

# 面向视频压缩的显著性协同检测 JND 模型<sup>①</sup>

李承欣<sup>1</sup>, 叶 锋<sup>1,2,3</sup>, 涂 钦<sup>3</sup>, 陈家祯<sup>1</sup>, 许 力<sup>1</sup>

<sup>1</sup>(福建师范大学 数学与计算机科学学院, 福州 350007)

<sup>2</sup>(福建星网锐捷通讯股份有限公司 通讯产品研究院, 福州 350002)

<sup>3</sup>(北京邮电大学 多媒体中心, 北京 100876)

**摘要:** 为了更好的将人眼感知特性用于视频压缩系统, 提出了一种改进的基于显著性协同检测的恰可察觉失真模型 (Just Noticeable Distortion, JND). 该模型通过像素域和变换域下联合建模计算得到的最优 JND 模型, 基于上下文感知的显著性算法得到相应的显著图, 并将检测结果用于 JND 模型权重分配. 提出的 JND 残差滤波器可以嵌入到 HEVC 视频编码框架中. 实验结果表明: 在全 I 帧配置下, 提出的算法编码结果与 HM16 相比, 在视觉主观感知质量一致的情况下, 平均码率可节省 10.7%.

**关键词:** 恰可察觉失真; 显著性检测; 视频编码; HEVC

## Improved Just Noticeable Distortion Model Based on Saliency Detection for Video Coding

LI Cheng-Xin<sup>1</sup>, YE Feng<sup>1,2,3</sup>, TU Qin<sup>2</sup>, CHEN Jia-Zhen<sup>1</sup>, XU Li<sup>1</sup>

<sup>1</sup>(School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China)

<sup>2</sup>(Fujian STAR-NET Communications Co. Ltd, Fuzhou 350007, China)

<sup>3</sup>(Multimedia Center, Beijing University of Posts and Telecommunication, Beijing 100876, China)

**Abstract:** A improved just noticeable distortion (JND) model based on saliency map detection algorithm is proposed for video coding in order to apply feature of visual perception to the video compression system. The optimal JND model is calculated by setting modle of pixel-based JND with the hybrid. The saliency map concluded by context aware saliency detection modle is used for the weight distribution of JND model. The proposed model and residual filter can be integrated into the framework of HEVC, which is useful for quantifying video data. The experimental results shows that in the case of visual subjective perceptual quality, the average bitrate reduction is 10.71%, compared with HM 16 (all Intra profile).

**Key words:** just noticeable distortion (JND); saliency detection; video coding; HEVC

## 1 引言

常见的混合视频编码技术框架(如 H.264/AVC、H.265/HEVC)主要针对空间域冗余、时间域冗余以及统计冗余进行压缩编码. 它们主要通过大幅增加计算复杂度来提高视频压缩性能, 其发展空间有限. 近年来, 研究人员将人类视觉系统(Human Visual System, HVS)引入传统的视频编码框架, 着重挖掘人眼的主观感知状态和人类视觉选择性注意机制等来消除人类主观视觉冗余以获得更高的编解码性能. 研究热点包括

恰可察觉失真(Just Noticeable Distortion, JND)模型、视觉感兴趣区(Region of Interest, ROI)模型和视觉显著图(Saliency Map, SM)模型等.

JND 模型主要考虑了人类在感知图像信息过程中的多通道、非线性及多种视觉掩蔽特性. 该模型通过一个阈值量化表现感知冗余, 低于这个阈值的改变, 人眼无法察觉. 若能在视频编码过程中充分利用上述掩蔽效应, 则可以在主观感觉图像失真相同的条件下允许更大的实际失真. 当前, 已提出的 JND 模型可以

<sup>①</sup> 基金项目:国家自然科学基金(61271190,U1405255);福建省教育厅项目(JA15136);福建省高校产学研合作重大项目(2014H61010105);福建师范大学科研创新团队(IRTL1207)

收稿时间:2016-03-10;收到修改稿时间:2016-04-14 [doi:10.15888/j.cnki.csa.005447]

分为两类: 像素域的 JND 模型<sup>[1-3]</sup>和变换域的 JND 模型<sup>[4-8]</sup>(如 DCT 域 JND 模型)。

基于像素域的 JND 模型直接给出具体每一像素的最小可察觉失真值。如文献[1]分别计算亮度自适应因子和纹理掩蔽效应因子, 然后将两个因子的最大值作为 JND 阈值。在此基础上, 文献[2]提出了扩展的 JND 模型——非线性叠加掩蔽模型(Nonlinear Additively Masking Model, NAMM), 将亮度掩蔽和纹理融合在一起用于运动估计等内容。Liu 等人<sup>[3]</sup>对图像进行分层处理, 将结构块和纹理块分别建模 JND 模型, 进一步提高了 JND 的精确性。

基于 DCT 域的 JND 模型通常包括对比敏感函数(CSF), 亮度自适应因子和各种掩蔽效应因子。Watson 等人提出了 DCTune 模型, 通过融合了亮度和对比度掩蔽效应到基本阈值中<sup>[4]</sup>。Ahumada 等人在提出的 DCT 域的 JND 模型中融合了 CSF 函数<sup>[5]</sup>。文献[6]提出一种新的 CM-JND 模型以及整合效应因子, 该模型利用快速 CU 分割算法得到图像块, 不同块大小建模 JND 更精确, 并使其应用在 H.265 编解码框架。文献[7]提出联合估计像素域和变换域的 JND 阈值用于图像压缩。Zheng 等人<sup>[8]</sup>提出图像纹理分离建模 JND, 改善了纹理块、结构块以及平坦块的划分。

人眼观看视频的过程就是一种选择重要信息的过程。显著性模型常应用于甄选视觉场景中的相关内容作为人眼视觉注意力区域。Goferman 等人<sup>[9]</sup>结合了图像的全局和局部特性提出了基于上下文注意模型的显著性探测方法。考虑到现实中大量的视频以压缩形式被获取, 文献[10]提出了一种面向压缩域的视频显著性检测算法, 利用一个移动窗口中的离散余弦变换(DCT)系数和运动信息修缮显著图模型, 取得了较好的效果。

本文提出一种改进的基于显著性协同检测的恰可察觉失真模型, 将代表视觉敏感度的 JND 模型与代表视觉注意机制的显著性模型用于视频编码。该模型采用基于上下文感知的显著性算法, 同时考虑像素域的 JND 模型和 DCT 域的 JND 模型的优点, 分别对图像纹理、边界和平坦区域分别估计出更为精确的显著性区域, 并将该模型用于视频编码的残差量化编码。

## 2 显著图协同检测JND模型

由上文可知: 当视频图像编码失真小于 JND 阈值时, 人眼无法感知; 而人类视觉一般会被图像的某些

显著性区域所吸引, 人类视觉系统容易对感兴趣的事物对象加以更多的视觉细节关注。利用上述特点, 在应用 JND 去除不可察觉的冗余系数, 利用生成的显著图协同引导 JND 阈值分配有助于进一步减少视频编码码率。在文献[7,10]基础上, 本文提出了改进的显著图协同检测 JND 模型(JND based on Saliency Detection, SD-JND), 通过显著性调整 JND 阈值的大小, 间接达到调整码率分配的目的。

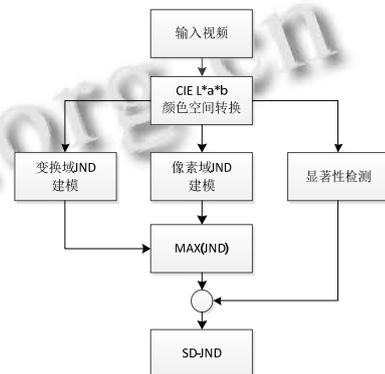


图 1 SD-JND 模型的计算流程图

SD-JND 模型的计算流程图如图 1 所示。系统使用非线性关系的 CIE L\*a\*b\* 颜色空间, 其可以有效模仿人类视觉的非线性响应, 从而较好地描述人类视觉系统可见的所有颜色。采集的 YUV 视频序列通过 CIE L\*a\*b\* 颜色空间转换, 分别计算其频域 JND 模型和像素域 JND 模型 SD-JND 模型并同时进行显著性检测。如式(1), 提出的 SD-JND 模型主要由两个影响因子组成: 最优 JND 估计模型和显著性检测因子。

$$SD-JND(x,y,i) = MAX\{JND_{PIXEL}(x,y,i), JND_{IDCT}(x,y,i)\} \times (Sa(x,y) - h) \quad (1)$$

其中,  $x$ 、 $y$  和  $i$  分别表示像素横纵坐标和视频帧序号,  $Sa(x,y)$  表示归一化的帧显著性阈值,  $h$  为折中参数, 用于调节主观质量与编码码率。实验结果表明, 对显著性阈值进行整体缩小比较容易与 JND 融合,  $h$  为 0.2 效果较佳。

$JND_{PIXEL}(x,y,i)$  是像素域 JND 模型, 主要由亮度自适应因子和对比度掩蔽因子组成。像素域 JND 模型在估计 JND 阈值时, 起主导作用的是对比度掩蔽因子, 这容易导致估计纹理等高频信息的 JND 阈值不准确; 而 DCT 域的 JND 模型能很好的体现低频、高频和中频信息, 尤其对高频信息(即纹理区域)的 JND 阈值估计更为有效。因此, 提出的最优估计 JND 模型通过

MAX函数计算两种像素域的JND模型,为不同区域分配不同的JND阈值的模型,可以动态的反映不同图像、不同区域所包含的信息.其中, $JND_{IDCT}(x,y,i)$ 表示从DCT域进行反变换得到的像素域JND模型.对亮度分量或色度分量分别进行计算DCT域下的JND阈值,公式如下:

$$JND_{IDCT}(x,y,i) = IDCT(sign(c_{LorC}(x,y,i)) \cdot JND_{DCT}(x,y,i)) \quad (2)$$

公式中的 sign 表示取得变换块系数的符号. $JND_{DCT}(x,y,i)$ 表示DCT域下的JND阈值.

提出JND建模主要考虑纹理、边缘、平坦区域和显著程度.除了综合变换域和像素域JND的优点.还考虑图像中的显著区域分布.即对显著性区域分配更少的JND阈值,对于人眼不感兴趣的区域适当增加JND阈值,具体细节描述见第2、3部分.整个模型将加入到HEVC框架的JND残差滤波器里.

### 3 像素域的JND建模

考虑到视频处理的计算复杂度,且人眼视觉系统对亮度更敏感使用计算量较小的NAMM模型<sup>[2]</sup>.并只对亮度通道掩蔽效应建模:

$$JND_v(x,y) = B^l(x,y) + M^l(x,y) - C_v \times \min\{B^l(x,y), M^l(x,y)\} \quad (3)$$

其中  $B^l(x,y)$ 和  $M^l(x,y)$ 分别是背景亮度自适应和对比度掩蔽的可觉察阈值.  $C_v$ 表示在通道  $v$  下的重叠效应. $B^l(x,y)$ 表示如下:

$$B^l(x,y) = \begin{cases} 17(1 - (\frac{I(x,y)}{127})^2) + 3 & I(x,y) \leq 127 \\ \frac{3}{128}(I(x,y) - 127) + 3 & I(x,y) > 127 \end{cases} \quad (4)$$

其中  $I(x,y)$ 表示平均背景亮度.  $M^l(x,y)$ 公式如下:

$$M^l(x,y) = \beta_v \cdot G_v(x,y) \cdot W_v(x,y) \quad (5)$$

$G_v(x,y)$ 表示像素梯度的最大加权平均,  $\beta_v$ 表示为每个颜色通道的控制参数,  $v$ 为亮度分量,具体的公式参照文献<sup>[11]</sup>.

### 4 变换域的JND建模

基于变换域的JND模型在亮度自适应因子和对比度掩蔽因子上还加入了CSF函数.在DCT域下,每个变换块系数分量能很好的体现低频、高频和中频信息.根据文献<sup>[12]</sup>,亮度和色度对应的变换域的JND模型可以由基本阈值和各种调节因子组成.亮度和色度的DCT域JND阈值可分别由式(6)和式(7)得到:

$$JND_{L\_DCT}(x,y) = T_{L\_Basic}(x,y,p) \times A_{lum} \times F_{L\_contrast}(x,y,p) \quad (6)$$

$$JND(x,y) = T(x,y,p) \times F(x,y,p) \quad (7)$$

其中,  $T_{L\_Basic}(x,y,p)$ 和  $T_{C\_Basic}(x,y,p)$ 对应于亮度和色度的基本阈值,基本阈值是由空域对比度敏感函数(Spatial Contrast Sensitivity Masking)生成.  $F_{L\_contrast}(x,y,p)$ 和  $F_{C\_contrast}(x,y,p)$ 分别对应亮度和色度的对比度掩蔽因子.  $A_{lum}$ 反映亮度自适应效应.

$$T(x,y,p) = s \cdot \frac{1}{\Phi_x \Phi_y} \cdot \frac{\exp(c \cdot \omega) / (a + b \cdot \omega)}{r + (1-r) \cdot \cos \vartheta} \quad (8)$$

$$\Phi_{x \text{ or } y} = \begin{cases} \sqrt{1/N} & x \text{ or } y = 0 \\ \sqrt{2/N} & x \text{ or } y > 0 \end{cases} \quad (9)$$

$$\vartheta_{xy} = \arcsin\left(\frac{2\omega_x \omega_y}{\omega_x^2 + \omega_y^2}\right) \quad (10)$$

其中的  $\Phi_x$ 和  $\Phi_y$ 分别表示变换系数的归一化因子.参数  $s$ 表示空域整合效应因子(Spatial Summation Effect)根据文献<sup>[13]</sup>将  $s$ 设置为0.25.  $\vartheta_{xy}$ 表示变换系数块的方向角,其他参数为常量.变换块的第  $(x,y)$ 坐标处对应的空间频率  $\omega_{xy}$ 计算公式如下:

$$\omega_{x,y} = \frac{1}{2N} \sqrt{(x/\theta_i)^2 + (y/\theta_j)^2} \quad (11)$$

$$\theta = \theta = 2 \cdot \arctan\left(\frac{1}{2 \cdot R \cdot Pic}\right) \quad (12)$$

参数  $\theta_i$ 和  $\theta_j$ 分别表示为像素的水平和垂直视角.变换系数块的维数是  $N$ .式(12)中的  $R$ 表示观看距离与图像高度的比值,  $Pic_{height}$ 表示图像高度.

通过拟合色度分量的CSF函数曲线可以得到色度的基本阈值公式如下:

$$T_{C\_Basic}(x,y,p) = \psi \cdot [(1-a)\exp(b \cdot f^c) + \alpha] \quad (13)$$

其中参数  $\alpha$ 为1/256.其他参数设置参考文献<sup>[14]</sup>.

最小感知亮度差异是随着图像背景亮度的增加而增加的:背景信号越强,人类的敏感度就越低,JND阈值就越高.大部分JND阈值都是在中等像素值(128)时生成的,对于其他的像素值,则要加上一个调节因子.因此,式(5)中的亮度自适应因子  $A_{lum}$ 表示如下:

$$A_{lum}(k) = \begin{cases} (60 - \bar{f}(k))/150 + 1 & \bar{f}(k) \leq 60 \\ 1 & 60 < \bar{f}(k) < 170 \\ (\bar{f}(k) - 170)/425 + 1 & \bar{f}(k) \geq 170 \end{cases} \quad (14)$$

$\bar{f}(k)$ 表示第  $k$ 个变换块的平均像素大小.

对比度掩蔽效应是指一个信号的存在影响人眼对另一个信号的敏感程度.不同区域的对比度掩蔽效应也是不同的:如图2所示,在纹理区域的噪声不容易被观看者发现,而在边界和平坦区域更容易察觉到.将视频帧块划分为三类:纹理块、平坦块、边缘块,式

(6)中亮度的对比度掩蔽因子计算如下:

$$F_{L\_contrast}(x,y,p)= \begin{cases} \Psi & (x^2+y^2)\leq 16 \text{ 边缘与平滑} \\ \Psi \cdot \min\left(4, \max\left(1, \left(\frac{C_L(x,y,p)}{T_{L\_Basic}(x,y,p)\times A_{lim}}\right)^{0.36}\right)\right) & \text{otherwise} \end{cases} \quad (15)$$

$$\Psi = \begin{cases} 1 & \text{边缘或平滑块} \\ 2.25 & (x^2+y^2)\leq 16 \text{ 纹理块} \\ 1.25 & (x^2+y^2)>16 \text{ 纹理块} \end{cases} \quad (16)$$

其中,  $C_L(x,y,p)$ 代表亮度变换块的系数. 式(16)说明了亮度对比度掩蔽效应在不同类型块所得到的掩蔽加权因子不同. 三类图像块通过计算边缘像素的密度值划分. 色度的对比度掩蔽因子可见参考文献[15].

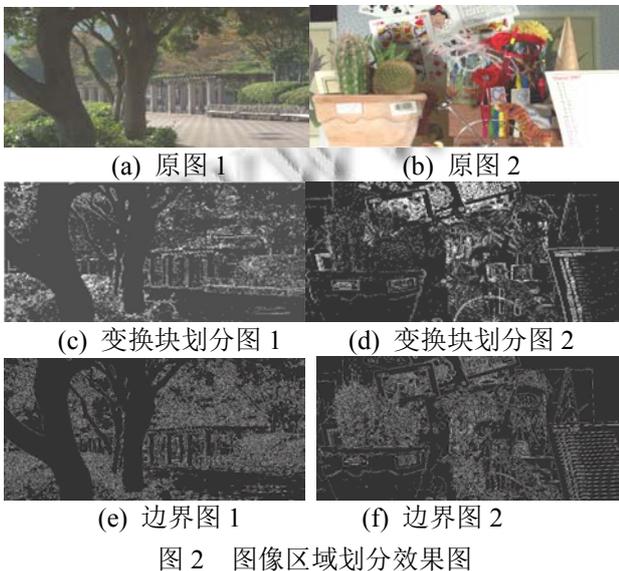


图2 图像区域划分效果图

(a)和(b)是原图, (c)和(d)为变换块划分图, 白色的是纹理块, 灰色是边界块, 黑色属于平面. (e)和(f)为Canny算子检测的边界图.

## 5 显著性检测算法

### 5.1 局部和全局相结合的显著性探测

当图像中的一个像素值所在的小块区域与其周围的小块区域的特征存在较大差异的时候, 当前的小区域块可以被认为是显著性区域. 本文采用Goferman等人提出的基于上下文感知的显著性算法<sup>[9]</sup>, 按照具有特定颜色的区域或图案应该获得较高的显著性. 该算法不仅考虑了多尺度还加入了上下文的相关信息, 这使得检测结果较为准确. 为减少计算量, 使用欧氏距离  $d(p_i, p'_j)$  寻找  $K$  个最接近的图像块为了估计周围的

块相似度, 这里的  $p_i$  和  $p'_j$  分别表示块和像素, 块之间的相异度计算公式如下.

$$d(p_i, p'_j) = \frac{d_{C\_ED}(p_i, p'_j)}{1 + c \cdot d_{P\_ED}(p_i, p'_j)} \quad (17)$$

背景图像块是分布在整个图像中的, 而显著块一般是聚集在一起的. 这意味着和该图像块越相似的图像块位置距离越近, 该块越显著. 因此, 欧氏距离  $d_{P\_ED}(p_i, p'_j)$  可以用来检测两块位置的远近程度,  $d_{C\_ED}(p_i, p'_j)$  表示 CIE  $L^*a^*b^*$  颜色空间中的欧氏距离. 如果当前显著的图像块  $p_i$  与图像块  $p'_j$  距离较远, 则认为  $p$  不是显著块. 这里的  $c$  为 3. 公式(17)结合了欧氏距离和位置距离来同时衡量块之间的相异度. 块之间的  $d_{C\_ED}(p_i, p'_j)$  和  $d_{P\_ED}(p_i, p'_j)$  越大, 则当前块越显著. 在单一的  $t$  尺度下, 像素域的显著性计算公式为:

$$S'_i = 1 - \exp\left\{-\frac{1}{k} \sum_{k=1}^k d(p'_i, (p'_k)^t)\right\} \quad (18)$$

### 5.2 多尺度显著性检测方法

多尺度的情况下, 背景像素块有许多相似的块. 显著图像块在一些尺度下有相似的图像块. 融合多尺度降低背景像素块的显著性, 从而改善显著性区域与非显著性区域的对比度效应. 在选择图像块时, 要把尺度因素考虑进来. 则式(18)可以改写为:

$$S'_i = 1 - \exp\left\{-\frac{1}{k} \sum_{k=1}^k d(p'_i, (p'_k)^t_k)\right\} \quad (19)$$

式(19)得到的显著值需要进行归一化.

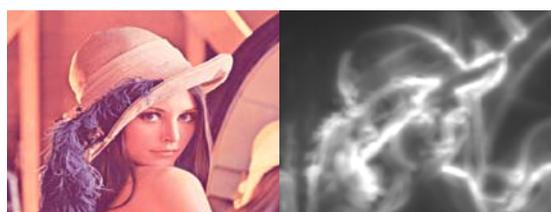
### 5.3 显著性上下文信息

依据 Gestalt 的理论<sup>[10]</sup>, 人类在观看某个场景时, 视觉会对该场景中的画面形成一个或多个注意力焦点. 距离注意力焦点近的区域比远的区域更重要. 注意力焦点区域的周边也同样会吸引人眼注意力, 因此该区域也应该是显著的. 为了利用这一原理, 像素  $i$  的显著性重新定义为:

$$S' = \frac{1}{M} \cdot S \cdot (1 - d(i)) \quad (20)$$

参数  $d'_{foci}(i)$  表示在尺度为  $t$  时, 像素  $i$  与最近的显著性像素的位置欧氏距离. 算法实验结果如图 3.





(c) 原图 2 (d) 映射图 2

图 3 显著性检测算法效果图, (a)和(c)是原图, (b)和(d)是显著性检测的映射图.

### 6 基于JND与显著性协同检测的残差滤波

为了对每个像素值  $i$  分配适当合理的比特值, 利用显著性阈值来调整 JND 模型. 提出的模型通过残差滤波器加入到 HEVC 框架里, 图 4 是嵌入了 SD-JND 模型的视频编码系统框架. 式(1)中的  $Sa(x,y)$ 需要进行反计算, 其中,  $S$  来至式(20), 参数  $L$  实验中设置为 1.67 时效果最佳.

$$Sa(x,y) = L - S(x,y) \quad (21)$$

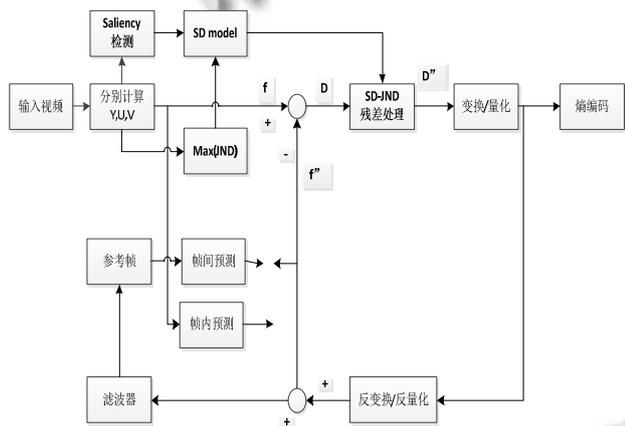


图 4 基于 SD-JND 模型的视频编码系统框架

如图 4 所示, 视频编码系统读入 YUV 视频序列后, 通过 SD-JND 建模, 然后导入到 JND 自适应残差滤波器中. 其中的  $f$ 、 $f'$ 、 $D$ 、和  $D''$  分别表示为原视频序列

帧、帧内预测后的视频帧、JND 滤波器后的残差数据和还没滤波的残差数据. 自适应残差滤波器加入到变换模块之前, 获得视频残差之后. 自适应残差滤波器的计算公式如下:

$$Re(x,y) = \begin{cases} D(x,y) + SD - JND(x,y) & D(x,y) - R < -(SD - JND(x,y)) \\ R & |D(x,y) - R| < SD - JND(x,y) \\ D(x,y) - SD - JND(x,y) & D(x,y) - R > SD - JND(x,y) \end{cases} \quad (22)$$

参数  $R$  表示平均残差值,  $Re$  表示滤波后的信号.

### 7 实验结果和讨论

为了评估提出的基于 SD-JND 模型的残差滤波器的编码性能, 该模型嵌入到 HEVC 视频编码框架软件 HM16.0 里. 以下主要使用 6 个标准测试序列来评估提出的系统性能: BasketballDrive、BQTerrace、Cactus、ParkScene、PeopleOnStreet. 序列 PeopleOnStreet 分辨率为 2560x1600, 其他都为 1920x1080, 本文与对比算法均使用全 I 帧配置, 分别测试计算了其码率大小和主观质量. 系统使用 DSCQS 方法进行主观测试<sup>[16]</sup>. 根据 DSCQS 方法, 两个视频序列将进行对比, 一个是原视频, 另一个是加入提出模块系统解码后的视频. 视频的显示顺序是: 视频序列 1, 等待 3 秒灰屏, 视频序列 2, 等待 3 秒的灰屏. 每个视频组播放 2 遍. 主观测试中采用 5 档的 MOS 平均主观打分机制.

表 1 分别描述了不同量化参数下, 提出的算法, 参考文献[10]算法和标准 HM16 的编码性能. 实验结果表明加入了 SD-JND 残差滤波模块的 HM16 要比标准代码有了更多的码率节省: 码率降低 2.51% ~ 28.57%. 提出的模型对比标准软件 HM16, 平均码率节省 10.7%, 文献[10]的模型平均码率节省为 8.56%. 随着量化参数 QP 的增加, 码率呈现出减少的趋势. 因为变换量化后非零系数减少了, 使得的码率下降有限.

表 1 各个模型的码率大小和码率节省

Video	Bitrate(Kb/s)					
	QP	提出的算法	文献[10]	HM16.0	提出的算法码率节省(%)	文献[10]算法码率节省(%)
Cactus	21	96425.9	100657.5	134987.7	28.57	25.43
	26	53252.6	54711.6	58828.1	9.48	7
	31	30090.6	30761.0	32140.7	6.38	4.29
	36	16405.95	16611.28	17294.0	5.14	3.95
BasketballDrill	21	18343.83	19183.98	22865.78	19.78	16.1
	26	10881.33	11248.95	12352.98	11.91	8.94

	31	6092.9	6227.65	6623.325	8.01	5.97
	36	3303.3	3378.325	3488.475	5.31	3.16
Basketball Drive	21	46055.43	46838.4	60203.03	23.5	22.2
	26	21599.4	21791.68	23356.63	7.52	6.7
	31	11884.45	11914.38	12190.54	2.51	2.27
	36	6462.88	6481.58	6735.029	4.04	3.76
PeopleOnStreet	21	106169.3	109099.3	115357.1	7.97	5.42
	26	63383.22	65398.99	67291.59	5.8	2.81
	31	36700.61	37639.77	38600.04	4.92	2.49
	36	21474.86	21826.88	22063.5	2.67	1.07
Park scene	21	43468.57	45286.33	55388.67	21.52	18.24
	26	25125.18	26068.93	30046.74	16.38	13.24
	31	13984.68	14504.87	15923.99	12.18	8.91
	36	7086.024	7285.2	7815.84	9.34	6.79
BQ Terrace	21	161823.8	167034.8	198802.8	18.6	15.98
	26	96566.07	98815.05	109458.8	11.78	9.72
	31	55004.34	56008.2	59331.94	7.29	5.6
	36	30405.96	30728.64	32531.28	6.53	5.54
平均值				10.7	8.56	

表 2 从主客观标准来评价算法的编码性能。可以看出,提出的模型在主观质量差不多的情况下,实际峰值信噪比(PSNR)比 HM16.0 参考软件平均下降了 2.599dB,与 Wang 等人<sup>[10]</sup>提出的模型 PSNR 平均下降了 0.129dB。如图 6 所示,提出的模型解码图和原始视

频帧的第 8 帧进行对比。当 QP 值增大时,PSNR 下降,而视频失真程度越高。QP 设置为 36 的情况下,提出模型的视觉质量也没有明显的下降。通过上面的分析说明提出的模型不仅仅能够容忍视频数据失真而且在视觉感知质量上没有明显的下降。

表 2 PSNR 大小和 MOS 评分

Video	PSNR(dB)				MOS	
	QP	HM16.0	提出的算法	[10]	提出的算法	[10]
Cactus	21	41.514	37.1415	37.3329	4.8	5
	26	38.2881	36.113	36.2482	4.8	4.8
	31	36.0445	34.6296	34.7144	4.6	4.7
	36	33.5786	32.6409	32.6754	4.7	4.6
BasketballDrill	21	42.4925	37.9718	38.2366	5	5
	26	38.9867	36.6884	36.8710	4.9	5
	31	35.9931	34.8671	34.9597	4.9	4.9
	36	33.3207	32.6799	32.7357	4.5	4.8
BasketballDrive	21	41.7544	39.2372	39.3494	4.7	4.8
	26	39.6737	38.0749	38.1676	4.6	4.7
	31	38.2634	36.8764	36.998	4.5	4.5
	36	36.5794	35.367	35.4064	4.3	4.4
PeopleOnStreet	21	43.9589	37.9718	38.3349	5	5
	26	40.4565	37.0595	37.2958	4.8	4.9
	31	37.3128	35.4716	35.5812	4.7	4.8
	36	34.4099	33.3813	33.4230	4.6	4.6
Park	21	42.1866	36.282	36.4065	4.9	5

scene	26	39.2747	35.5929	35.6878	4.7	4.8
	31	36.4221	34.27	34.3266	4.6	4.6
	36	33.6073	32.3638	32.3847	4.5	4.6
BQ Terrace	21	43.7052	36.9224	37.2225	5	5
	26	38.6144	35.4272	35.6257	4.9	4.9
	31	35.1101	33.3734	33.4761	4.8	4.9
	36	32.0818	30.8226	30.8563	4.5	4.6
平均值		38.067	35.468	35.597	4.72	4.79

图 5 所示为 R-D 曲线, 横坐标为比特率, 纵坐标为 PSNR 值. R-D 曲线通常用于衡量压缩算法的性能优劣, 在不同的量化值 QP 下编码得到 PSNR 和码率. 从曲线图可以直观地看出, 我们的模型和参考模型对比位于下方. 当码率相同时, 我们的模型拥有更低的 PSNR, 进一步说明了提出的模型能够容忍更多的数据失真.



(a) 原始视频 (b) 我们的模型解码结果

图 6 Traffic 视频序列解码图对比, QP 设置为 36

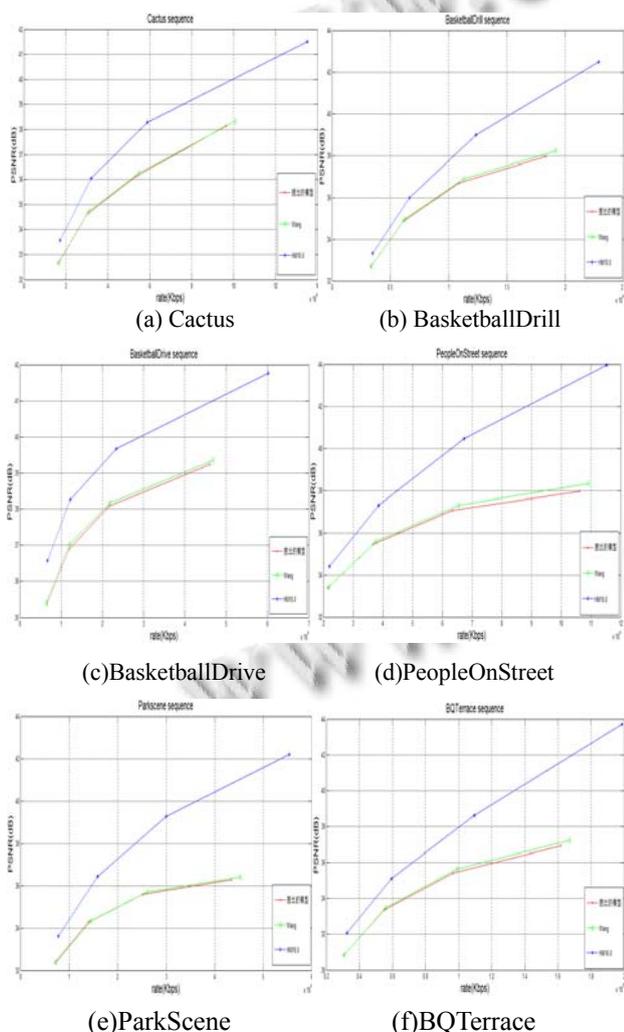


图 5 6 个测试序列的 R-D 曲线对比

### 8 结语

本文提出了一种改进的基于显著性协同检测的综合估计 JND 阈值的模型. 该模型利用像素域和变换域下建模得到的最优 JND 模型进行协同估计, 对图像不同区域分别计算得到不同的 JND 阈值. 同时, 利用显著性检测得到图像中的重要区域, 对重要区域给予低的 JND 阈值. 从实验结果可以得出, 主观感知没有下降的情况下, 提出的模型与现有的模型对比更好的压缩性能, 能够容忍更多的视频数据失真. 提出的模型在 DCT 域建模 JND 还存在分块不合理等问题, 在未来的工作中, 可以进一步改善.

### 参考文献

- 1 Chou CH, Li YC. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. IEEE Trans. on Circuits and Systems for Video Technology, 1995, 5(6): 467-476.
- 2 Yang XK, Lin WS, Lu Z, ed. Just-noticeable distortion profile with nonlinear additivity model for perceptual masking in color images. IEEE International Conference on Acoustics, Speech, and Signal Processing. 2003. 609-612.
- 3 Liu A, Lin W, Paul M, ed. Just noticeable difference for images with decomposition model for separating edge and textured regions. IEEE Trans. on Circuits and Systems for Video Technology, 2010: 1648-1652.
- 4 Watson AB, Yang GY, Solomon JA, ed. Visibility of wavelet

- quantization noise. *IEEE Trans. on Image Processing*, 1997, 6(8): 1164–1175.
- 5 Ahumada AJ, Peterson HA. Luminance-mode based DCT quantization for color image compression. *Human Vision, Visual Processing, and Digital Display III*, San Jose: SPIE Press, 1992: 365–374.
- 6 Bae SH, Kim M. A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images. *IEEE Trans. on Image Processing*, 2014, 23(8): 3227–3240.
- 7 郑明魁,苏凯雄,王卫星.一种基于像素域与变换域联合估计的 JND 改进模型. *福州大学学报*, 2014, 2: 225–230, 241.
- 8 Zheng MK, Kaixiong S, Wang WX, ed. Enhanced subband JND model with textural image. *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*. 2013. 1–4.
- 9 Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 1915–1926.
- 10 Wang HQ, Wang L, Hu XL, ed. Perceptual video coding based on saliency and just noticeable distortion for H.265/HEVC. *International Symposium on Wireless Personal Multimedia Communications (WPMC)*. 2014. 106–111 .
- 11 Yang XK, Lin W, Lu ZK, ed. Just noticeable distortion model and its applications in video coding. *Signal Process: Image Commun*, 2005, 20(7): 662–680.
- 12 Jia Y, Lin WS, Kassim AA. Estimating just-noticeable distortion for video. *IEEE Trans. on Circuit and system for video technology*, 2006, 16(7).
- 13 Wei Z, Ngan KN. Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. *IEEE Trans. on Circuits and Systems for Video Technology*, 2009, 19(3): 337–346.
- 14 Nadenau MJ, Reichel J. Compression of color images with wavelets considering the HVS. *Electronic Imaging. International Society for Optics and Photonics*, 1999: 129–140.
- 15 Chou CH, Liu KC. Color image compression using adaptive color quantization. *2004 International Conference on Image Processing. ICIP. 2004*. 2331–2334.
- 16 Recommendation I.500–11. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union. Geneva, Switzerland*. 2002.