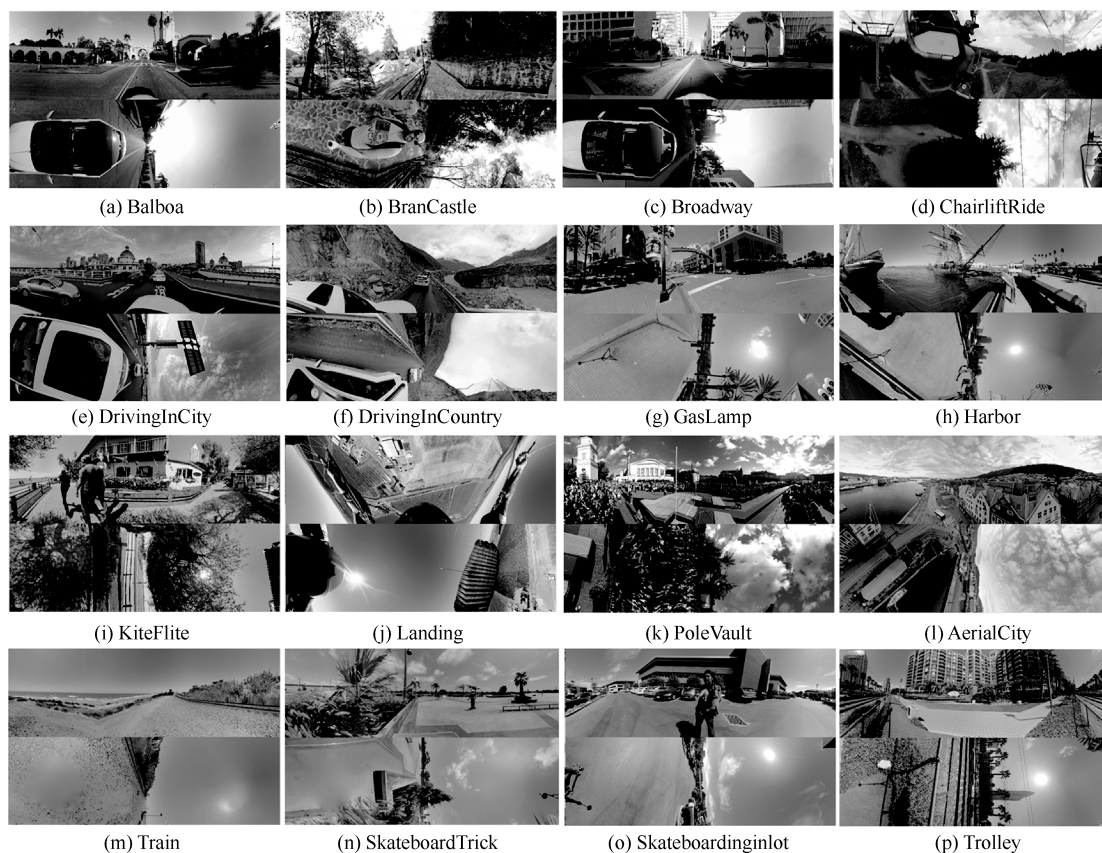


图 9 全景视频测试序列

法相对于基准的码率节省百分比, 负值表示率失真性能提升, 正值表示率失真性能降低。表 2 为基于带权重的空域映射模型的球域率失真优化 (WM-SRDO, weighted mapping model based spherical rate-distortion optimization) 算法相较于基准的 BD-rate。从表 2 可以看出, 在 8K、6K 和 4K 测试序列上, 分别使用 SPSNR、CPP-PSNR 和 WS-PSNR 作为客观质量评价指标, WM-SRDO 算法的 BD-rate 总体平均为 -2.2%、-2.1% 和 -2.6%, 该算法对所有全景视频序列都是有效的。实验结果表明, 该算法在一定程度上能有效改善投影过程中的过采样问题, 而且运动较小的视频序列的编码性能整体提升更高, 比如 GasLamp、PoleVault 等。

表 3 为 STD-SRDO 算法相较于基准的 BD-rate。从表 3 可以看出, 以 S-PSNR、CPP-PSNR、WS-PSNR 为客观质量评价指标, BD-rate 总体平均为 -7.4%、-7.4%、-7.9%。STD-SRDO 算法对所有测试序列均有效, 特别是在 8K 全景视频序列上, 亮度分量的平均码率节省超过了 12%, 色度分量码率节省超过了 16%。对单个视频序列而言, 序列 Train 的亮度分量码率节省高达 22.6%, Harbor 的色度分量的码率节省达到

30%以上。整体来看, 大部分 8K 视频序列具有图像稳定、背景静止的特征, 而考虑失真时域传播的率失真优化算法对这类视频十分有效。对于运动变化较大的视频序列来说, 由于帧间预测技术的使用频率相对降低, 帧内编码的使用频率相对增高, 因此时域依赖程度也会降低, 考虑时域依赖的率失真优化算法的编码性能也会相对较低。这一点在 6K 视频序列上体现得尤其明显, 6K 视频大都具有运动镜头, 图像变化快、运动速度高, 因而时域传播链相对较短, 提升当前帧的编码质量对后续帧的编码质量的改善影响有限。这一点也可以从编码结果上看出, STD-SRDO 算法在 6K 视频序列上的平均码率节省为 1.9%。4K 视频序列的图像相对 6K 视频变换缓慢, 整体上有 3.4% 的码率节省, 这也比较符合图像的运动特征。

表 4 为 STD-SRDO 算法在 LDB 配置下相较于基准的 BD-rate。从表 4 可以看出, 与 LDP 配置下编码结果类似, 在 SPSNR、CPP-PSNR、WS-PSNR 客观质量评价指标下, STD-SRDO 算法平均码率节省达到 7.0%、7.0%、7.6%。同样, 在时域相关性较强的视频序列上的编码性能提升比较明显。

表 2 WM-SRDO 算法相较于基准的 BD-rate

类别	序列	SPSNR			CPP-PSNR			WS-PSNR		
		SPSNR-Y	SPSNR-U	SPSNR-V	CPPPSNR-Y	CPPPSNR-U	CPPPSNR-V	WSPSNR-Y	WSPSNR-U	WSPSNR-V
8K	Trolley	-2.7%	-4.0%	-3.6%	-2.8%	-4.1%	-3.7%	-3.0%	-4.0%	-3.7%
	GasLamp	-4.8%	-6.0%	-6.1%	-4.8%	-6.0%	-6.1%	-4.9%	-6.0%	-6.1%
	Skateboardinginlot	-1.2%	-2.1%	-0.8%	-1.2%	-2.1%	-0.8%	-1.6%	-2.2%	-1.0%
	ChairliftRide	-1.5%	-0.5%	-0.5%	-1.5%	-0.4%	-0.5%	-1.9%	-0.4%	-0.5%
	KiteFlite	-2.9%	-4.0%	-3.5%	-2.9%	-4.0%	-3.4%	-3.3%	-4.1%	-3.4%
	Harbor	-3.1%	-5.4%	-5.9%	-3.2%	-5.5%	-5.9%	-3.5%	-5.6%	-5.8%
	SkateboardTrick	-1.6%	-0.8%	-0.7%	-1.3%	-0.6%	-0.6%	-1.9%	-0.8%	-0.6%
	Train	-3.4%	-3.4%	-4.2%	-3.4%	-3.3%	-4.2%	-3.4%	-3.5%	-4.2%
平均	-2.7%	-3.3%	-3.2%	-2.6%	-3.3%	-3.2%	-2.9%	-3.3%	-3.2%	
6K	Balboa	-0.9%	0.3%	0.8%	-0.8%	0.3%	1.0%	-1.6%	0.3%	1.0%
	Broadway	-0.4%	-0.3%	-0.4%	-0.5%	-0.2%	-0.4%	-1.5%	-0.3%	-0.4%
	Landing	-1.3%	-0.5%	-0.6%	-1.3%	-0.4%	-0.5%	-1.6%	-0.6%	-0.6%
	BranCastle	-1.3%	-0.8%	0.4%	-1.3%	-0.8%	0.5%	-1.8%	-0.8%	0.5%
	平均	-1.0%	-0.3%	0.1%	-1.0%	-0.3%	0.2%	-1.6%	-0.4%	0.1%
4K	PoleVault	-4.7%	-3.5%	-1.9%	-4.6%	-3.5%	-1.6%	-5.4%	-3.8%	-1.8%
	AerialCity	-1.8%	-9.4%	-8.4%	-1.5%	-9.3%	-8.1%	-2.0%	-9.3%	-8.1%
	DrivingInCity	-2.5%	-5.2%	-4.7%	-2.3%	-5.3%	-4.6%	-2.6%	-5.3%	-4.7%
	DirvingInCountry	-1.3%	-4.4%	-3.5%	-1.0%	-4.2%	-3.1%	-1.7%	-4.3%	-3.2%
	平均	-2.6%	-5.6%	-4.6%	-2.4%	-5.6%	-4.4%	-2.9%	-5.7%	-4.5%
总体平均	-2.2%	-3.2%	-2.8%	-2.1%	-3.1%	-2.7%	-2.6%	-3.2%	-2.6%	

表 3 STD-SRDO 算法相较于基准的 BD-rate (LDP 配置)

类别	序列	SPSNR			CPP-PSNR			WS-PSNR		
		SPSNR-Y	SPSNR-U	SPSNR-V	CPPPSNR-Y	CPPPSNR-U	CPPPSNR-V	WSPSNR-Y	WSPSNR-U	WSPSNR-V
8K	Trolley	-18.9%	-21.4%	-15.2%	-18.3%	-21.3%	-15.3%	-18.7%	-21.4%	-15.4%
	GasLamp	-17.1%	-21.6%	-21.5%	-17.1%	-21.6%	-21.5%	-18.1%	-21.8%	-21.5%
	Skateboardinginlot	-1.7%	-4.5%	-6.2%	-1.7%	-4.4%	-6.2%	-2.0%	-4.6%	-6.5%
	ChairliftRide	-1.5%	-5.8%	-5.7%	-1.6%	-5.7%	-5.6%	-2.0%	-5.6%	-5.6%
	KiteFlite	-8.6%	-8.0%	-5.6%	-8.6%	-8.0%	-5.5%	-9.0%	-8.1%	-5.6%
	Harbor	-19.7%	-30.6%	-31.2%	-19.3%	-30.6%	-31.0%	-21.4%	-30.9%	-31.1%
	SkateboardTrick	-7.6%	-17.0%	-17.5%	-8.2%	-16.9%	-17.4%	-8.3%	-17.2%	-17.6%
	Train	-22.1%	-23.4%	-27.8%	-22.2%	-23.5%	-28.0%	-22.3%	-23.7%	-28.1%
平均	-12.1%	-16.5%	-16.3%	-12.1%	-16.5%	-16.3%	-12.7%	-16.7%	-16.4%	
6K	Balboa	-1.4%	-3.7%	-4.2%	-1.4%	-3.6%	-4.0%	-2.0%	-3.6%	-4.0%
	Broadway	-0.9%	-3.7%	-5.0%	-1.0%	-3.6%	-5.0%	-1.8%	-3.7%	-5.0%
	Landing	-3.2%	-4.1%	-4.5%	-3.4%	-3.9%	-4.4%	-3.8%	-4.1%	-4.5%
	BranCastle	-2.3%	-0.8%	-1.4%	-2.5%	-0.8%	-1.4%	-2.8%	-0.8%	-1.4%
	平均	-1.9%	-3.1%	-3.8%	-2.1%	-3.0%	-3.7%	-2.6%	-3.0%	-3.7%
4K	PoleVault	-5.4%	-10.7%	-8.9%	-5.2%	-10.4%	-8.5%	-5.9%	-10.9%	-8.7%
	AerialCity	-1.3%	-14.0%	-15.4%	-1.0%	-13.7%	-15.2%	-1.1%	-13.9%	-15.3%
	DrivingInCity	-4.4%	-8.8%	-8.6%	-4.4%	-8.8%	-8.5%	-4.8%	-8.8%	-8.6%
	DirvingInCountry	-2.6%	-1.8%	-3.4%	-2.4%	-1.6%	-3.2%	-2.8%	-1.7%	-3.3%
	平均	-3.4%	-8.8%	-9.1%	-3.2%	-8.6%	-8.8%	-3.7%	-8.8	-9.0%
总体平均	-7.4%	-11.3%	-11.4%	-7.4%	-11.2%	-11.3%	-7.9%	-11.3%	-11.4%	

表 4 STD-SRDO 算法相较于基准的 BD-rate (LDB 配置)

类别	序列	SPSNR	CPP-PSNR	WS-PSNR
8K	Trolley	-18.2%	-17.6%	-18.1%
	GasLamp	-16.7%	-16.7%	-17.7%
	Skateboardinginlot	-1.6%	-1.7%	-1.9%
	ChairliftRide	-1.2%	-1.2%	-1.6%
	KiteFlite	-8.7%	-8.7%	-9.0%
	Harbor	-18.6%	-18.2%	-20.3%
	SkateboardTrick	-6.5%	-7.5%	-7.4%
	Train	-21.9%	-22.1%	-22.3%
6K	Balboa	-0.9%	-0.9%	-1.6%
	Broadway	-0.6%	-0.7%	-1.6%
	Landing	-3.2%	-3.4%	-3.8%
	BranCastle	-2.3%	-2.5%	-2.8%
4K	PoleVault	-5.4%	-5.2%	-5.9%
	AerialCity	-0.2%	0.3%	0.1%
	DrivingInCity	-4.3%	-4.2%	-4.7%
	DirvingInCountry	-2.4%	-2.4%	-2.8%
总体平均	-7.0%	-7.0%	-7.6%	

表 5 和表 6 分别为本文算法与同类算法在 LDP 和 LDB 配置下相较于基准的 BD-rate。为简便起见, 将 SPSNR、CPPPSNR 及 WSPSNR 分别缩写为 S、CPP 及 WS。文献[8]的边界补偿算法有 0.5% 的码率节省, 该算法对运动较大的全景视频比较有效, 而运动较小甚至静止图像存在较少的跨边界运动, 编码性能提升较少。文献[10]针对 CMP 边界进行拓展, 以保持图像边界的连续性, 性能提升达 1.2%。文献[11]在边界补偿的基础上还提出了球域运动模型, 平均有 2.1% 的码率节省。文献[12]提出基于熵平衡

的比特分配算法来提升球域率失真性能, 整体平均有 2.6% 的码率节省。文献[13]提出带权重的客观质量评价指标 WS-PSNR, 并用于调整量化参数, 平均有 3.2% 的码率节省。本文算法在 SPSNR 或 WSPSNR 评价指标下分别有 8.1% 或 8.6% 的码率节省, 远高于同类算法, 特别是在时域依赖强的视频序列上, 本文算法效果更加明显, 码率节省最高达 20%。

为了更直观地观察率失真性能对比情况, 图 10 给出了部分测试序列在不同算法下的率失真曲线对比。TD-SRDO 和 STD-SRDO 算法分别仅考虑球域失真时域传播以及同时考虑球域失真空域映射和时域传播。从图 10 可以看出, 本文算法 STD-SRDO 率失真性能更优, 特别是在 Trolley、Gaslamp、Harbor 等时域依赖性较强的视频序列上效果更加明显。

### 5.2 编码时间对比

编码时间是衡量算法编码复杂度的重要指标, 为了弄清时域依赖算法是否会增加编码复杂度, 表 7 对比了本文算法 TD-SRDO 与 STD-SRDO 以及基准编码所有测试序列所用时间, 比率为算法与基准的编码时间比值。从表 7 可以看出, TD-SRDO 算法的编码时间与基准的编码时间几乎持平, 观察具体视频序列的编码时间可以看出, 运动较大的视频序列的编码时间有所增加, 原因为增加了时域传播权重的计算量, 比如 Skateboardinginlot、ChairliftRide 等; 而运动较小的视频序列的编码时间有所降低, 原因为帧级自适应 QP 调整策略使时域依赖强的视频序列的量化参数增大, 比如 GasLamp、Train 等。STD-SRDO 算法相较于基准的编码时间下降 9%;

表 5 本文算法与同类算法在 LDP 配置下相较于基准的 BD-rate

类别	序列	文献[8]	文献[10]	文献[13]	本文算法 DTD-SRDO		
		WS	WS	WS	S	CPP	WS
8K	Trolley	0.2%	-0.1%	-2.6%	-18.9%	-18.3%	-18.7%
	GasLamp	0.0%	0.0%	-2.4%	-17.1%	-17.1%	-18.1%
	Skateboardinginlot	1.2%	-1.8%	-4.0%	-1.7%	-1.7%	-2.0%
	ChairliftRide	-2.4%	-2.9%	-3.8%	-1.5%	-1.6%	-2.0%
	KiteFlite	0.2%	-0.1%	-1.6%	-8.6%	-8.6%	-9.0%
	Harbor	1.5%	0.0%	-2.4%	-19.7%	-19.3%	-21.4%
4K	PoleVault	-0.2%	-0.1%	-3.8%	-5.4%	-5.2%	-5.9%
	AerialCity	-1.6%	-2.5%	-3.8%	-1.3%	-1.0%	-1.1%
	DrivingInCity	-0.8%	-0.7%	-3.0%	-4.4%	-4.4%	-4.8%
	DirvingInCountry	-3.1%	-3.3%	-4.8%	-2.6%	-2.4%	-2.8%
总体平均	-0.5%	-1.2%	-3.2%	-8.1%	-8.0%	-8.6%	

表 6 本文算法与同类算法在 LDB 配置下相较于基准的 BD-rate

类别	序列	文献[10]	文献[11]	文献[12]	本文算法 STD-SRDO		
		WS	S	S	S	CPP	WS
8K	Trolley	-0.1%	-0.50%	-3.1%	-18.2%	-17.6%	-18.1%
	GasLamp	0.0%	-0.2%	-0.7%	-16.7%	-16.7%	-17.7%
	Skateboardinginlot	-1.6%	-2.8%	2.1%	-1.6%	-1.7%	-1.9%
	ChairliftRide	-3.0%	-4.7%	-3.9%	-1.2%	-1.2%	-1.6%
	KiteFlite	-0.1%	0.0%	-0.5%	-8.7%	-8.7%	-9.0%
	Harbor	0.0%	-0.1%	0.7%	-18.6%	-18.2%	-20.3%
4K	PoleVault	-0.2%	-0.3%	-4.2%	-5.4%	-5.2%	-5.9%
	AerialCity	-2.6%	-5.5%	-5.4%	-0.2%	0.3%	0.1%
	DrivingInCity	-0.9%	-1.3%	-1.2%	-4.3%	-4.2%	-4.7%
	DirvingInCountry	-3.4%	-6.0%	-5.5%	-2.4%	-2.4%	-2.8%
总体平均		-1.2%	-2.1%	-2.6%	-7.7%	-7.6%	-8.2%

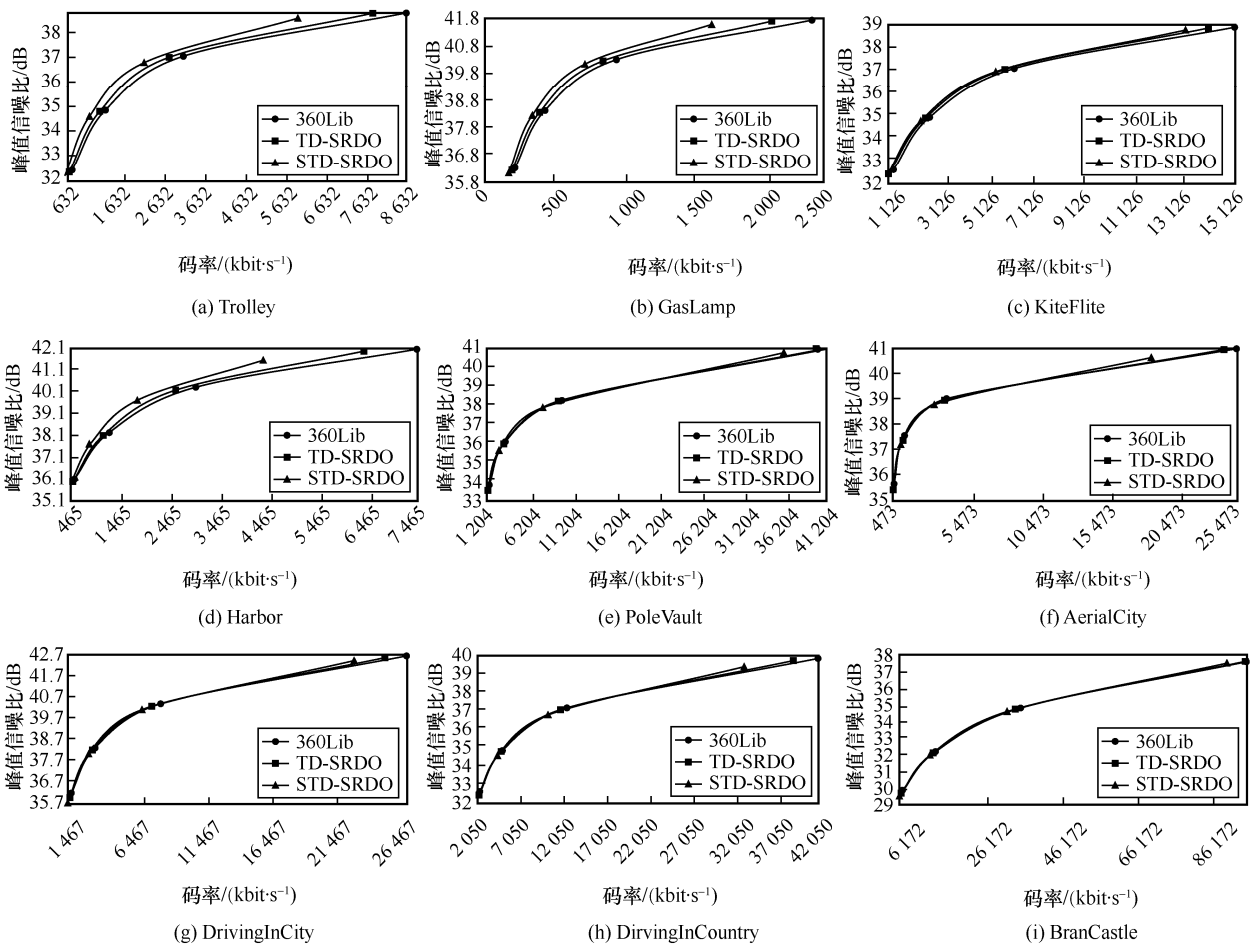


图 10 率失真曲线对比

表 7 不同算法编码时间对比

类别	序列	基准/h	TD-SRDO		STD-SRDO	
			编码时间/h	比率	编码时间/h	比率
8K	Trolley	135	135	100%	114	84%
	GasLamp	67	62	93%	55	82%
	Skateboardinginlot	401	410	102%	381	95%
	ChairliftRide	243	256	105%	235	97%
	KiteFlite	167	165	98%	152	91%
	Harbor	129	122	95%	98	76%
	SkateboardTrick	393	371	94%	316	80%
	Train	269	256	95%	219	82%
6K	Balboa	584	571	98%	546	93%
	Broadway	615	615	100%	576	94%
	Landing	489	490	101%	467	95%
	BranCastle	593	595	101%	567	96%
4K	PoleVault	335	337	101%	302	90%
	AerialCity	247	257	104%	215	87%
	DrivingInCity	320	321	101%	293	92%
	DirvingInCountry	376	381	101%	350	93%
总计	5 370	5 352	99%	4 894	91%	

在 TD-SRDO 算法基础上，编码时间进一步降低，编码时间下降的主要原因是增加的空域映射模型使高纬度区域的量化参数进一步增加，从而节省编码时间。

## 6 结束语

本文针对球域失真空-时域依赖的全景视频编码进行了研究。首先，在空域上建立球域失真到压缩失真的映射模型，提升了单帧编码的率失真性能。其次，根据运动补偿预测编码，当前编码单元的失真不仅影响自身的编码性能，还会影响后续帧中与之具有参考关系的编码单元的编码性能，以此建立球域失真时域传播链并量化时域影响权重，再根据球域失真映射权重和时域影响权重来调整拉格朗日乘子和量化参数。实验结果显示，与 VTM14.0 基准相比，本文算法实现了 7.4% 的码率节省和 9% 的编码时间节省，显著提升了全景视频的编码性能。

## 参考文献:

[1] VISHWANATH B, NANJUNDASWAMY T, ROSE K. Rotational motion model for temporal prediction in 360 video coding[C]//Proceedings of 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). Piscataway: IEEE Press,

2017: 1-6.  
 [2] CHIARIOTTI F. A survey on 360-degree video: coding, quality of experience and streaming[J]. Computer Communications, 2021, 177: 133-155.  
 [3] XU M, LI C, ZHANG S Y, et al. State-of-the-art in 360° video/image processing: perception, assessment and compression[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 5-26.  
 [4] YOUVALARI R G, AMINLOU A, HANNUKSELA M M, et al. Efficient coding of 360-degree pseudo-cylindrical panoramic video for virtual reality applications[C]//Proceedings of IEEE International Symposium on Multimedia (ISM). Piscataway: IEEE Press, 2017: 525-528.  
 [5] SULLIVAN G J, OHM J R, HAN W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.  
 [6] BROSS B, WANG Y K, YE Y, et al. Overview of the versatile video coding (VVC) standard and its applications[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3736-3764.  
 [7] CHEN Z, LI Y, ZHANG Y. Recent advances in omnidirectional video coding for virtual reality: projection and evaluation[J]. Signal Processing, 2018, 146: 66-78.  
 [8] SAUER J, SCHNEIDER J, WIEN M. Improved motion compensation for 360° video projected to polytopes[C]//Proceedings of 2017 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2017: 61-66.  
 [9] XU M, SONG Y H, WANG J Y, et al. Predicting head movement in panoramic video: a deep reinforcement learning approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019,

- 41(11): 2693-2708.
- [10] LI L, LI Z, MA X, et al. Co-projection-plane based 3-D padding for polyhedron projection for 360-degree video[C]//Proceedings of 2017 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2017: 55-60.
- [11] LI L, LI Z, MA X, et al. Advanced spherical motion model and local padding for 360° video compression[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2342-2356.
- [12] ZHOU Y M, TIAN L, ZHU C, et al. Video coding optimization for virtual reality 360-degree source[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 118-129.
- [13] SUN Y L, YU L. Coding optimization based on weighted-to-spherically-uniform quality metric for 360 video[C]//Proceedings of IEEE Visual Communications and Image Processing (VCIP). Piscataway: IEEE Press, 2018: 1-4.
- [14] LI Y M, XU J Z, CHEN Z Z. Spherical domain rate-distortion optimization for omnidirectional video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(6): 1767-1780.
- [15] GAO Y B, ZHU C, LI S, et al. Temporally dependent rate-distortion optimization for low-delay hierarchical video coding[J]. IEEE Transactions on Image Processing, 2017, 26(9): 4457-4470.
- [16] GAO Y B, ZHU C, LI S, et al. Source distortion temporal propagation analysis for random-access hierarchical video coding optimization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(2): 546-559.
- [17] LI S, ZHU C, GAO Y B, et al. Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(1): 117-129.
- [18] YANG T W, ZHU C, FAN X J, et al. Source distortion temporal propagation model for motion compensated video coding optimization[C]//Proceedings of 2012 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2012: 85-90.
- [19] 郭红伟, 朱策, 杨栩, 等. 基于失真反向传播的时域依赖率失真优化[J]. 通信学报, 2022, 43(12): 222-232.
- GUO H W, ZHU C, YANG X, et al. Temporal dependent rate-distortion optimization based on distortion backward propagation[J]. Journal on Communications, 2022, 43(12): 222-232.
- [20] GUO H, ZHU C, YE M, et al. Pre-encoding based temporal dependent rate-distortion optimization for HEVC[J]. Signal Processing: Image Communication, 2023, 115: 116957.
- [21] ZHAO T S, WANG Z, CHEN C W. Adaptive quantization parameter cascading in HEVC hierarchical coding[J]. IEEE Transactions on Image Processing, 2016, 25(7): 2997-3009.
- [22] LI B, XU J Z, ZHANG D, et al. QP refinement according to lagrange multiplier for high efficiency video coding[C]//Proceedings of 2013 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE Press, 2013: 477-480.
- [23] ALSHINA E, BOYCE J, ABBAS A, et al. JVET common test conditions and evaluation procedures for 360 degree video: JVET document, JVET-G1030[S]. 2017.
- [24] JIANG L, XU M, LIU T, et al. DeepVS: a deep learning based video saliency prediction approach[C]//European Conference on Computer Vision. Cham: Springer, 2018: 625-642.

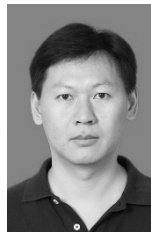
## [作者简介]



杨栩 (1983- ), 男, 四川苍溪人, 重庆邮电大学博士生, 成都师范学院讲师, 主要研究方向为全景视频编码与通信。



朱策 (1969- ), 男, 四川自贡人, 博士, 电子科技大学教授、博士生导师, 主要研究方向为视频编码与通信、视频分析和处理。



郭红伟 (1980- ), 男, 彝族, 云南金平人, 博士, 红河学院教授, 电子科技大学在站博士后, 主要研究方向为视频编码与通信。



罗雷 (1986- ), 男, 重庆人, 博士, 重庆邮电大学副教授, 电子科技大学在站博士后, 主要研究方向为视频编码与通信。