DOI: 10.19650/j.cnki.cjsi.J2108588

基于视觉感知的率失真优化算法*

魏宏安[1,2],刘嘉棋[1,2],林丽群[1,2],杨 静[1,2],陈炜玲[1,2]
(1. 福州大学物理与信息工程学院 福州 350108; 2. 福建省媒体信息智能处理与无线传输重点实验室 福州 350108)

摘 要:率失真优化在视频编码中起着关键作用,其目的是在压缩效率和视频质量失真之间取得平衡。现有的率失真算法主要针对视频中时间和空间冗余的消除,未充分考虑视频内容的主观感知冗余。本文提出了一种基于视觉感知的率失真优化算法,通过数据驱动的 JND 预测模型推导拉格朗日乘数因子,并使用显著性模型优化拉格朗日乘子权重系数,最终融合应用于率失真优化,并采用 SW-SSIM 评估视频质量,实现视频编码的感知优化。实验结果表明,与 AVS3 标准率失真算法相比,本文所提算法平均节省 12.15% 的码率,SW-SSIM 提高了 0.004 3,有效降低了视频内容中的感知冗余,提高了视频感知质量和编码性能。

关键词: 率失真优化;恰可察觉失真;显著性模型;第三代国家数字音视频编码技术标准
中图分类号: TN762 TN919.81 TH69 **文献标识码**: A **国家标准学科分类代码**: 510.99

A rate-distortion optimization algorithm based on visual perception

Wei Hongan[1,2], Liu Jiaqi[1,2], Lin Liqun[1,2], Yang Jing[1,2], Chen Weiling[1,2]
(*1. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China;*
2. Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, Fuzhou 350108, China)

Abstract: The rate-distortion optimization plays a key role in video coding, which aims to achieve a tradeoff between compression efficiency and video quality distortion. The existing rate-distortion optimization algorithms mainly aim to eliminate time and space redundancy, which ignore subjective perception of video content and result in a large amount of perceptual redundancy in video. To address these issues, a rate distortion algorithm based on visual perception is proposed in this article. Firstly, the Lagrangian multiplier factor is obtained based on the data-driven just noticeable distortion prediction mode, which is more in line with the perception of the human eye. Secondly, the Lagrangian multiplier weight coefficient is based on salient model. Finally, the fusion of two models is applied to rate-distortion optimization, and SW-SSIM is used to evaluate video quality and achieve perceptual video coding optimization. Compared with the third generation audio and video coding standard algorithm, experimental results show that the proposed algorithm reduces bitrate by 12.15% averagely, and the salience weighted-structural similarity index metric increases by 0.004 3. Furthermore, the proposed algorithm reduces the perceptual redundancy in video content, and improves the video perceptual quality and coding performance.

Keywords: rate distortion optimization; just noticeable distortion; saliency model; the third generation audio and video coding standard

0 引 言

多媒体技术和现代通信技术的飞速发展,给视频技术的发展带来了前所未有的机遇。视频编码算法在降低视频比特率的同时,也不可避免的带来了失真,导致视频质量降低。如何在有限的比特率条件下,尽可能地减少失真,已成为近年的研究热点。现有的率失真优化(rate-distortion optimization, RDO)算法主要针对时间和空间冗余的消除[1],忽略了视频内容中的主观感知冗余。在视频编码中引入人的视觉特性,可以消除感知冗余,进一步提高编码性能[2]。

收稿日期:2021-09-16 Received Date: 2021-09-16
*基金项目:福建省教育厅中青年教师(JAT200024,JAT200007)、国家青年基金(61901119)项目资助

1 研究现状

近年来,人类视觉系统(human visual system, HVS)特性被广泛应用于视频编码,其方法大致可以分为两种,第 1 种是恰可察觉失真(just noticeable distortion, JND)模型,该模型可以消除视频中人眼无法察觉的感知冗余。第 2 种是视觉显著性模型,该模型可以为视频不同区域合理分配码率。

JND 模型通过模拟人类视觉系统的感知特性,获得人眼的最小视觉阈值,当视频中的变化低于该阈值时,人眼将察觉不到此变化。将 JND 模型应用于视频编码,可以更准确地估计视觉感知冗余,提高编码效率[2]。近年来,学者们在视频编码中 JND 模型应用方面做了许多工作。文献[3]通过 JND 阈值自适应量化编码器残差,提高了率失真性能和整体压缩效率。文献[4]引入了基于感知视频编码(perceptual video coding, PVC)的 JND 模型以优化视频压缩性能。文献[5]提出了一种基于像素级 JND 模型的拉格朗日因子调节,可以平衡比特率和感知失真。然而,现有的 JND 模型大多是通过数学推导构建的,不能准确描述人眼视觉特性,无法充分消除视频中的感知冗余。

视觉显著性模型考虑了 HVS 的注意力特点,可以区分视频内容的重要性[6],近年来收到广泛关注。文献[7]在率失真代价中使用显著分布的相对权重来增加显著区域的失真权重。文献[8]基于一组互补感知特征对拉格朗日值进行编码树单元级调整。文献[9]中提出了一种基于感兴趣区域编码块的快速分割方法。

此外,在率失真算法中,压缩引起的视频失真可以通过不同的方法进行评估,其中率失真度量将直接影响编码性能[10]。目前,在中国第三代国家数字音视频编码技术标准(the third generation audio and video coding standard, AVS3)中经常使用的误差平方和算法和绝对误差和算法与人类视觉系统的相关性不高。有研究指出结构相似度指标(structural similarity index metric, SSIM)与人类视觉具有的高度相似性,其不仅可以作为质量评价指标,还可以应用于视频编码。文献[11]提出了一种基于 SSIM 的视频编码感知 RDO 方案,该方案在帧级和宏块级都能自适应地选择拉格朗日乘子。文献[12]建立了基于 SSIM 的感知失真模型 dSSIM,并对失真度量和拉格朗日因子进行了调整,大大提高了编码性能。

本文提出了一种基于视觉感知的率失真算法,通过数据驱动的 JND 预测模型推导拉格朗日乘数因子,以有效降低视频中的感知冗余;并使用显著性模型优化拉格朗日乘子权重系数,以合理实现显著区域和非显著区域编码比特的分配;最后,融合应用于率失真优化,以充分发挥两种模型的优势,全面利用 HVS 特性,并设计和采用显著加权的相似度指标(salience weighted-structural similarity index metric, SW-SSIM)评估视频感知质量。

2 算法描述

率失真优化的目标是在码率 R 受到约束时,获得最小失真 D,公式为:

$$\min\ D \ \text{s.t.}\ R < R_C \tag{1}$$

其中,R_C 是码率约束。通常,这个约束优化问题是通过引入拉格朗日乘子 λ,转换成一个新的目标函数 J,如式(2)所示。

$$\min\{J \mid J = D + \lambda R\} \tag{2}$$

目前,对率失真优化算法中拉格朗日因子 λ 的研究忽略了视觉感知冗余和视觉显著性特征。本文所提算法步骤如图 1 所示。首先,为了在编码宏块(large coding block, LCU)上调整比特率,将视频帧分成 64×64 的 LCU 块;其次,基于像素域 JND 模型推导出拉格朗日乘数调节因子 JND_{LCU},基于显著性模型得到拉格朗日乘数的权重系数 P_i;最后,为弥补 JND 模型忽略了人眼对不同区域的关注度,而显著性模型未考虑人眼可感知的最小阈值的不足,本文将结合这两个模型得到改进的基于感知编码的拉格朗日因子 λ_{ipc},实现视频编码优化。

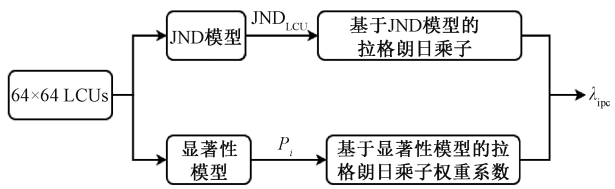


图 1 本文率失真算法流程

Fig. 1 The flowchart of proposed rate-distortion optimization algorithm

2.1 基于 JND 模型的拉格朗日乘子调节因子

在视觉感知系统中,人眼对不同区域可察觉到的失真阈值是不同的,而 JND 模型可以获得人眼的最小视觉阈值,将其应用于视频编码,可以更准确地估计视频中的感知冗余。因此,本文提出了一种基于 JND 模型的 RDO 算法,并通过改进的感知失真度量得到拉格朗日乘子。

在之前的工作[13]中,本文作者提出了一种基于数据驱动的 JND 预测模型,方案如图 2 所示。该方案使用预处理后的 Video Set 视频帧图像块及视频 JND 阈值作为训练集,为提高训练效率,以原始 VGG 网络为基础,去掉了用于分类的 softmax 激活层,改进并训练了 J-VGGNet;最后,使用该网络可预测待编码视频的像素域 JND 阈值,应用于感知视频编码。

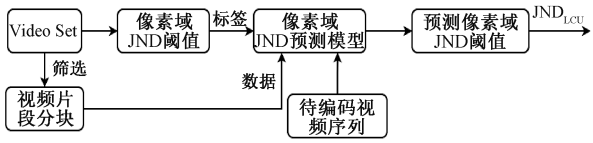


图 2　数据驱动的 JND 预测模型

Fig. 2　The data-driven JND prediction model

本文将上述模型中的 JND_{LCU} 作为客观失真的感知敏感度调节因子,将率失真公式中的客观失真修正为更符合人眼特性的感知失真 D_p, 如式(3)所示。

$$D_p = \frac{SSE}{JND_{LCU} + c} \tag{3}$$

其中,c 为常数,取经验值为 1。由式(3)可知,在客观失真 SSE 相同的情况下,编码块的 JND 阈值越小,感知失真 D_p 就会越大,人眼越容易察觉到失真。

对于每个 LCU 块,使用式(3)所提的感知失真指标 D_p 替换原有失真指标后得到新的率失真方程:

$$J = SSE + (JND_{LCU} + c)\lambda_p R \tag{4}$$

其中,λ_p 是基于感知失真的拉格朗日乘子。当以 D_p 作为率失真公式中的失真指标时,式(4)可以转换为以下公式:

$$\min\ J = \sum_{i=0}^{N} \frac{d_i}{JND_{LCU} + c} + \sum_{i=0}^{N} \lambda_{pi} r_i = \sum_{i=0}^{N} \left(\frac{d_i}{JND_{LCU} + c} + \lambda_{pi} r_i \right) \tag{5}$$

其中,d_i 和 r_i 分别表示第 i 个 LCU 块的失真和速率,N 为 LCU 的块数。

为求解式(5)率失真代价的最优解,将每个编码单元对 r_i 求导,则有:

$$\frac{\partial J}{\partial r_i} = \frac{1}{JND_{LCUi} + c} \cdot \frac{\partial d_i}{\partial r_i} + \lambda_{pi} = 0 \tag{6}$$

根据编码速率 R 与失真 D 之间存在对数关系:

$$R(D) = \alpha \ln\left(\frac{\delta^2}{D}\right) \tag{7}$$

其中,δ^2 表示编码位移帧差,α 为常数。求解式(6)的率失真优化方程,得到最优解为:

$$r_i^* = \alpha \log\left(\frac{\delta_i^{\,2}}{\alpha(JND_{LCUi} + c)\lambda_{pi}}\right) \tag{8}$$

其中,r_i^* 为第 i 个编码单元在率失真最优编码模式下对应的编码比特。因此整个视频图像帧消耗的编码比特可以表示为:

$$R_p = \alpha \sum_{i=0}^{N} \log\left(\frac{\delta_i^{\,2}}{\alpha(JND_{LCUi} + c)\lambda_{pi}}\right) \tag{9}$$

类似的,对以 SSE 为失真指标时的原始率失真公式进行求导,求解最优模式下消耗的编码比特,得到

原始率失真算法中整个视频图像帧消耗的编码比特如式(10)所示。

$$R_{SSE} = \alpha \sum_{i=0}^{N} \log\left(\frac{\delta_i^{\,2}}{\alpha\lambda_{SSEi}}\right) \tag{10}$$

进而可以推导出第 i 个 LCU 的拉格朗日乘子计算公式为:

$$\lambda_{pi} = \frac{JND_{LCUi} + c}{\exp\left[\frac{1}{N}\sum_{j=1}^{N} \log(JND_{LCUj} + c)\right]} \lambda_{SSEi} \tag{11}$$

其中,N 表示当前帧的 LCU 个数,JND_{LCUi} 表示第 i 个 LCU 的 JND 阈值,λ_{SSEi} 表示 AVS3 原始率失真公式的拉格朗日乘子。由上述公式可得到第 i 个 LCU 块基于 JND 的自适应拉格朗日乘子调节因子:

$$\eta_i = \frac{JND_{LCUi} + c}{\exp\left[\frac{1}{N}\sum_{j=1}^{N} \log(JND_{LCUj} + c)\right]} \tag{12}$$

2.2　基于显著性模型的拉格朗日乘子权重系数

显著性检测方法是一种视觉注意力建模方法,本文通过显著性模型推导出拉格朗日乘数的权重系数,实现显著区域和非显著区域编码比特的合理分配。

近年来,基于深度神经网络的显著性检测方法降低了对人工提取特征的依赖,已成为显著性检测的主流方向。本文采用文献[14]中提出的基于端对端神经网络的显著性检测模型,相对于其它基于神经网络的显著性模型来说更轻量,性能表现更佳。模型显著图检测效果如图 3 所示。

图 3　显著图检测效果图

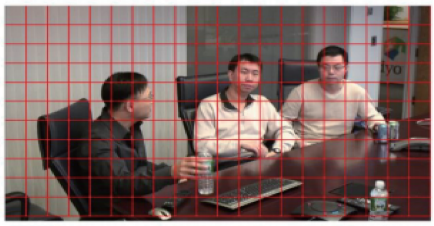
Fig. 3　Saliency map detection

首先,从显著图全局出发,为每个 LCU 块确定显著程度。由图 4 可知,显著性具有区域性特性,存在一部分 LCU 属于显著区域,而另一部分 LCU 则属于非显著区域。

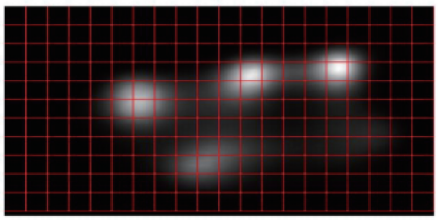
其次,将显著图中每个像素点值对应帧图像中各像素的显著性值,则可将 LCU 的显著程度定义为:

$$s_i = \frac{p_i}{p_{avg}} \tag{13}$$

其中,p_i 为显著性图中第 i 个 LCU 的像素点平均值。p_{avg} 为整帧显著性图的像素点平均值。s_i 为第 i 个 LCU 在整帧中的显著性比重,如式(14)和(15)所示。



(a) 原图块划分
(a) Original block division



(b) 显著图块划分
(b) Salient block division

图 4 块划分

Fig. 4 Block division

$$p_i = \frac{1}{N \times N}\sum_{x=1,y=1}^{N\times N} f(x,y) \tag{14}$$

$$p_{\text{avg}} = \frac{1}{L \times H}\sum_{x=1,y=1}^{L\times H} f(x,y) \tag{15}$$

其中,$f(x,y)$ 为显著图中对应的相对坐标像素值,N 为 LCU 边长,L 和 H 分别表示该图像帧的长度和高度。

本文将 s_i 大于 1 的 LCU 块的集合定义为显著区域,而 s_i 小于 1 的 LCU 块的集合定义为非显著区域。并通过多次实验,根据显著程度分布情况确定对应 LCU 的显著权重系数,如式(16)所示。

$$\omega_i = a + bs_i \tag{16}$$

其中,a 和 b 取经验值 0.5。

对每个 LCU 的客观失真进行基于显著性的调整:

$$D_{si} = \omega_i \times D_{\text{SSE}i} \tag{17}$$

其中,D_{SSE_i} 和 D_{s_i} 分别表示第 i 个 LCU 的客观失真和感知失真。

为平衡全局码率和失真,对率失真公式中的拉格朗日乘子基于显著性的加权:

$$\lambda_{si} = \frac{1}{\omega_i} \times \lambda_{\text{SSE}i} \tag{18}$$

其中,λ_{s_i} 代表第 i 个 LCU 基于显著性加权的拉格朗日乘子,λ_{SSE_i} 表示原始率失真公式中的拉格朗日乘子。

由式(18)可知,位于感兴趣区域的 LCU 块的拉格朗日乘子会增大,编码器会倾向于选择客观失真较小,编码比特多的编码模式;而对于非感兴趣区域的 LCU,编码器会倾向于选择编码比特少的编码模式。

2.3 总体算法

为充分利用人眼视觉系统特征,本文将 2.1 中得到的基于 JND 模型拉格朗日调节因子,和 2.2 中推导的基于显著性模型的拉格朗日乘子权重系数结合,提出了一种基于视觉感知的率失真优化算法:

$$J = D_{\text{SSE}i} + \frac{\eta_i}{\omega_i}\lambda_{\text{SSE}i}R_i \tag{19}$$

其中,η_i 是基于 JND 的拉格朗日乘子调节因子,如式(12)所示,ω_i 是基于显著性的拉格朗日乘子权重系数,如式(16)所示。

此外,为了防止率失真优化过程中出现数值异常波动而影响到编码质量,需要对以上调整系数进行范围限制[15]:

$$\frac{\eta_i}{\omega_i} = \begin{cases} 2.5, & \dfrac{\eta_i}{\omega_i} > 2.5 \\ \dfrac{\eta_i}{\omega_i}, & 0.5 \leqslant \dfrac{\eta_i}{\omega_i} \leqslant 2.5 \\ 0.5, & \dfrac{\eta_i}{\omega_i} < 0.5 \end{cases} \tag{20}$$

3 实验验证及分析

为更好地体现本文算法对率失真性能的优化,本文将所提算法集成到 AVS3 标准编码平台 HPM5.0 中进行实验,并提出了基于显著性加权的 SSIM 算法来评估重建视频的质量。

3.1 基于显著性加权的结构相似度算法

现有的主流视频质量客观评价指标峰值信噪比(peak signal-to-noise ratio, PSNR)未能考虑到人眼视觉特征,故不能准确评价感知视频质量。感知失真指标 SSIM 在一定程度上更符合 HVS 特性,但该指标未将不同图像区域的视觉显著性考虑在内。为了更好地对结合 JND 和显著性改进的感知率失真优化算法进行性能分析,提出基于显著性加权的结构相似度指标 SW-SSIM 来评估视频编码质量。

首先,根据全局显著性图计算 SSIM 每个采样窗口的显著性:

$$p(\bar{x}) = \frac{1}{M \times M}\sum_{x=1,y=1}^{M\times M} p(i,j) \tag{21}$$

其中,$p(i,j)$ 像素坐标,M 为 SSIM 采样窗口的边长,AVS3 中取 11。则每一帧的 SW-SSIM 计算公式如下:

$$\text{SW-SSIM} = \frac{\sum_{i=1}^{N} p(\bar{x})_i \text{SSIM}_i}{\sum_{j=1}^{N} p(\bar{x})_j} \tag{22}$$

其中,N 为当前帧被采样窗口划分的总块数,$p(\bar{x})_i$ 为第 i 个采样窗口块的显著性,$p(\bar{x})_j$ 为第 j 个采样窗口块的显著性,SSIM_i 为第 i 个采样窗口块的 SSIM 评分。

为了验证所提评价指标的性能,本文在 LIVE VQA 数据库[16]上评估了 SW-SSIM 的性能,如表 1 所示。由表 1 可以看出,斯皮尔曼等级相关系数(Spearman rank ordered correlation coefficient, SRCC),皮尔逊相关系数(Pearson′s correlation coefficient, PLCC),肯德尔相关系数(Kendell rank order correlation coefficient, KRCC)均有所提升。其中,KRCC 提升效果最好,提升了 0.002 6,进一步表明所提算法在加入感知质量度量时,与人眼的视觉感知更加一致。

表 1　SW-SSIM 的性能

Table 1　The performance of SW-SSIM

相关性	SSIM	SW-SSIM
PLCC	0.582 0	**0.583 4**
SRCC	0.566 8	**0.568 1**
KRCC	0.405 1	**0.407 7**

3.2　感知率失真优化算法性能

为了验证所提的感知率失真优化算法的有效性,本文将所提 RDO 算法与 AVS3 标准 RDO 算法比较,其中所有的测试序列都是 AVS3 标准测试序列,包括 4 种分辨率共 12 个序列,测试序列如表 2 所示。

表 2　测试序列信息

Table 2　Summarization of test sequences

序列名称	分辨率	帧率/fps	位宽/bit
BasketballDrive	1 920×1 080	50	8
Cactus	1 920×1 080	50	8
City	1 280×720	60	8
Crew	1 280×720	60	8
Vidyo1	1 280×720	60	8
Vidyo3	1 280×720	60	8
BasketballDrill	832×480	50	8
BQMall	832×480	60	8
PartyScene	832×480	50	8
RaceHorses	832×480	30	8
BasketballPass	416×240	50	8
BQSquare	416×240	60	8

表 3 为所提算法基于 SW-SSIM 的感知率失真性能,其中 BD-Rate(SW-SSIM)为负值代表码率下降,BD-SWSSIM 为正值代表失真较少即质量提升。实验结果表明,与 HPM5.0 相比,本文所提算法在保证视频质量的前提下,可以平均节省 12.15%的码率;在相同码率情况下,SW-SSIM 平均提升 0.004 3。

表 3　所提总体算法感知率失真性能

Table 3　The performance of the proposed overall algorithm

序列名称	BD-Rate(SW-SSIM)/%	BD-SWSSIM
BasketballDrive	−20.49	0.007 5
Cactus	−9.31	0.004 6
City	−7.90	0.003 7
Crew	−25.58	0.009 9
Vidyo1	−12.59	0.002 6
Vidyo3	−11.10	0.001 7
BasketballDrill	−20.13	0.005 4
BQMall	−11.33	0.004 3
PartyScene	−11.09	0.004 5
RaceHorses	−10.86	0.005 8
BasketballPass	−6.98	0.003 1
BQSquare	−3.95	0.002 3
Average	**−12.15**	**0.004 3**

为更加直观地表示所提算法的感知率失真性能,本文选取 4 个不同分辨率的测试序列,根据它们的编码码率和 SW-SSIM 值描绘各个序列不同量化参数(quantization parameter, QP)对应的率失真曲线,如图 5 所示,横坐标是编码码率,纵坐标是 SW-SSIM 值,虚线代表 HPM 原始率失真算法,实线代表本文所提的率失真算法。由图 5 可知,基于视觉感知特性改进的率失真优化算法曲线始终在标准算法之上,说明本文所提率失真算法可不同程度地提高视频压缩性能。

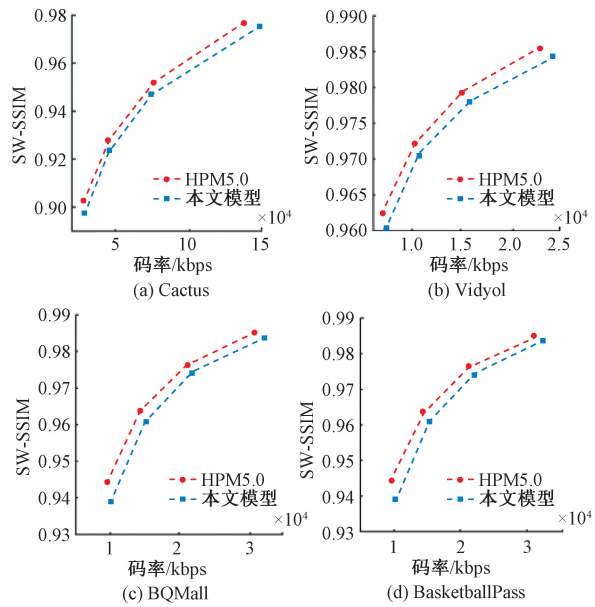
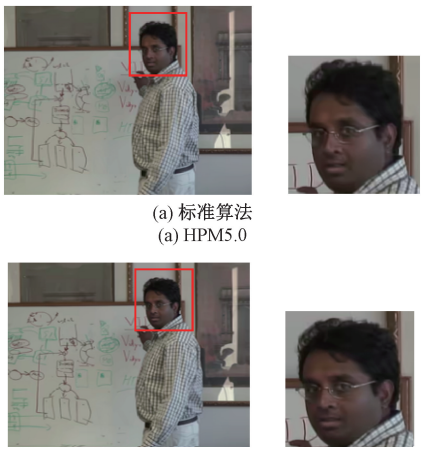


图 5　总体算法与 HPM5.0 在率失真性能上的比较

Fig. 5　The comparison between our proposed overall algorithm and HPM5.0 in rate distortion performance

为进一步验证所提算法在主观感知方面的有效性，本文展示了部分测试序列的解码图像。图 6、7 和 8 分别代表部分序列的解码图像。从图中可以看出，本文算法的解码图像与 HPM 标准算法的解码图像在全局上察觉不到差异，但在人眼容易注意到的局部显著区域，如图 6 局部图的嘴唇部分和眼镜框，图 7 局部图的整个面部五官，图 8 局部图的嘴部区域，所提率失真方案的解码图像都比 AVS3 原率失真方案的解码图像更加清晰，纹理细节更加丰富，一定程度上提升了人眼感知效果。

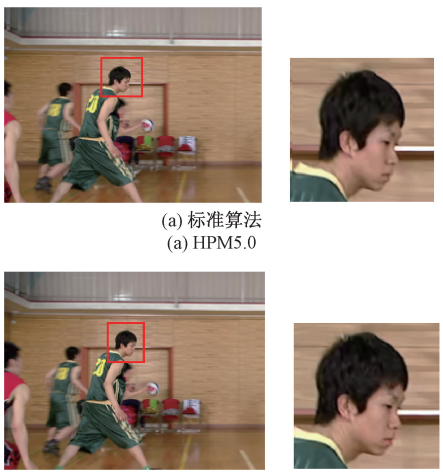


(a) 标准算法
(a) HPM5.0

(b) 本文算法
(b) Proposed algorithm

图 6 Vidyo3 序列主观对比图

Fig. 6 Subjective quality comparison of Vidyo3



(a) 标准算法
(a) HPM5.0

(b) 本文算法
(b) Proposed algorithm

图 7 BasketballDrill 序列主观对比图

Fig. 7 Subjective quality comparison of BasketballDrill

此外，考虑到视频显著区域对人眼主观感知影响较大，图 9 给出了部分测试序列显著区域 Y-PSNR 逐帧对比图，由图可知，显著区域的 Y-PSNR 均有所提升，说明本文所提算法在提高编码效率的同时，可以改善视频质量。

(a) 标准算法
(a) HPM5.0

(b) 本文算法
(b) Proposed algorithm

图 8 Crew 序列主观对比图

Fig. 8 Subjective quality comparison of Crew

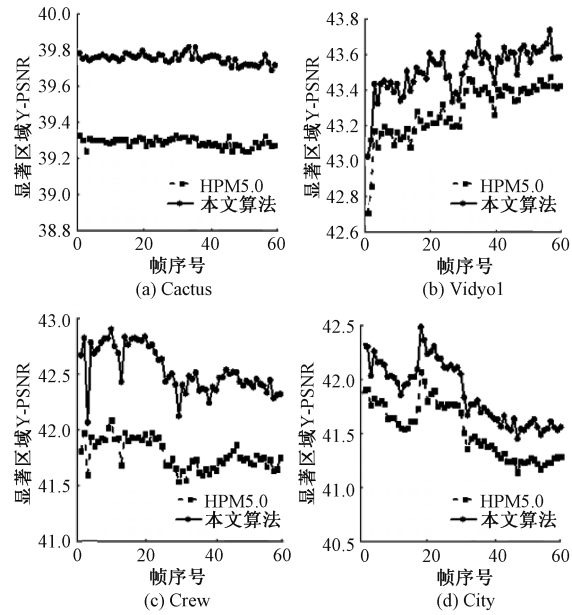


图 9 显著区域 Y-PSNR

Fig. 9 Salient area Y-PSNR

为进一步验证本文算法能消除视觉感知冗余的有效性，表 4 将所提算法与文献[17]中所提基于层级的时域率失真优化和文献[18]中所提基于参考结构决定拉格朗日乘子算法分别进行比较。实验中以 Y-PSNR 为指标，将这三种算法分别与 AVS 标准率失真算法进行比

较。数据显示,本文算法平均可节省 6.13%的码率,与文献[17]所提和文献[18]的平均码率节省相比,本文的实验结果远远优于对比方法。原因是本文算法充分考虑了人眼可见的最小阈值和视觉注意力,能够更好地消除感知冗余,并对码率资源进行了非均等的分配,可以更有效地节省编码码率,实现感知编码优化。

表 4　现有算法与 AVS 标准算法的 BD-Rate 比较

Table 4　BD-rate comparison of the existing algorithms and the AVS standard algorithm　%

分辨率	文献[17]	文献[18]	所提算法
1080p	-1.20	-0.87	**-8.89**
720p	-2.05	-0.86	**-5.50**
WVGA	-0.93	-0.56	**-5.37**
WQVGA	-0.96	-0.59	**-4.75**
Average	-1.23	-0.72	**-6.13**

4　结　　论

本文提出了一种结合 JND 模型和显著性模型的率失真优化算法。该算法通过对人眼视觉系统的研究,充分考虑了人眼可察觉的最小阈值和视觉注意力,对拉格朗日因子进行调整,以消除视频编码中的感知冗余及自适应分配编码比特。实验结果表明,本文所提算法与 HPM5.0 标准 RDO 算法相比,在保证感知视频质量的情况下,平均能节省 12.15% 的码率,SW-SSIM 平均提升 0.004 3。本文的工作将有助于消除视频中感知冗余,优化混合视频编码器,改善视频感知质量,促进感知视频编码方案的研究。

参考文献

[1] 李晓辉, 吴小培. 基于率失真最佳的视频编码码率控制方法[J]. 仪器与仪表学报, 2006, 27(3): 327-330.
LI X H, WU X P. Video coding rate control method based on rate distortion[J]. Chinese Journal of Scientific Instrument, 2006, 27(3): 327-330.

[2] YUAN D, ZHAO T, XU Y, et al. Visual JND: A perceptual measurement in video coding[J]. IEEE Access, 2019, 7(99): 29014-29022.

[3] NACCAR M, PEREIRA F. Decoder side just noticeable distortion model estimation for efficient H.264/AVC based perceptual video coding[C]. In Proceedings of the International Conference on Image Processing (ICIP), 2010: 2573-2576.

[4] KI S, KIM M, KO H. Just noticeable quantization distortion based preprocessing for perceptual video coding[C]. In Proceedings of the Visual Communications and Image Processing (VCIP), 2017.

[5] NAMI S, PAKDAMAN F, HASHEMI M. Juniper: A JND-based perceptual video coding framework to jointly utilize saliency and JND[C]. In Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2020.

[6] 卢笑, 曹意宏, 周炫余, 等. 基于深度强化学习的两阶段显著性目标检测[J]. 电子测量与仪器学报, 2021, 35(6): 34-42.
LU X, CAO Y H, ZHOU X Y, et al. Two-stage saliency object detection based on deep reinforcement learning[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6): 34-42.

[7] ZHU P, XU Z. Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network[J]. Neurocomputing, 2018, 275: 511-522.

[8] ROUIS K, LARAB MI, TAHAR J. Perceptually adaptive lagrangian multiplier for HEVC guided rate-distortion optimization[J]. IEEE Access, 2018, 6: 33589-33603.

[9] SONG R, ZHANG Y. High efficiency video coding intra prediction optimization algorithm based on region of interest[J]. Journal of Electronics and Information Technology, 2020, 42(11): 2781-2787.

[10] ZHOU M, WEI K, WANG S. SSIM-Based global optimization for CTU-Level rate control in HEVC[J]. Transactions on Multimedia, 2019, 21(8): 1921-1933.

[11] WANG S, REHMAN A, WANG Z, et al. SSIM-Motivated rate-distortion optimization for video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(4): 516-529.

[12] YEO C, TAN H, TAN Y. On rate distortion optimization using SSIM[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(7): 1170-1181.

[13] 李兰兰, 刘晓琳, 吴珂欣, 等. 数据驱动的 AVS3 像素域最小可觉差预测模型[J]. 数据采集与处理, 2021, 36(1): 53-62.
LI L L, LIU X L, WU K X, et al. Data-driven AVS3 pixel-wise JND prediction model[J]. Journal of Data Collection and Processing, 2021, 36(1): 53-62.

[14] LIU J, HOU Q, CHENG M, et al. A simple pooling-based design for real-time salient object detection[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 3912-3921.

[15] 王浩. 基于视觉特性的 HEVC 视频编码优化研究[D]. 上海:海交通大学, 2019.
WANG H. Research on optimization of HEVC video coding based on visual characteristics[D]. Shanghai: Shanghai Jiao Tong University, 2019.

[16] SESHADRINATHAN K, SOUNDARARAJAN R, BOVIK A C. Study of subjective and objective quality assessment of video[J]. IEEE Transactions on Image Processing, 2010, 19(6): 1427-1441.

[17] 毛敏. 帧内视频编码及率失真优化算法研究[D]. 四川:电子科技大学,2018.
MAO M. Research on video coding and rate distortion optimization algorithm [D]. Sichuang: University of Electronic Science and Technology of China, 2018.

[18] 王宏宇. 视频编码结构与优化算法研究[D]. 四川:电子科技大学, 2019.
WANG H Y. Research on video coding structure and optimization algorithm [D]. Sichuang: University of Electronic Science and Technology of China, 2019.

作者简介

魏宏安,1999 年于上海同济大学获得学士学位,2007 年于福州大学通信与信息系统获得硕士学位,现为福州大学高级实验师,硕士生导师,主要研究方向为智能视频编码与通信、图像处理与计算机视觉、全媒体融合应用。

E-mail: weihongan@fzu.edu.cn

Wei Hongan received his B. Sc. degree from Shanghai Tongji University in 1999, and received his M. Sc. degree in Communication and Information System from Fuzhou University in 2007. He is currently a senior laboratory technician and master instructor at Fuzhou University. His research interests include intelligent video coding and communication, image processing and computer vision, all-media convergence applications.

刘嘉棋,2020 年于上饶师范学院电子信息科学与技术专业获得学士学位,现为福州大学电子与通信工程硕士研究生,主要研究方向为智能视频编码。

E-mail: 201127045@fzu.edu.cn

Liu Jiaqi received her B. Sc. degree in Electronic Information Science and Technology from Shangrao Normal University in 2020. She is currently a master student in Electronic and Communication Engineering at Fuzhou University. Her research interest is intelligent video coding.

林丽群(通信作者),分别于 2007 年和 2019 年于福州大学通信与信息系统专业获得硕士学位和博士学位,现为福州大学副教授,硕士生导师,主要研究方向为图像/视频信号处理,视觉质量评估、视频编码。

E-mail: lin_liqun@fzu.edu.cn

Lin Liqun (Corresponding author) received her M. Sc. and Ph. D. degrees in Communication and Information System from Fuzhou University in 2007 and 2019, respectively. She is currently an associate professor and a master advisor at Fuzhou University. Her research interests include image/video signal processing, visual quality assessment, and video coding.