

基于全景感知的码率控制优化编码

杜宝祯¹ 陈文明¹

DU Baozhen CHEN Wenming

摘要

提出一种基于全景感知的码率控制优化编码方法, 首先基于全景视频内容特点, 得出了面向全景视频的感知失真阈值模型, 然后在此基础上提出了码率控制感知权重计算方法, 从而对 LCU 层级码率控制模型进行了优化。实验结果表明, 所提出的方法能在保证编码质量的同时, 有效提升率失真性能, 编码效率明显提升。

关键词

视频编码; 全景视频; 码率控制; 视觉感知; 感知失真

doi: 10.3969/j.issn.1672-9528.2024.01.012

0 引言

近年来, 多媒体视频系统以其“身临其境”的视听体验被广泛运用于各个领域, 具有广阔的应用前景^[1]。在视频编码算法框架中, 码率控制技术在视频编码中起着重要作用。该技术通过调整量化参数来控制输出码流的总数据量, 以适应网络传输环境, 在保证一定视频质量前提下, 尽可能减少码率误差, 即编码码率与目标码率误差。特别是在有限带宽条件下, 对保证视频的质量和稳定性非常重要^[2]。对于全景视频系统, 由于其显示方式、数据量与普通视频存在显著差异, 因此如何使传统码率控制算法很好地适用于全景视频系统, 提升全景视频编码效率, 成为该系统应用亟待解决的问题^[3]。

目前已有大量相关研究, 如文献[4]提出了一种基于 tile 的全景视频自适应比特率算法, 通过强化学习可以根据网络和客户端状态对视频中的每个 tile 做出自适应的比特率决策, 该方法能够兼顾视频质量和视频流畅性, 同时可以为用户提供更高的 QoE 质量。文献[5]提出了一种简单而有效的视口自适应流算法, 该算法可以跨视频段提供稳定的视口质量, 从而增强观看全景视频时的用户体验。文献[6]提出了一种交替预测和更新步骤的递归贝叶斯估计的帧级恒比特率控制方法, 不仅可以精确估计码率, 而且可以根据先前编码帧的失真变化来分配目标比特, 同时兼顾了码率和失真。然而上

述视频编码方法, 未充分考虑人眼视觉感知特性, 编码性能难以进一步提升^[7]。因此, 如何结合 HVS 人眼视觉感知特性, 指导码率控制编码参数选择, 就成为研究的一个重要突破方向。本文通过研究全景视频内容特征, 提出了全景感知失真阈值模型, 并在此基础上对 LCU 层级码率控制模型进行修正。实验结果表明, 该方法能有效提升算法性能, 降低编码码率误差。

1 全景码率控制分配权重模型

1.1 全景感知失真阈值

研究发现, 人眼视觉存在的感知冗余, 使得图像的轻微降质并不会被察觉。文献[8]进一步定量地测量了这种视觉冗余度量, 具体为:

$$QP_p = \alpha \times QP_{org} + \beta \quad (1)$$

式中: QP_{org} 表示视频编码原始的量化参数, 参数 α 和 β 取值分别为 -0.235 8 和 9.39。

上述模型是基于平面视频主观实验基础上得出的, 而对于 ERP 格式的全景视频, 由于各个纬度采样程度不同, 这类图像的高纬度区域特别是两极区域被极度拉升, 因此存在大量感知冗余。由文献[9]可知, 对于 ERP 投影格式全景视频, 其图像内任意像素点的维度权重为:

$$\omega_{ERP}(i, j) = \cos \frac{(j+0.5-H/2)\pi}{H} \quad (2)$$

式中: W 为视频宽度, H 为视频高度, (i, j) 为视频中像素点所在的位置。由式(2)可知, 图像中越接近两极和高纬度的区域, 维度权重越小, 可容忍失真越大; 反之则维度权重越大, 可容忍失真越小。由此, 图像中 LCU 的全景维度均值可计算为:

$$\omega_p(m, n) = \frac{\gamma_1}{1 + \omega_{LCU}(m, n)} + \gamma_2 \quad (3)$$

1. 宁波职业技术学院电子信息工程学院 浙江宁波 315800
[基金项目] 浙江省教育厅一般科研项目“基于视觉感知的立体全景视频编码方法研究”(项目编号: Y202147539); 宁波职业技术学院 2021 年下半年校级人才引进培养专项课题“基于机器学习的屏幕内容高效编码理论与关键技术研究”(项目编号: NZ21RC005)

式中: (m, n) 为当前 LCU 左上角顶点在整帧图像中的坐标值, N 表示 LCU 尺寸。

定义 ω_p 为当前待编码 LCU 的全景维度感知因子, 以表征全景视频图像对维度失真的敏感程度, ω_p 越小表示图像区块对维度因素失真敏感度越低, 可感知失真容忍程度越高; 反之, 敏感度越高, 失真容忍程度越低。

$$\omega_p(m, n) = \frac{\gamma_1}{1 + \omega_{LCU}(m, n)} + \gamma_2 \quad (4)$$

式中: (m, n) 为当前 LCU 左上角顶点在整帧图像中的坐标值, $\omega_{LCU}(m, n)$ 表示当前 LCU 块全景维度均值。通过反复实验, 参数 γ_1 和 γ_2 最终取值分别为 0.4 和 0.8。

结合式 (1) 和式 (4), 全景感知 Q_p 阈值为:

$$Q_{GZ}^P = Q_P \times \omega_p(m, n) \quad (5)$$

研究表明, 失真与编码量化步长相关。文献 [10] 给出了定量模型, 失真与量化步长 Q 满足如下关系:

$$D = \frac{\delta^2 Q^2}{12\delta^2 + Q^2} \quad (6)$$

式中: δ^2 表示 LCU 未量化前的 DCT 系数方差, Q 表示量化步长。由于量化参数 QP 和量化步长 Q 满足 $Q = 2^{(QP-4)/6}$, 因此失真与量化参数 QP 的关系满足:

$$D = \frac{\delta^2 2^{(QP-4)/3}}{12\delta^2 + 2^{(QP-4)/3}} \quad (7)$$

由失真 D 与 MSE 的关系, 对于当前 LCU, 可以进一步得到:

$$MSE_{LCU} = \frac{\delta^2 2^{(QP-4)/3}}{(12\delta^2 + 2^{(QP-4)/3}) \times N^2} \quad (8)$$

式中: QP 为编码量化参数, δ^2 表示当前 LCU 未量化前的 DCT 的系数方差。由于当前 LCU 尚未进行编码, 考虑到时域的相关性, δ^2 值由与 LCU 相对位置前一帧已编码的 LCU 相关参数计算得到。

由此可得到全景感知的失真阈值模型 MSE_p , 具体计算为:

$$MSE_p = \frac{\delta^2 2^{(QP_{GZ}-4)/3}}{12\delta^2 + 2^{(QP_{GZ}-4)/3}} \times \frac{1}{N^2} \quad (9)$$

1.2 码率控制模型修正

由经典码率控制理论可知, LCU 层的目标分配比特为:

$$Bit_{LCU} = \frac{Bit_{Pic} - Bit_{Header} - C_{Pic}}{\sum_{NotCodedLCUs} \omega_i} \cdot \omega_{LCU} \quad (10)$$

式中: Bit_{LCU} 表示当前 LCU 的目标比特数, Bit_{Pic} 表示当前帧总的目标比特, Bit_{Header} 表示当前帧预测头信息比特数, C_{Pic} 表示当前帧中已编码编码单元消耗的比特数, ω_{LCU} 表示当前 LCU 的比特分配权重, ω_i 表示尚未编码的第 i 个 LCU 比特分配权重且每个 LCU 的权重 ω_{LCU} 计算公式为

$\omega_{LCU} = MAD^2_{LCU}$, 其中 MAD 为原始像素与预测像素值的平均绝对差值。由于 MAD 仅从信号处理角度简单计算了原始信号与预测信号的失真, 并未考虑人眼视觉系统以及全景视频特点, 因而由此得到的码率分配权重对视觉感知和视频内容特征的特征有限, 并不能更有效地去除感知冗余。

本文采用全景感知的失真阈值计算感知权重, 定义 ω_p 为当前 LCU 的码率分配感知权重, 计算公式为:

$$\omega_p = \frac{MSE_p}{N^2} \quad (11)$$

进一步对式 (11) 进行修正, 最终得到 LCU 层的全景码率控制分配方案, 计算方法为:

$$Bit'_{LCU} = \frac{Bit_{Pic} - Bit_{Header} - C_{Pic}}{\sum_{NotCodedLCUs} \omega_{p_i}} \cdot \omega_p \quad (12)$$

式中: ω_p 为当前 LCU 码率分配感知权重, ω_{p_i} 表示尚未编码的第 i 个 LCU 码率分配感知权重。

2 实验结果分析

编码实验选取视频编码标准参考软件 HM16.20 作为测试平台, 所提出算法基于该平台实现。原始算法采用 HM16.20 默认原始算法标准, 且开启码率控制模式。实验硬件配置为 Intel(R) Core(TM) i7-10700 CPU, 主频 2.9 GHz, 内存为 64 GB 的 64 位 Win10 操作系统, 开发工具 VS2013。为了更好地比较所提出的算法与原始算法性能, 我们使用以下方法生成目标码率。首先基于全景视频编码通用测试条件 (CTC) 运行原始算法, 实验选取全景视频序列 AerialCity、DrivingInCity、Balboa、BranCastle2、Landing2 作为测试序列^[11]。测试序列的具体参数如表 1 所示。

表 1 测试序列参数

序列	分辨率
AerialCity	3840×1920
DrivingInCity	3840×1920
Balboa	6144×3072
BranCastle2	6144×3072
Landing2	6144×3072

实验中, 将生成的码率四舍五入, 并用作所提出算法的目标码率。这样, 由于所提出的算法与原始算法的码率相似, 因此可以使性能比较更加简洁。进一步采用, 更适合全景视频的质量指标 WS-PSNR、S-PSNR、CPP-PSNR 作为质量评价指标^[12-14], 分别计算出各视频序列在该质量评价方法 WS-PSNR、S-PSNR、CPP-PSNR 下的率失真性能指标 $BDBR_{WS-PSNR}(\%)$ 、 $BDBR_{S-PSNR}(\%)$ 、 $BDBR_{CPP-PSNR}(\%)$, 以综合评估所提出的算法性能。表 2 列出了所提出算法相比于原始算法的各个序列的率失真性能。

表 2 所提出的算法相比于原始算法的率失真性能

视频序列	BDBR _{WSPSNR} /%	BDBR _{SPSNR} /%	BDBR _{CPPSPNR} /%
AerialCity	-1.8	-1.9	-1.8
DrivingInCity	-0.2	-0.1	-0.1
Balboa	-1.2	-1.1	-1.1
BranCastle2	-0.9	-0.8	-0.9
Landing2	-0.4	-0.4	-0.5
平均	-0.9	-0.9	-0.9

从表 2 中可以看出，所提出的算法采用 BDBR_{WSPSNR}、BDBR_{SPSNR}、BDBR_{CPPSPNR} 率失真性能均值为 -0.9%、-0.9%、-0.9%，根据各指标含义，表明在相同 WS-PSNR、S-PSNR、CPP-PSNR 质量评价指标条件下，相比于原始平台算法，采用所提出的算法编码码率节省均值在 0.9% 左右。从实验结果来看，针对各不同场景、不同运动状态的全景视频序列，所提出的算法率失真性能均有不同程度提升。但由于各测试序列运动状态、场景等视频特性各异，所提出的算法在各序列测试的具体提升程度有所差异。其中 AerialCity 序列性能提升最为明显，各指标下的算法性能提升分别为 -1.8%、-1.9%、-1.8%，该序列为远景拍摄，视频运动比较平缓。DrivingInCity 序列性能提升最低，各指标下的算法性能提升分别为 -0.2%、-0.1%、-0.1%，该序列为车辆在行进过程中的全景拍摄视频，视频运动比较剧烈，周围场景变换程度更高，率失真性能提升比较有限。但从总体上来看，所提出的算法相比于原始算法在各质量指标情况下，其率失真性能均有所提升。因此相比于原始算法，所提出的算法性能更优。

3 结论

本文在研究全景视频内容特点基础上，将已有失真阈值模型向全景视频领域进行了推广，进一步提出了全景感知的失真阈值计算感知权重，并在此基础上提出了全景感知的码率控制优化编码算法。实验测试表明，所提出的方法能有效提升编码的率失真性能。针对各全景视频序列，所提出的算法编码率失真性能有所差异。下一步，将进一步研究图像内容特征对全景失真的影响，建立基于图像内容特征的编码失真模型并应用于视频编码。

参考文献:

- [1] WEI X, ZHOU M, KWONG S, et al. A hybrid control scheme for 360-degree dynamic adaptive video streaming over mobile devices[J]. IEEE transactions on mobile computing, 2021, 21(10): 3428-3442.
- [2] YANG S, HU J, JIANG K, et al. Hybrid - 360: an adaptive bitrate algorithm for tile - based 360 video streaming[J]. Transactions on emerging telecommunications technologies, 2022, 33(4): 4430.
- [3] LI L, YAN N, LI Z, et al. λ -domain perceptual rate control

for 360-degree video compression[J]. IEEE journal of selected topics in signal processing, 2019,14(1):130-145.

- [4] YANG S,HU J,JIANG K,et al. Hybrid-360: an adaptive bitrate algorithm for tile - based 360 video streaming[J]. Transactions on emerging telecommunications technologies, 2022(4):33.
- [5] DUC N, HOANG H L D, HUONG T T, et al.Smooth viewport bitrate adaptation for 360 video streaming[C]//The 6th NAFOSTED Conference on Information and Computer Science (NICS).Piscataway:IEEE,2019:512-517.
- [6] HYUN M H, LEE B, KIM M. A frame-level constant bit-rate control using recursive bayesian estimation for versatile video coding[J].IEEE access, 2020(99):1.
- [7] KI S, BAE S H, KIM M, et al. Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding[J]. IEEE transactions on image processing, 2018, 27(7): 3178-3193.
- [8] 杜宝祯,陈文明,姜浩,等. 基于视频内容编码感知失真模型研究 [J]. 信息与电脑（理论版）,2022,34(24):199-202.
- [9] 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.
- [10] XU L, JI X Y, GAO W, et al. Laplacian distortion model (LDM) for rate control in video coding[C]//Advances in multimedia information processing-PCM 2007: 8th Pacific Rim Conference on Multimedia.Berlin:K.C. Wong Foundation, 2007:11-14.
- [11] BOYEE J,ALSHINA E,ABBAS A.,et al. JVET common test conditions and evaluation procedures for 360 Video[C]// Document JVET-D1030.Piscataway:IEEE,2016:10-20.
- [12] YE Y, ALSHINA E, BOYCE J. Algorithm descriptions of projection format conversion and video quality metrics in 360Lib[J]. Joint video exploration team of ITU-T SG, 2017, 16.
- [13] CHEN S, ZHANG Y, LI Y, et al. Spherical structural similarity index for objective omnidirectional video quality assessment[C]//2018 IEEE international conference on multimedia and expo (ICME). Piscataway:IEEE, 2018: 1-6.
- [14] TRAN H T T, NGOC N P, BUI C M, et al. An evaluation of quality metrics for 360 videos[C]//2017 ninth international conference on ubiquitous and future networks (ICUFN). Piscataway: IEEE, 2017: 7-11.

【作者简介】

杜宝祯（1985—），男，甘肃陇西人，博士，副教授，研究方向：计算机视觉、图像处理、图像视频编码。

（收稿日期：2023-10-17）