

基于支持向量回归的无参考 MS-SSIM 视频质量评价模型

卓 力, 张美娜, 王贯瑶, 李嘉锋

(北京工业大学计算智能与智能系统北京市重点实验室, 北京 100124)

摘要: 多尺度结构相似度 (multi-scale structural similarity, MS-SSIM) 是一种常用的全参考视频质量评价准则, 由于评价时需要原始视频作为参考, 因此无法用于实时的网络视频质量评价中, 故提出一种基于 H.264 码流的无参考 MS-SSIM 视频质量评价模型. 该模型从 H.264 码流中提取出 I 帧和 P 帧的编码模式、运动矢量等参数, 然后对这些参数进行统计分析, 来表征视频的纹理丰富程度和运动剧烈与复杂程度; 结合量化参数等信息构成码流特征参数集, 使用支持向量回归 (support vector regression, SVR) 方法建立码流特征参数和 MS-SSIM 之间的映射关系模型, 用于预测 H.264 码流的 MS-SSIM 视频质量度量. 该模型只使用从 H.264 码流中提取的编码参数, 无须原始的参考视频, 也无须对视频进行解码. 与现有的无参考码流预测模型相比, 该模型可以获得更高的预测精度.

关键词: H.264; 码流参数; 无参考; 支持向量回归; 多尺度结构相似度

中图分类号: TP 391

文献标志码: A

文章编号: 0254-0037(2018)12-1486-08

doi: 10.11936/bjtxb2018020028

No-reference MS-SSIM Video Quality Assessment Model Based on Support Vector Regression

ZHUO Li, ZHANG Meina, WANG Guanyao, LI Jiafeng

(Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology,
Beijing 100124, China)

Abstract: Multi-scale structural similarity (MS-SSIM) is a commonly used full-reference quality assessment metric. Since the original video is required for reference, it is not suitable to be applied in real-time assessment of network video quality. In this paper, a no-reference MS-SSIM video quality assessment model based on H.264 bitstream was proposed. First, I-frame and P-frame encoding mode and motion vector parameters from H.264 bitstream were extracted from H.264 bitstream, which then were statistically analyzed to characterize the richness of the texture and the intensity and complexity of the motion of the video. Second, the parameters were combined with quantization parameter to form the bitstream feature parameter set, and support vector regression (SVR) was finally applied to the relationship between the bitstream feature parameters and MS-SSIM to predict the video quality metric of MS-SSIM for H.264 bitstream. The proposed model only uses the parameters extracted from the video bitstream and does not demand to decode the video bitstream completely. Compared with a state-of-the-art no-reference bitstream prediction model, the proposed model can achieve higher prediction accuracy.

收稿日期: 2018-02-25

基金项目: 国家自然科学基金资助项目(61531006)

作者简介: 卓 力(1971—), 女, 教授, 主要从事图像/视频的编码与传输、多媒体大数据处理方面的研究, E-mail: zhuoli@bjut.edu.cn

Key words: H. 264; bitstream parameters; no-reference; support vector regression (SVR); multi-scale structural similarity (MS-SSIM)

随着网络 and 多媒体技术的不断发展,实时视频传输越来越普及. 对于视频内容提供商来说,如何实时、精确地测量传输的视频质量就成为保障视频服务质量的关键.

根据所参考的比对信息,视频质量评价方法可以分为全参考(full-reference, FR)质量评价^[1]、部分参考(reduced-reference, RR)质量评价^[2]和无参考(no-reference, NR)质量评价^[3]3种方式. FR质量评价以原始图像作为参考,待评价图像的质量通过与原始图像之间的误差来进行评估. 均方误差(mean square error, MSE)和峰值信噪比(peak signal noise ratio, PSNR)是最常用的全参考质量评价方法,但是两者与人类的视觉感知没有很好的相关性. Wang等^[4]考虑人眼视觉系统(human visual system, HVS)的特点,提出了结构相似性(structural similarity, SSIM),它与主观质量感知的相关性优于PSNR和MSE,因此被广泛用作全参考质量评价方法. RR在评估过程中,同样需要原始图像作为参考,但是在比较的过程中,使用的是提取的图像特征信息,而不是完整的图像信息. 在使用FR和RR评估方法进行评估的过程中,均需要将原始参考视频同时进行传送,不仅需要消耗大量的网络资源,而且无法满足实时性的需求,因此FR和RR往往无法满足对实时性有严格要求、带宽受限的应用需求.

NR质量评价方法无须以原始图像作为参考,因此实时性好,也无须消耗额外的网络资源,是目前主流的实时视频质量评价方法. NR质量评价可以进一步分为基于像素域(pixel-based)的方法、基于码流(bitstream-based)的方法以及混合方法3种. 其中基于像素域的方法利用解码后的视频(在空间域和/或其他域)进行质量评价;而基于码流的方法则无须解码,直接从视频的压缩码流中提取参数进行质量评价;混合方法则是将二者结合起来进行质量评价. 文献[5]利用解码像素的系数来分析和估计视频质量. 该方法可以区分2种类型(H. 264和MPEG-2)的视频,估计I帧中使用的量化级别,并利用这些信息来评估视频质量. 文献[6]从H. 264压缩码流中提取量化参数(quantization parameter, QP)、运动矢量在 x 和 y 方向上的最大最小值、编码比特率、网络适应性(network adaptation layer, NAL)

单元大小等参数,从像素域提取灰度共生矩阵的平方和、垂直灰度梯度、图像熵、对比度等作为参数,将这些参数结合起来形成特征参数,使用线性回归的方法建立特征参数与视频质量之间的关系模型. 文献[7]提取H. 264码流中的比特率、帧数、场景复杂度、视频运动等参数,结合从完全解码后的像素域中选取模糊、块效应、运动强度和噪声功率组成特征参数,输入到受限玻尔兹曼机中进行训练,建立视频质量度量(video quality metric, VQM)质量评价指标与码流特征参数之间的映射模型. 深度模型往往比较复杂,需要更多的计算资源和时间,实时性较差.

总的来说,基于像素域的方法和混合方法,都需要对视频进行完全解码才能提取参数,无法满足实时性要求严格的应用需求. 为此,很多学者开展了基于码流的视频质量评价方法研究,即无须解码的情况下直接估计码流对应的视频质量. 文献[8]提出了一种基于H. 264的无参考视频质量评价方法. 该方法将PSNR作为视频质量度量指标,从码流中提取的预测残差的变换系数来估计PSNR. 文献[9]从H. 264码流中提取量化参数和预测残差的整数余弦变换(integer cosine transform, ICT)系数方差来估计客观质量PSNR,以此表示压缩失真视频质量. 文献[10]从H. 264码流中选取QP、运动矢量 x 方向和 y 方向上的平均值和方差、子块高度的均值和方差、子块宽度的均值和方差作为特征参数,利用支撑向量回归(support vector regression, SVR)来预测视频质量度量指标SSIM. 上述方法所使用的码流参数未能很好地表征视频的纹理复杂程度和运动复杂程度,因此模型的预测精度难以令人满意.

本文提出了一种基于H. 264码流的无参考多尺度SSIM(multi-scale SSIM, MS-SSIM)视频质量评价模型,仅使用码流中提取的参数来实现视频质量评价. 本文的主要贡献是:1)使用运动矢量的模大小和方向统计直方图来表征视频的运动剧烈和复杂程度. 本文将运动矢量的方向以 30° 为间隔划分成12个区间,统计运动矢量方向直方图. 运动矢量方向越丰富,则表明视频的运动越复杂. 2)综合使用I帧中每个 4×4 预测残差块的非零ICT系数个数、I帧和P帧中帧内 4×4 块与 16×16 块所占的比例、skip编码宏块所占的比例等来表征视频的纹理丰富

程度. 研究表明,采用本文提出的纹理丰富程度、运动剧烈与复杂程度表征方法,可以获得高精度的 MS-SSIM 质量评价结果.

1 视频质量度量指标 MS-SSIM 的评估

本文提出的无参考 MS-SSIM 视频质量评价模型的建立过程如图 1 所示,共包括模型训练和模型预测 2 个部分. 在模型训练部分,首先构建训练所需的数据样本集,每个样本包括 H. 264 视频码流及其对应的 MS-SSIM 值;然后从 H. 264 码流中提取出 I 帧编码模式、运动矢量等参数,对这些参数进行统计分析来表征视频的纹理丰富程度和运动剧烈程度,结合量化参数构成码流特征参数集;最后利用 SVR 建立码流特征参数与该码流对应的 MS-SSIM 值之间的映射关系,得到质量评估模型. 在模型预测部分,将从输入的视频码流中提取的码流特征参数输入到训练阶段得到的映射模型,模型的输出则是 MS-SSIM 预测值.

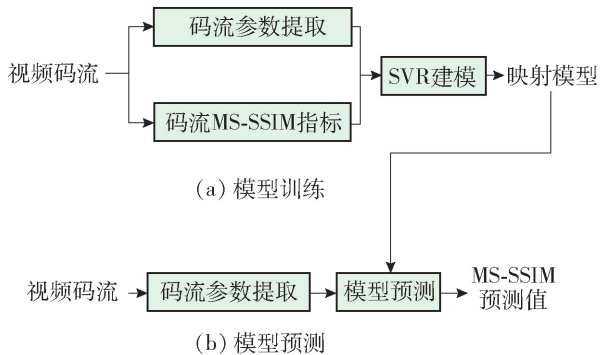


图 1 本文提出的质量评估模型建立框图

Fig. 1 Block diagram of the proposed quality assessment model

可以看出,整个建模过程包括 3 个核心部分:码流参数的提取与统计分析、MS-SSIM 评价指标的确定和 SVR 映射模型的建立. 下面分别详细介绍这 3 个部分的实现过程.

1.1 码流参数提取与统计分析

根据人眼的视觉特性,人眼对纹理丰富程度和运动剧烈程度不同的视频序列具有不同的主观视觉感受. 因此,如果只是片面地考虑单一的视频质量损伤因素,而不考虑视频本身的特性,显然无法获得与人眼感知相符的客观评价结果. 为了保证针对压缩码流进行的客观质量评估兼具全面性以及准确性,本文按以下原则提取码流参数,并对其进行统计分析:1) 选择最能体现视频压缩编码损伤程度的参

数;2) 选择与视频序列自身的纹理丰富程度和运动复杂度有关的参数.

根据以上原则,本文首先提取了以下的 H. 264 码流参数.

1) QP:量化是视频编码过程中引入视频质量损伤的唯一来源. QP 用来直接控制视频输出比特率和视频质量损伤程度. QP 越大,视频失真程度越大,用户的主观感受越差. 量化参数可以为视频质量评估提供直接的依据.

2) skip 编码宏块数:H. 264 采用的帧间预测编码技术支持在图像的平坦区域使用 skip 宏块技术,这些宏块不需要进行编码. skip 宏块的数目则能在一定程度上反映视频的纹理复杂度以及帧间运动程度,有利于进行视频质量评估.

3) I 帧 4×4 编码块中非零 ICT 系数的个数: ICT 系数反映了预测残差的能量, ICT 系数中的非零个数越多,则表明该编码块所对应的区域细节纹理越丰富.

4) 4×4 帧内预测模式的块数量: 4×4 帧内预测模式适用于带有大量细节的区域,采用 4×4 帧内预测模式的块越多,说明视频中包含的细节纹理越丰富.

5) 16×16 帧内预测模式的块数量: 16×16 帧内预测模式适用于平坦区域,采用 16×16 帧内预测模式的块越多,说明视频中包含的平坦区域越多.

6) 运动矢量:运动矢量大小可以直接刻画运动的剧烈程度;而运动矢量的方向越丰富,则表明视频的运动复杂程度越高.

接下来,本文将上述码流参数进行统计分析,用于表示视频的纹理丰富程度和运动复杂程度. 具体包括以下几个方面.

1) Ave_NC_I: I 帧中平均每个 4×4 块包含的 ICT 变换系数非零个数

$$\text{Ave_NC_I} = \frac{\sum_B \text{CoffeToken}}{B} \quad (1)$$

式中: CoffeToken 为 4×4 块中非零系数的个数; B 为 I 帧中包含的 4×4 块数目.

2) $I_{\text{ratio}_{16 \times 16}}$: I 帧中采用 16×16 编码模式的宏块所占的比例

$$I_{\text{ratio}_{16 \times 16}} = \frac{I_{L_{16 \times 16}}}{I_{L_{MB}}} \quad (2)$$

式中: $I_{L_{16 \times 16}}$ 为采用 16×16 帧内预测模式的块的数量; $I_{L_{MB}}$ 为 I 帧的宏块数.

3) $I_{\text{ratio_I}_{4 \times 4}}$: I 帧中采用 4×4 编码模式的块所占的比例

$$I_{\text{ratio_I}_{4 \times 4}} = \frac{I_{\text{I}_{4 \times 4}}}{I_{\text{I_MB}}} \quad (3)$$

式中: $I_{\text{I}_{4 \times 4}}$ 为采用 4×4 帧内预测模式的块的数量; $I_{\text{I_MB}}$ 为 I 帧的宏块数。

4) $I_{\text{ratio_P_skip}}$: skip 编码宏块所占的比例

$$I_{\text{ratio_P_skip}} = \frac{\sum_{i=1}^{14} I_{i_P_skip}}{\sum_{i=1}^{14} I_{i_P_MB}} \quad (4)$$

式中: $I_{i_P_skip}$ 为第 i 个 P 帧中 skip 宏块的数目; $I_{i_P_MB}$ 为第 i 个 P 帧中宏块数目。

5) $I_{\text{ratio_P}_{4 \times 4}}$: P 帧中采用 4×4 帧内编码模式的块所占的比例

$$I_{\text{ratio_P}_{4 \times 4}} = \frac{I_{\text{I}_{4 \times 4}}}{I_{\text{P_MB}}} \quad (5)$$

式中: $I_{\text{I}_{4 \times 4}}$ 为采用 4×4 帧内预测模式的块的数量; $I_{\text{P_MB}}$ 为 P 帧的宏块数。

6) $I_{\text{ratio_P}_{16 \times 16}}$: P 帧中采用 16×16 帧内编码模式的宏块所占的比例

$$I_{\text{ratio_P}_{16 \times 16}} = \frac{I_{\text{I}_{16 \times 16}}}{I_{\text{P_MB}}} \quad (6)$$

式中: $I_{\text{I}_{16 \times 16}}$ 为采用 16×16 帧内预测模式的块的数量; $I_{\text{P_MB}}$ 为 P 帧的宏块数。

7) 运动的复杂程度: 对于提取出来的运动矢量, 本文定义了运动方向直方图来表示运动的复杂程度。

根据

$$f_x = \arctan \frac{MV_y}{MV_x} \quad (7)$$

对于提取出来的运动矢量 (MV_x , MV_y) 进行计算。式中 f_x 表示运动矢量的方向, 取值范围为 $[-\pi, \pi]$ 。

以 30° 为间隔将 360° 划分成 12 个区间, 然后统计各帧的运动矢量方向在各个区间出现的次数, 有

$$I_{MV_i} = \frac{MV_i}{\sum_{i=1}^{12} MV_i} \quad (8)$$

式中: I_{MV_i} 表示在第 i 个方向区间运动矢量出现次数占所有运动方向的百分比; MV_i 表示在第 i 个方向区间运动矢量出现的次数。

接下来, 按照式 (9) ~ (11) 计算得到各个方向的运动出现次数的平均值 Ave_{MV} 、标准差 Std_{MV} 和运动矢量模长的平均值 Ave_Len_{MV} :

$$Ave_{MV} = \frac{\sum_{m=1}^M \sum_{i=1}^I MV_{i,m}}{M \times I} \quad (9)$$

$$Std_{MV} = \sqrt{\frac{1}{M \times I} \sum_{m=1}^M \sum_{i=1}^I (MV_{i,m} - Ave_{MV})^2} \quad (10)$$

$$Ave_Len_{MV} =$$

$$\frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N \sqrt{MV_{(x,n,m)}^2 + MV_{(y,n,m)}^2} \quad (11)$$

式中: M 为 GOP 的长度, 本文设备为 15; I 为方向区间编号, 本文设置为 12; $MV_{i,m}$ 为第 m 帧中第 i 个区间内运动矢量的数目; N 为第 m 帧内宏块总数; $MV_{(x,n,m)}$ 和 $MV_{(y,n,m)}$ 分别表示第 m 帧中第 n 个宏块水平方向和垂直方向上 MV 的值。

对上述的统计结果分别进行归一化处理, 级联后得到码流的特征参数矢量, 共计 22 维, 具体包括:

1) 量化参数 QP。

2) I 帧中平均每个 4×4 块包含的 ICT 系数非零个数。

3) I 帧帧内 4×4 块所占的比例。

4) I 帧帧内 16×16 块所占的比例。

5) skip 编码宏块所占的比例。

6) P 帧帧内 4×4 块所占的比例。

7) P 帧帧内 16×16 块所占的比例。

8) ~ 19) 12 个方向的运动矢量出现的百分比。

20) 各个方向的运动矢量出现次数平均值。

21) 各个方向的运动矢量出现次数的标准差。

22) MV 模长的平均值。

其中, 2) ~ 7) 表征了视频纹理丰富程度, 8) ~ 22) 表征了视频的运动剧烈和复杂程度。

这些特征均是在帧级别上进行计算的, 而每个特征的平均值均是以一个图像组 (group-of-pictures, GOP) 为单元计算得到的。

1.2 视频质量度量 MS-SSIM

SSIM 是 Wang 等^[4]提出的一种客观视频质量评价方法。SSIM 考虑到图像的结构信息在人的感知上的模糊变化, 分别从亮度、对比度、结构 3 个方面度量图像相似性, 与 PSNR、MSE 等评价指标相比, 更符合人眼视觉感知, 因此被广泛应用于视频质量评价中。SSIM 方法中亮度、对比度、结构相似性的定义分别为

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (12)$$

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (13)$$

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \quad (14)$$

SSIM 则定义为

$$\text{SSIM}(X, Y) = l(X, Y)c(X, Y)s(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (15)$$

式中: μ_X, μ_Y 分别表示图像 X 和 Y 的均值; σ_X, σ_Y 分别表示图像 X 和 Y 的标准差; σ_{XY} 代表图像 X 和 Y 协方差, C_1, C_2 和 C_3 为常数. 通常取 $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, $C_3 = \frac{C_2}{2}$, 一般地, $K_1 = 0.01, K_2 = 0.03$, $L = 255$ (L 指的是像素的取值范围, 一般取 255).

SSIM 是一种单尺度的图像质量评价方法, 它对一定尺度的图像有很好的评价效果, 但它的缺点是不能适应多种尺度的变化, 缺乏灵活性. 为此, Wang 等^[11]对 SSIM 进行了扩展, 提出了 MS-SSIM, 利用迭代方法对图像进行低通滤波和降采样, 在每一个尺度上分别计算 SSIM 指标, 综合后作为最终的评价结果. MS-SSIM 在每一个尺度 i 上都会进行对比度比较和结构比较, 分别计算 $C_i(X, Y)$ 和 $S_i(X, Y)$, 亮度比较则只在最终的尺度 M 上进行计算. MS-SSIM 的计算公式为

$$\text{MS-SSIM}(X, Y) = [I_m(X, Y)]^{\alpha_m} \prod_{i=1}^M [C_i(X, Y)]^{\beta_i} [S_i(X, Y)]^{\gamma_i} \quad (16)$$

通常取 $M=5, \alpha_m = \beta_i = \gamma_i, \sum_{i=1}^M \gamma_i = 1$.

研究表明, 在进行视频质量评价时, MS-SSIM 比 SSIM 更符合人眼的主观感受, 所以本文采用 MS-SSIM 作为视频质量的评价准则.

1.3 SVR 预测模型

SVR 是 Drucker 等^[12]在 1996 年提出的一种机器学习方法, 该方法能够将预测特征映射到高维空间, 并利用它们进行预测. 本文采用 ϵ -SVR 来建立码流特征参数与 MS-SSIM 之间的映射模型^[13].

假设有样本 $\{(X_j, y_j)\}$, 其中 $j = 1, 2, \dots, M$, M 为样本的个数, X_j 为是 $n \times 1$ 的特征向量 ($n = 22$, 如 1.1 节所示), y_j 为第 j 个样本相对应的 MS-SSIM.

对于给定的 $C > 0$ 和 $\epsilon > 0$, SVR 优化问题可以

表述为

$$\begin{aligned} \min_{W, b, \xi, \xi^*} & \frac{1}{2} W^T W + C \sum_{j=1}^M \xi_j + C \sum_{j=1}^M \xi_j^* \\ \text{s. t: } & W^T \phi(X_j) + b - y_j \leq \epsilon + \xi_j \\ & y_j - W^T \phi(X_j) - b \leq \epsilon + \xi_j^* \\ & \xi_j, \xi_j^* \geq 0, j = 1, 2, \dots, M \end{aligned} \quad (17)$$

其中 $\phi(X_j)$ 将 X_j 映射到更高维度的空间. 本文使用的核函数是径向基函数 (radial basis function, RBF), 有

$$K(x_j, x_k) \equiv \phi(X_j)^T \phi(X_k) = \exp(-\gamma \|X_j - X_k\|^2), \gamma > 0 \quad (18)$$

将上述问题转换成其对偶问题, β 是其最优解, 使用 β 来预测 MS-SSIM 值. 对于输入的码流特征参数 X , MS-SSIM 预测值的计算公式为

$$\hat{\text{MS-SSIM}} = y(X) = \sum_{j=1}^M \beta_j K(X_j, X) + b \quad (19)$$

2 实验结果与分析

为了验证所提出的视频质量评价模型的准确性, 本文利用公共的视频测试数据集进行了实验.

2.1 视频测试序列

本文共选取了 15 个不同分辨率的视频序列, 这些视频的详细信息如表 1 所示, 其中空间感知信息 (spatial perceptual information, SI) 表示空间感知信息, 用于表征视频序列空间细节量, SI 值越高, 空间场景越复杂; 时间感知信息 (temporal perceptual information, TI) 表示时间感知信息, 用于表征视频序列的时间变化量, TI 值越高, 则运动程度越高. 图 2 所示的是部分视频帧, 这些序列具有不同的纹理丰富程度和运动剧烈程度. 本文使用 H. 264/AVC 标准的 JM8.6^[14] 参考软件按照表 2 的参数设置进行编码, 得到压缩码流.

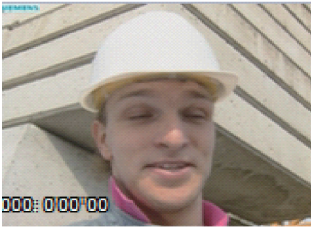
2.2 模型预测性能分析

本文选用皮尔森线性相关系数 (Pearson's linear correlation coefficient, PLCC)、斯皮尔曼秩相关系数 (Spearman rank-order correlation coefficient, SROCC)、MSE 作为衡量预测模型准确性的评价指标. PLCC 和 SROCC 能说明评估方法的预测单调性, 取值范围 $[0, 1]$, 其值越大越好; MSE 则能说明评估模型的预测准确性, 其值越小越好. 具体计算公式为

$$\text{SROCC} = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{S_i - \bar{S}}{D_i} \right) \left(\frac{(S_{pi} - \bar{S}_P)}{D_{pi}} \right) \quad (20)$$

表1 视频序列信息
Table 1 Video sequence information

序号	名称	分辨率	帧率	SI	TI
1	Foreman	176 × 144	30	105.66	35.89
2	Hall	176 × 144	30	127.09	8.27
3	Salesman	176 × 144	30	76.23	6.39
4	Akiyo	176 × 144	30	79.39	5.18
5	Highway	352 × 288	30	62.95	14.81
6	Mobile	352 × 288	30	173.51	33.03
7	Coastguard	352 × 288	30	121.14	34.91
8	Bus	352 × 288	30	150.63	38.33
9	Flower vase	416 × 240	30	104.14	5.14
10	RaceHorses	416 × 240	30	103.52	34.70
11	BasketballPass	416 × 240	50	83.54	30.08
12	BQSquare	416 × 240	60	160.76	17.93
13	PartyScene	832 × 240	50	105.44	17.95
14	BasketballDrill	832 × 240	50	76.84	21.29
15	BQMall	832 × 240	60	109.34	33.65



Foreman



Mobile



Flower vase



BasketballDrill



Coastguard



BQMall

图2 部分视频帧

Fig. 2 Some of the video frame

表2 H. 264 编码参数设置

Table 2 H. 264 encoding parameter settings

编码参数	数值或类型
ProfileIDC	基本档
LevelIDC	50
GOP 长度	15
GOP 结构	IPPP
帧率	30/50/60
QP	24,30,36,39,42,48,51
编码比特率	可变

$$PLCC = \frac{\sum_{i=1}^N (S_i - \bar{S})(S_{pi} - \bar{S}_p)}{\sqrt{\sum_{i=1}^N (S_i - \bar{S})^2 \sum_{i=1}^N (S_{pi} - \bar{S}_p)^2}} \quad (21)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (S_i - S_{pi})^2 \quad (22)$$

式中: S_i, S_{pi} 分别为实际测量值和预测值; \bar{S}, \bar{S}_p 分别为实际测量值和预测值的平均值; D_i, D_{pi} 分别为实际测量值和预测值之间的标准差; N 为样本个数。

本文选取表1中的15个视频序列用于训练模型,每个序列分别使用7个不同的QP值进行编码,共得到105个码流。因为码流特征参数是以GOP为

单元计算的,所以最终共使用了 1 103 个 GOP 数据用于训练模型. 剩余的 3 个视频序列共包括 210 个 GOP,用于验证该模型的准确性.

为了验证本文模型的性能,将提出的模型与 Lin 等^[10]提出的模型进行了对比. 该模型采用平均子块宽度、平均子块高度、子块宽度方差、子块高度方差、 x 方向运动矢量的均值和方差、 y 方向运动矢量的均值和方差,以及量化参数(QP)等 9 个特征,使用 SVR 建立特征参数和 SSIM 之间的映射关系来预测视频质量度量 SSIM.

图 3 所示的分别是采用 Lin 等^[10]提出的模型和本文模型对 MS-SSIM 值进行预测的结果对比. 从图 3 可以看出,本文提出方法的预测曲线与真实值曲线更加吻合. 也就是说,采用本文方法可以得到更精确的预测结果.

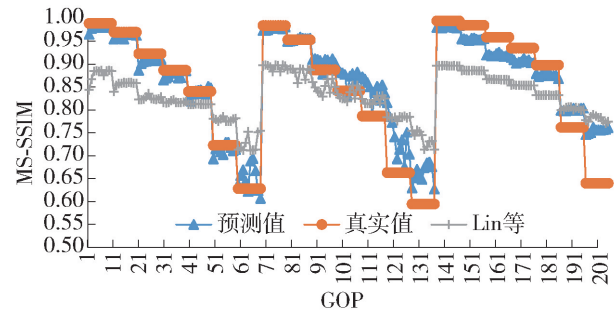


图 3 Lin 等提出的方法、本文方法对 MS-SSIM 的预测值和真实值之间的对比

Fig. 3 MS-SSIM comparison of the true values and the prediction values obtained by the methods proposed by Lin et al and in this paper, respectively

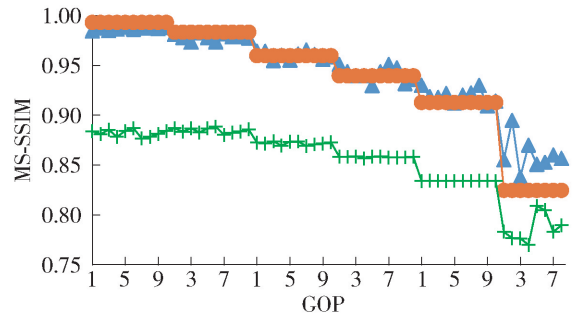
表 3 所示的是本文方法的预测性能和 Lin 等^[10]所提方法预测性能的对比结果. 可以看出,本文提出的模型得到的评估结果有如下特点:1) 具有较高的 PLCC 和 SROCC 值,这说明该模型具有较好的预测单调性;2) MSE 值较小,说明该模型具有较高的预测准确性.

表 3 本文方法的预测性能和 Lin 等提出方法预测性能的对比结果

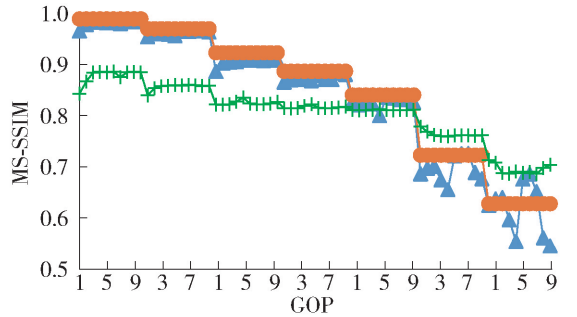
Table 3 Prediction performance comparison results of the methods proposed by Lin et al and in this paper

方法	PLCC	SROCC	MSE
Lin 等 ^[10]	0.876 7	0.931 3	0.007 1
本文	0.901 6	0.945 2	0.006 5

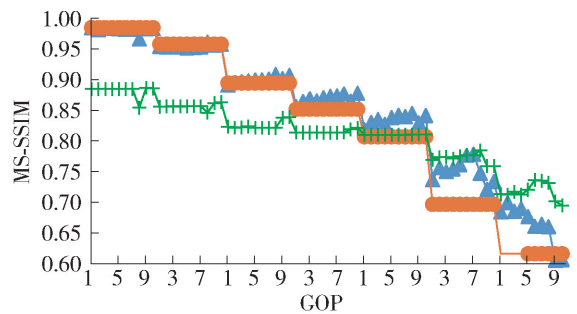
图 4 所示的分别是 Akiyo、Bus、Racehorse、BasketballDrill 等 4 个具有不同纹理复杂程度和运



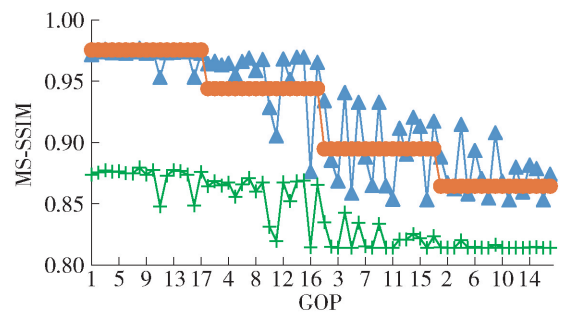
(a) Akiyo | 176x144



(b) Bus | 352x288



(c) Racehorse | 416x240



(d) BasketballDrill | 832x480

图 4 MS-SSIM 实际测量值与模型预测值之间的对比
Fig. 4 Comparison of the actual measurements of the MS-SSIM and the model prediction

动剧烈程度的视频序列,其实际测量 MS-SSIM 值与模型预测值之间的对比结果. 从图 4 可以看出,对于具有不同纹理复杂程度和运动剧烈程度的视频序列,本文模型获得的 MS-SSIM 预测值都可以与实际测量值更好地吻合.

表 4 所示的是本文方法的时间复杂度和 Lin 等^[10]提出方法时间复杂度的对比. 本文的仿真环

境是3.30 GHz的CPU、4.00 GB内存、i5处理器、64位操作系统的PC机和版本为R2014a的MATLAB。从表4可以看出,虽然本文提出模型的特征数目比Lin等^[10]提出模型的数量多,但是本文提出方法的时间复杂度和Lin等提出方法的时间复杂度大致相同。由此可见,本文的模型能准确地预测视频的MS-SSIM,同时时间复杂度并不高。

表4 本文方法的时间复杂度和Lin等提出方法时间复杂度的对比

Table 4 Time complexity comparison of the methods proposed by Lin et al and in this paper

方法	特征数目	时间复杂度/s
Lin等 ^[10]	9	0.60
本文	24	0.62

3 结论

1) 本文提出了一种基于H.264的无参考MS-SSIM评价模型,该模型从码流中选取可以表征视频纹理信息、运动信息的参数,通过支持向量回归的方法建立码流参数与MS-SSIM之间的映射模型。该模型能有效地利用H.264码流直接实时地预测MS-SSIM指标,无须对码流进行解码。

2) 接下来的工作中,可以使用其他的视频质量度量指标如VQM和视频质量感知评估(perceptual evaluation of video quality,PEVQ)进行无参考视频质量评价方法的研究。同时,也可以通过采用更多不同的参数来提高视频质量的预测性能。

参考文献:

[1] ITU-T. Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference: J 341 [S/OL]. [2018-01-31]. <https://www.itu.int/en/ITU-T/publications/Pages/structure.aspx>.

[2] ITU-T. Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal: J 342 [S/OL]. 2011. [2018-01-31]. <https://www.itu.int/en/ITU-T/publications/Pages/structure.aspx>.

[3] NACCARI M, TAGLIASACCHI M, TUBARO S. No-reference video quality monitoring for H. 264/AVC coded video [J]. IEEE Transactions on Multimedia, 2009, 11

(5): 932-946.

[4] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2004, 13(4): 600.

[5] SØGAARD J, FORCHHAMMER S, KORHONEN J. No-reference video quality assessment using codec analysis [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2015, 25(10): 1637-1650.

[6] WEI B W, ZHANG Y. No-reference video quality assessment with frame-level hybrid parameters for mobile video services [C] // 2016 2nd IEEE International Conference on Computer and Communications (ICCC). Piscataway: IEEE, 2016: 490-494.

[7] VEGA M T, MOCANU D C, FAMAHEY J, et al. Deep learning for quality assessment in live video streaming[J]. IEEE Signal Processing Letters, 2017, 24(6): 736-740.

[8] EDEN A. No-reference estimation of the coding PSNR for H. 264-coded sequences [J]. IEEE Transactions on Consumer Electronics, 2007, 53(2): 667-674.

[9] CHEN Z B, LIAO N, GU X D, et al. Hybrid distortion ranking tuned bitstream-layer video quality assessment [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(6): 1029-1043.

[10] LIN T L, YANG N C, SYU R H, et al. NR-bitstream video quality metrics for SSIM using encoding decisions in AVC and HEVC coded videos [J]. Journal of Visual Communication & Image Representation, 2015, 32(C): 257-271.

[11] 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. Piscataway: IEEE, 2003: 9-12.

[12] DRUCKER H, BURGESS C J C, KAUFMAN L, et al. Support vector regression machines [J]. Advances in Neural Information Processing Systems, 1996, 28(7): 779-784.

[13] CHANG C -C, LIN C -J. LIBSVM: a library for support vector machines [J/OL]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27 [2018-02-06]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

[14] SÜHRING K. H. 264/AVC reference software [Z/OL]. [2018-01-31]. <http://iphome.hhi.de/suehring/tml/download/>.