

# 显著度和运动特性加权的视频质量评价方法

周 莺<sup>1,2</sup> 张基宏<sup>2</sup> 梁永生<sup>1</sup>

(1. 深圳信息职业技术学院可视媒体处理与传输深圳市重点实验室, 深圳, 518172; 2. 深圳大学信息工程学院, 深圳, 518060)

**摘要:** 进行客观视频质量评价时, 为了与主观评价结果尽可能一致, 需要考虑视频的动态特性和人眼观看的视觉特性, 因此本文提出一种基于显著区域和运动特性加权的视频质量评价方法。该评价指标基于传统的结构相似性指数 (Structural similarity index measurement, SSIM) 方法并在此基础上作了改进。首先通过频谱分析得到空域显著度, 通过视觉注意模型并结合运动特性获取时域显著度, 并根据时、空显著度动态融合得到帧级显著度。以帧级显著度加权 SSIM 指数, 便可得到整个视频帧的质量评价指标。在 LIVE VQA 标准数据集上的实验结果表明, 该评价指标更加接近于人眼对视频质量的主观评价价值。

**关键词:** 人眼视觉系统; 显著区域; 时空显著度; 运动特性; 结构相似性指数

**中图分类号:** TP37      **文献标志码:** A

## Saliency and Motion Weighed Video Quality Assessment

Zhou Ying<sup>1,2</sup>, Zhang Jihong<sup>2</sup>, Liang Yongsheng<sup>1</sup>

(1. Shenzhen Key Laboratory of Visual Media Processing and Transmission, Shenzhen Institute of Information Technology, Shenzhen, 518172, China; 2. College of Information Engineering, Shenzhen University, Shenzhen, 518060, China)

**Abstract:** To evaluate accurately the video quality and make it consistent with the subjective evaluation result, a video quality assessment method based on saliency region and motion characteristics weighting is proposed in this paper. This method is based on traditional structural similarity index measurement (SSIM) and makes some improvement. The spatial and temporal saliency is firstly extracted by spectrum analysis and visual attention model combined with motion characteristics, respectively. Then the frame saliency can be extracted by dynamic fusion of temporal and spatial saliency. Finally, the quality assessment index for entire video frame can be got by frame saliency weighted SSIM. The results of experiment on LIVE VQA standard data set show that this index is more close to objective assessment on video quality from human visual system.

**Key words:** human visual system; region of saliency; spatio-temporal saliency; motion characteristics; structural similarity index

## 引 言

随着视频编解码、传输等技术的进步以及播放终端的不断改进,用户对于视频播出质量的要求越来越高<sup>[1]</sup>。对于 IP 网络的视频流媒体应用而言,视频流在采集、压缩、传输和播出过程中,由于采集设备、压缩方法、传输介质和播出终端的限制,视频质量往往有所降低。为了把握视频播出质量的下降程度,及时反馈给编解码端和视频传输端,通过改进算法和传输机制来提高用户的视觉感知,就需要对视频播出质量进行准确的评价。

根据结构相似性理论,Wang 等<sup>[2]</sup>提出一种基于结构相似性度量 (Structural similarity index measurement, SSIM) 的视频质量评价方法,分别计算图像的亮度、对比度和结构的失真度,并求均值获得整幅图像的失真度。该方法由于计算及实现简单,准确性较好,受到了国内外学者的广泛关注。但它只对单帧图像计算准确度较高,而忽略了视频的动态信息,并且当图像严重模糊时,评价效果并不理想。近年来许多学者在此基础上进行改进,提出了一系列基于 SSIM 的改进算法。Wang 还提出了多尺度结构相似性度量方法 (Multi-scale SSIM, MS-SSIM)<sup>[3]</sup> 和信息内容加权的结构相似性度量方法 (Information weighted SSIM, IW-SSIM)<sup>[4]</sup>; 文献[5]提出一种基于结构边缘信息的图像质量评价方法 (Edged-based SSIM, ESSIM),改善了对模糊图像的失真分析;文献[6]提出一种基于运动信息和结构信息的视频质量评价方法 (Motion and edged-based SSIM, MESSIM),依据人眼视觉系统 (Human visual system, HVS) 特性对场景结构和运动信息进行加权;文献[7]提出基于 HVS 的结构相似性视频质量评价方法 (HVS-based SSIM, HSSIM),利用 HVS 对不同频率分量和不同区域的敏感程度不同这一特性,引入亮度因子、纹理因子和空间位置因子对宏块的权重加以调节,并利用运动矢量来计算序列中单个帧的权重;文献[8]基于视频时域统计特性,提出显著度和失真加权的视频质量评价方法 (Saliency and distortion weighted SSIM with temporal pooling strategy, SDTW-SSIM);文献[9]利用密度、颜色和运动矢量取平均得到显著图,并以此为权重来加权 SSIM;文献[10]首先通过多尺度方法提取视频运动特征得到时域显著图,密度、色度、对比度的加权和得到空域显著图,并对时空显著图通过高斯滤波得到最终显著度,作为 SSIM 的加权重。

然而目前大多数视频质量评价方法还是基于单帧图像质量的计算,再取所有帧质量的均值或者加权均值作为整个视频的质量评分。但视频区别于图像的最显著特性就是其动态性,人眼在观察视频序列时,能够明显感受到画面中运动物体的方向性和连贯性。此类方法恰恰忽略了视频最重要的运动特性。人眼观察视频时,会迅速转移到感兴趣区域 (Region of saliency, ROS),这些区域大都是纹理较丰富、细节信息较多或者运动较剧烈的区域,而对于一幅画面的背景,通常很少关注。同样的误码失真如果发生在 ROS,带来的误码扩散失真会比发生在非 ROS 的更严重。随着时间的推移,人眼的视觉焦点也会跟着转移,对运动目标的关注度会逐渐降低,进而转向注意画面中新出现的物体或者显著度较低的区域。其次,HVS 对具有不同运动特征的视频反映也不相同<sup>[11]</sup>。发生在同一视频场景内的运动越剧烈,人眼观察的失真感知就越差;对于快速运动变化的场景,人眼很难分别其细节信息,因此发生在快速运动变化的视频帧中的失真更难被察觉。当发生场景切换时,人眼对接下来的几帧场景切换图像内发生的失真几乎察觉不到,因此,对场景切换帧内发生的失真应赋予极小的权重。

基于以上考虑,本文在视频动态性描述和融合视频运动特性的显著区域提取的基础上,提出一种显著度和运动特性动态加权的 SSIM 指数 (Saliency and motion characteristics weighted SSIM, SMW-SSIM),其核心思想就是权衡视频帧的运动变化幅度及发生在 ROS 的失真对视频质量的影

响,对于发生在同一帧视频 ROS 内的失真赋予较大的权值,对于运动剧烈的视频帧内的失真赋予较小的权值。在对视频进行质量评价时,充分考虑人眼对于视觉感兴趣区域和运动变化物体的主观感受,且该视频质量评价方法不依赖于视频格式和编码方式,对于下一代视频编码标准 HEVC 仍具有实用性<sup>[1]</sup>。该视频质量评价方法的整体框架如图 1 所示。

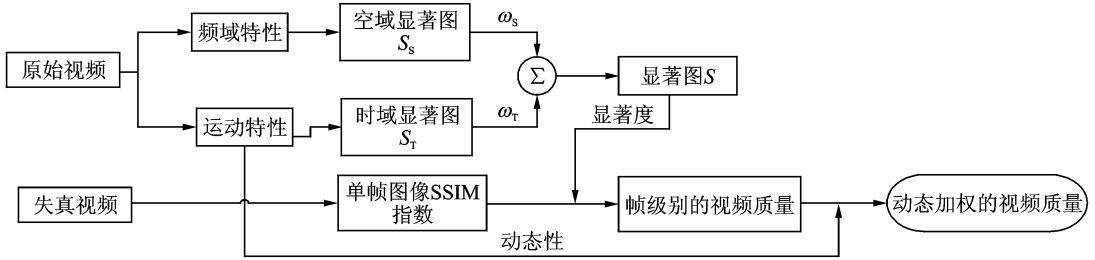


图 1 视频质量评价方法框图

Fig. 1 Block diagram of video quality assessment

## 1 基于显著度和运动特性动态加权的视频质量评价方法

### 1.1 融合视频运动特性的显著度计算

由于 HVS 的视觉感知特性,人眼通常对移动、变化或者色彩、纹理上与周围反差较大的目标反应敏感<sup>[12]</sup>。本文首先分别提取空域特征和时域特征,以人眼视觉感知特性为依据,动态加权融合时空显著度,最终得到图像显著性区域。

根据图像的统计不变性,在频域仍能够保持空域特性,因此可以用频域的能量谱来刻画图像的空域特性,避免了传统空域显著图提取时特征选取的问题。其中,幅度谱可以表示图像信息量的变化情况,而相位谱可以表明信息变化的位置。本文采用文献[13]的谱冗余(Spectral residual, SR)方法,首先通过傅里叶变换得到图像的频域对数谱,通过滤波去除大部分冗余信息,最后经过傅里叶反变换得到空域显著图<sup>[12]</sup>。具体步骤为<sup>[14]</sup>

$$A(f) = \log |\hat{F}(f)| \times h_n(f) \quad (1)$$

$$I(A(f)) = A(f) \times L(f, k) \quad (2)$$

$$R(f) = A(f) - I(A(f)) \quad (3)$$

$$S_s(f) = g(f) \times \hat{F}^{-1}[\exp(R(f) + P(f))]^2 \quad (4)$$

式中:假设输入图像为  $f$ ,则  $\hat{F}(f)$  为  $f$  的傅里叶变换; $A(f)$  为  $f$  的对数振幅谱; $h_n(f)$  为  $n \times n$  的均值滤波器; $P(f)$ ,  $R(f)$  分别为输入图像的相位谱和频谱冗余;对  $f$  高斯滤波得到输出  $g(f)$ ,  $S_s(f)$  即为最终的空域显著图; $L(f, k) = \frac{1}{\sqrt{2\pi}2^k\sigma} \times \exp\left[-\frac{u^2 + v^2}{(2^k\sigma)^2}\right]$  为高斯核函数; $k = 1, 2, \dots, K$  为尺度参数; $K = \log_2 \min\{H, W\}$  由图像尺寸决定,  $H, W$  分别为图像的高度和宽度。

显著区域提取的好坏决定于尺度参数  $k$ ,若  $k$  过小,则冗余信息不能有效抑制;若  $k$  过大,则只能突出显著区域的边界信息,最优尺度参数  $k_e$  满足

$$k_e = \operatorname{argmin}\left(-\sum_{i=1}^n P_s(k)_i, \log(P_s(k)_i)\right) \quad (5)$$

式中: $P_s(k)_i$  为尺度  $k$  时的显著图的熵。通过寻找具有最小熵值的显著图,便可得到最优尺度参数  $k_e$ 。

视频区别于图像的最大特征在于其包含的丰富运动信息。根据人眼视觉特性和视觉注意机制, HVS 观察视频时更容易被移动、变化的目标所吸引, 并且对运动特性不同的视频, 人眼主观感知也不一样<sup>[12]</sup>。据统计, 在实际观察视频时, 当运动物体相对于周围对象的运动强度较大或运动方向不同时, 人眼对其注意力将速度提升。本文通过相关性实验, 选取最能反映视频显著区域运动特性且便于计算的运动强度、运动密度和运动方向 3 个时域特征。首先采用文献[11]中的方法提取出运动目标, 然后对目标对象的运动矢量图进行高斯滤波, 再根据目标对象的运动强度、方向和密度 3 个特征通道, 利用视觉注意模型, 便可得到视频当前帧的时域显著图。

随着注视时间的推移, 人眼对于运动物体的关注度也会发生变化<sup>[11]</sup>。运动目标持续出现时, 人眼对其敏感程度会有所降低, 此时人眼会转而关注之前显著度较低区域或在画面中新出现的运动物体<sup>[15]</sup>, 因此, 最终时域显著图  $S_T(i)'$  可表达为

$$S_T(i)' = \text{norm}\{S_T(i) + \int_{t-\Delta t}^t [(S_T(i) - S_T(i - \Delta t)) \times g] dt\} \quad (6)$$

式中:  $g$  为高斯函数,  $(i - \Delta t)$  为与  $i$  帧间隔  $\Delta t$  的另一帧,  $S_T(i)'$  即为修正之后的时域显著图。

根据人眼视觉感知特性自适应融合时空显著图, 获得视频序列每一帧的时空显著图  $S(n)$  为<sup>[12]</sup>

$$S(n) = \frac{\sum_{i=1}^N \omega_S \times S_S(i) + \sum_{i=1}^N \omega_T \times S_T(i)}{\sum_{i=1}^n \omega_S + \sum_{i=1}^n \omega_T} \quad (7)$$

式中:  $S_S(i)$  代表第  $i$  帧视频序列的空域显著图;  $S_T(i)$  代表对应的时域显著图;  $S$  即为加权融合后的时空显著图;  $\omega_S, \omega_T$  为权重。

由于人眼具有向心性, 在考虑显著区域分布与人眼关注度的关系时, 显著区域离中心点越近, 分布越集中, 则人眼对其越敏感, 显著性权重也越高。假设图像中心点位置为  $(x, y)$ , 显著区域任一点像素位置为  $(x_i, y_i)$ <sup>[12]</sup>, 则

$$\omega_S = \|(x_i, y_i) - (x, y)\| \times \frac{1}{2\pi\sigma} \exp\left[-\frac{(x_i - x)^2 + (y_i - y)^2}{2\sigma^2}\right] \quad (8)$$

式中:  $\omega_S$  为高斯加权的欧氏距离。

根据视频的运动特性对  $\omega_T$  进行动态调节需要同时考虑运动的空域分布  $\omega_1$ 、运动强度  $\omega_2$  和运动复杂度  $\omega_3$  这 3 个因素<sup>[14]</sup>, 令

$$\omega_T = \omega_1 \times \omega_2 \times \omega_3 \quad (9)$$

$$\omega_1 = \frac{N_0}{N(s)} \quad (10)$$

式中:  $N_0$  代表目标的运动矢量非零的宏块个数,  $N(s)$  代表该帧总的宏块数, 则  $\omega_1$  为运动空域分布值, 其值越大, 说明运动分布越集中。

$$\omega_2 = \frac{\sum_{i=1}^{N_{kb}} (\|v_x\| + \|v_y\|)}{\sum_{i=1}^{N(s)} (\|v_x\| + \|v_y\|)} \quad (11)$$

式中:  $v_x$  和  $v_y$  分别代表目标运动矢量的横、纵坐标,  $N_{kb}$  代表背景中运动矢量非零的宏块个数, 则  $\omega_2$  为运动能量的大小, 其值越大表明视频包含的运动信息越丰富。

$$\omega_3 = \frac{-\left[\sum_{i=1}^m \frac{N(s_i)}{N(s)} \times \log\left(\frac{N(s_i)}{N(s)}\right)\right]}{\lg(36)} \quad (12)$$

式中:  $s_i$  代表目标中运动矢量方向直方图中各非空的维度;  $N(s_i)$  代表各个维度内运动矢量非零的宏块个数,  $i \leq 36$ ;  $\omega_s$  即为各宏块运动矢量在各维度的分布, 其值越大表明运动对象在方向上的分布越复杂<sup>[12]</sup>。

## 1.2 视频运动特性加权因子

在视频序列中, 用  $v(i, j, n)$  表示第  $n$  帧视频  $(i, j)$  处的运动矢量, 则  $\|v\|_2$  表示其运动矢量强度。视频帧的运动矢量场由以下 3 种运动矢量组成: 绝对运动  $v_a$ 、背景运动  $v_b$  和相对运动  $v_r$ <sup>[16]</sup>。3 种运动矢量满足

$$v_r = v_a - v_b \quad (13)$$

其中  $v_r$  的求解可采用文献[13]的全局运动估计法。则第  $n$  帧视频画面的运动剧烈程度  $M_v$  可用运动矢量强度  $\|v_r\|$  和  $\|v_b\|$  的加权和表示, 即

$$M_v = (1 - \omega_b) \|v_r\| + \omega_b \|v_b\| \quad (14)$$

式中:  $\omega_b$  由运动的空域分布  $\omega_1$ 、运动强度  $\omega_2$  和运动复杂度  $\omega_3$  3 种特征构成, 即

$$\omega_b = \omega_1 \times \omega_2 \times \omega_3 \quad (15)$$

由于人眼对视觉信号刺激的反映表现为非线性, 且视频运动变化越剧烈, 人眼对其失真越不容易察觉。根据 Weber 法则, 人眼视觉感知信号强度与物理信号强度近似服从对数变换, 则基于视频动态性的权重可表示为

$$\omega_M = \log\left(\frac{M_{\max}}{M_v^\alpha}\right) \quad (16)$$

式中:  $M_{\max}$  为视频序列运动矢量强度加权和的最大值,  $\alpha$  为一调节常数且  $\alpha > 1$ , 通过非线性拟合和数值分析, 文本算法中设定  $\alpha = 2$ 。

## 1.3 基于显著度和运动特性动态加权的视频质量评价方法

假设视频序列由  $n = 1, 2, \dots, N$  帧组成,  $(i, j)$  表示视频帧对应像素点上的位置坐标,  $(H, W)$  表示视频帧画面的高度和宽度,  $SSIM_{ij}(n)$  表示第  $n$  帧视频  $(i, j)$  位置上的结构相似性指数。利用 Wang 提出的 SSIM 指数评价方法<sup>[2]</sup>逐帧计算失真视频与参考的原始视频之间的 SSIM 指数  $SSIM_{ij}(n)$ ; 利用每帧图像的显著度对 SSIM 指数进行加权, 得到改进后的单帧图像质量 (Frame quality metric, FQM) 为

$$FQM(n) = \frac{\sum_{i=1}^H \sum_{j=1}^W [SSIM_{ij}(n) \times S(n)]}{\sum_{i=1}^H \sum_{j=1}^W S(n)} \quad (17)$$

根据 HVS 特性, 人眼感兴趣区域及运动剧烈的场景能明显引起人眼的注意, 因此在单帧图像质量考虑了显著性区域的基础上, 整个视频序列的质量评估还需要考虑到视频的运动特性。根据运动剧烈程度对单帧图像质量进行加权, 最终得出整体视频质量 SMW-SSIM, 其表达式为

$$SMW-SSIM = \frac{\sum_{n=1}^N \omega_M \cdot FQM(n)}{\sum_{n=1}^N \omega_M} \quad (18)$$

## 2 仿真实验与结果分析

为了验证本文提出的显著度和运动特性动态加权的视频质量评价方法更加接近于人眼的主观评

价, 对其在公共数据集 LIVE VQA 上进行了实验验证<sup>[17]</sup>, 并选取若干个典型客观评价方法与之进行比较, 其中包括: 峰值信噪比 (Peak signal noise ratio, PSNR)、视觉对比度 (Visual signal noise ratio, VSNR)、基于运动信息的视频整体质量估计指数 (Motion-based video integrity evaluation, MOVIE)、视频质量评价指数 (Video quality metric, VQM)、视觉互信息 (Visual information fidelity, VIF)、结构相似性指数 (Structural similarity index measurement, SSIM)、多尺度结构相似性指数 (Multi-scale structural similarity index measurement, MS-SSIM)、基于视频显著度的结构相似性指数 (Visual saliency structural similarity index measurement, VS-SSIM)<sup>[18]</sup> 以及显著度和失真加权的结构相似性指数 (Saliency distortion weighted structural similarity index measurement, SDW-SSIM)<sup>[8]</sup>。

根据 VQEG 提供的报告<sup>[19]</sup>, 客观质量评价指标与主观质量评价指标之间的关系为非线性, 这种映射关系可以通过下式进行非线性回归处理

$$f(x) = \alpha_1 + \frac{\alpha_2 - \alpha_1}{1 + e^{-\frac{x - \alpha_3}{\alpha_4}}} \quad (19)$$

式中:  $x$  代表客观评价指标,  $f(x)$  为处理后的算法结果。本文采用 Matlab 中的 nlinfit 函数进行非线性最小二乘优化。

为了比较各客观视频质量评价方法, 从而证明本文提出的方法更接近于主观评价, 选用斯皮尔曼秩等级相关系数 (Spearman rank order correlation coefficient, SROCC)、线性相关系数 (Linear correlation coefficient, LCC) 和离出率 (Outlier ratio, OR) 3 个指标来评判各种评价方法的性能。SROCC 和 LCC 值越大, OR 值越小, 说明客观评价结果越好。用本文提出的加权思想对视频序列每一帧的 MS-SSIM 进行加权和, 然后再在帧级别利用运动特性权重进行加权和, 还可得到改进的 SMW-MSSSIM 方法。各质量评价方法的结果比较如表 1 所示。

从表 1 可以看出, 本节提出的 SMW-SSIM 以及 SMW-MSSSIM 指数, 与原始的 SSIM 以及其他改进算法 MS-SSIM、VS-SSIM 和 SDW-SSIM 相比, 具有较高的精确度, 较好的单调性和较低的离出率, 显示了该

评价指标的优越性。表 1 中的 MOVIE 指数明显优于其他算法, 是因为前者充分利用了视频的运动信息, 分别在时域和空域进行失真估计, 增加了评估有效性。但由于 MOVIE 指数用光流场来表示运动信息, 计算复杂度大幅增加, 因此不太具有实用性。而本文提出的评价指标在有效降低计算复杂度的前提下仍能够得到较好的评价结果。

为了更直观地对改进算法进行比较, 实验用散点测试图 (如图 2 所示) 加以说明。客观预测值与主观评测值越接近, 拟合出的曲线越接近于直线。由图 2 可以看出, 本章提出的视频质量评价算法的散点图更加收敛到一条近似于直线的曲线上, 说明这种加权求和的质量评价方法更有效, 更加接近于人眼的

表 1 LIVE VQA 数据集质量评价结果

Tab. 1 Quality evaluation results of LIVE VQA database

质量评价方法	评价指标		
	LCC	SROCC	OR
PSNR	0.398 4	0.349 2	99.63
VSNR	0.684 2	0.671 9	62.69
MOVIE	0.806 7	0.779 5	40.94
VQM	0.716 2	0.698 1	56.90
VIF	0.567 0	0.559 4	80.37
SSIM	0.537 8	0.516 3	86.71
MS-SSIM	0.724 3	0.714 6	57.66
VS-SSIM	0.630 8	0.618 7	73.26
SDW-SSIM	0.730 5	0.709 8	57.21
SMW-SSIM	0.730 1	0.718 2	56.65
SMW-MSSSIM	0.737 2	0.731 0	55.387 4

主观感受。但由于该算法需要获取视频的显著度特性指标和运动特性指标,故相比于原始的 SSIM 算法,计算复杂度略高。

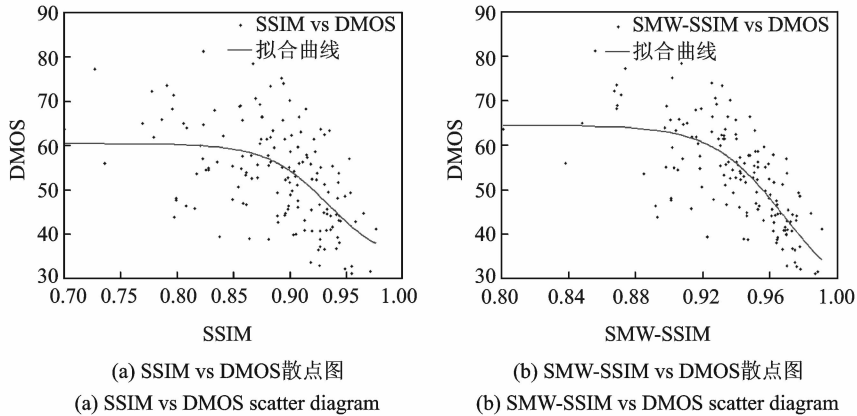


图 2 视频评价方法散点图比较

Fig. 2 Scatter diagram comparison of different VQA on LIVE VQA database

### 3 结束语

视频质量评价对于视频编码优化算法、视频传输方案的改进都有至关重要的作用,为了准确、快速衡量视频播出质量,并与人眼主观判断尽可能一致,本文提出一种全参考客观视频质量评价方法,用以模拟人眼视觉系统对视频质量的主观感知。该方法以 Zhou Wang 提出的结构相似性指数 SSIM 为基础,并结合视频的运动特性和显著度加以优化。首先,根据人眼视觉特性提取视频图像的显著图,并将时空显著度作为单帧图像的 SSIM 指数因子,对每一帧图像进行加权;然后再根据视频序列的动态性,求每帧 SSIM 指数的加权和,作为整个视频序列的播出质量评价指标;通过 LIVE VQA 标准数据集的实验,将该评价指标与其他常用视频质量评价方法进行比较,实验数据表明该方法与传统的 SSIM 指数相比,评价结果更加接近于人眼的主观评判,且算法简便快捷,易于实现。

### 参考文献:

- [1] 赵耀,黄晗,林春雨,等.新一代视频编码标准 HEVC 的关键技术[J].数据采集与处理,2014,29(1):1-10.  
Zhao Yao, Huang Han, Lin Chunyu, et al. Key techniques in new video coding standard HEVC[J]. Journal of Data Acquisition and Processing, 2014,29(1):1-10.
- [2] Wang Z, Bovik A C, Sheikh H R. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004,13(4):600-612.
- [3] Wang Z, Simoncelli E P, Bovik A C. Multi-scale structural similarity for image quality assessment[C]//Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers. Pacific Grove: CA, 2003,2(2):1398-1402.
- [4] Wang Z, Li Q. Information content weighting for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2011,20(5):1185-1198.
- [5] Chen Guanhao, Yang Chunling, Po Laiman, et al. Edge-based structural similarity for image quality assessment[C]//IEEE International Conference on Acoustics Speech and Signal Processing. Toulouse, France: IEEE,2006:933-935.
- [6] 卢国庆,李均利,陈刚,等.基于运动信息和结构信息的视频质量评价方法[J].计算机仿真,2010,27(6):262-267.  
Lu Guoqing, Li Junli, Chen Gang, et al. Video quality assessment measurement based on motion information and structural distortion[J]. Computer Simulation, 2010,27(6):262-267.
- [7] 汪志兵,廖煜鹏,汪博,等.基于 HVS 的结构相似性的视频质量评价[J].通信技术,2010,43(2):77-81.  
Wang Zhibing, Liao Yupeng, Wang Bo, et al. A novel HVS-based SSIM on video quality assessment[J]. Communications

Technology, 2010,43(2):77-81.

- [8] Zhu L, Su L, Huang Q, et al. Visual saliency and distortion weighting based video quality assessment[C]//Proceedings of the 13th Pacific-Rim Conference on Advances in Multimedia Information Processing. Singapore: Springer Berlin Heidelberg, 2012:546-555.
- [9] Gao X, Liu N, Lu W, et al. Spatio-temporal salience based video quality assessment[C]//IEEE International Conference on Systems Man & Cybernetics. Istanbul, Turkey: IEEE, 2010:1501-1505.
- [10] Fu B, Lu Z, Wen X, et al. Visual attention modeling for video quality assessment with structural similarity[C]//16th International Symposium on Wireless Personal Multimedia Communications. Atlantic City: IEEE, 2013:1-5.
- [11] 周莺,柳伟,张基宏.基于内容感知的可分级视频码流排序方法[J].信号处理,2013,29(8):1012-1018.  
Zhou Ying, Liu Wei, Zhang Jihong. Content-aware based sorting approach of scalable video bit stream[J]. Journal of Signal Processing, 2013,29(8):1012-1018.
- [12] 周莺,张基宏,梁永生,等.基于视觉运动特性的视频时空显著性区域提取方法[J].计算机科学,2015,42(11):118-122.  
Zhou Ying, Zhang Jihong, Liang Yongsheng, et al. Motion characteristics based video salient region extraction method[J]. Computer Science, 2015,42(11):118-122.
- [13] Hou X, Zhang L. Saliency detection: A spectral residual approach[C]//IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007:1-8.
- [14] Zhou Y, Zhang J, Liang Y, Liu W. Video motion characteristics based spatial-temporal salient region extraction method[J]. Journal of Information Hiding and Multimedia Signal Processing, 2015,6(2):225-233.
- [15] 陈旭,张基宏,柳伟,等.基于视觉注意的的视频可伸缩 ROI 算法[J].山东大学学报(工学版),2013,43(1):1-8.  
Chen Xu, Zhang Jihong, Liu Wei, et al. New scalable ROI algorithm based on visual attention[J]. Journal of Shandong University, 2013,43(1):1-8.
- [16] Gopalakrishnan V, Hu Yiqun, Rajan D. Salient region detection by modeling distributions of color and orientation[J]. IEEE Transactions on Multimedia, 2011,2(5):892-905.
- [17] Wang Z, Li Q. Video quality assessment using a statistical model of human visual speed perception[J]. Journal of the Optical Society of America, 2007,24(12):61-69.
- [18] Ma L, Li S, Ngan N. Motion trajectory based visual saliency for video quality assessment[C]//18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE, 2011:233-236.
- [19] Antkowiak J, Baina T J. Final report from the video quality experts group on the validation of objective models of video quality assessment[R]. ITU-T Standards Contribution COM 9-80-E, 2000.

#### 作者简介:



周莺(1982-),女,博士,讲师,研究方向:多媒体通信、信号处理与网络仿真, E-mail: zhouying722 @ 163.com.



张基宏(1964-),男,教授,博士生导师,研究方向:图像处理、神经网络与多媒体通信, E-mail: zhjh @ szu.edu.cn.



梁永生(1971-),男,教授,研究方向:计算机网络与数据通信、信号处理与模式识别, E-mail: liangys @ sziit.edu.cn.