

中图法分类号: TP37 文献标志码: A 文章编号: 1006-8961(2012)01-0054-08

论文引用格式: 俞力克, 代锋, 张勇东, 林守勋. 基于纹理和亮度感知特性的率失真优化策略 [J]. 中国图象图形学报, 2012, 17(1): 54-61

基于纹理和亮度感知特性的率失真优化策略

俞力克^{1,2}, 代锋¹, 张勇东¹, 林守勋¹

1. 中国科学院计算技术研究所, 北京 100190; 2. 中国科学院研究生院, 北京 100190

摘要: 率失真优化(RDO)策略在视频编码体系中对编码效率有着重要影响。当前主流的率失真优化策略以MSE或类似的方法描述失真,不能很好反映人眼的主观感受。为了提高视频编码的主观感知质量,首先建立一种新的感知失真模型。该模型利用人眼对纹理和亮度的敏感特性,使评价结果能够更好地反映主观质量。在此基础上,提出一种基于纹理和亮度感知特性的率失真优化策略,简称TL-RDO(texture and luminance based RDO)策略。该策略对不同的区域自适应地调整拉格朗日乘子,使得编码结果更好地符合人眼的观察特性。实验结果表明,TL-RDO相比现有最常用的QP-RDO方法,编码效率显著提高;与一些典型的基于感知失真特性的率失真优化策略相比,TL-RDO策略计算复杂度较低,适合于实时编码系统。

关键词: 感知质量;感知失真模型;视频编码;人类视觉系统;H.264/AVC;率失真优化

Novel rate distortion optimization strategy based on perceptual properties of texture and luminance

Yu Like^{1,2}, Dai Feng¹, Zhang Yongdong¹, Lin Shouxun¹

1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100190, China

Abstract: RDO (rate-distortion optimization) plays an important role in video coding systems and has a great effect on coding efficiency. The most widely used RDO strategy uses MSE or other similar metrics for distortion modeling, which is not a good metric for subjective evaluation. In order to improve the perceptual quality, a novel perceptual distortion model is firstly proposed which takes the perceptual properties of texture and luminance into consideration. Based on the perceptual distortion model, a TL-RDO (texture and luminance based RDO) strategy is proposed which adjusts the Lagrangian multiplier dynamically according to visual perception. The simulation result shows that TL-RDO gets higher coding efficiency than the famous QP-RDO. Moreover, it has low computational consumption compared to other perceptual RDO strategies, which is suitable for real-time systems.

Key words: perceptual quality; perceptual distortion model; video coding; HVS; H.264/AVC; RDO

0 引言

视频编码标准发展至今,编码性能取得了长足

的进步,最新的编码标准 H.264 及 AVS 取得了空前的编码性能。同时,视频编码中所采用的编码技术种类也越来越多。最新的 H.264/AVC 视频标准允许一个 16×16 像素的宏块可选择 7 种不同的分块

收稿日期: 2011-03-14; 修回日期: 2011-04-19

基金项目: 国家重点基础研究发展计划(973)项目(2007CB311100); 国家自然科学基金项目(60802028); 北京市科技新星计划项目(2007B071)

第一作者简介: 俞力克(1983—),男,2011年于中国科学院计算技术研究所获计算机应用技术专业博士学位,主要研究方向为视频编码及处理。E-mail: yulike@ict.ac.cn

编码模式; 在帧间参考中, 最多可选择 16 个参考帧。因此, 在如此多种可选编码方式集合中, 如何完成最合理的模式类型组合, 便是率失真优化 (rate-distortion optimization) 技术所要解决的重要问题。

率失真优化策略在现有的混合编码框架中, 负责权衡码率 (R) 和失真 (D) 之间的选择关系, 对编码器的最终编码效率有着至关重要的影响。率失真优化的目标是在整体目标码率的限制要求下, 取得最小的失真编码结果, 用表达式可以写为

$$\min\left\{\sum_{i=1}^N D_i\right\} \quad \text{s. t.} \quad \sum_{i=1}^N R_i \leq R_{\text{total}} \quad (1)$$

式中, N 表示视频编码单元的数量, D_i 和 R_i 分别表示第 i 个编码单元的失真与码率消耗, R_{total} 表示总体的码率限制要求。当前主流的率失真优化方法是利用拉格朗日优化方法将式 (1) 中的非约束求极值问题转化为约束性求极值问题, 可以表达为

$$\min\{J = D + \lambda R\} \quad (2)$$

式中, λ 是拉格朗日乘子, D 和 R 分别指图像的失真和码率。通过计算不同编码参数下代价函数 J 的值, 选取最小值所对应的模式, 便可得到最优解。

假设率失真函数 $D(R)$ 处处可导, 可以通过对代价函数 J 对 R 求导的方式获得 J 的极值点, 即

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0 \quad (3)$$

从而得到

$$\lambda = -\frac{dD}{dR} \quad (4)$$

可见 λ 的最优值是率失真函数的负斜率, 即 λ 的值与图像本身的率失真特性相关。但图像本身的率失真特性只有在编码之后才能统计获得; 而为了使编码效率达到最优, 需要在编码之前就得到图像的率失真曲线。这便是率失真优化领域中一个经典的蛋鸡悖论问题。

文献 [1] 利用了 R-D 建模, 提出一种基于量化参数的率失真优化方法 (简称 QP-RDO)。该方法在编码领域得到了广泛的认可, 被很多经典编码器所采纳使用。该文献作者也同时指出, 代价函数中的拉格朗日乘子 λ 的取值, 对率失真优化的效率有着非常重要影响。

尽管 QP-RDO 方法对视频编码领域的技术发展有着深远的影响, 但同时也应该看到, 该方法也存在不足之处。从式 (4) 中可以看到拉格朗日乘子 λ 的取值跟失真模型 D 有关。当前绝大多数率失真优化方法 (包括 QP-RDO 在内) 在对失真建模时往往

采用了传统的 MSE (mean square error, 均方误差) 或者类似的 SSD (sum of square distortion, 平方误差和)、MAD (mean absolute distortion, 均值误差) 方式。这类方法虽然计算简单且数学意义明确, 但是它们不能很好反映主观视觉质量 [2]。因此, 以此为基础的率失真优化结果, 必然也无法在主观感知层面上达到最优。为了解决这个问题, 也有学者试图用基于人眼感知的质量评价模型代替 MSE 进行率失真优化研究 [3-4]。但总体来说, 目前这个领域的研究工作非常少, 原因主要是因为现有的基于人眼感知评价模型 [5-8] 大多计算比较复杂, 不能很好地适应视频编码的计算实时性要求。

基于以上对目前率失真优化研究现状的分析, 本文从失真模型优化角度入手, 利用更准确的感知失真建模达到优化 RDO 策略的目的。人眼视觉有两项基本特性: 一是人眼对纹理复杂区域的失真敏感度较低, 对纹理平坦区域的失真敏感度较高; 二是人眼对失真的敏感度受到背景亮度的刺激影响。根据这两项来自人类视觉系统 (HVS) 领域的研究结论, 我们利用这两个感知特性, 首先建立一种新的感知失真模型, 新模型由于加入了人眼感知特性, 因此能够更好地反映主观视觉感受; 然后利用新的感知模型代替传统的 MSE 失真模型用于率失真优化的推导, 得到一种基于感知特性的率失真优化策略, 简称 TL-RDO (texture and luminance based RDO) 策略。新算法能够对图像不同内容特性的区域采取不同的率失真选择策略, 通过自适应调整拉格朗日乘子, 达到合理分配编码资源, 提高整体视觉质量的目的。与当前主流的率失真优化方法相比, 新方法利用了人眼的感知特性, 能够明显改善视频的主观感知质量, 提高编码效率。同时, 该算法与其他基于感知失真模型的率失真优化方法相比, 计算复杂度较低, 能够更方便地用于实时编码系统。

1 基于人眼观察特性的失真评价方法

研究结果表明, 人眼对于不同场景, 不同情况下的图像失真, 其敏感度是不同的。比如, 纹理复杂的区域中, 人眼不易察觉到失真; 而在平坦的图像内容中, 失真就相对容易被感知 [9]。再比如在背景亮度比较高的区域, 人眼对图像失真的敏感度相对较低 [6]。传统率失真优化方法由于缺乏基于感知特性的失真评价手段, 所以在建模推导过程中缺乏对

感知特性的考虑,也就不能达到感知层面的最优化设计。因此,本文的工作分为两部分:1)是建立基于感知特性的图像失真模型;2)是结合新模型,推导新的率失真优化策略。

1.1 图像内容带来的失真掩蔽效应

1.1.1 图像纹理的掩蔽效应

现有的研究表明,复杂的纹理对图像失真有着更好的掩蔽效应^[9]。也就是说,在纹理复杂的区域,人眼对于失真的敏感度较低;反之,则敏感度较高,失真更容易被察觉。文献[10]给出了图像MSE失真、图像纹理和人眼主观失真三者之间的关系模型。可以表达为

$$D_{\text{perceptual}} = k D_{\text{MSE}} \quad (5)$$

式中, $D_{\text{perceptual}}$ 指图像质量, D_{MSE} 为图像MSE失真, k 是与图像纹理复杂度相关的函数,即

$$k = 0.03585 \exp(-0.02439 G_{\text{EdgeStrength}}) \quad (6)$$

$G_{\text{EdgeStrength}}$ 是利用Sobel算子计算得到的图像纹理复杂度。该文献给出了两点重要启示:1)在使用相同的编码策略条件下,人眼感知失真与图像本身的MSE失真呈线性关系;2)利用人眼感知特性分析,可以对式(1)中的加权系数 k 进行准确的预估。实验表明,这样的加权预估方法有效性非常高。

1.1.2 环境对比度对观察者的影响

人类对于观察对象所发生变化的敏感程度与对象本身变化的幅度有关。举个例子,一斤重和两斤重的两个铁球,人们很容易将它们区分开。而50斤重和51斤重的两个铁球,区分它们就相对比较困难。这就是19世纪Weber提出的Weber原则,可以用表达式写为

$$\Delta S = KS \quad (7)$$

式中, ΔS 表示能感知到的变化程度, S 是观察对象所在环境的背景信号幅值, K 是观察对象发生变化的度量参数。和1.1.1节中的建模方式非常类似的是,Weber认为,感知变化程度与背景刺激呈线性关系,同样可以用加权形式来表达。

1.2 利用掩蔽效应对感知失真与统计失真的关系建模

通过1.1节从两个不同角度对人眼感知特性的描述,将这两点人眼感知因素融合进行考虑,根据它们可加权表达的特性,提出一种新的感知失真模型,新模型同时考虑了纹理复杂度和图像背景亮度对人眼感知的影响,可写为

$$D_{\text{perceptual}} = k_{\text{Texture}} k_{\text{Luminance}} D_{\text{MSE}} \quad (8)$$

也可写为简单的加权形式

$$D_{\text{perceptual}} = \varepsilon D_{\text{MSE}} \quad (9)$$

把 ε 称为视觉掩蔽效应参数,该参数由代表亮度掩蔽效应参数的 $k_{\text{Luminance}}$ 和代表纹理掩蔽效应参数的 k_{Texture} 的乘积得到。新模型利用了基于纹理的感知特性参数和基于亮度的感知特性参数进行加权,使结果能够更好地反映观察特性,符合人眼的实际感受。

1.2.1 亮度权重系数

图1是通过人眼观测实验数据采集获得的亮度敏感度曲线示意图。横坐标表示背景亮度(8位灰度值范围 $0 \leq L_i < 255$)。纵坐标表示人眼的感知门限值,该值越小,代表人眼的敏感度越高。从图1中可以看到,人眼对亮度的敏感度特征函数类似带通滤波器的形态:在亮度较高或较低的情况下,敏感度相对较低;而在亮度值居中时,人眼对变化(或失真)的敏感度最高。

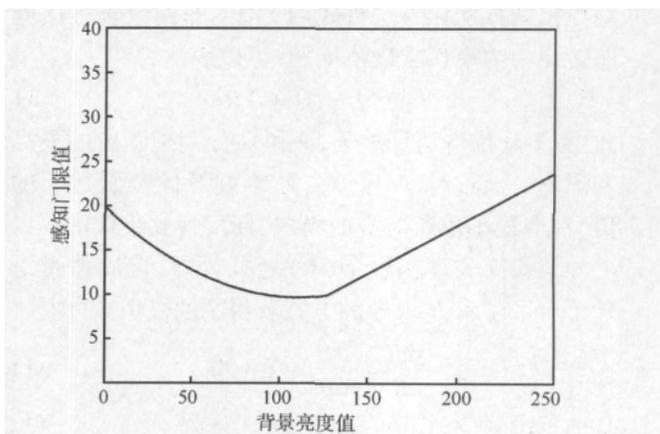


图1 视觉亮度敏感度函数

Fig. 1 Visual sensitivity function of luminance stimulus

用分段函数对该特征曲线进行近似表达

$$T_i = \begin{cases} 20 - 0.133 \times L_i & L_i < 75 \\ 0.108 \times L_i - 3.462 & L_i > 125 \\ 10 & \text{其他} \end{cases} \quad (10)$$

式中, L_i 表示当前像素 i 周围 5×5 像素范围内的平均亮度值。 T_i 表示敏感度门限值。

取敏感度门限值 T_i 的倒数设为敏感度系数,并将人眼在亮度为 $75 \leq L_i < 125$ 的区间内(即敏感度最高的区间内)的敏感度系数置为1,那么就可以推算出其他不同亮度下的敏感度系数

$$k_{\text{Luminance}} = \begin{cases} \frac{1}{2 - 0.0133 \times L_i} & L_i < 75 \\ \frac{1}{0.0108 \times L_i - 0.3462} & L_i > 125 \\ 1 & \text{其他} \end{cases} \quad (11)$$

1.2.2 纹理权重系数

文献[10]给出了一种纹理权重系数的获取方法。通过参考此方法,结合亮度敏感系数,对纹理复杂度系数进行拟合,得到适用于表达式(9)的纹理复杂度系数形式。具体方法如下:

1) 选择 6 个标准测试序列,使用 H.264 官方参考编码器,分别将量化参数固定设置为 12,18,24,30,36,42,48 进行编码。我们招募了 12 名测试人员,对编码结果遵循 ITU-BT.500 中制定的规范进行主观打分。

2) 根据式(9),将各个序列的主观打分结果 DMOS(difference mean opinion score)与 MSE 值对应排列,用线性拟合方法得到各个序列的视觉掩蔽参数 ε (如图 2 所示)。表 1 给出了各序列的拟合结果。

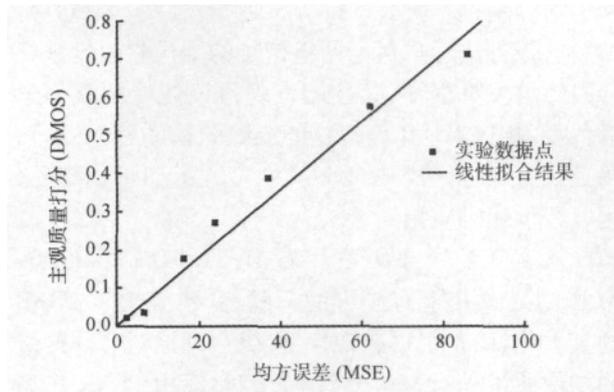


图 2 DMOS 与 MSE 之间的线性拟合
Fig. 2 Linear fitting of DMOS and MSE

3) 将各个序列的 ε 除以相应的 $k_{Luminance}$ (可由式(11)计算得到,见表 1),得到相应的 $k_{Texture}$ 。

表 1 ε 拟合结果及 $k_{Texture}$ 的计算结果

序列	ε 拟合值	$k_{Texture}$ 的估计值
Foreman	0.016 94	0.028 686
Bus	0.006 84	0.007 312
Coastguard	0.010 55	0.010 55
Flower	0.005 15	0.010 834
Football	0.018 18	0.018 18
Mother&Daughter	0.032 26	0.032 26

4) 将得到的 $k_{Texture}$ 与各个序列计算得到的纹理复杂度值对应排列,通过对数拟合得到最终的表达式(如图 3 所示)。最终得到纹理掩蔽效应参数的计算方式为

$$k = 0.0573 \times \exp(-0.01434 \times G_{EdgeStrength}) \quad (12)$$

式中, $G_{EdgeStrength}$ 由 Sobel 算子计算获得。

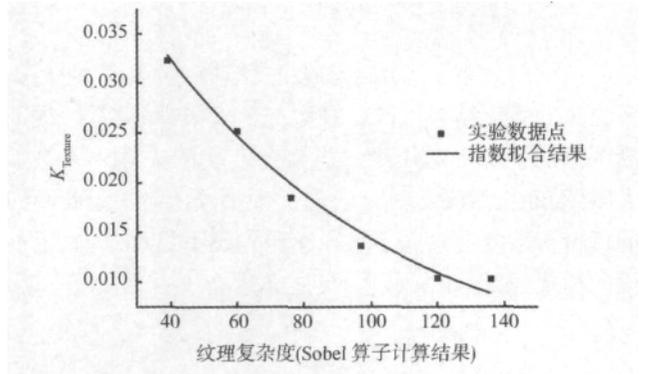


图 3 $k_{Texture}$ 与纹理复杂度之间的指数拟合

Fig. 3 Exponential fitting of $k_{Texture}$ and texture complexity

2 基于新失真模型的 TL-RDO 策略

从式(1)中可以看到,率失真优化的本质是在同样的整体带宽限制下尽可能提高视频的质量。目前往往采用图像所有单元的失真总和来描述整体图像失真。从人眼的观察特性来看,在追求所有单元失真总和尽可能小的同时,相邻单元或区域的图像视觉质量也应该尽可能接近,这样能更好地提高主观视觉感受,因此视频编码的一条基本规则便是保持帧内和帧间的编码重建质量相近^[11]。以式(9)中的感知失真评价模型作为评价指标,我们希望每个单元的感知质量相同,用表达式可以写为

$$\varepsilon_1 D_{MSE}^1 = \varepsilon_2 D_{MSE}^2 = \dots = \varepsilon_i D_{MSE}^i = \varepsilon_{all} D_{MSE}^{all} \quad (13)$$

式中, ε_i 表示第 i 个宏块的掩蔽效应参数。 ε_{all} 是整帧图像的掩蔽效应参数,可以用每个宏块的系数取平均获得。 D_{MSE}^i 表示第 i 个宏块的 MSE 失真。这样,第 i 个宏块的拉格朗日系数可以写为

$$\lambda_i = - \frac{dD_{MSE}^i}{dR} = - \frac{dD_{MSE}^{all}}{dR} \times \frac{\varepsilon_{avg}}{\varepsilon_i} \quad (14)$$

从这个表达式可以看出,新的拉格朗日乘子可以写成 $\varepsilon_{avg}/\varepsilon_i$ 与 $-dD/dR$ 乘积的形式。而后者,正是一般意义上需要求解的拉格朗日乘子。因此,我们对于这部分的取值,可以直接利用已有的率失真优化策略选值方案。这里采用文献[12]提供的用于 H.264 标准的参考值 $0.85 \times 2^{(QP-12)/3}$,该值也是目

前大多数 H. 264 编码器中所采用的经验值。于是, 便得到用于 H. 264 标准的拉格朗日乘子自适应取值方法, 写为

$$\lambda_i = \frac{\varepsilon_{avg}}{\varepsilon_i} \times 0.85 \times 2^{(QP-12)/3} \quad (15)$$

新的拉格朗日乘子考虑了图像的纹理掩蔽效应, 使得率失真的优化结果可以更好地符合人眼观察特性, 提高图像的主观视觉质量。当某个宏块的失真不易被人眼感知时, 意味着 $\varepsilon_i < \varepsilon_{avg}$, λ_i 比原始值大, 因此新的代价函数便会趋向于选择比原先码率更小, 失真更大的模式(如图 4 所示)。反之亦成立。

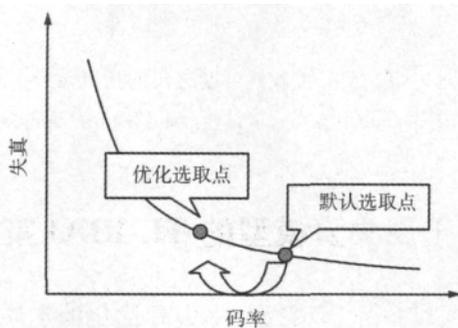


图 4 率失真优化改进策略

Fig. 4 Optimization for RDO Strategy

把式(15)中的自适应拉格朗日乘子选择方法称为 TL-RDO 策略。这种新的率失真优化策略通过人眼的感知特性对图像不同区域采取自适应调整。在人眼对失真敏感度低的区域, 代价函数会倾向于选择码率消耗较小, 失真稍大的模式。相反则选择码率消耗偏大, 但失真较小的模式。如此, 低敏感度区域图像的码率被节约下来用于提高高敏感度区域图像的质量。从人眼观察的角度来看, 整体的视觉质量会更加平衡, 主观感受得到提高。

3 实验结果与分析

选用 H. 264 的官方参考代码 JM 作为实验平台, 将 TL-RDO 和当前主流的方法进行对比。实验环境与基本设置如下: JM 软件版本选用 15.1 版, 编码器采用帧级码率控制算法, 使用 Baseline 编码等级, 测试对象为 6 个标准序列, 分辨率为 CIF (352 × 288), 每个序列编码 90 帧, 序列结构为 IPPPP。

实验主要分为两个部分, 首先讨论与 TL-RDO 相比 JM 中使用的 QP-RDO 的性能提升; 然后, 将 TL-RDO 与另一个基于感知特性的率失真优化策略

进行对比, 讨论新算法在性能、计算复杂度方面的表现。

3.1 与 QP-RDO 策略性能对比

首先采用客观质量评价方法对 TL-RDO 的编码效率进行测试。除了最常用的 PSNR 外, 还加入了 SSIM 评价方法^[8]作为测试结果。SSIM 是一种基于人眼视觉特性的图像质量评价方法, 其计算方式为

$$SSIM(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \times \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \times \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (16)$$

式中: x 和 y 分别代表原始图像和失真图像信号。 μ_x 和 μ_y 分别是 x 和 y 的平均值; σ_x^2 和 σ_y^2 分别是 x 和 y 的方差; σ_{xy} 表示 x 和 y 的协方差。 C_1, C_2, C_3 是常数。该方法通过对比两组信号的结构相似度来评价图像的失真程度, 其性能表现被该领域的众多学者认可, 相比 PSNR 而言更能反映图像的感知质量, 并且该方法也已经被 H. 264 的官方组织正式接受, 用以评价编码质量。

从表 2 的结果中可以看出, TL-RDO 相比 QP-RDO 在 PSNR 上有少许的下降, 并不是追求 PSNR 的最大化, 而是根据感知特性, 更合理地对不同区域的图像质量进行率失真选择, 因此仅用 PSNR 来衡量最后的结果是不完全准确的。相比之下, 当采用另一个质量评价方式 SSIM 时, 结果出现了明显的改观, 在所有的测试序列上, 新算法都取得了明显的改进, 最大可以达到 7% 以上。

为了进一步证明 TL-RDO 对图像视觉质量改善的有效性, 图 5 给出了实际编码后图像的对比。图 5(a) 是关于 Flower 序列的第 90 帧图像对比, 可以看到本文算法虽然降低了花丛部分的编码质量, 但从实际观感来说, 由于人眼对该区域的失真敏感度较低, 几乎感受不到质量下降。因此节省下的码率被用于增强其他区域的图像质量, 例如房屋以及树木部分。相比之下, 右图中 QP-RDO 的编码结果在一些视觉敏感区域, 质量下降非常明显, 使得整体的感知质量不佳。图 5(b) 是 Bus 序列的第 105 帧图像。类似的, 在一些人眼敏感度高的区域(如树林部分), QP-RDO 算法能够尽可能保证编码质量, 而对于公交车车身的图案, 轿车车身与围栏等感知敏感度较低的区域, 虽然编码失真较大, 但对视觉感知影响较小。

表 2 综合测试结果对比
Table 2 Comparison of coding results

序列	码率/(kbit/s)	SSIM		提高/%	PSNR		提高/%
		TL-RDO	QP-RDO		TL-RDO	QP-RDO	
Flower	200	0.633 1	0.608 7	4.01	24.66	24.68	-0.08
	300	0.671 3	0.645 4	4.01	26.41	26.42	-0.04
Coastguard	200	0.724 7	0.713 6	1.56	28.22	28.17	0.18
	300	0.784 1	0.772 5	1.50	29.37	29.3	0.24
Bus	200	0.749 8	0.702 9	6.67	26.59	26.66	-0.26
	300	0.829 8	0.777 4	6.74	28.46	28.44	0.07
Foreman	200	0.830 1	0.811 5	2.29	34.07	34.34	-0.79
	300	0.867 7	0.849 4	2.15	35.76	36.05	-0.80
Container	200	0.768 9	0.712 1	7.98	35.57	35.97	-1.11
	300	0.815 8	0.758 3	7.58	36.76	37.30	-1.45
Stefan	200	0.822 4	0.802 4	2.49	27.25	27.32	-0.26
	300	0.861 5	0.836 5	2.99	29.47	29.51	-0.14

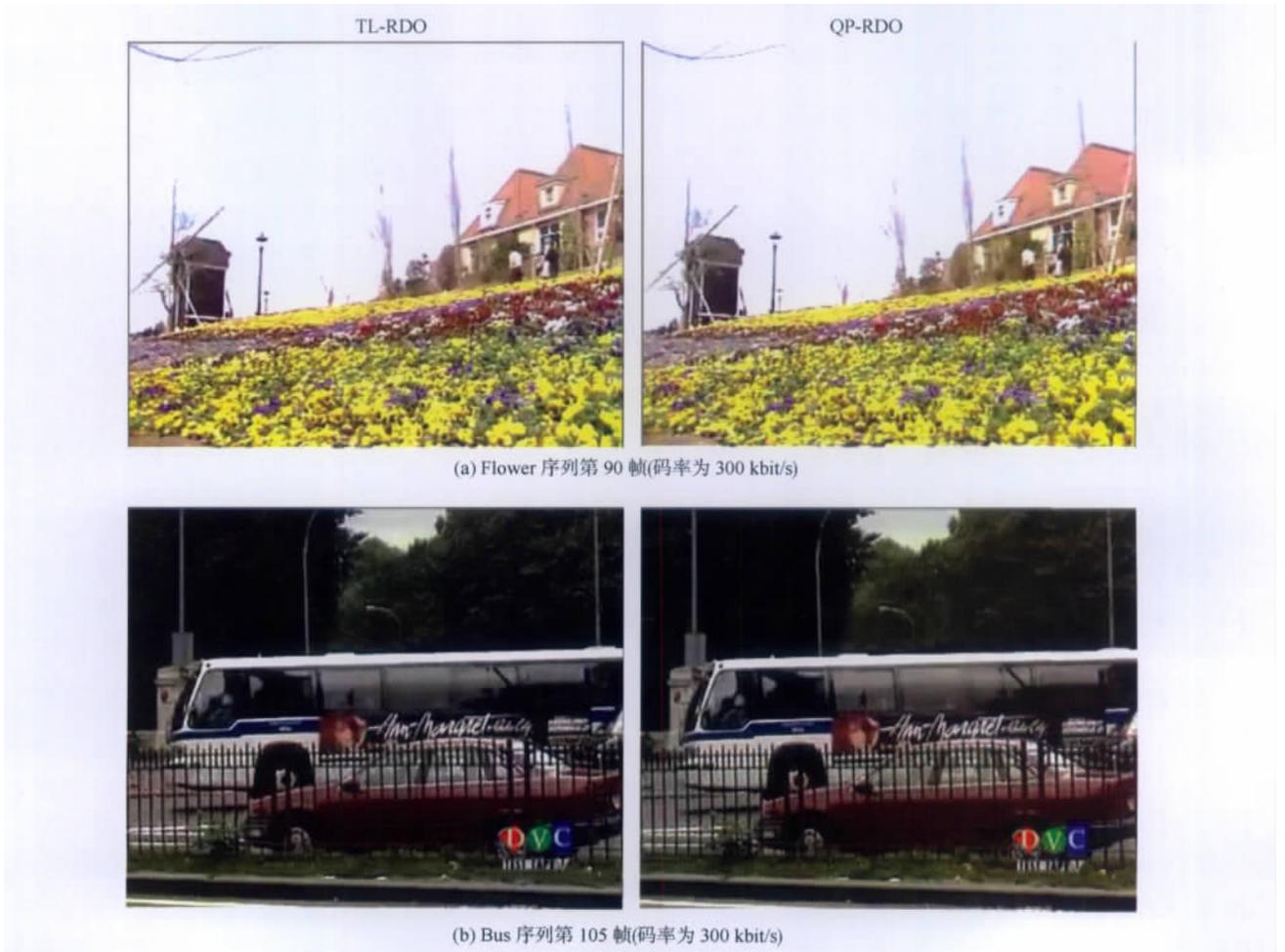


图 5 解码图像主观质量对比

Fig. 5 Subjective comparison of decoded frames

3.2 TL-RDO 与 SSIM-RDO 策略性能对比

TL-RDO 建立在一种基于人眼感知特性的失真模型的基础上。因此,将它和另外一种基于感知失真评价模型发展而来的率失真优化算法进行比较。这里选择了文献[3]提出的 SSIM-RDO 算法作为比较对象。前文中我们就提到 SSIM 方法在图像质量评价领域获得了较高的认可,因此有学者提出基于 SSIM 失真模型的率失真优化方法。文献[3]提出的方法比较具有典型代表性,从该方法的工作结果来

看,相对传统的 QP-RDO 方法也有了明显的进步。所以我们将它作为参照,进行对比测试。

图 6 是两种算法的编码效率对比,由于两种方法都强调主观感知质量,因此仅采用 SSIM 来评价性能。可以看到,我们的算法在码率比较高的情况下和 SSIM-RDO 算法互有优劣,性能差别不大。在码率比较低的情况下,TL-RDO 算法甚至略有胜出。总体而言,两种算法都以感知质量改进为目标,相比原先的 QP-RDO,性能有明显的提高。

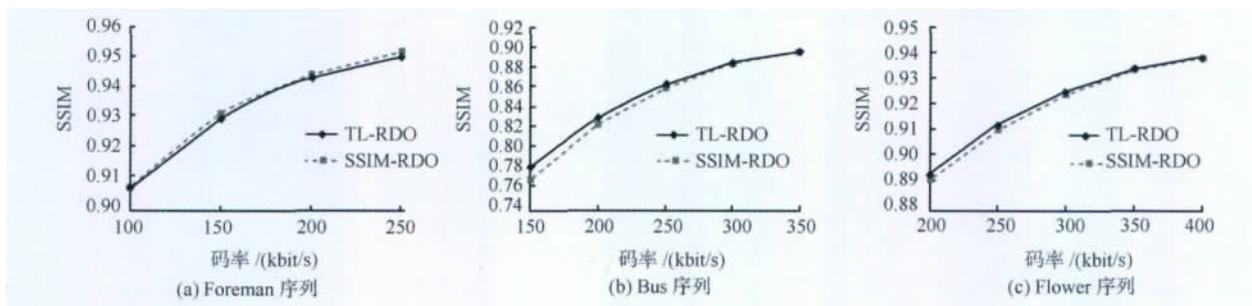


图 6 编码效率对比

Fig. 6 Comparison of coding efficiency

除了编码效率,由于视频编码常常有实时性需求,因此,视频编码算法的计算复杂度也是大家非常关心的问题。表 3 列出了 QP-RDO、TL-RDO 和 SSIM-RDO 这 3 者的计算量分析。这 3 种方法都基于拉格朗日的代价函数式(2)计算,不同的是其中 D 和 λ 的计算方式略有差异。以 QP-RDO 为例,在代价函数中,以 MSE 的计算为主, λ 的值一般通过查表获得。本文提出的 TL-RDO 相比 QP-RDO 方法,计算量增加主要集中在纹理复杂度权重系数 k_i 的计算。该权重系数在每一帧编码前只需要以宏块为单位计算一次,计算主要以 Sobel 算子计算为主。Sobel 算子可以由简单的移位和加法构成,计算量比较小。因此 TL-RDO 算法相对于 QP-RDO 来说,计算量增加并不多。

相比较之下,SSIM-RDO 方法的计算复杂度增加比较多。原因是该方法需要在每次计算代价函数时进行 SSIM 运算。从式(16)可以看出,相比 MSE 计算,SSIM 需要更多的乘法运算。以一个 16×16 的宏块为例,计算一次 MSE 大约需要 512 次加法和 256 次乘法;而计算一次 SSIM 大约需要 1 280 次加法和 768 次乘法,其计算量是 MSE 计算的 3 倍左右。此外,SSIM-RDO 的 λ 值获取也需要进行预编码估算^[3],因而整体来看,SSIM-RDO 相对于

QP-RDO,计算量增加达到 3 倍以上。

表 3 不同算法计算复杂度对比

Table 3 Comparison of different algorithms about computational complexity

使用算法	计算方式	与 QP-RDO 相比
QP-RDO	失真计算: MSE	
	拉格朗日乘子 λ 计算: $0.85 \times 2^{(QP-12)/3}$	
TL-RDO	失真计算: MSE	每一帧计算一次 k_i , 计算量增加较小
	拉格朗日乘子 λ 计算: $(k_{avg}/k_i) \times 0.85 \times 2^{(QP+2)/3}$	
SSIM-RDO	失真计算: SSIM	每一次都需要计算一次 SSIM, 且 λ 值的获取也需要进行预编码统计, 计算量增加幅度大
	拉格朗日乘子 λ 计算: 通过两次预编码获得	

为了具体测试计算量增加带来的实际影响,对 CIF 分辨率序列 Foreman 进行了编码速度测试。我们的计算平台采用 Intel Q9550 CPU,内存 3 GB, Windows XP 操作系统。测试结果表明,相对于 QP-RDO,TL-RDO 每帧计算需要多花费约 4 ms 时间,SSIM-RDO 则需要 28 ms。在一些对于实时性要求很高的场合,每帧的编码时间不能超过 30 ms。如果编码算法计算复杂度太高,势必会严

重影响算法的实用性。从这点来说,TL-RDO 策略计算复杂度较低,完全可以使用于实时编码系统中。

4 结 论

利用人眼对图像内容失真敏感度特性,建立起感知失真与传统数值统计失真之间的关系模型。利用新的感知失真模型,对视频编码器中的率失真优化策略进行推导改进,使得新策略能够有针对性地对不同特性的图像区域采取自适应调整,最终达到优化编码效率,提高视觉质量的目的。实验结果表明,本算法能够很好地用于 H.264 标准,与当前主流的QP-RDO方法相比,新算法能够显著改善视频图像的主观视觉质量,提高编码效率。与主流的基于感知失真模型的SSIM-RDO方法相比,计算复杂度较低,更适合用于实时编码系统。

参考文献(References)

- [1] Wiegand T ,Girod B. Lagrange multiplier selection in hybrid video coder control [C] //Proceedings of IEEE International Conference on Image Processing. New Jersey ,USA: IEEE Press , 2001: 542-545.
- [2] Girod B. What ' s Wrong With Mean-Squared Error [M] //Digital Images and Human Vision. Cambridge ,MA: MIT Press ,1993: 207-220.
- [3] Huang Y H ,Ou T S ,Su P Y ,et al. Perceptual rate-distortion optimization using structural similarity index as quality metric [J]. IEEE Transactions on Circuits and Systems for Video Technology 2010 20(11) : 1614-1624.
- [4] Yang C L ,Leung R K ,Po L M ,et al. An SSIM-optimal H.264/AVC inter frame encoder [C] //Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems. New Jersey USA: IEEE Press 2009: 291-295.
- [5] Winkler S. A perceptual distortion metric for digital color video [EB/OL]. [2011-4-11]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.3079&rep=rep1&type=pdf>.
- [6] Lubin J ,Fibush D. Sarnoff JND Vision Model [R]. New Jersey: TI Standards Committee ,1997.
- [7] Watson A B ,Hu J ,McGowan III J F. Digital video quality metric based on human vision [J]. Journal of Electronic Imaging 2001 , 10(1) : 20-29.
- [8] Wang Z ,Bovik A C. A universal image quality index [J]. IEEE Signal Processing Letters 2002 9(3) : 81-84.
- [9] Ong E P ,Lin W ,Lu Z K ,et al. Perceptual quality metric for H.264 low bit rate videos [C] // Proceedings of IEEE International Conference on Multimedia and Expo. New Jersey , USA: IEEE Press 2006: 677-680.
- [10] Bhat A ,Richardson I ,Kannangara S. A new perceptual quality metric for compressed video [C] //Proceedings of IEEE International Conference on Acoustics , Speech and Signal Processing. New Jersey ,USA: IEEE Press 2009 933-936.
- [11] Jayant N ,Noll P. Digital Coding of Waveforms [M]. New Jersey: Englewood Cliffs ,1994.
- [12] Wiegand T ,Schwarz H ,Joch A ,et al. Rate-constrained coder control and comparison of video coding standards [J]. IEEE Transactions on Circuits and Systems for Video Technology 2003 , 13(7) : 688-703.