

# 基于时域加权处理的 MPEG 视频质量评价方法

程德强<sup>1\*\*</sup>, 许超<sup>1\*</sup>, 李岩<sup>1</sup>, 游大磊<sup>1,2</sup>, 陈刚<sup>1</sup>

<sup>1</sup>中国矿业大学信息与控制工程学院, 江苏 徐州 221008;

<sup>2</sup>河南应用技术职业学院信息工程学院, 河南 开封 475000

**摘要** 结合人眼对显著物体及运动目标的感知特性,提出了一种基于时域加权处理的视频质量客观评价方法。首先利用经典的结构相似性算法得出单帧图像质量分数,然后依据视频中图像组(GoP)不同类型帧失真对视频质量的影响程度对每一帧图像质量进行加权得到每个图像组的质量分数,最后分别利用显著性和时间感知信息对所有的图像组进行时域加权处理,并均衡二者的权重得出最终的视频质量客观评价价值。在 LIVE 视频数据库上进行性能测试,实验结果表明,该方法与主观评价结果有较好的一致性,评价效果优于现有的典型模型。

**关键词** 图像处理; 视频质量评价; 显著性; 时间感知信息; 图像组

中图分类号 TP919.8

文献标识码 A

doi: 10.3788/LOP55.081011

## MPEG Video Evaluation Method with Time-Domain Weighting

Cheng Deqiang<sup>1\*\*</sup>, Xu Chao<sup>1\*</sup>, Li Yan<sup>1</sup>, You Dalei<sup>1,2</sup>, Chen Gang<sup>1</sup>

<sup>1</sup>School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221008, China;

<sup>2</sup>School of Information Engineering, Henan Vocational College of Applied Technology, Kaifeng, Henan 475000, China

**Abstract** In combination of the perceiving characteristics of human eyes for salient objects and moving targets, a video quality evaluation method based on time-weighted processing is proposed. Firstly, the quality score of single frame is obtained by using the classical structural similarity algorithm, and then the quality score of each group of pictures (GoP) is weighted to obtain the GoP mass score according to the influence of different types of frame distortions on the video quality in the video. Finally, all GoPs are time-weighted with saliency and temporal perceptual information, and the weights of both are equalized to get the final objective video quality evaluation value. The proposed metric is tested on the LIVE video database. The experimental results show that this method has good consistency with the subjective evaluation results, and the evaluation results are better than the existing classical models.

**Key words** image processing; video quality assessment; saliency; temporal perceptual information; group of pictures

**OCIS codes** 100.2960; 110.3000

## 1 引言

数字视频在诸如采集、处理、压缩、存储、传输和再现等各个阶段遭受各种各样的破坏和噪声,导致视频的视觉质量下降,从而影响了人眼的感官或造成机器处理的误差。因此,需要利用视频质量评价方法来估计视频的质量。质量评价可以是主观的或客观的<sup>[1-3]</sup>。客观质量评价的数学模型利用图像或视频的属性来模仿主观评价,相较于主观视频质量

评价更省时省力,适合实时应用。基于可参考信息的多少,客观质量评价方法分为全参考、半参考和无参考。在三种方法中,全参考质量评价能计算出最准确的值。在视频流的帧质量评价中,提取来自参考视频和失真视频的帧,捕获和比较来自两个视频的对应该帧,并采用全参考标准来估计两个视频的两个对应该帧之间的质量差异,以此获得所有帧的质量值,组合成整个视频的质量分数。

目前,国内外相关学者和研究机构在该领域做

收稿日期: 2018-02-08; 修回日期: 2018-03-21; 录用日期: 2018-03-28

基金项目: 国家自然科学基金(51774281)、江苏省“六大人才高峰”高层次人才培养项目(2015-ZBZZ-009)

\* E-mail: xccumt@126.com; \*\* E-mail: chengdq@cumt.edu.cn

了大量的研究,并已提出一些评价模型。如基于像素域的评价方法峰值信噪比(PSNR)和均方误差(MSE)<sup>[4]</sup>,该类方法没有考虑到人眼的视觉特性,评价结果与主观评价结果没有很好的一致性。为了解决这一问题,出现了基于结构相似性(SSIM)指标<sup>[5]</sup>的方法,以及 Wang 等<sup>[6]</sup>在该方法上的改进——多尺度结构相似性指标(MS-SSIM)。此外,还有基于自然场景统计的方法,例如,Sheikh 等<sup>[7]</sup>通过对自然场景图像在小波域进行特征描述,提出了视觉信息置信度方法(VIF)。刘河潮<sup>[8]</sup>根据失真帧类的型不同,提出了对单帧加权计算图像组(GoP)质量的考虑丢帧类型的质量评价算法。Seshadrinathan 等<sup>[9]</sup>通过三维 Gabor 滤波对视频进行时空分解,并计算运动矢量,提出了基于运动调谐的自然视频时空质量评价方法 MOVIE。

以上大部分研究都先计算各个视频帧的质量评价分数,然后简单合并这些值,其中,考虑丢帧类型的质量评价算法已经考虑到利用 GoP 帧结构对单帧质量进行加权,然而依然没有顾及到视频内容在时域上的各种变化对人眼观测起到的影响。因此本文在此基础上对帧结构加权方法进行改进,并考虑到视频时域变化的影响,通过改变分配给每个 GoP 质量分数的权重来进行组合以获得最终质量。首先权衡计算复杂度和计算准确度,选取合适的图像质量评价算法求得单帧图像质量分数,再利用 MPEG 编码格式视频的 GoP 结构对单帧质量分数进行加权得出 GoP 的质量分数。考虑到人眼视觉的关注特性,再利用显著性和时间复杂性对所有的 GoP 质量分数进行加权,从而得出视频质量客观评价价值。

## 2 改进的 GoP 质量分数求取

### 2.1 单帧图像质量分数

由于视频由大量帧图像组成,单帧的质量分数计算精确度和算法复杂度会影响到整段视频质量分数的计算精确度和用时,所以需要权衡单帧图像质量评价方法的准确度以及复杂度,选择出最合适的算法。本文选取经典的全参考质量评价算法结构相似性指标计算单帧图像质量分数。

### 2.2 不同类型帧失真对视频质量的影响

本文以 MPEG 编码格式的视频作为实验的对象。其编码标准的视频流中包含有 I 帧、P 帧和 B 帧三种类型的帧,具体如图 1 和图 2 所示,图 1 表示一个结构为 15:2 的 GoP,其参数 15 指 GoP 图像组的长度,参数 2 指 I 帧和 P 帧之间 B 帧的数量。图 2 表示

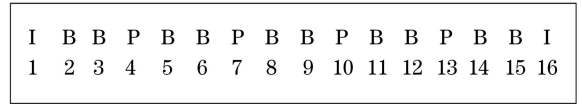


图 1 15:2 GoP 的结构

Fig. 1 Structure of a 15:2 GoP

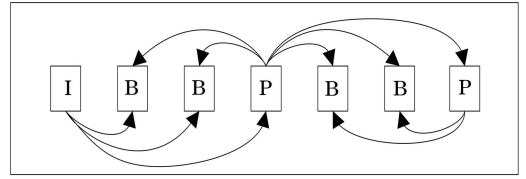


图 2 I 帧、P 帧和 B 帧的参考关系

Fig. 2 Reference relationship of I frame, P frame and B frame

同一个 GoP 中 I 帧、P 帧和 B 帧之间的参考关系。

在解码器中,I 帧的解码是参考其本身,与其他帧无关。同一个 GoP 中的其他帧解码时都直接或间接地以 I 帧为参考帧。若一个 GoP 中只有一个 I 帧,则当该帧丢失或受损时,会影响到该 GoP 中所有帧的质量。

在 P 帧中,宏块的编码方式有帧内预测和帧间预测,P 帧以其本身或同一个 GoP 中帧号排列在其前面的 I 帧或 P 帧为参考帧,如图 2 所示。如果 P 帧的帧内容信息丢失或受损,则该 P 帧解码后的图像会受到损伤,该损伤主要会影响对应 GoP 中该 P 帧与前一个 P 帧间的 B 帧以及其后所有 P 帧和 B 帧的质量。

MPEG 格式中 B 帧不会作为其他帧的参考帧,若丢包发生在 B 帧上,则其只会影响该 B 帧本身的质量。

### 2.3 优化的 GoP 质量分数

在获取整个 GoP 的质量分数时,不能对 GoP 中各帧图像质量分数进行简单的平均加权处理,而应该分配不同的权重,I 帧分配的权值最大,其他的 P 帧根据在 GoP 中位置的先后顺序(影响后面 P 帧与 B 帧的多少)来分配权重比例,因为 B 帧的图像对其他帧的图像没有影响,可以舍弃,不进行评价,减少计算量。对 GoP 中 I 帧和 P 帧质量分数加权得出整个 GoP 的质量评价分数  $Q_G$ :

$$Q_G = aQ_1 + \sum_{i=1}^n b_i Q_{P_i}, \quad (1)$$

式中  $Q_1$  和  $Q_{P_i}$  分别为采用 SSIM 方法计算出的 I 帧和 P 帧质量分数, $n$  为该 GoP 中 P 帧的数量。 $a$ 、 $b_i$  分别为 I 帧和 P 帧质量分数的权重。以 15:2 的 GoP 为例,其中 I 帧图像有 1 帧,P 帧图像有 4 帧,则这 5 帧图像的权重分别为  $a$ 、 $b_1$ 、 $b_2$ 、 $b_3$ 、 $b_4$ 。本文

根据单个 GoP 中各帧失真时所影响的帧数来分配权重。I 帧影响 15 帧, P1 帧影响 14 帧, P2 帧影响 11 帧, P3 帧影响 8 帧, P4 帧影响 5 帧, 影响的总帧数为 53 帧, 则权值  $a$ 、 $b_1$ 、 $b_2$ 、 $b_3$ 、 $b_4$  分别为 15/53、14/53、11/53、8/53、5/53。

### 3 GoP 质量分数的时域加权

考虑到视频内容在时域上的各种变化对人眼

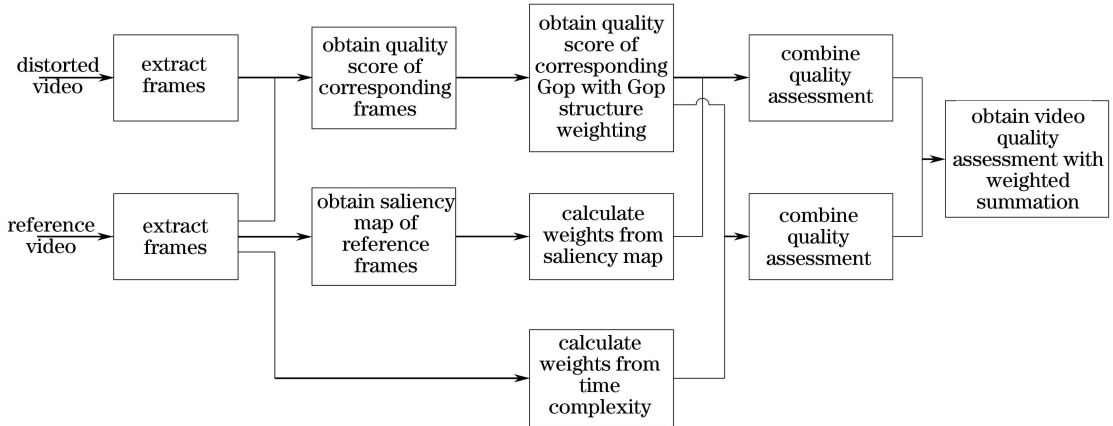


图 3 算法流程图

Fig. 3 Algorithm flow chart

#### 3.1 GoP 显著值求取

视觉显著性是独特的主观感知标准,使图像中一些物体从相邻环境中脱颖而出。自动显著性检测方法计算出视觉显著(VS)图,其给每个像素一个对应于感知显著性的值,一般给出的感知值越大则代表其像素点越显著。所以,本文利用显著图求取对应帧图像的显著值来代表图像的显著性。Zhang 等<sup>[10]</sup>提出的结合简单先验的新型显著性检测(SDSP)模型已被证明具有较高的显著性预测精度,其优点是计算复杂度低。显著性模型 SDSP 是一种简单的显著区域检测方法,通过组合频率、色彩和区域三个简单因子获得。其中,频率因子利用 Log-Gabor 滤波器模拟人眼视觉系统对显著物体进行检测;色彩因子利用人眼更易被暖色调吸引的视觉特性;区域因子考虑了人眼更有可能关注图像的中心

区域的视觉特性。SDSP 模型为

$$S_{SDSP} = S_F(x) \cdot S_C(x) \cdot S_D(x), \quad (2)$$

式中  $S_{SDSP}$  为对应的显著图,  $S_F(x)$  为频率因子,  $S_C(x)$  为色彩因子,  $S_D(x)$  为区域因子。

由于一个 GoP 中的其他帧都直接或者间接地以 I 帧为参考帧, I 帧的显著度可以近似地替代整个 GoP 的显著度,因此本方法计算参考视频中 GoP 的 I 帧图像的显著值作为整个 GoP 的显著值,大大减少了计算量。首先利用 SDSP 算法求出 I 帧图像的显著图,然后对其显著图进行二值化处理,即利用最大类间方差法得出合适的阈值,将显著图矩阵上的元素值设置为 0 或 255,得出二值化显著矩阵。最后求取二值化显著矩阵的信息熵作为 I 帧图像的显著值,即该 I 帧所在 GoP 的显著值。求取过程的图像如图 4 所示。



图 4 求取显著值过程。(a)原始视频帧;(b)视觉显著图;(c)二值化显著图

Fig. 4 Process of obtaining saliency. (a) Original video frame; (b) visual saliency map; (c) binary saliency map

信息熵描述了信源的不确定性,熵值越大,则表示显著度越大,计算公式如下:

$$S_G = S_I = -p \cdot \log_2 p, \quad (3)$$

式中  $S_G$  和  $S_I$  分别为 GoP 和 I 帧图像的显著值,  $p$  为二值化显著矩阵中出现值为 255 的元素的概率。

### 3.2 GoP 时间感知信息求取

时间感知信息可以用来衡量视频中每一帧图像的运动剧烈程度。同样,对于 GoP 而言,一个 GoP 的时间感知信息可以近似由其 I 帧图像的时间感知信息替代。考虑到视频序列第一个 GoP 的 I 帧没有前向帧,先求参考视频 I 帧与其后一帧的帧差:



$$D_n(x, y) = I_{n+1}(x, y) - I_n(x, y), \quad (4)$$

式中  $D_n(x, y)$  为帧差图像,  $I_n(x, y)$  和  $I_{n+1}(x, y)$  分别为 I 帧图像和其后一帧图像。

然后将帧差图像转化为灰度图,如图 5(b) 所示,显示出了运动中的行人、单车,计算其标准差作为 I 帧的时间感知信息。计算参考视频 GoP 中 I 帧时间感知信息:

$$T_G = T_I = \text{std}(D_{\text{Gray}}), \quad (5)$$

式中  $T_G$  和  $T_I$  分别为 GoP 和 I 帧图像的时间感知信息,  $\text{std}(D_{\text{Gray}})$  为帧差图像对应灰度图的标准差。



图 5 求取时间感知信息的过程。(a)原始图像帧;(b)帧差图像

Fig. 5 Process of obtaining time complexity. (a) Original video frame; (b) frame difference image

### 3.3 基于显著性和时间感知信息的 GoP 质量分数加权平均处理

分别利用显著值和时间感知信息对每一个 GoP 加权求平均,并对两个结果分配不同的权重。假设有  $n$  个 GoP,那么最终得到的视频序列的质量分数为

$$Q_V = \omega \cdot \frac{\sum_{i=1}^n (S_{G_i} \cdot Q_{G_i})}{\sum_{i=1}^n S_{G_i}} + (1 - \omega) \cdot \frac{\sum_{i=1}^n (T_{G_i} \cdot Q_{G_i})}{\sum_{i=1}^n T_{G_i}}, \quad (6)$$

式中  $S_{G_i}$  为第  $i$  个 GoP 的显著值,  $T_{G_i}$  为第  $i$  个 GoP 的时间感知信息,  $Q_{G_i}$  为利用(1)式计算出的第  $i$  个 GoP 的质量分数,  $\omega$  为加权系数。

## 4 实验结果与分析

为了验证本文提出的基于 GoP 的时域加权视频质量评价方法的优越性,选用了 LIVE 视频库中 MPEG 压缩的失真视频进行实验。在 LIVE 视频数据库<sup>[11-12]</sup>中,一共有 10 组视频序列,每一组又包含 16 种视频序列,其中有一个为原始视频,其余则为 15 种不同失真的视频序列,失真种类分为 MPEG 压缩编码失真(4 个视频)、无线传输失真

(4 个视频)、IP 传输失真(3 个视频),以及 H.264 压缩失真(4 个视频)。本文实验选取的是经过 MPEG 压缩失真的视频序列,共  $4 \times 10$  个视频序列,外加 10 个原始视频序列,分辨率为  $768 \times 432$ ,播放速度为  $25 \text{ frame} \cdot \text{s}^{-1}$  和  $50 \text{ frame} \cdot \text{s}^{-1}$  两种。本文共进行了 40 个视频序列的质量评价,文中的主观评价数据参考的是 LIVE 视频库中提供的主观结果。

将文中方法得出的质量分值与数据库中的视频主观评价分数(DMOS)值进行四参数拟合;另外选用性能参数皮尔森线性相关系数(PLCC)、斯皮尔曼秩相关系数(SROCC)和均方根误差(RMSE)评价算法的好坏,PLCC 和 RMSE 用于预测算法的准确性,SROCC 用于预测算法的单调性。PLCC 与 SROCC 的范围值是  $0 \sim 1$  之间,得到的值越大表明与主观结果越接近;RMSE 得到的值越小表明预测结果越准确。在确定(6)式中的  $\omega$  值时,遍历值为  $0 \sim 1$  的多个权重  $\omega = \{0, 0.01, 0.02, 0.03, \dots, 0.98, 0.99, 1.00\}$ ,实验中发现  $\omega = 0.23$  时,算法性能最佳,因此选取  $\omega = 0.23$  作为公式中的权重。

选取了不同的视频质量评价方法对上述 LIVE 视频库进行评价,通过 PLCC、SROCC 和 RMSE 这三种性能参数对得到的评价结果进行对比,同时以每处理 30 帧的时间消耗来描述各模型的计算复杂



性。选取现有经典算法与本文算法进行对比,包括 PSNR、SSIM、VIF、基于小波模型的自然图像视觉信噪比(VSNR)<sup>[13]</sup>、MOVIE、GoP+SSIM(即所有 GoP 质量分数的平均值)和本文方法。结果如表 1 所示。

由表 1 可知,本文方法的 PLCC 和 SROCC 值都比其他方法高,且 RMSE 结果均低于其他算法,说明本文方法比其他算法有更高的预测准确性和更优的预测单调性。其中,GoP+SSIM 方法的准确性和单调性较 SSIM 算法也有较大的提升,可见本文提出的改进的 GoP 帧结构加权方法有效地提高了 SSIM 算法的性能。同时,本文方法经过对 GoP+SSIM 方法进行时域加权处理,其性能也有了较大的提升,说明利用显著性和时间感知信息对 GoP 时域加权较好地提高了算法的性能。

从参数计算复杂性来看,本文方法的计算复杂性较低,由于省略了对 B 帧的评价,计算耗时少于未加权的 SSIM 算法,且远少于使用光流场进行运

动估计的 MOVIE 方法,可见,本文方法在具有较高评价准确性的同时也有较低的复杂度。

表 1 本文方法与其他方法在 LIVE 视频库上的实验结果

Table 1 Performances of the proposed method and the existing methods in LIVE video database

Method	PLCC	SROCC	RMSE	Complexity /ms
PSNR	0.3887	0.3921	8.7847	407.7
SSIM	0.5871	0.5771	7.7181	2173.6
VSNR	0.5960	0.5816	7.6943	5259.9
VIF	0.6504	0.6197	7.2419	3393.4
MOVIE	0.6579	0.6532	7.1531	6934.5
GoP+SSIM	0.6359	0.6427	7.3167	612.2
Proposed method	0.6665	0.6827	7.1076	1069.97

图 6 为 SSIM 与本文方法质量评价结果与主观评价结果的拟合图。分析可知,经过对 SSIM 算法的时域加权,本文算法的拟合图中散点分布明显更为集中,更加贴近于拟合曲线。本文算法有着更好的拟合效果。

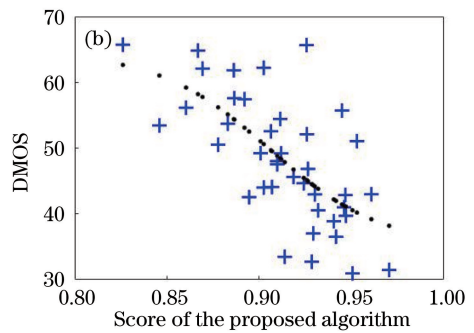
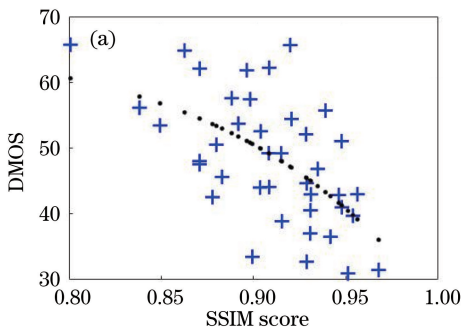


图 6 本文算法与 SSIM 算法在 LIVE 视频库上的拟合曲线图。(a) SSIM 与 DMOS 的拟合图;  
(b) 本文方法与 DMOS 的拟合图

Fig. 6 Fitting curves of the proposed algorithm and SSIM on LIVE video database. (a) Fitting plots of SSIM and DMOS;  
(b) fitting plots of proposed method and DMOS

## 5 结 论

利用 GoP 的帧结构特性对单帧质量分数进行加权,并使用人眼视觉系统时域与空域的特征值对 GoP 质量分数进行加权求和,提出了一种基于时域 GoP 加权的视频质量评价算法。为了评估本文算法的性能,采用 LIVE 数据库中测试视频验证了算法的单调性和准确性。实验结果表明,本文算法能够较准确地评估不同失真类型的视频。其性能优于一些主流的算法且比直接求取 GoP 质量分数平均值有更好的效果,可见对 GoP 质量分数进行显著性和时间感知信息加权有效地提高了算法的性能。本文方法较好的克服了不同帧图像质量分数在人眼视觉系统中的有着不同重要性的问题。

## 参 考 文 献

- [1] Xue X B, Yu M, He M L. Stereoscopic image-quality-assessment method based on visual cell model [J]. Laser & Optoelectronics Progress, 2016, 53(4): 041004.  
薛小波, 郁梅, 何美伶. 基于仿视觉细胞模型的立体图像质量评价方法[J]. 激光与光电子学进展, 2016, 53(4): 041004.
- [2] Zhang Y, Jin W Q. Assessment method of fusion image quality in wavelet domain structural similarity [J]. Chinese Journal of Lasers, 2012, 39(s1): s109007.  
张勇, 金伟其. 小波域结构相似度融合图像质量评价方法[J]. 中国激光, 2012, 39(s1): s109007.
- [3] Zhao W Z, Qin S Y. Image quality assessment and

- some solving approaches to current issues[J]. *Laser & Optoelectronics Progress*, 2010, 47(4): 041002.
- 赵文哲, 秦世引. 图像质量评价的研究进展和若干问题的解决途径[J]. *激光与光电子学进展*, 2010, 47(4): 041002.
- [4] Wang Z, Bovik A C. A universal image quality index [J]. *IEEE Signal Processing Letters*, 2002, 9(3): 81-84.
- [5] Wang Z, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity [J]. *IEEE Transaction on Image Processing*, 2004, 13(4): 600-612.
- [6] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment [C]//*Proceedings of Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, 2004, 2: 1398-1402.
- [7] Sheikh H R, Bovik A C. Image information and visual quality [J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 430-444.
- [8] Liu H C. No reference quality assessment of videos affected by packet losses [D]. Xi'an: Xidian University, 2013: 51-70.
- 刘河潮. 网络丢包的无参考视频质量评估方法研究 [D]. 西安: 西安电子科技大学, 2013: 51-70.
- [9] Seshadrinathan K, Bovik A C. Motion tuned spatio-temporal quality assessment of natural videos [J]. *IEEE Transactions on Image Processing*, 2010, 19(2): 335-350.
- [10] Zhang L, Gu Z Y, Li H Y. SDSP: a novel saliency detection method by combining simple priors [C]//*Proceedings of IEEE International Conference on Image Processing*, 2013: 171-175.
- [11] Seshadrinathan K, Soundararajan R, Bovik A C, *et al.* Study of subjective and objective quality assessment of video[J]. *IEEE Transactions on Image Processing*, 2010, 19(6): 1427-1441.
- [12] Seshadrinathan K, Soundararajan R, Bovik A C, *et al.* A subjective study to evaluate video quality assessment algorithms [C]. *SPIE*, 2010, 7527: 75270H.
- [13] Chandler D M, Hemami S S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images [J]. *IEEE Transactions on Image Processing*, 2007, 16(9): 2284-2298.