融合空时感知特性的无参考视频质量评估算法

吴泽民[1],彭韬频[1],田 畅[1],胡 磊[1],王露萌[2]
(1. 中国人民解放军理工大学通信工程学院, 江苏南京 210007; 2. 电子科技大学外国语学院,四川成都 611731)

摘 要: 本文通过简化视频质量评估中人眼感知模型的复杂性,提出了一种新的无参考视频质量评估模型. 首先通过分别抽取视频的空间域和时间域特征,然后按照视频局部块、视频帧、视频段等从细到粗的不同粒度,模拟人眼感知特性进行多重加权汇聚,最终得到整段视频的特征向量描述. 本方法以支持向量回归器为评估模型训练工具,通过有监督的视频样本库训练,以无参考方式完成未知视频的质量评估. 实验结果表明,该评估算法的性能不但要优于当前已知最经典的无参考评估算法 Video BLLINDS,而且与部分参考评估算法相当.

关键词: 视频质量评估; 人眼视觉系统; 离散余弦变换; 支持向量回归器
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)03-0557-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2017.03.008

Blind VQA Pooling Spatial-Temporal Perceptual Characteristics

WU Ze-min[1], PENG Tao-pin[1], TIAN Chang[1], HU Lei[1], WANG Lu-meng[2]
(1. *College of Communications Engineering, PLA University of Science and Technology, Nanjing, Jiangsu* 210007, *China*;
2. *School of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, Sichuang* 611731, *China*)

Abstract: This paper proposes a no-reference video quality assessment model by reducing the complexity of the human visual system(HVS). The characteristics of spatial domain and temporal domain of the videos are firstly extracted. Then multi-weight convergence is conducted by simulating visual perception according to different granularity from fine-grained to coarse-grained of video local block, video frame, video segment, etc. Finally the feature vector of the whole video is achieved. The support vector regression(SVR) is taken as quality assessment tool in this algorithm. The quality assessment of the unknown video is obtained without reference after supervised training. The experiments we have done show that the algorithm is not only superior to all of the other no-reference quality assessment algorithms, but also can be compared to part-reference algorithms.

Key words: video quality assessment; human visual system; discrete cosine transform(DCT); support vector regressor(SVR)

1 引言

随着通信技术的发展,视频业务通过有线和无线网络的发送信息量都呈指数级增长. 在承受巨大的负荷的同时,需要保证视频通信的服务质量,使视频最终接受者对视频的感知质量能维持在可接受的水平[1]. 实时地获取网络传输视频的质量是现阶段视频通信质量保障领域的迫切需求. 通过监控网络视频质量,反馈视频质量情况,可以调节视频编码参数和视频传输信道参数,从而改善网络视频的服务质量[2]. 因此,视频质量评估也就成为了视频处理研究领域的热点.

视频质量评估包含主观和客观两类模型. 主观视频质量是人的主观感觉,需要在采集大量观察者对相同视频的主观评分后,再通过统计分析的方法获得. 这种主观质量评估方法所得到的评估值就是 MOS(Mean Opinion Score)值. 该方法不但繁琐,而且代价很大,不能满足视频通信系统实际控制的需要. 因此,设计计算机自动预测视频质量的客观评估模型就显得非常必要. 视频质量评估的研究者一直致力于探索视频损伤的来源,并通过对损伤建模形成人眼感知视频质量模型. 目前,客观视频质量评估模型按是否需要无损视频的信息,可以分为全参考、部分参考和无参考三种类型.

视频的受损主要取决于编码和信道,其中信道是最难掌控的. 信道中视频受损主要受到丢包、丢帧以及视频

收稿日期:2015-03-02;修回日期:2015-05-25;责任编辑:梅志强
基金项目:光电控制技术重点实验室和航空科学基金联合资助项目(No. 20125286005);国家自然科学基金青年基金(No. 61501509)

空间复杂度的影响. 对视频的质量评估可分为解码前即基于码流分析的质量评估[3,4]和解码后即基于像素分析的质量评估. 基于码流分析的质量评估主要收集视频在传输信道中的比特率、帧率、丢包率、及时延等信息对视频进行质量评估,这表现为一种统计信息,很难考虑到视频内容的复杂性以及视频空间的复杂性,很难模拟人眼视觉系统感知特性. 基于像素分析的视频质量评估需要对视频码流进行完全解码,得到像素域信息,评估时更好的考虑到了影响人眼视觉系统的感知特性.

全参考视频质量评估要求有无损的像素级视频作为参考信号. 典型的全参考视频质量评估包括 VSSIM[5]、MC-SSIM[6]、MOVIE[7]、ST-MAD[8]、Vis3[9]. VSSIM 在 SSIM[10]的基础上,结合视频的运动矢量信息,采用多重区域加权的方式构造视频质量指标. MC-SSIM 采用块运动信息作为时间域信息,结合 SSIM 对空间域的评估系数,构造出视频段的质量系数. MOVIE 模仿人类大脑皮层的视觉感知机制,用 Gabor 滤波器组对空间失真进行建模,用光流场空时 Gabor 滤波器组进行空时失真建模. ST-MAD 从连续的视频帧中抽取空间域图像和时间域图像,然后提取参考视频与被评估视频空时图像的差异特征,通过失真量分析视频质量. Vis3 也采用了与 ST-MAD 相似的时空建模思路,但是它的特征抽取和汇聚过程更加复杂.

然而在实际应用中,作为参考的完整无损视频是很难得到的,这也极大地限制了全参考视频质量评估模型的应用. 而部分参考视频质量评估作为补充方法,只需要基于无损视频的参数级信息作为参考. 目前比较典型的部分参考视频质量评估有 VQM[11]、STRRED[12]. VQM 算法需提取源视频的部分特征信息,包括视频帧的灰度图梯度特征、运动信息特征和色彩空间特征. STRRED 算法则采用 Gaussian 尺度模型对视频帧和相邻帧差的小波系数进行建模,通过计算参考视频和有损视频的空时熵差来评估视频质量.

无参考视频质量评估则不需要任何无损视频作为参考,因此应用条件最宽松. 通过模拟人眼视觉系统的特性,无参考评估方法一般在像素级兴趣区域和特征级兴趣区域进行建模[13~15]. 早期的无参考视频质量评估算法要求区分特定的视频受损类型,如块效应、振铃效应、模糊等. 而在进行评估时如何由计算机自动判定视频的受损类型这个问题本身就有待解决,所以无参考评估算法性能不好. 随着自然(场景)统计模型 NSS(Natural Scene Statistics)[16]的发展,以支持向量回归器(SVR)[17]为代表的方法的应用,可以不区分受损类型而估计视频质量. 当前的无参考评估方法侧重于寻找视频特征的描述方式,用机器学习的方法完成估计器的训练和测试.

融合 BLIINDS-II[18]和 NIQE[19]两个图像质量评估算法特点的 Video BLIINDS[20]算法,是目前最好的无参考视频质量评估算法. BLIINDS-II 首先对图像分块,然后通过 DCT 变换和广义 Gaussian 分布模型抽取图像的形状参数、频率变化系数、子带能量比和各向异性等 4 种特征,级联形成每个图像块的特征,最后经全图的平均汇聚和百分比汇聚构造图像特征序列以评估图像质量. NIQE 只使用无损图像形成归一化对比度图(MSCN),然后随机采样多个 MSCN 图像块作为正样本. NIQE 利用了 MSCN 的方差偏离作为图像信息量的测度,用 SVR 建立无参考视频质量估计器.

现有无参考视频评估算法的问题是:(1)精确性还不高,与主观质量评估值的相关性还有提高空间;(2)算法复杂性高,实时性差. 为改进现有无参考评估算法的缺陷,本文通过模拟人的视觉感知特性,提出了融合时空感知特性的无参考视频质量评估方法 STEP(Spatial-TEmporal Perception). STEP 利用人眼对时间和空间的感知特性,对 DCT 变换域提取的频域特征与视频的运动信息和 Renyi 熵信息进行加权,获得了目前最好的视频质量无参考评估性能.

2 STEP 视频质量评估算法

视频质量是对连续多帧数据的综合评价. 视频的特征主要可以分为空间域特征和时间域特征,分别抽取最能反映人眼视觉特性的特征. 本文在构造评估模型时,更多考虑人眼视觉系统的感知特性,通过模拟时空敏感性完成对视频信息的汇总. 在时间域以运动显著性为特征,模拟人眼对运动区域的不敏感性;在空间域以纹理复杂性为特征,模拟人眼对纹理复杂区域的敏感性;通过多重加权汇聚形成视频质量的评估值. 尽管人眼视觉特征还有很多其它的特性,但是通过实验验证,在多帧的视频信息加权汇总过程中,它们的效应并不是很明显.

本文在进行特征抽取时,选择了局部块运动张量特征、Renyi 熵特征、DCT 系数特征. 经研究,局部块 DCT 变换系数的统计特征与视频帧的受损程度直接相关[21]. 与 Video-BLIINDS 算法不同的是,本文采用了不同的 DCT 系数条带划分方式和参数汇聚方法,从而降低了局部块的特征向量维数.

2.1 局部块划分

对视频片段中的每个视频帧,首先划分为相互不覆盖的矩形局部块. 假设单帧的宽、高分别为 W 和 H 个像素,每个局部块的大小为 B,则视频帧可以划分为 $\lfloor W/B \rfloor \times \lfloor H/B \rfloor$ 个局部块,其中符号 $\lfloor \rfloor$ 表示下取整运算. 剩余的帧边缘参与像素信息不在特征抽取使用.

采用该种划分方法和边缘像素信息处理技巧的原因如下:(1)由于视频编码时是将视频帧划分为大量的

局部块,所以视频在编码中不仅存在空间压缩受损,还存在编码局部块边缘受损,且局部块边缘受损在视频编码中是不可避免的. 本文采用的基于局部块的特征提取方法,并且考虑到视频编码中的分块效用产生的边缘受损信息. 视频编码时宏块划分一般采用 8、16 等像素值,所以本文选择区别于视频编码时的局部块尺寸 17 像素;(2)若按照该种局部块划分方法,每帧图像的边缘都会存在部分残余像素信息. 人们观看视频时,在没有特意的强调的情况下,注意力一般都放在视频的中间区域附近,而不会把注意力集中在视频的边缘区域,尤其是视频的下边缘. 这也是图像和视频显著性计算的基本假设. 本文从人眼观看视频的一般情况出发,重点考虑了人眼观看视频的一般区域.

2.2　空间特征抽取

首先,提取局部块各方向的纹理特征[22],利用纹理的分布规律表示每个局部块的重要性. 虽然直接对局部块进行二维 Gabor 变换可以获得更好的纹理系数,但计算速度慢,且纹理方向控制难度较大. 本文采用了一维 Psudo-Wigner 滤波,其主要原因是一维 Psudo-Wigner 滤波不但计算速度更快,而且一维 Psudo-Wigner 滤波可以精确控制纹理方向,具有良好的方向性,且能更好的反应视频局部块的纹理特征. 一维 Psudo-Wigner 滤波的表达式为:

$$P[m,n,k]=2\sum_{b=-N/2}^{N/2-1} z_{m,n}(b)z_{m,n}^{*}(b)\mathrm{e}^{-2\mathrm{i}(2\pi bk)/N} \tag{1}$$

其中,$z_{m,n}(\cdot)$是处于滤波器窗口中的局部块像素灰度值,$z_{m,n}^{*}(\cdot)$是其共轭,m,n 是当前局部块中心像素对应的帧内空间位置下标.

滤波器窗口中的局部块像素位置 m,n 可以由滤波器方向和窗口宽度计算得到. k 是空间频率下标,代表不同的空间频率. 对 $P[m,n,k]$ 按能量归一化后,得到各个方向的 Psudo-Wigner 分布归一化形式:

$$\bar{P}_{\theta}[m,n,k]=\frac{P_{\theta}[m,n,k]^2}{\sum_{k}P_{\theta}[m,n,k]^2} \tag{2}$$

其中,下标 θ 表示不同滤波器窗口的方向,本文共使用四个方向,分别是 0°、45°、90°、135°. 在方向固定的情况下,局部块的归一化 Psudo-Wigner 系数是空间频率的函数. 不同方向的滤波输出构成 Wigner 分布.

熵是数据复杂性的度量方式. 局部块不同方向 Wigner 分布的差异可以反映纹理复杂性. 根据局部块的归一化 Wigner 分布,局部块在特定方向的 Renyi 熵可以表示为:

$$R_{\theta}[m,n]=-\frac{1}{2}\log\Big(\sum_{k}\bar{P}_{\theta}[m,n,k]^3\Big) \tag{3}$$

局部块纹理的复杂性由不同方向的 Renyi 熵的差异性来表示. 把纹理复杂性表示为 θ 各个方向 Renyi 熵的标准差 Std 与均值 Mean 的比值:

$$\eta_{m,n}=\frac{\mathrm{Std}\{R_{\theta}[m,n]\}}{\mathrm{Mean}\{R_{\theta}[m,n]\}} \tag{4}$$

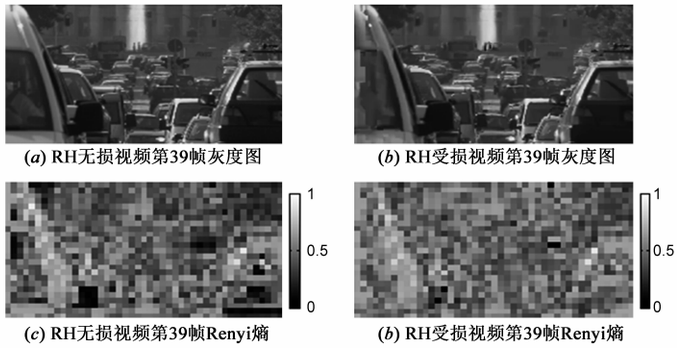
其中 $\mathrm{Std}\{\cdot\}$ 表示标准差,$\mathrm{Mean}\{\cdot\}$ 表示均值.

把每个局部块的纹理复杂性信息按局部块在帧内的位置进行排列,构成单帧的空间显著图 $\Delta=\{\eta_{m,n}\}$. 视频的纹理各向异性信息如图 1. 图 1(a)和(b)分别是 LIVE 视频库中 Rushhour 视频段中某帧的无损和受损图像;图 1(c)和(d)中局部方块的亮度的明暗表示纹理复杂性,局部块亮表示纹理复杂性高,局部块暗表示纹理复杂性低.

对每个局部块的灰度分量进行 DCT 变换,形成 DCT 变换域系数 $d_{u,v}$. 对变换后的系数 $d_{u,v}$ 非直流系数进行条带划分,从低频向高频划分为 $\{C_0,C_1,C_2,\cdots,C_n\}$ 条带,如图 2 所示. 其中,每个条带的取值是相同频率的所有系数的算术和,可以表示为:

$$C_i=\sum_{u+v=i}d_{u,v} \tag{5}$$

在实际的人眼视觉感知中,人眼对低频信号敏感,对频率过高的信号不敏感. 所以,对 DCT 变换后的频率条带,STEP 算法根据人眼视觉系统 HVS 的掩码效应特性,去除直流和高频信息,保留低频、中频、次高频的频率信号构成向量 $\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_{17}\}$. 把这些成分级联形成 17 维的向量,用于表示局部块的空间域特征.



(a) RH无损视频第39帧灰度图　(b) RH受损视频第39帧灰度图

(c) RH无损视频第39帧Renyi熵　(b) RH受损视频第39帧Renyi熵

图1　Rushhour(RH)视频帧的Renyi熵示意图

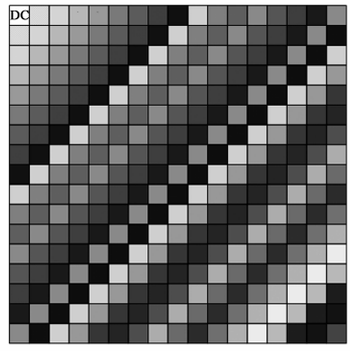


图2　DCT变换系数条带划分方法

在设计条带取值的汇聚方式时,本文测试了使用可感知模型(JND,Just Noticeable Distortion)进行频率条带归一化的方案,但是测试效果不明显. 主要的原因还是局部块的 JND 加权效果,在经过帧级和多帧的更高层汇聚后变得不再显著,对最后估计器的影响很小. 相反,直接进行频率条带的幅度的累积算数平均汇聚,这样的粗粒度汇聚能在多帧汇聚后保持特征不变.

2.3 时间特征抽取

视频与图像的最大的区别是时间因素的加入,时间域的感知机理比空间域的感知机理更为复杂. 通过对人眼视觉系统的研究,人们发现了人眼视觉系统 HVS 的运动掩蔽效应:对运动速度较快的区域,人眼可以容忍更大的失真.

Video-BLIINDS 是对视频的相邻帧亮度信息差做 DCT 变换,求取视频的 DCT 条带系数的均值信息和方差信息、全局运动信息和运动一致性信息、广义 Gaussian 分布拟合的形状参数信息作为视频的时间域特征. STEP 抽取了视频的运动张量信息,Renyi 熵信息,DCT 系数低频、中频、高频信息三种特征. 本文提出的 STEP 算法与 Video-BLIINDS 最大的区别是前者将分别抽取得到的特征信息进行级联,最后输入 SVR 训练估计器,后者是将抽取得到的信息通过空间域与时间域加权合并的方式构造出空时特征,再输入 SVR 训练估计器进行训练以评估未知视频的质量. Video-BLIINDS 和 STEP 算法各有优劣,在下文表 4 的实验对比中将对两个算法的视频质量评估性能进行对比和优劣分析.

STEP 算法需要用时域的运动信息对空域特征进行加权. 为了简化计算,本文没有采用光流等复杂的运动信息提取算法,而是采用了基于块的三步搜索法搜索局部运动向量. 为了进一步减少搜索计算量,本文规定三步搜索的范围大小是 3 个局部块,精度是半个局部块. 局部运动向量用 $\boldsymbol{M}_x$ 和 $\boldsymbol{M}_y$ 表示.

由于搜索获得的局部运动向量存在误差,在进行视觉质量的运动掩蔽效应模拟时,需要在更大范围内对局部块的运动信息进行平滑和汇聚. 本文利用搜索获得的运动向量构造得到局部运动张量以平滑和汇聚局部块运动信息. 运动张量 $\boldsymbol{T}$ 示如下:

$$\boldsymbol{T}=\begin{bmatrix} f(\boldsymbol{M}_x) & f(\boldsymbol{M}_{xy}) \\ f(\boldsymbol{M}_{xy}) & f(\boldsymbol{M}_y) \end{bmatrix} \tag{6}$$

其中:

$$f(\boldsymbol{M}_x)=\sum_{i,j} w(i,j)\boldsymbol{M}_x^2(i+m,j+n) \tag{7}$$

$$f(\boldsymbol{M}_y)=\sum_{i,j} w(i,j)\boldsymbol{M}_y^2(i+m,j+n) \tag{8}$$

$$f(\boldsymbol{M}_{xy})=\sum_{i,j} w(i,j)\boldsymbol{M}_x(i+m,j+n)\boldsymbol{M}_y(i+m,j+n) \tag{9}$$

$w(i,j)$是加权窗函数. 本文采用 Gaussian 窗函数,方差为 1.5,选择在 7×7 个局部块范围内考察局部块运动一致性. 通过矩阵分解可有运动张量 $\boldsymbol{T}$ 的特征值为 λ_1 和 λ_2,如式:

$$\boldsymbol{T}=\boldsymbol{P}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\boldsymbol{P}^{\mathrm{T}} \tag{10}$$

其中 $\boldsymbol{P}$ 是 $\boldsymbol{T}$ 的特征向量构成的矩阵,$\boldsymbol{P}^{\mathrm{T}}$表示 $\boldsymbol{P}$ 的转置.

所以有,局部的运动一致性定义为:

$$\gamma_{m,n}=\left(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}\right)^2 \tag{11}$$

显然,运动一致性越明显的局部块,它的两个特征值差值越大,获得的运动一致性测度越趋近于 1. 利用一致性信息对运动向量进行加权,获得准确的显著运动区域:

$$\mu_{m,n}=\frac{1}{\gamma_{m,n}+c}(\boldsymbol{M}_x^2+\boldsymbol{M}_y^2)^{1/2} \tag{12}$$

其中 c 是个极小常数,取值 0.001,是为避免出现数值计算异常. 把每个局部块的运动显著信息按局部块在帧内的位置进行排列,构成单帧的运动显著图集合$\Gamma=\{\mu_{m,n}\}$.

2.4 多尺度空时特征的汇聚

根据人眼对运动信息感知的特性,对快速运动目标的视觉信号可以有较强的损伤容忍度,而对慢速运动目标的视觉信号的损伤容忍度较差,轻微的失真也能被感知到. 另外,人眼对纹理复杂的区域更敏感,能分辨纹理区域中的失真细节,而对纹理简单区域可以容忍较大失真而不影响主观评价. 由此,STEP 算法融合纹理复杂性、运动显著性信息,用它们对 DCT 频域条带幅度进行加权. 每个局部块的感知权重值表示为:

$$W_{m,n}=\eta_{m,n}(1-\mathrm{e}^{-5\mu_{m,n}/\max(\Gamma)}) \tag{13}$$

其中 $\max(\Gamma)$是对局部块帧内运动显著图求取最大值,$\eta_{m,n}$表示局部块的纹理复杂性. 纹理越复杂、运动越不显著的区域,对最终的视频质量影响越大. 根据感知权重 $W_{m,n}$对每个局部块内的 DCT 条带进行加权:

$$g_{m,n,i}=W_{m,n}C_i \tag{14}$$

其中 C_i 表示局部块不同 DCT 条带信息的汇聚,本文根据人眼对频率的敏感性强弱,最终将条带划分为 17 条,即有$\{C_1,C_2,\cdots,C_{17}\}$;向量 $\boldsymbol{G}_{m,n}$由 17 维$\{g_{m,n,1},g_{m,n,2}\cdots,g_{m,n,15},g_{m,n,16},g_{m,n,17}\}$组成.

加权后形成的是单帧局部块时空质量指标图 $\psi=\{G_{m,n}\}$. 根据单帧时空质量指标图 ψ,对单帧内所有局部块的空时质量指标的特征向量求算术均值,汇聚形成的单帧质量特征向量,可以表示为:

$$\boldsymbol{Q}_k=\frac{1}{m\times n}\sum_{m,n}\boldsymbol{G}_{m,n} \tag{15}$$

对视频序列的每帧分别计算单帧质量特征向量,

由此获得视频所有帧的时空特性指标向量序列集合$\{\boldsymbol{Q}_k, k=1,\cdots,L\}$,其中$k$表示当前帧,$L$表示视频帧的总数. 根据视频序列对应的单帧质量特征向量序列$\{\boldsymbol{Q}_k\}$,对视频片段的所有单帧质量特征向量求算术均值,汇聚形成视频平均质量特征向量,可以表示为:

$$\boldsymbol{V}=\frac{1}{L}\sum_{k=1,\cdots,L}\boldsymbol{Q}_k \tag{16}$$

$\boldsymbol{V}$在逐级平均汇聚过程中,丢弃了大量损伤相关的信息,单纯依靠$\boldsymbol{V}$不能达到区分视频质量的目的. 为此,STEP算法再对视频质量特征向量序列集合$\{\boldsymbol{Q}_k\}$的相邻单帧质量特征向量求差,获得相邻帧的质量差值特征向量. 质量差值特征向量代表了视频序列中大的视觉效应突变. 然后,把视频片段所有质量差值特征向量进行算术平均,形成视频差值特征向量,可以表示为:

$$\boldsymbol{U}=\frac{1}{L-1}\sum_{k=2,\cdots,L}|\boldsymbol{Q}_k-\boldsymbol{Q}_{k-1}| \tag{17}$$

上式中的绝对值运算表示对向量的每个元素求绝对值. 视频平均质量向量和视频差值向量级联,构成视频片段质量特征向量$\boldsymbol{Z}=[\boldsymbol{V}^{\mathrm{T}},\boldsymbol{U}^{\mathrm{T}}]^{\mathrm{T}}$最终的视频片段向量维数是34维.

2.5　视频质量估计器的训练和测试

利用带主观视频质量评分值的训练视频库,完成客观视频质量估计器的训练. STEP算法采用与Video-BLIINDS相同的SVR作为估计器[23,24].

视频评估模型的建立尤为重要,所以训练阶段视频的选择也就显得十分重要. 在训练时要选择视频库中的主观评分相对准确的视频,而判断标准就是依据视频库中主观得分方差小的视频. 在样本充分的情况下,方差小说明主观评估一致性强,则判分更为准确. 对训练视频库中的每段视频,计算它的视频片段质量特征向量$\boldsymbol{Z}_r$,其对应的主观质量评分S_r值已知. 把视频片段质量特征向量集合$\{\boldsymbol{Z}_r, r=1,\cdots,R\}$和对应的主观质量评分集合$\{S_r, r=1,\cdots,R\}$送入SVR,完成视频质量估计器的训练. 当对未知质量的视频片段进行测试时,根据相同的步骤计算该视频片段的特征向量$\boldsymbol{Z}$,然后输入训练完成的SVR,输出客观视频质量值.

SVR是基于结构风险最小化原理的一种预测器. SVR的预测性能受到多种因素的影响,最主要的有:(1)误差惩罚参数的选取,该参数可以根据PSO寻优算法[25]得到;(2)核函数的形式及其参数,不同核函数对分类性能有影响,相同核函数不同参数也有影响. 由于视频质量的MOS值限定在[0,100],在有限的区间选用多项式核函数具有明显的优势;(3)损失函数,SVR回归预测需要考虑到最优分类面,损失函数即对选择最优分类面时的容错程度,调整损失函数的大小,对预测性能将会有明显的影响. 在STEP算法的测试过程中,本文考虑了上述影响因素,取得了满意的效果,下面将对算法设计中的实验数据进行比对和分析.

3　实验结果及其分析

为了验证STEP算法的评估性能,本文采用Austin大学的标准视频库Live Video Database进行实验. 图3是Live库中各视频的某一帧. 该视频库包含丰富的数据信息,为进行视频质量评估验证提供了很好的环境. 该视频库的视频格式为YUV420,分辨率为768×432像素,并提供了每段视频的DMOS值以及主观评估方差. 视频库中包含10段参考视频和150段受损视频,受损视频分别包括四种不同类型的特定受损,分别是MPEG压缩受损、H.264压缩受损、无线信道受损、IP受损,每种受损类型分别有多段不同程度的受损视频.

图3　Live Video Database视频库中部分帧

表1　不同损失函数取值的训练样本的性能比较

损失函数	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26	0.28
PLCC	0.922	0.892	0.861	0.838	0.797	0.756	0.700	0.660	0.614	0.578
SROCC	0.916	0.881	0.851	0.822	0.771	0.728	0.674	0.619	0.553	0.513

表 2 不同损失函数取值的测试样本性能比较

损失函数	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26	0.28
PLCC	0.473	0.507	0.551	0.712	0.820	0.779	0.729	0.734	0.703	0.673
SROCC	0.484	0.478	0.517	0.685	0.792	0.780	0.752	0.744	0.697	0.667

分析表 1 和表 2,损失函数的取值对视频质量的预测性能有明显的影响. 对于训练样本,在损失函数递增时,用训练好的模型来对测试样本做预测,PLCC 和 SROCC 随着损失函数的增大而减小;对于测试样本,在损失函数因子 $P<0.18$ 时,PLCC 和 SROCC 随呈单调递增趋势,在损失函数因子 P 大于 0.18 时,PLCC 和 SROCC 随 P 增大而减小. 通过对仿真数据的分析,选择 P 取值 0.18 时,算法的预测性能最佳. 图 4 是 P 取 0.18 时,评估模型对训练视频和测试视频的主观质量预测散点图.

视频库总共提供了四种不同类型的受损视频. 但其实视频的受损类型不是绝对独立的,不同视频受损类型之间具有较强的相关性,可以相互影响. 实验中每次采用其中三种受损类型的视频作为训练样本,另外一种受损视频作为预测样本. 最后视频质量预测值与 DMOS 的相关性比较见表 3,可见算法对不同受损类型的视频的评估性能存在差异,这不仅与视频的训练样本有关,同时也与抽取的视频特征有关. 针对不同的受损类型,能够明确反映该种受损的特征是有差异的.

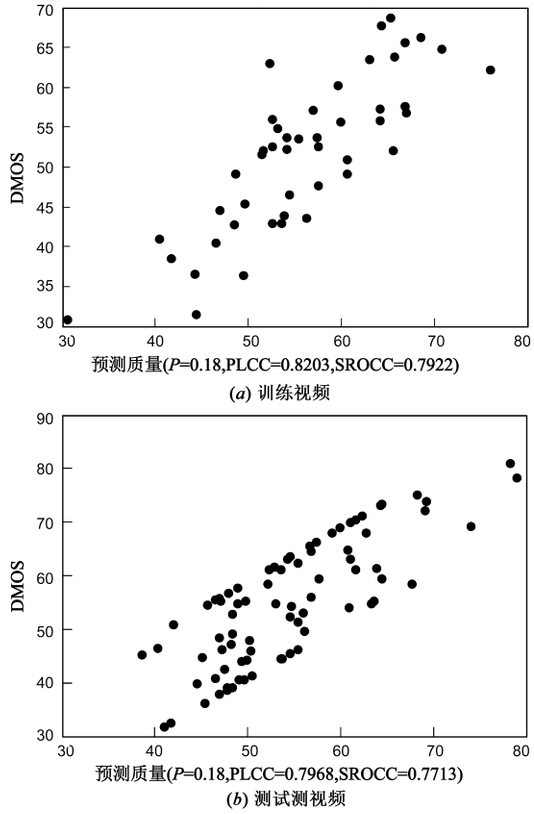


图4 P=0.18时,STEP模型对训练视频和测试视频的预测值与DMOS散点图

表 3 不同受损类型的测试样本

受损类型	MPEG-2 受损	H.264 受损	Wireless 受损	IP 受损	所有受损
PLCC	0.813	0.781	0.829	0.725	0.813
SROCC	0.820	0.742	0.813	0.686	0.788

现阶段无参考视频质量评估算法较少,比较经典的只有 Video BLIINDS,所以很难找到对应的参考算法与本文算法 STEP 进行比较. 本文最终选择典型的 PSNR、VQM、MOVIE、Vis3 全参考算法,STRREED 部分参考算法和 Video BLIINDS 无参考算法,与本文提出的 STEP 算法进行视频质量预测性能比较,测试结果如表 4 所示. 全参考和部分参考视频质量评估算法由于有参考信息可以使用,比无参考算法有先天的优势. 从表 4 可以看出,STEP 算法的性能远远超过早期的全参考算法 PSNR 和部分参考算法 VQM,与部分参考算法 STRREED 算法性能相当,比复杂的全参考算法 MOVIE 和 Vis3 相差不超过 3%. 而 Video-BLIINDS 的算法代码是原作者提供的,我们的测试性能比它也有较大提高.

表 4 不同视频质量评估算法性能比较

算法	PSNR	MOVIE	Vis3	VQM	STRREED	Video BLIINDS	STEP
PLCC	0.561	0.852	0.829	0.733	0.826	0.764	0.813
SROCC	0.539	0.807	0.816	0.703	0.725	0.760	0.788

与 Video BLIINDS 相比 STEP 的计算复杂度低优势在于:(1)局部图像块尺度 Video BLIINDS 为 5×5 像素,STEP 为 17×17 像素,相比较 STEP 需统计的图像块数量少,特征的抽取时间缩短;(2)STEP 抽取视频片段特征维数 32 维,Video BLIINDS 特征为 46 维,随着特征维数的增加,模型的预测复杂度会呈指数级增长,所以 STEP 较 Video BLLINDS 预测时间缩短. 表 5 是 STEP 和 Video BLIINDS 视频质量估计器建立后,对同一段未知质量的视频进行特征抽取和质量预测的时间比较,该视频帧率 25 帧/s,长度 250 帧,分辨率 768×432 像素.

表 5 STEP 和 Video BLLINDS 视频特征抽取和质量预测时间比较

算 法	STEP	Video BLIINDS
特征抽取时间/秒	926.1	1016.8
质量预测时间/秒	0.009842	0.054201

4 结束语

本文提出了一种融合局部和全局时空特性的无参考视频质量评估算法,进一步提升了无参考视频质量评估算法的性能. STEP 算法从人眼视觉系统的感知效果出发,对空间特征和时间特征进行模拟,完成了从局部块到整段视频的汇聚过程,形成的视频特征向量作为 SVR 的输入完成视频质量评估器的训练和估计. STEP 算法不但比典型的无参考算法 Video-BLIINDS 性能更优,而且评估结果与视频库中的主观视频质量相关性强、鲁棒性强.

参考文献

[1] S Chikkerur, V Sundaram, M Reisslein, L Karam. Objective video quality assessment methods: A classification, review, and performance comparison [J]. Broadcasting, IEEE Transactions on, 2011, 57(2): 165 – 182.

[2] A C Bovik. Automatic prediction of perceptual image and video quality[J]. Proceedings of the IEEE, 2013, 101(9): 2008 – 2024.

[3] YANG Fuzheng, WAN Shuai, XIE Qingng, et al. No-reference quality assessment for networked video via primary analysis of bit-stream[J]. Circuits and Systems for Video Technology, IEEE Transaction on, 2010, 20(11): 1544 – 1554.

[4] Paulikas S. Estimation of video quality of H. 264/AVC video streaming [A]. EUROCON [C]. Vienna: IEEE, 2013. 694 – 700.

[5] Z Wang, L Lu, A C Bovik. Video quality assessment based on structural distortion measurement [J]. Signal Process Image Commun, 2004, 19(2): 121 – 132.

[6] Moorthy A K, Bovik A C. Efficient video quality assessment along temporal trajectories[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2010, 20(11): 1653 – 1658.

[7] Seshadrinathan K, Bovik A C. Motion tuned spatio-temporal quality assessment of natural videos[J]. Image Processing, IEEE Transactions on, 2010, 19(2): 335 – 350.

[8] Phong V Vu, Cuong T Vu, Damon M Chandler. A spatio-temporal most-apparent-distortion model for video quality assessment[A]. Image Processing (ICIP), 18th IEEE International Conference on [C]. Brussels, Belgium: IEEE, 2011. 2505 – 2508.

[9] Vu P V, Chandler D M. Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices[J]. Journal of Electronic Imaging, 2014, 23(1): 013016 – 013016.

[10] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. Image Processing, IEEE Transactions on, 2004, 13(4): 600 – 612.

[11] Pinson M H, Wolf S. A new standardized method for objectively measuring video quality[J]. Broadcasting, IEEE Transactions on, 2004, 50(3): 312 – 322.

[12] Soundararajan R, Bovik A C. Video quality assessment by reduced reference spatio-temporal entropic differencing [J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2013, 23(4): 684 – 694.

[13] Jia L, Zhong X, Tu Y. No reference video quality assessment model based on eye tracking data[A]. 2nd International Conference on Information, Electronics and Computer (ICIEAC 2014)[C]. Bandung, Indonesia: Atlantis Press, 2014. 97 – 100.

[14] You J, Korhonen J, Perkis A. Spatial and temporal pooling of image quality metrics for perceptual video quality assessment on packet loss streams[A]. Acoustics Speech and Signal Processing, IEEE International Conference on [C]. Dallas, TX: IEEE, 2010. 1002 – 1005.

[15] Rodriguez D Z, Rosa R L, Bressan G. Video quality assessment in video streaming services considering user preference for video content[A]. Consumer Electronics, IEEE International Conference on[C]. Las Vegas, NV: IEEE, 2014. 570 – 571.

[16] A K Moorthy, A C Bovik. Statistics of natural image distortions [A]. IEEE International Conference AcousticsSpeech Signal Process [C]. Dallas, TX, USA: IEEE, 2010. 962 – 965.

[17] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27.

[18] Saad M A, Bovik A C, Charrier C. Blind image quality assessment: A natural scene statistics approach in the DCT domain [J]. Image Processing, IEEE Transactions on, 2012, 21(8): 3339 – 3352.

[19] Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer[J]. Signal Processing Letters, IEEE, 2013, 20(3): 209 – 212.

[20] Saad M A, Bovik A C, Charrier C. Blind prediction of natural video quality[J]. IEEE Transactions on Image Processing, 2014, 23(3): 1352 – 1365.

[21] Saad M A, Bovik A C, Charrier C. A DCT statistics-based blind image quality index[J]. Signal Processing Letters, IEEE, 2010, 17(6): 583 – 586.

[22] Gabarda S, Cristóbal G. Blind image quality assessment through anisotropy [J]. Journal of Software, 2007, 24(12): 42 – 51.

[23] Seshadrinathan K, Soundararajan R, Bovik A C, et al. Study of subjective and objective quality assessment of

video[J]. Image Processing, IEEE Transactions on, 2010, 19(6): 1427 – 1441.
[24] Seshadrinathan K, Soundararajan R, Bovik A C, et al. A subjective study to evaluate video quality assessment algorithms[A]. SPIE Electronic Imaging. International Society for Optics and Photonics[C]. California: SPIE, 2010. 7527 – 7532.
[25] 陈金环,王冠,王东云. 粒子群算法及特性的研究[J]. 中原工学院学报,2007,18(4):13 – 15.

作者简介

吴泽民 男,1973 年生,四川成都人,2002 年毕业于解放军理工大学通信工程学院,获博士学位. 现为解放军理工大学通信工程学院副教授,研究方向为数据链技术、雷达数据融合、图像视频分析.

彭韬频 男,1989 年生,四川内江人,2012 年毕业于电子科技大学新信息对抗技术系,获学士学位. 现为解放军理工大学硕士研究生,研究方向为视频质量分析.

田 畅(通讯作者) 男,1963 年生,山东青岛人,2001 年毕业于解放军理工大学通信工程学院,获博士学位. 现为解放军理工大学通信工程学院教授,研究方向为数据链技术、视频编码与传输优化.
E-mail: changtian@163.com