



研究与开发

基于球形加权结构相似度的全向视频编码率失真优化方法

周瑜锋, 郁梅, 杜宝祯, 宋洋, 蒋刚毅

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 针对传统视频编码中编码率失真模型的不足, 考虑等矩形全向视频的特征, 提出一种基于球形加权结构相似度 (WS-SSIM) 的全向视频编码率失真优化方法。通过考虑投影平面内部结构相似度的失真情况, 以及球面失真与投影平面失真的关系, 提出“球形加权结构相似度”来描述平面全向图像局部块相对于观看球面的失真程度, 将其应用于全向视频编码的率失真优化过程并自适应选择量化参数, 以提高基于视觉的编码效率。实验结果表明, 与 HEVC 视频编码标准的 HM16.9 测试平台的方法相比, 所提出方法在相同视觉质量下可平均节省 24.48% 的码率, 说明该方法对于提高全向视频编码率失真性能具有显著的效果。

关键词: 全向视频编码; 球形加权结构相似度; 率失真优化; 量化参数自适应选择

中图分类号: TN919

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019025

Omni directional video coding rate distortion optimization method based on weighted-to-spherically-uniform structural similarity

ZHOU Yufeng, YU Mei, DU Baozhen, SONG Yang, JIANG Gangyi

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Aiming at the shortcomings of the coding rate distortion model in traditional video coding, considering the characteristics of equal rectangular omnidirectional video, a method of omnidirectional video coding rate distortion optimization based on spherical weighted structural similarity (WS-SSIM) was proposed. By considering the distortion of the internal structural similarity of the projection plane and the relationship between the spherical distortion and the projection plane distortion, the “spherical weighted structural similarity” was proposed to describe the degree of distortion of the planar omnidirectional image local block relative to the viewing sphere, which was applied to the rate-distortion optimization process of omnidirectional video coding and adaptive selection of quantization parameters to improve vision-based coding efficiency. The experimental results show that compared with the HEVC video coding standard HM16.9 test platform, the proposed method can save an average of 24.48% code rate under the same visual quality, which indicates that the method has significant performance for improving the omnidirectional video coding rate distortion performance.

Key words: omnidirectional video coding, weighted-to-spherically-uniform structural similarity, rate distortion optimization, adaptive quantization parameter selection

收稿日期: 2018-07-16; 修回日期: 2019-01-21

基金项目: 国家自然科学基金资助项目 (No.61671258, No.61871247); 浙江省自然科学基金资助项目 (No.LY15F010005)

Foundation Items: The National Natural Science Foundation of China (No.61671258, No.61871247), The Natural Science Foundation of Zhejiang Province of China (No.LY15F010005)



1 引言

全向视频（也称为 360° 视频）是通过专门的全向相机或者多个普通相机拍摄后拼接形成的具有大视场、高分辨率以及沉浸式特点的视频^[1]。通过佩戴头戴式显示器（head-mounted display, HMD），观看者可以移动他们的头部和身体来自由地选择要观看的内容^[2]。全向视频通信系统通常包括成像、投影、编码压缩、逆投影和交互显示等环节。由于当前编码器的限制，通常 360° 的球形全向视频需要通过不同的投影方式转换到 2D 平面，经过编码传输，再逆投影回球形视频，最后通过头戴式显示器进行显示^[3]。经过投影的全向视频分辨率较高，通常可达 4K、8K 甚至更高^[4]，所以提高编码效率是实现全向视频系统应用的关键环节。

传统的视频编码器，如 H.264/AVC、H.265/HEVC 等都可直接对投影后的平面全向视频进行编码，但编码效率往往比较低，原因是忽略了全向视频的特点。不同投影方式下全向视频有不同的特点：等矩形投影^[5]，随着纬度升高，插值的像素增加，引入的冗余也增加；立方体投影^[6]，不同的面之间互不相关，且每个面具有相同的像素插值情况；八面体投影^[7]，由于面数量的增加，投影平面的不连续性上升。

现有的全向视频编码方案可分为两类：基于用户观看视口（viewport）的实时编码方案以及针对投影方式进行高效压缩的编码方案。由于全向视频观看设备的限制，现有头戴式显示器的显示视场通常为 96°~110°^[8]，这决定了不可见区域占 360° 视频播放消耗的大部分网络带宽，造成了网络带宽的极大浪费。Qian 等^[9]将等矩形全向图像在编码端分成若干个大小相等的 tile（矩形编码区域），通过对观看者的头部运动进行预测，对包含下一时刻可见视口的 tile 以及周边的 tile 进行编码传输，但由于其预测方式仅采用简单的加权线性

回归预测，可能造成较大的预测误差。Skupin 等^[10]提出基于 tile 变化分辨率的全向视频编码方法，通过降低可见视口外 tile 的分辨率来降低传输码率，但是提出的方法仍需要传输整个视频帧，对于用户完全不可见的 tile 的码流来说仍造成了不必要的开销。

不同于视口的实时编码方案，针对投影方式进行高效压缩的编码方案因投影方式的不同而各有差异。对于各个面互不相关的立方体投影，Tang 等^[11]提出根据不同的运动强度，对立方体投影的不同面分配不同的时间分辨率（即帧率），从而降低码率的消耗。对于等矩形投影两极的像素插值情况，Youvalari 等^[12]提出区域下采样方法，将等矩形视频帧分为不同的条带，根据纬度信息进行下采样并编码，但是考虑到复杂度，该方法只分为 3 个条带，其中上下两个条带 1/2 下采样，中间条带不采样，该过程不符合等矩形投影像素插值规律，不能很好地去除两极冗余。同样针对等矩形投影，Hendr 等^[13]提出基于纬度因素的自适应量化参数（quantization parameter, QP）编码，根据纬度因素对于高纬度区域采取较高的量化参数值去除像素的拉伸冗余，但是该方法仅考虑了 QP 与纬度因素的关系，而没有考虑率失真模型中失真定义的不合理性。

针对全向视频的观看空间（即球面）与投影平面的失真非线性的特点，目前已提出几种全向视频的客观质量评价模型，包括 WS-PSNR^[14]、S-PSNR^[15]以及 CPP-PSNR^[16]。这 3 种全向视频客观质量评价模型，都以峰值信噪比（PSNR）评价模型为基础，针对全向视频特点进行改进，使之适合评价全向视频的质量。

由于 PSNR 评价模型与人眼感知存在一定差异，为了使全向视频的客观质量评价模型更符合人眼的感知，本文给出面向全向视频的球形加权结构相似度（weighted-to-spherically-uniform structural similarity, WS-SSIM）的客观质量评价

模型。WS-SSIM 评价模型在传统结构相似度评价模型^[17]的基础上, 结合不同投影方式中投影平面与球面的失真对应关系, 使用户观看的球面失真与平面失真线性对应。

考虑到传统视频编码中率失真模型的失真评估不合理以及等矩形全向视频的特征, 本文针对等矩形全向视频, 提出一种基于 WS-SSIM 的全向视频编码率失真优化方法。在考虑投影平面结构失真的基础上, 结合球面失真与等矩形投影平面失真的对应关系, 根据 WS-SSIM 模型对率失真过程进行调整并自适应选择量化参数。实验结果表明, 与 HEVC 编码标准 HM16.9 测试平台的原始编码方法相比, 所提出方法可以获得明显的编码增益。

2 基于球形加权结构相似度的全向视频编码率失真优化方法

对于全向视频的基于球形加权结构相似度的率失真优化编码, 本文先给出 WS-SSIM 客观质量评价模型, 再以此指导等矩形投影方式下全向视频编码的率失真优化过程。本文提出的率失真优化方法包括 3 部分, 首先根据投影平面内结构相似度局部图像块的失真情况, 修改原始率失真模型的拉格朗日因子以考虑结构相似度失真; 其次, 针对等矩形全向视频的特征, 对率失真模型中的结构相似度失真度量进行优化, 使不同区域的失真具有不同的权重; 最后, 为了达到最优的编码效果, 根据修改的 WS-SSIM 率失真优化模型自适应地调整局部块的量化参数, 从而提升全向视频的编码效率。

2.1 面向全向视频图像的 WS-SSIM 客观质量评价模型

不同于 PSNR 评价模型简单地比较两幅图像的差值, 结构相似度 (SSIM)^[17]从图像组成的角度将结构信息定义为独立于亮度、对比度, 反映场景中物体结构的属性, 被认为是一种更符合人

眼感知的质量评价模型^[18]。因此, 本文通过考虑等矩形全向视频的特征, 改进 SSIM 模型使之适用于评价等矩形全向视频。

SSIM 作为图像客观质量评价模型, 综合亮度、对比度以及结构比较因子, 度量图像的失真情况, 其计算式定义如下:

$$SSIM = \left(\frac{2\mu_f\mu_d + c_1}{\mu_f^2 + \mu_d^2 + c_1} \right) \left(\frac{2\sigma_{fd} + c_2}{\sigma_f^2 + \sigma_d^2 + c_2} \right) \quad (1)$$

其中, μ_f 和 μ_d 分别表示参考图像块与失真图像块的均值, σ_f^2 和 σ_d^2 表示参考图像块与失真图像块的方差, σ_{fd} 表示协方差, c_1 和 c_2 为常数。

球面投影为等矩形平面时不同纬度区域存在不同的像素插值情况, 即不同的面积比例, 并且低纬度区域的观看概率明显高于高纬度区域^[19], 所以将普通平面图像的评价模型直接用于评价全向视频图像不合适。根据等矩形平面在逆投影时不同区域的面积比例关系, 引入不同区域的失真权重 $w(u, v)$:

$$w(u, v) = \cos \left(\left(v - \frac{H}{2} + \frac{1}{2} \right) \times \frac{\pi}{H} \right) \quad (2)$$

其中, u 、 v 表示当前像素所在图像中的横、纵坐标, H 表示图像的高度。

$w(u, v)$ 随着纬度的升高逐渐变小, 并在两极区域达到最小值。WS-SSIM 模型通过对不同区域添加权重使结构相似度评价模型适合对等矩形全向视频图像进行评价。在 WS-SSIM 评价模型中, 以一个 $k \times k$ 大小的滑动窗口, 在参考图像与失真图像之间以步长 1 进行滑动, 根据式 (3) 得到第 n 个滑动窗口的基于局部权重的结构相似度值 $S_{WSSIM}(n)$:

$$S_{WSSIM}(n) = \left(\frac{2\mu_f\mu_d + c_1}{\mu_f^2 + \mu_d^2 + c_1} \right) \left(\frac{2\sigma_{fd} + c_2}{\sigma_f^2 + \sigma_d^2 + c_2} \right) \times w_n \quad (3)$$

其中, w_n 为第 n 个滑动窗口中心点像素根据式 (2) 得到的权重。

滑动窗口依次在参考图像与失真图像之间滑



动, 当遍历完整幅图像之后, 进行归一化处理, 得到整幅图像的 WS-SSIM 质量评价值 $S_{WS-SSIM}$:

$$S_{WS-SSIM} = \left(\sum_{n=1}^N S_{WSSIM}(n) \right) / \left(\sum_{n=1}^N (w_n) \right) \quad (4)$$

其中, N 表示窗口滑动的次数。

WS-SSIM 评价指标通过对每一帧的评价值求平均得到。给出的 WS-SSIM 客观质量评价模型由于结合了结构相似度评价模型以及等矩形全向视频的特征, 在评价等矩形全向视频图像的客观质量上表现出了良好的性能。这里, 只针对最常用的等矩形投影给出了 WS-SSIM 评价模型; 事实上, 对于不同的投影方式, 例如立方体投影、八面体投影等, 在逆投影回球形视频时都存在面积比例关系; 因此, 该评价模型也可以通过修改权重用于评价其他全向平面视频的客观质量, 即式 (4) 的 WS-SSIM 客观质量评价模型具有良好的适应性。

2.2 基于 WS-SSIM 的全向视频编码率失真优化方法

SSIM 相比 PSNR 是一种考虑结构信息的、更符合人眼感知的客观质量评价模型, 但 HEVC 编码中率失真函数的失真定义仅仅考虑了像素级失真, 所计算的率失真代价并不能很好地符合人眼的感知。在等矩形全向视频系统中, 平面上不同区域逆投影回观看球面时存在不同的面积比例关系, 所以在编码过程中不同区域的失真应分别对待。

HEVC 编码标准的目的是在保证一定视频质量的条件尽可能地减少编码比特数, 或者在一定编码比特数的条件下尽可能地减小失真^[20], 它通过比较率失真代价选择最优的编码模式实现:

$$\min \{J = D + \lambda \times R\} \quad (5)$$

其中, J 为当前编码模式的率失真代价, D 是失真, λ 和 R 分别表示拉格朗日因子和编码所产生的码率。

令 x 表示原始编码宏块, y 表示 x 的编码重建宏块, 假设编码所带来的误差是线性的, 即

$y = x + e$ 。 e 表示编码重建后所带来的误差, 且 e 的均值为 0, 方差为 σ_e^2 。那么均方误差 (mean square error, MSE) 失真可由式 (6) 计算得到:

$$MSE = \frac{1}{M} \sum_{i=1}^M (y_i - x_i)^2 = \frac{1}{M} \sum_{i=1}^M e_i^2 \quad (6)$$

其中, M 表示当前编码宏块中像素的个数。根据大数定律, 当 M 变大时, $MSE \rightarrow \sigma_e^2$ 。

当采用较高量化精度时, 存在以下关系: $u_x \approx u_y$, $\sigma_y^2 \approx \sigma_x^2 + \sigma_e^2$, $\sigma_{xy} \approx \sigma_x^2$ 。将此关系代入式 (1) 得:

$$SSIM = \frac{2\sigma_x^2 + c_2}{2\sigma_x^2 + \sigma_e^2 + c_2} \quad (7)$$

由于 $0 < SSIM < 1$, 并且 SSIM 值越大表示重建图像块与原始图像块的相似度越高, 根据式 (8) 定义基于 SSIM 的失真度量值 dSSIM:

$$dSSIM = \frac{1}{SSIM} = 1 + \frac{\sigma_e^2}{2\sigma_x^2 + c_2} \approx 1 + \frac{MSE}{2\sigma_x^2 + c_2} \quad (8)$$

其中, $dSSIM > 0$, 并且 dSSIM 越大, 重建图像的结构相似度失真越大。式 (8) 表明 dSSIM 与 MSE 存在的关系, 由此可以定义当前编码宏块的新的率失真代价模型为:

$$J = M \times dSSIM + \lambda \times R \approx M \left(1 + \frac{MSE}{2\sigma_x^2 + c_2} \right) + \lambda \times R = M + \frac{1}{2\sigma_x^2 + c_2} \times (D_{SSE} + (2\sigma_x^2 + c_2) \times \lambda \times R) \quad (9)$$

其中, D_{SSE} 表示当前编码模式下的原始误差平方和 (sum of square error, SSE) 失真。相似地, 为每个编码宏块定义如下率失真模型:

$$J = D_{SSE} + (2\sigma_x^2 + c_2) \times \lambda \times R \quad (10)$$

相比原始率失真优化模型, 式 (10) 中的拉格朗日因子 λ 乘以了一个与局部宏块方差相关的系数。通过引入该系数将率失真模型的失真度量由 SSE 转化为 dSSIM, 从而在编码率失真优化过程中考虑图像的结构信息。将 dSSIM 作为率失真优化模型的失真度量之后, 为了保持帧的整体编

码速率不变^[21], 通过式 (11) 得到第 i 个编码宏块基于 dSSIM 的拉格朗日因子:

$$\lambda_{\text{SSIM}i} = w_{\text{SSIM}i} \times \lambda = \frac{2\sigma_{xi}^2 + c_2}{\exp\left(\frac{1}{S} \sum_{j=1}^S \log(2\sigma_{xj}^2 + c_2)\right)} \times \lambda \quad (11)$$

其中, λ 为原始拉格朗日因子, S 表示当前编码帧中宏块的个数, σ_{xi}^2 和 σ_{xj}^2 分别表示第 i 个和第 j 个编码原始宏块的方差。

由于全向视频最终以球面的形式在头戴式显示器上显示, 而编码过程中的等矩形平面失真与观看的球面失真存在非线性关系, 即由投影时的像素插值造成, 考虑到此失真非线性关系, 根据逆投影时的面积比例将失真权重引入率失真代价计算中, 建立第 i 个编码宏块的基于 WS-SSIM 的率失真优化模型:

$$\min\{J = D_{\text{SSE}} \times w_{ci} + w_{\text{SSIM}i} \times \lambda \times R\} \quad (12)$$

其中, D_{SSE} 、 R 和 λ 分别为原始 SSE 失真情况、原始码率以及原始拉格朗日因子, 考虑到计算复杂度, 使用 w_{ci} (当前编码宏块中心点像素的失真权重) 代表整个宏块的失真权重, 其中 w_{ci} 通过式 (2) 计算所得。式 (12) 的率失真优化模型也可以等效为:

$$\min\{J = D_{\text{SSE}} + \lambda_i \times R = D_{\text{SSE}} + \frac{w_{\text{SSIM}i}}{w_{ci}} \times \lambda \times R\} \quad (13)$$

其中, 第 i 个宏块的基于 WS-SSIM 的拉格朗日因子计算如式 (14):

$$\lambda_i = \frac{w_{\text{SSIM}i}}{w_{ci}} \times \lambda = \frac{2\sigma_{xi}^2 + c_2}{\exp\left(\frac{1}{S} \sum_{j=1}^S \log(2\sigma_{xj}^2 + c_2)\right)} \times w_{ci} \quad (14)$$

分析式 (14) 可知: 在相同纬度下, 当前编码宏块的纹理越复杂 (即 σ_{xi}^2 越大), 该块的拉格朗日因子越大, 其失真掩蔽性越好; 在相同方差的情况下, 当前编码宏块的纬度越高 (即 w_{ci} 越小), 则该块的拉格朗日因子也越大, 其所能容忍的失真也就越大。

2.3 基于 WS-SSIM 的自适应 QP 分配

为了达到最优的编码效果, 根据局部宏块的拉格朗日因子的变化相应地调整量化步长是必须的, 即相应地调整量化参数 QP。不同的宏块根据 WS-SSIM 率失真优化后的拉格朗日因子的变化设置不同的量化调整参数 $\text{QP}_{\text{offset}}$ 。HEVC 视频编码中, 原始拉格朗日因子 λ 与 QP 存在以下关系:

$$\lambda = \beta \times 2^{(\text{QP}-12)/3} \quad (15)$$

其中, β 为与 QP 无关的常数。根据 WS-SSIM 率失真优化后拉格朗日因子的变化, 得到式 (16):

$$\text{QP}_{\text{offset}i} = \text{QP}_i - \overline{\text{QP}_i} = 3 \times \text{lb}(w_{\text{SSIM}i} / w_{ci}) \quad (16)$$

其中, $\overline{\text{QP}_i}$ 为第 i 个宏块的原始 QP 值, QP_i 表示 WS-SSIM 率失真优化后根据 λ_i 得到的新的 QP 值。根据式 (16) 得到基于 WS-SSIM 的率失真优化方法的 QP 调整策略:

$$\text{QP}_{\text{offset}i} = \text{QP}_i - \overline{\text{QP}_i} = 3 \times (P_i - \frac{1}{S} \sum_{j=1}^S P_j) - 3 \times \text{lb}w_{ci} \quad (17)$$

$$P_i = \text{lb}(2\sigma_{xi}^2 + c_2) \quad (18)$$

式 (17) 和式 (18) 表明了 WS-SSIM 客观质量评价模型指导率失真优化过程中每个宏块的量化调整参数 $\text{QP}_{\text{offset}}$ 。在实际编码过程中, 为了降低计算复杂度, 以 64×64 的 CTU (coding tree unit, 编码树单元) 为单位进行操作。在编码每帧图像之前, 先计算每个 CTU 的方差以及方差归一化因子。在编码每个 CTU 前, 根据当前 CTU 方差和中心点像素的失真权重调整拉格朗日因子以及 $\text{QP}_{\text{offset}}$, 使该 CTU 达到最佳的编码结果。

3 实验结果与分析

本文实验分为两个部分, 首先通过主观感知质量和客观评价指标的比较验证所提出的基于球形加权结构相似度 (WS-SSIM) 的全向视频客观质量评价模型的有效性; 其次为了验证所提出的基于 WS-SSIM 的全向视频编码率失真优化方法的有效性, 将所提出方法与 HEVC 编码标准的



表 1 全向视频序列信息

序列名称	分辨率	位深	场景是否固定
Ballooning_vr_25p_2880×2160.yuv	4 096 dpi×2 048 dpi	8	否
Dancing360.yuv	4 096 dpi×2 048 dpi	8	是
DrivingInCountry_3840×1920_30fps_8bit_420_erp.yuv	3 840 dpi×1 920 dpi	8	否
Drive360.yuv	4 096 dpi×2 048 dpi	8	否
PoleVault_le_3840×1920_30fps_8bit_420_erp.yuv	3 840 dpi×1 920 dpi	8	是
Paramotor_Training_vr_25p_2880×2160.yuv	4 096 dpi×2 048 dpi	8	是

HM16.9 测试平台进行了对比。实验中所使用的全向视频测试序列来自于 Nokia^[22]和 Letin VR^[23]等, 包含镜头移动和镜头固定两种情况, 相关情况见表 1。其中, Balloning 序列镜头移动且每一帧下半部分纹理较复杂; Dancing 序列纹理简单且镜头固定, 运动区域所占比例小; DrivingInCountry 和 Drive 序列镜头移动且高纬度区域纹理较简单; PoleVault 序列场景复杂度高, 纹理较丰富; Training 序列由于前几帧对象距离太近, 存在拼接过缝现象。图 1 给出了每个序列第一帧的情况。

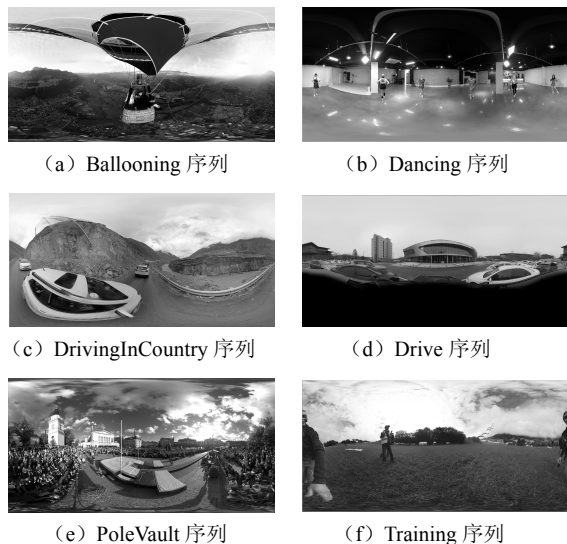


图 1 全向视频测试序列第一帧示意图

3.1 WS-SSIM 评价模型性能验证

所提出的 WS-SSIM 全向视频客观质量评价模型在传统 SSIM 评价模型的基础上, 结合投影平面不同区域的面积比例关系, 使 SSIM 评价模型适合评价全向视频的客观质量。本文对全向视

频序列加入不同类型的噪声 (高斯噪声和均匀噪声), 在不同主观感知质量的基础上对比 WS-PSNR 与 WS-SSIM 评价结果, 以验证 WS-SSIM 客观质量评价模型的有效性。DrivingInCountry 序列和 Drive 序列受不同噪声污染后的局部区域放大图如图 2 所示。

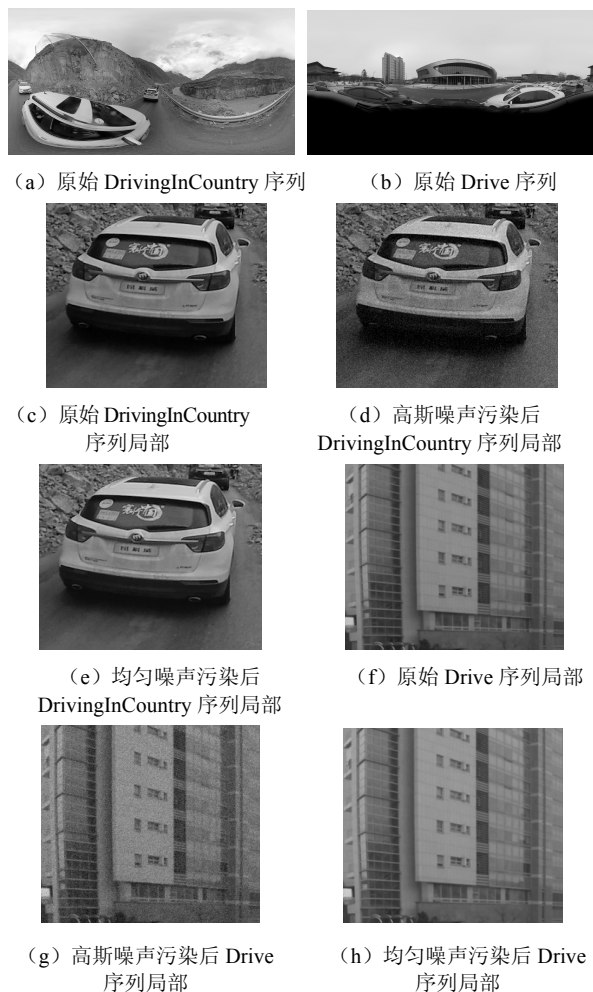


图 2 DrivingInCountry 序列和 Drive 序列受不同噪声污染局部区域放大图

观察图 2 可以看出, DrivingInCountry 序列和 Drive 序列在加入高斯噪声和均匀噪声后, 主观质量上呈现出较大的差异, 并且图 2 中的局部放大区域均位于全向视频的低纬度区域, 与用户直接观看的视口内容具有较强一致性。对于 DrivingInCountry 序列, 对比图 2 (c) 和图 2 (d)、图 2 (e), 可以看出受高斯噪声污染后的全向视频具有较差的主观质量, 噪声明显, 然而采用 WS-PSNR 进行客观质量评价所得结果分别为 29.99 dB 和 30.07 dB。评价结果显示两个全向视频的质量所差无几, 与人眼的主观感知产生了较大的差异。当采用所提出的 WS-SSIM 模型进行评价时, 受高斯和均匀噪声污染后的全向视频的客观质量分别为 0.645 5 和 0.996 9。评价结果显示两个全向视频的质量存在较大的差异, 并且质量评价结果与人眼主观感知质量一致。同样对于 Drive 序列, 受高斯和均匀噪声污染后的全向视频在采用 WS-PSNR 模型进行评价时, 客观评价结果分别为 30.04 dB 和 30.07 dB, 而采用所提出的 WS-SSIM 模型进行评价的所得结果分别为 0.556 2 和 0.978 5。通过对全向视频加入不同类型噪声, 验证了所提出的 WS-SSIM 是一种更符合人眼主观感知的全向视频客观质量评价模型。

3.2 基于 WS-SSIM 的全向视频编码率失真优化结果与分析

所提出的基于 WS-SSIM 的全向视频编码率失真优化方法在 HM16.9 平台上进行改进。编码方式采用典型的低时延 P 帧 (low-delay P, LDP), 量化参数分别选取 22、27、32 和 37。LDP 编码方式中, 每个序列的第一帧为 I 帧, 其余都为单向参考 P 帧。I 帧作为参考帧, 其质量对后续 P 帧的编码质量有较大影响, 所以在实际编码中对 I 帧不作处理, 仅从第二帧开始使用本文所提出的方法。在 WS-PSNR^[14]和 WS-SSIM 评价模型下比较所提出方法和原始 HM16.9 平台的率失真性能, 实验结果见表 2。BD-WS-PSNR 和 BD-WS-SSIM 表示在相同码率情况下, 所提出方法相比原始 HM16.9 平台提升的质量, 数值越大表示质量提升得越多; BD-Rate 表示在相同质量情况下, 所提出方法相比原始 HM16.9 平台节省的码率, 数值越小表示码率节省得越多。

分析表 2, 可以看到使用 WS-PSNR 进行评价时, 在相同的码率下, 所提出方法相比 HM16.9 平台中方法质量最多可以提升 0.84 dB、平均提升 0.45 dB; 在相同质量的情况下, 码率最多可以节省 26.5%、平均节省 13.61%。使用 WS-SSIM 评价时, 在相同码率下, 质量最多提升 0.010 1、平均提升 0.005 5;

表 2 所提出方法与原始 HM16.9 平台结果对比

序列名称	WS-PSNR 评价模型		WS-SSIM 评价模型	
	BD-WS-PSNR/dB	BD-Rate	BD-WS-SSIM	BD-Rate
Balloning	0.84	-26.50%	0.008 5	-40.16%
Dancing	0.60	-14.97%	0.003 2	-27.26%
DrivingInCountry	0.41	-14.73%	0.010 1	-24.10%
Drive	0.44	-10.79%	0.002 5	-19.34%
PoleVault	0.31	-10.99%	0.005 2	-21.13%
Training	0.10	-3.65%	0.003 6	-14.90%
平均值	0.45	-13.61%	0.005 5	-24.48%



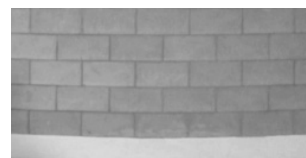
在相同质量下, 码率最多节省 40.16%、平均节省 24.48%。特别指出, Balloning 序列相比其他序列在图像下半部分区域纹理复杂度较高且相机移动, 在原始平台编码中需要分配大量的码流, 然而这些区域失真掩蔽性较强, 所以本文提出方法会对这些区域分配相对较少的比特, 在保证感知质量一致的情况下所提出方法相比原始平台节省较多的码流。所提出的 WS-SSIM 的率失真优化编码方法, 在率失真优化过程中考虑的是结构相似度失真, 而与 WS-PSNR 只简单地考虑 PSNR 失真不同, 所以使用 WS-PSNR 评价模型评价提出的基于球形加权结构相似度 (WS-SSIM) 的全向视频编码率失真优化方法的性能已不合适。WS-SSIM 评价模型既考虑了等矩形全向视频的特征, 也结合了 SSIM 评价模型的本质, 适合评价所提出方法。对实验结果分析可得, 虽然用 WS-PSNR 评价所提出方法的结果优于原始平台, 但是总体效果远没有 WS-SSIM 评价效果好, 这也验证了本文提出的基于 WS-SSIM 率失真优化方法的有效性。值得一提的是, 无论是场景固定还是场景移动的序列, 所提出方法都可以取得比较好的编码效果。

Dancing 序列第 2 帧原始图以及局部区域放大图如图 3 所示。为了更加全面地评判所提出方法, 本文比较了 2 个不同码率级别下重建帧的质量。首先比较重建图像整帧质量, 所提出方法重建帧的码率均低于原始编码平台的码率, 并且用 WS-SSIM 评价重建帧所得结果均优于原始平台, 实现相比原始平台更低码率和更高质量的编码效果; 分析局部区域放大图, 所提出方法局部区域相比原始平台在纹理结构上保留得更好, 主观质量更佳, 并且从局部区域的结构相似度评价指标可以看到, 本文评价结果明显优于原始平台。

为进一步说明所提出方法提高全向视频编码效果的显著有效性, 将所提出方法和参考文献[13]的纬度自适应 QP 方法、参考文献[21]的传统 SSIM 率失真优化方法进行了比较。参考文献[13]对等



(a) 原始视频第 2 帧



(b) 原始图像局部放大区域

(c) 码率= 2 375.184 kbit/s, 重建帧 WS-SSIM= 0.979 7, 放大区域 SSIM=0.963 804



(d) 码率= 2 183.904 kbit/s, 重建帧 WS-SSIM= 0.980 838, 放大区域 SSIM=0.972 421



(e) 码率= 644.995 kbit/s, 重建帧 WS-SSIM= 0.962 814, 放大区域 SSIM=0.908 015



(f) 码率= 568.906 kbit/s, 重建帧 WS-SSIM= 0.966 969, 放大区域 SSIM=0.936 007

图 3 Dancing 序列第 2 帧原始图以及局部区域放大图

矩形投影下的平面全向视频帧的不同纬度区域分配不同的量化参数以去除像素冗余, 并且随着纬度的升高, 量化参数随之增大; 参考文献[21]的率失真优化方法主要针对传统视频, 相比本文所提出方法缺少了对全向视频特征的考虑。编码性能的比较在所提出的更符合人眼主观感知的全向视频评价模型 WS-SSIM 下实现。

所提出方法与纬度自适应 QP 方法、传统 SSIM 率失真优化方法的编码结果比较结果见表 3。从表 3 中可以看出,所提出的方法相比纬度自适应 QP 方法、传统 SSIM 率失真优化方法均实现了更好的编码增益。相比纬度自适应 QP 方法,所提出方法实现了在相同质量下平均多节省 9.5% 的码率。其中,对于纹理较为平坦的序列,如 Dancing 序列,纬度自适应 QP 方法提升的效果较差,原因是在 P 帧中这些平坦区域几乎都是跳过 (skip) 模式,产生的码流很少并且与量化参数无关,绝大多数码流由低纬度非 skip 区域预测残差量化产生,而本文所提出方法对低纬度非 skip 区域(纹理相对复杂、运动且掩蔽性较好区域)分配相对较大的量化参数,实现相同感知质量下更多码流的节省。相比传统 SSIM 率失真优化方法,所提出的方法实现了在相同质量下平均多节省 16.78% 的码率。应该指

出,对于纹理平坦的序列,如 Dancing 和 Drive 序列,传统 SSIM 率失真优化方法表现出了良好的编码性能,甚至在 Dancing 序列上要略优于本文所提出方法。但是对于纹理较为复杂的序列,如 DrivingInCountry 和 PoleVault,则表现出了较差的编码性能。所提出方法通过结合全向视频特征,在纹理较为复杂的全向视频序列上较大幅度地提升编码性能。另外,图 4 给出了 Balloning 序列和 Drive 序列在不同编码方法下的率失真性能。通过图 4 可以看出,所提出方法相比传统 SSIM 率失真优化方法和纬度自适应 QP 方法具有更好的率失真性能。

4 结束语

本文针对传统视频编码率失真模型中失真评估不合理以及等矩形投影下全向视频的特征,提出了一种基于球形加权结构相似度的全向视频编

表 3 所提出方法与纬度自适应 QP 方法、传统 SSIM 率失真优化方法编码结果对比

序列名称	纬度自适应 QP		传统 SSIM 率失真优化	
	BD-WS-SSIM	BD-Rate	BD-WS-SSIM	BD-Rate
Balloning	0.002 1	-8.87%	0.004 5	-22.18%
Dancing	0.002 6	-21.49%	-0.000 6	5.85%
DrivingInCountry	0.001 8	-4.38 %	0.018 5	-36.35%
Drive	0.001 0	-8.31%	0.000 3	-3.72%
PoleVault	0.001 7	-6.33%	0.020 8	-35.13%
Training	0.002 4	-7.60%	0.002 1	-9.18%
平均值	0.001 9	-9.50 %	0.007 6	-16.78%

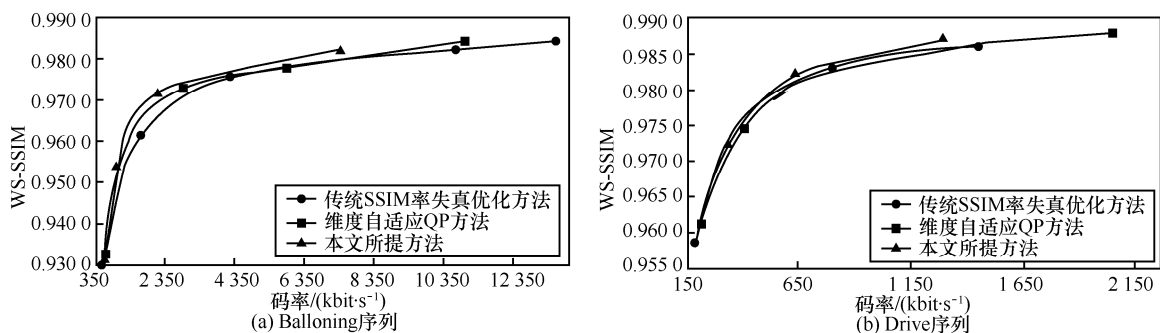


图 4 所提出方法与纬度自适应 QP 方法、传统 SSIM 率失真优化方法部分序列率失真性能比较



码率失真优化方法,在考虑投影平面内部纹理结构失真情况的同时,结合投影平面失真与球面失真的关系,在编码过程中指导率失真模型的优化以及量化参数的自适应选择。实验结果表明,在相同质量下,所提出方法能够明显节省码率。在未来工作中,计划研究基于 WS-SSIM 评价模型的全向视频编码的码率分配策略,以进一步提升全向视频的编码效果。

参考文献:

- [1] 罗莹, 宋利, 解蓉, 等. 全景媒体的系统架构研究综述[J]. 电信科学, 2018, 34(2): 88-98.
LUO Y, SONG L, XIE R, et al. Review of system architecture of omnidirectional media[J]. Telecommunications Science, 2018, 34(2): 88-98.
- [2] RAI Y, CALLET P L, GUILLOT P. Which saliency weighting for omni directional image quality assessment? [C]//Ninth International Conference on Quality of Multimedia Experience, May 31-June 2, 2017, Erfurt, Germany. Piscataway: IEEE Press, 2017: 1-6.
- [3] CHEN Z Z, LI Y M, ZHANG Y X. Recent advances in omnidirectional video coding for virtual reality: projection and evaluation[J]. Signal Processing, 2018(146): 66-78.
- [4] 罗传飞, 孔德辉, 刘翔凯, 等. 智慧家庭的 VR 全景视频业务实现[J]. 电信科学, 2017, 33(10): 185-193.
LUO C F, KONG D H, LIU X K, et al. A business realization of VR panoramic video in smart home[J]. Telecommunications Science, 2017, 33(10): 185-193.
- [5] SIMONE F D, FROSSARD P, WILKINS P, et al. Geometry-driven quantization for omnidirectional image coding[C]//Picture Coding Symposium (PCS), Dec 4-7, 2016, Nuremberg, Germany. Piscataway: IEEE Press, 2016: 1-5.
- [6] LI L, LI Z, BUDAGAVI M, et al. Projection based advanced motion model for cubic mapping for 360-degree video[C]//2017 IEEE International Conference on Image Processing (ICIP), Sept 17-20, 2017, Beijing, China. Piscataway: IEEE Press, 2017: 1427-1431.
- [7] LIN H C, HUANG C C, LI C Y, et al. AHG8: an improvement on the compact OHP layout[C]//Joint Video Exploration Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0056, January 12-20, 2017, Geneva, Switzerland. [S.l.:s.n.], 2017.
- [8] ZARE A, AMINLOU A, HANNUKSELA M, et al. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications[C]// ACM International Conference on Multimedia, October 15-19, 2016, Amsterdam, The Netherlands. New York: ACM Press, 2016: 601-605.
- [9] QIAN F, GOPALAKRISHNAN V, et al. Optimizing 360 video delivery over cellular networks[C]//The Workshop on All Things Cellular: Operations, Applications and Challenges, October 3-7, 2016, New York, USA. New York: ACM Press, 2016: 1-6.
- [10] SKUPIN R, SANCHEZ Y, POBORSKI D, et al. HEVC tile based streaming to head mounted displays[C]// 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Jan 8-11, 2017, Las Vegas, NV, USA. Piscataway: IEEE Press, 2017: 613-615.
- [11] TANG M, ZHANG Y, WEN J, et al. Optimized video coding for omnidirectional videos[C]//IEEE International Conference on Multimedia and Expo, July 10-14, 2017, Hong Kong, China. Piscataway: IEEE Press, 2017: 799-804.
- [12] YOUVALARI R G, AMINLOU A, HANNUKSELA M. Analysis of regional down-sampling methods for coding of omnidirectional video[C]//Picture Coding Symposium (PCS), Dec 4-7, 2016, Nuremberg, Germany. Piscataway: IEEE Press, 2016: 1-5.
- [13] HENDR Y, MUHAMMED C, MARTA K, et al. AHG8: adaptive QP for 360° video ERP projection[C]//Joint Video Exploration Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JVET-F0049, March 31-April 7, 2017, Hobart, Australia. [S.l.:s.n.], 2017.
- [14] SUN Y, LU A, YU L. Weighted-to-spherically-uniform quality evaluation for omnidirectional video[J]. IEEE Signal Processing Letters, 2017, 24(9): 1408-1412.
- [15] YU M, LAKSHMAN H, GIROD B. A framework to evaluate omnidirectional video coding schemes[C]// IEEE International Symposium on Mixed and Augmented Reality, Sept 29-Oct 3, 2015, Fukuoka, Japan. Piscataway: IEEE Press, 2015: 31-36.
- [16] ZAKHARCHENKO V, ALSHINA E, SINGH A, et al. AHG8: suggested testing procedure for 360-degree video[C]//Joint Video Exploration Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. JVET-D0027, October 15-21, 2016, Chengdu, China. [S.l.:s.n.], 2016.
- [17] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [18] YEO C, TAN H L, TAN Y H. SSIM-based adaptive quantization in HEVC[C]// IEEE International Conference on Acoustics, Speech and Signal Processing, May 26-31, 2013, Vancouver, BC, Canada. Piscataway: IEEE Press, 2013: 1690-1694.
- [19] CORBILLON X, DE SIMONE F, SIMON G. 360-degree video

head movement dataset[C]//ACM on Multimedia Systems Conference, June 20-23, 2017, Taipei, China. New York: ACM Press, 2017: 199-204.

- [20] ZHAO P, LIU Y, LIU J, et al. Low-complexity content-adaptive Lagrange multiplier decision for SSIM-based RD-optimized video coding[C]//IEEE International Symposium on Circuits and Systems, May 19-23, 2013, Beijing, China. Piscataway: IEEE Press, 2013: 485-488.
- [21] YEO C, TAN H L, TAN Y H. On rate distortion optimization using SSIM[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(7): 1170-1181.
- [22] SEBASTIAN S, ALIREZA A, IGOR D, et al. AHG8: tampere pole vaulting sequence for virtual reality video coding[C]//Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11. JVET-D0143, October 15-21, 2016, Chengdu, China. [S.l.:s.n.], 2016.
- [23] SUN W B, GUO R D, MEN X H, et al. AHG8: test sequences for virtual reality video coding from LetinVR[C]//Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11. JVET-D0179, October 15-21, 2016, Chengdu, China. [S.l.:s.n.], 2016.

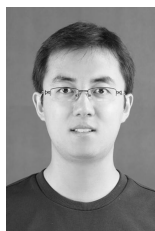
[作者简介]



周瑜锋（1995-），男，宁波大学信息科学与工程学院硕士生，主要研究方向为全向视频系统信号处理与编码。



郁梅（1968-），女，博士，宁波大学信息科学与工程学院教授，主要研究方向为三维视频系统信号处理、编码，视觉感知质量评价与系统优化。



杜宝祯（1985-），男，宁波大学信息科学与工程学院博士生，主要研究方向为视觉感知模型研究与三维视频感知编码。



宋洋（1989-），男，宁波大学信息科学与工程学院博士生，主要研究方向为图像处理与视觉通信、视频图像质量评价。



蒋刚毅（1964-），男，博士，宁波大学信息科学与工程学院教授，主要研究方向为全向视频/三维视频系统信号处理、编码与传输，视觉感知质量评价与系统优化。