注意力分布机制下的全景图像质量评价

安平，刘欣，丁文欣，孟春丽
（上海大学通信与信息工程学院，上海 200444）

摘要：全景图像是虚拟现实媒体内容最主要的形式之一，可以为观看者提供 360 度自由观看的效果。全景图像在采集、拼接、压缩、传输和播放等环节都可能对图像质量造成损害，严重影响观看者的体验，准确地评价全景图像的质量对推进其应用有重要的意义。本文从观看者注意力分布角度，提出一种基于显著性检测的无参考全景图像质量评价方法，可减少弱纹理区域对质量计算的干扰；此外，从全局质量感知并结合注意力分布特性，提出一种基于多特征融合的无参考全景图像质量评价方法，能更好地反映全景图像的颜色及清晰度质量。
关键词：全景图像；图像质量评价；显著性检测；多特征融合
中图分类号：TP37　　文献标识码：A

Panoramic image quality assessment based on attention distribution mechanism

Ping An, Xin Lin, Wenxin Ding, Chunli Meng

(School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China)

Abstract: Panoramic image, as one of the most important forms of virtual reality media content, can provide a 360 degree free view. Acquisition, splicing, compression, transmission and playback may compromise the quality of the image, greatly influencing the viewer's experience. An accurate assessment of panoramic image quality is of great significance to promote its application. From the perspective of viewer attention distribution, this paper proposes a no-reference panoramic image quality assessment method based on saliency detection, which can reduce the interference of weak texture areas to quality evaluation. In addition, from the global quality perception and combined with the characteristics of attention distribution, a no-reference panoramic image quality assessment method based on multi-feature fusion is proposed, which can better reflect the color and definition quality of panoramic image.
Key words： panoramic image; image quality assessment; saliency detection; multi-feature fusion

1 引言

虚拟现实(Virtual Reality, VR)作为新兴的媒体，融合了新型显示、计算机仿真、人机交互、图像处理、人工智能等多个领域技术，可以创建一个符合现实世界规则的虚拟环境，也可以构建一个与现实相悖的完整假设环境，给人们带来沉浸式体验。虚拟现实在教育、游戏、房地产、汽车、军事等众多领域都发挥着重要作用，应用前景巨大。

头戴式设备(Head Mount Display, HMD)用来显示全景和立体图像或视频，为用户提供模拟的沉浸式环境。全景图像作为VR媒体内容最主要的形式之

基金项目：国家自然科学基金项目 62020106011，上海市科委研发平台专项 20DZ2290100
作者简介：安平（1968-），女（回族），上海大学教授，anping@shu.edu.cn。

一，可以为观看者提供360度自由观看的效果。自然场景的全景图像通常有两种拍摄方式：一种是利用全自动的全景相机拍摄得到，这种全景相机内部安装有多个不同方向的镜头，拍摄后利用相机自带的图像拼接算法得到全景图像；另一种是使用单反相机配合鱼眼镜头和云台，将相机固定在云台上，拍摄多张四周以及上下有部分重叠的图像，再使用拼接算法进行拼接。

全景图像在拼接时可能会因为视差产生重影或者模糊；在编码时会降低图像的质量；在进行播放时，可能会因为头戴式设备的硬件条件不足，使得呈现的画质差或者画面卡顿，也可能因为观看者对于画面或者设备不适应，产生眩晕感。采集、拼接、压缩、传输和播放等环节都可能对全景图像质量造成损害，严重影响观看者的体验。准确地评价全景图像的质量可以有效帮助拼接、压缩算法以及播放设备的改进。

与普通图像的质量评价类似，全景图像质量评价也包括主观评价和客观评价两个分支。主观评价结果相对可靠，可以作为客观评价模型的真值；而客观评价模型具有批处理和结果可再现的优点。大多数客观评估模型都基于自然场景统计和模拟人类视觉系统的数学模型。根据其对参考图像的依赖性，客观评价包括三类：完全参考(full reference, FR)、半参考(reduced reference, RR)、无参考(no reference, NR)[1]。FR评价方法充分利用参考图像的完整信息，通常更可靠和准确；RR评估方法通过提取参考图像的部分统计特征来计算图像质量；NR评价模型只使用测试图像，具有很高的灵活性，最具实用价值。由于全景图像是从球面投影到平面格式进行编码传输的，传统的图像质量评价模型并不适用于全景图像。因此，建立一个有效的客观质量评价模型对全景图像的发展具有重要意义。

现有的全参考全景图像质量评价方法大多是基于峰值信噪比(Peak Signal to Noise Ratio, PSNR)或者结构相似性(Structural Similarity, SSIM)，通过反投影、增加权重策略的方式来扩展到全景图像质量评价。如 S-PSNR(Spherical PSNR)[2]、WS-PSNR (Weighted-to-Spherically-uniform PSNR)[3]、CPP-PSNR(Craster Parabolic Projection PSNR)[4]、S-SSIM(Spherical SSIM)[5]、WS-SSIM(Weighted-to-Spherically-uniform PSNR)[6]、USS-PSNR (Uniformly Sampled Spherical PSNR)[7]。这些方法在计算上较为方便，但未能考虑到受试者在观看全景图像时，对不同区域的感兴趣程度和观看时长不同，甚至有超过 1/3 的区域没有被观看[8]。无参考方法多为基于深度学习的方法，Kim 等[9]提出了一种基于生成对抗网络的模型，Li 等[10]提出了基于视口的卷积神经网络方法，Xu 等人提出了一种面向视口的图卷积网络方法[11]。深度学习的方法需要对图像进行分块操作，利用网络对每个分块单独进行打分，然后使用权重分配策略对分块的质量进行融合。因此，这种方式十分依赖分块的质量分数，并且无法大范围地感知全景图像的质量。

由于360°图像的观看范围以及观看方式与2D图像完全不同，因此，在人类视觉感知方面， 2D图像和360°图像最大的区别在于视觉注意力[12]。针对当前全景图像质量评价研究存在的问题，在设计客观质量评价方法时应考虑受试者在观看全景图像的注意力分布，同时从全局的角度感知全景图像的质量。本文重点讨论两种无参考的全景图像质量评价方法：（1）针对全景图像中不同区域分块受到编码失真的影响不同以及受试者对不同区域分块关注度不同的特点，提出基于显著性检测的无参考全景图像质量评价方法；（2）兼顾受试者对全景图像的全局感知质量和局部注意力，提出基于多特征融合的全景图像质量评价方法。

2 全景图像处理过程及显著图特性

2.1 全景图像处理过程

一个完整的全景图像处理过程包括图像采集、拼接、投影变换、压缩编码传输、图像解码、反投影，如图 1 所示。其中常用的投影格式有：等矩形投影(Equi-Rectangular Projection, ERP)、立方体投影(Cubmap Projection, CMP)、等面积投影(Equal-area Projection, EAP)、八面体投影(Octahedron Projection, OHP)、正二十面体投影(Icosahedron Projection, ISP)[13]。由于不同投影格式在压缩比和画质重现上，有各自的优缺点，因此面对不同的使用场景，便出现了不同的选择方案。全景视频技术发展迅速，至今仍未有统一的标准。

图 1 全景图像处理过程

在已有的投影类型中，从球面到平面的采样密度在每个像素位置是不均匀的。因此，直接应用传统图像质量评价方法会使得不同像素位置对质量分数的贡献存在偏差。

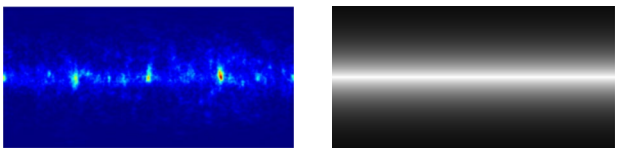
2.2 全景图像显著图的特性

Sitzmann等[14]对受试者观看全景视频时的注视点区域做了定量分析，发现平均显著图在纬度上有一种“赤道偏差(Equator Bias, EB)”的现象。如图2所示，平均显著图的分布情况可以很好地用拉普拉斯分布来描述。拉普拉斯分布的概率密度函数为：

$$f(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}} \tag{1}$$

其中，μ和λ为常数，分别表示位置参数和尺度参数。平均显著图在纬度方向上的分布具体为位置参数μ=91.3°、尺度参数λ=18.58°。

EB现象说明，受试者在赤道区域的观测频率远高于其他区域，因此对赤道区域的图像质量也更为敏感。受试者的观看方向从统计上来说更偏向赤道前方区域，但对于特定图像内容，观看方向又会有所不同[15]，比如强纹理的区域。



(a) (b)

图 2 全景图像的平均显著图及其分布规律：(a)平均显著图；(b)平均显著图按纬度的拉普拉斯分布

3 显著性检测的全景图像质量评价

人们在观看图像时会注意到图像的显著区域，尤其当观看全景图像时这一现象更加突出。此外，大尺寸的全景图像中往往存在许多受失真影响很小的弱纹理平坦区域。考虑这些特性，我们提出一种利用全景图像显著性信息的质量评价模型。

3.1 基于显著性检测的全景图像质量评价方法框架

图 3 为方法框图，包括显著性预测网络、显著信息筛选模块和质量评价网络三个部分。

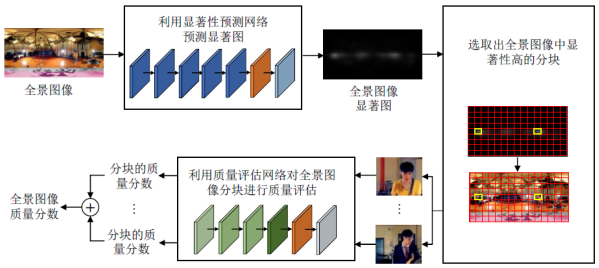


图 3 基于显著性检测的全景图像质量评价框架

（1）全景图像的显著性预测网络

首先，将全景图像输入显著性检测网络，得到全景图像的显著图。由于受试者在观看全景图像时会重点观看显著目标及其附近区域，因此，全景图像的主观质量受到显著目标及其附近区域的影响较大。本文采用多级网络结构 ML-Net[16]提取全景图像的显著图，网络结构如图 4 所示。其中，特征提取网络是在 VGG-16[17]的基础上改进的，将不同卷积层的特征抽取出来并叠加得到多级特征图，这种多级特征图可以更好地表达图像的显著性特征；编码网络负责对多级特征图进行编码，得到显著图；先验学习网络结合了前述的全景图像拉普拉斯分布平均显著图特性，在编码网络输出的显著图的基础上，进一步提高了整体网络的性能，使得网络最终的输出与人类注视点图更加接近。

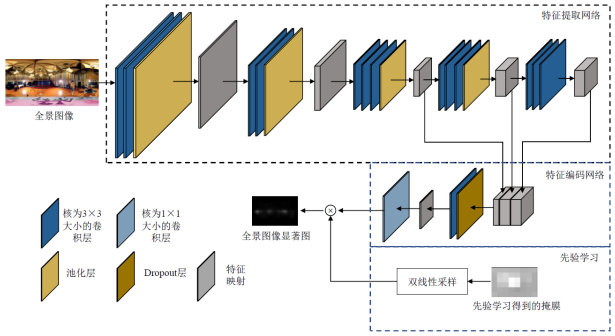


图 4 全景图像显著性预测网络

（2）全景图像的显著信息筛选模块

在图像输入到分数预测网络之前，需要将图像均匀分割成小块，每个小块单独输入到网络中进行训练。小块的质量分数与完整全景图像的质量分数相同。为了满足所有小块质量一致且能代表整个图像质量的要求，鉴于显著区域是受试者重点关注的区域，我们以显著性检测网络生成的显著图为依

据，将每个小块按照显著性从大到小排序，选取显著性较高的部分小块输入到质量评价预测网络中。

（3）全景图像的质量评价分数预测网络

考虑到 ResNet-50[18]在质量评价任务上的良好性能[19]，本方法用其作为质量评估主干网络。将上述筛选模块输出的高显著性全景图像小块输入到质量评估网络中进行训练和质量预测，得到小块的质量分数。最后计算这些小块的平均得分作为整个全景图像的质量分数。

3.2 实验结果

实验在 ERP 格式的全景视频上进行。对于显著性检测任务，使用 SALION 数据库[20]训练和测试多级特征网络。兼顾可训练图像的数量和利于表达图像质量，筛选模块中的图像小块尺寸为 128*128，筛选出前 40%的高显著小块。对于质量评价预测网络，采用 VQA-ODV 数据库[8]进行训练和预测；学习率设置为 0.0005。质量评价指标采用常用的斯皮尔曼秩序相关系数(Spearman Rank-order Correlation Coefficient, SROCC)、皮尔逊线性相关系数(Pearson Linear Correlation Coefficient, PLCC)和根均方误差(Root Mean Squared Error, RMSE)。其中，SROCC、PLCC 的值越接近 1 表示算法的效果越好，RMSE 的值越接近 0 表示算法的效果越好。

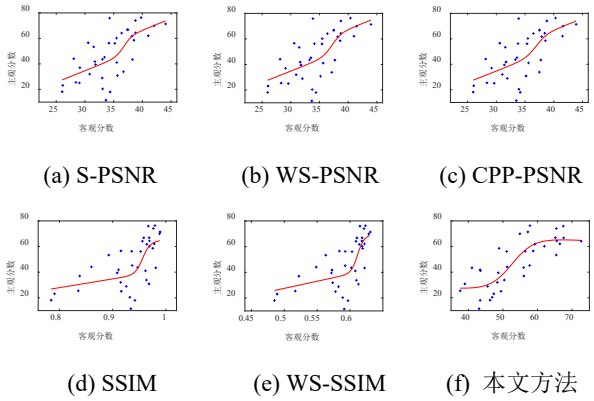
本方法与平面图像评价算法 SSIM 以及全景图像质量评价算法 S-PSNR[2]、WS-PSNR[3]、CPP-PSNR[4]、WS-SSIM[6]、BP-QAVR[21]、Li[8]的比较结果如表 1 所示，对应的散点图和拟合曲线如图 5 所示。可见本方法具有更好的性能，预测分数的散点图更加集中地分布在拟合曲线附近，证明了本方法对于全景图像质量分数的预测更加准确。

4 多特征融合的全景图像质量评价

人们在观看全景图像时，一方面，注意力往往

表 1 基于显著性检测评价方法对比实验结果

Metrics	SROCC↑	PLCC↑	RMSE↓
S-PSNR	0.6803	0.6943	13.3526
WS-PSNR	0.6821	0.6939	13.3598
CPP-PSNR	0.6798	0.6888	13.4489
SSIM	0.6803	0.6940	13.3569
WS-SSIM	0.7176	0.7345	12.5893
BP-QAVR	0.6801	0.6588	-
Li	0.7953	0.7821	-
本文提出的方法	**0.8064**	**0.8152**	**10.7447**



(a) S-PSNR (b) WS-PSNR (c) CPP-PSNR

(d) SSIM (e) WS-SSIM (f) 本文方法

图 5 MOS 与客观质量评价算法预测分数的散点图及拟合曲线

分布在赤道区域；另一方面，纹理明显的区域也更容易获得受试者的关注。此外，失真也会对颜色造成一定的影响，而颜色也是我们主观感知全景图像质量的重要因素。为此，我们提出一种基于多特征融合的全景图像质量评价方法。

4.1 方法框架

多特征融合的全景图像质量评价方法框图如图 6。主要步骤包括全景图像预处理、特征提取、模型训练与测试等。

（1）图像预处理

图像预处理主要有两项内容，一项是根据平均显著区域的范围，对输入图像进行裁剪，获取平均显著性区域的图像，用于后续提取纹理特征。考虑到人眼的单眼舒适区域为 60°，以及 ERP 投影的变换方式，本文未严格按照图 2 的平均显著图来截取显著区域，而是截取南纬 30°至北纬 30°这一范围作为平均显著区域。另一项是对图像进行多次下采样，得到不同尺度图像，用于后续提取自然场景统计(Natural Scene Statistics, NSS)特征。

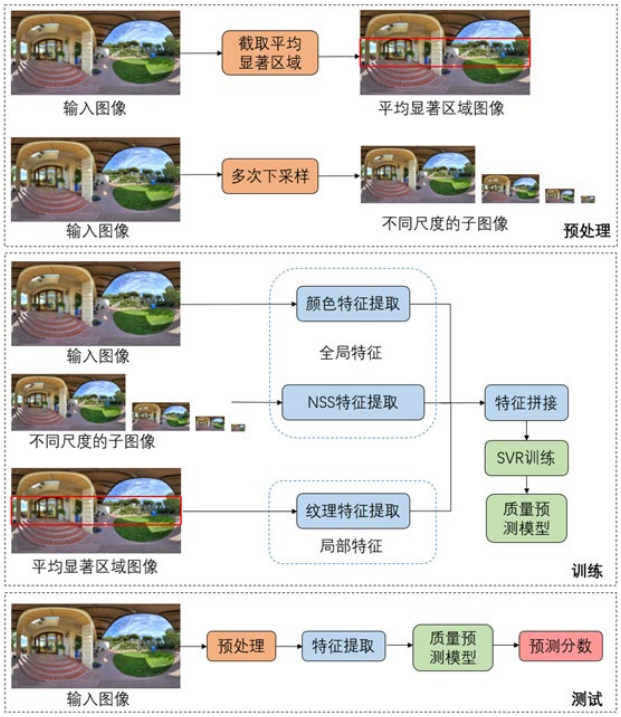


图 6 多特征融合的全景图像质量评价方法框图

（2）特征提取

为了从全局角度感知全景图像的质量，并考虑观看者的注意力分布，本文同时提取全局特征和局部特征。

a) 全局特征提取

我们提取颜色和多尺度 NSS 两种全局特征。

通过头戴式设备观看全景图像时，2D ERP 图像被解码成 3D 球形图像，这是一个从低分辨率转为高分辨率的过程，大量的插值使得颜色仿佛被“稀释”了一样。所以受试者在佩戴 HMD 观看全景图像时，往往会发现图像的色彩观感不如直接在显示屏上看到的 ERP 图像。因此，我们把颜色特征作为一个基础视觉特征。具体地，将图像的 RGB 颜色通道转换成更接近于人眼视觉系统的 HSV（Hue：色调；Saturation：饱和度；Value：明度）空间，并分别计算这三个分量的平均值，作为图像的色彩特征。

NSS 特征提取步骤参见文献[22]。不同于原方法对输入图像仅做一次下采样、最后得到一个维度为 36 的特征向量，本文考虑到全景图像具有高分辨率的特性，对全景图像进行了四次下采样。算上原图像，共有 5 个尺度图像，得到一个 90 维的特征向量，作为最终的多尺度 NSS 特征。同一全景图像在不同的头戴式设备上播放时，受试者主观感知的清晰度会有所不同。不同的尺度可以用于表征不同的清晰度，大的尺度对应着较高的清晰度，小的尺度对应着较低的清晰度。因此，提取多尺度 NSS 特征不仅扩充了特征的维度，使得特征包含更多的信息，还能更好地表征图像的清晰度。

b) 局部特征提取

纹理特征通过刻画图像中反复出现的局部模式及其排列规则来反映物体表面的特性，具备旋转不变性以及良好的抗噪声性能。对于全景图像，无论是从投影对纹理造成拉伸的角度，还是从受试者观看全景图像的注意力分布特点，在赤道附近提取全景图像的纹理特征都比全局的纹理特征更为合理。如图 6 所示，我们在赤道区域进行纹理特征提取。

提取纹理特征有多种方法，本文选取基于统计的灰度共生矩阵(Gray Level Co-occurrence Matrix，GLCM)方法[23]，得到 80 维纹理特征。该方法易于实现，并且能够体现一幅图像中灰度的有关方向、相邻间隔和幅度变化的综合信息。

（3）模型训练及测试

将上面计算得到的三类特征拼接在一起，作为输入图像的整体特征。本文使用 OIQA 数据库[24]进行模型的支持向量回归(Support Vector Regression, SVR)训练与测试。由于 SVR 训练和测试的结果具有随机性，为了保证结果的可靠性，每次训练时随机选择 12 个场景中对应的原始图像及失真图像，测试时选择剩下的 4 个场景对应的原始图像及失真图像。这样重复 1000 次交叉验证，取所有实验中 SROCC 的中位数及其对应的其他指标作为最终的实验结果。

4.2 实验结果

本文提出的方法在 OIQA 数据库上得到的散点图如图 7 所示，观察拟合曲线可以发现，模型预测的分数和主观分数拟合的较为完好。本方法与其他方法的对比实验结果如表 2 所示，可见本方法在 SROCC、PLCC 和 RMSE 三个指标上均优于现有的一些全景图像质量评价方法。

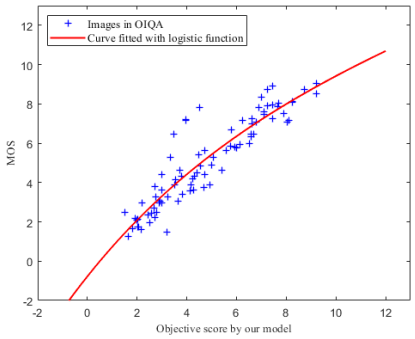


图 7 多特征融合质量评价方法的散点图

表 2 多特征融合方法对比实验结果

方法	SROCC↑	PLCC↑	RMSE↓
S-PSNR	0.4533	0.4901	1.8906
WS-PSNR	0.4705	0.5309	1.8383
CPP-PSNR	0.4623	0.4891	1.8918
SSIM	0.3743	0.2577	2.0956
WS-SSIM	0.3829	0.2837	2.0798
BRISQUE	0.8806	0.8897	1.0307
本文提出的方法	**0.9171**	**0.9290**	**0.7915**

5 结论

全景图像在采集、拼接、编码、传输、播放等过程中都可能引起图像失真，破坏观看者的体验。本文利用受试者注意力分布特点，提出了基于显著性检测的全景图像质量评价方法：首先将全景图像输入显著性检测网络，得到全景图像的注视点预测图；接着通过显著信息筛选模块，将全景图像均匀分割为小块，选择显著性高的小块输入到质量评价网络中进行训练和预测，提高了质量评价网络的性能。同时结合受试者全局质量感知特性，提出了基于多特征融合的无参考全景图像质量评价方法：首先，从全局的角度，提取颜色特征和多尺度 NSS 特征；其次，从注意力机制的角度，提取平均显著区域的纹理特征；最后，将多特征融合输入到 SVR 中训练并得到回归模型，获得了较好的质量预测性能。

参考文献

[1] Wang Z, and Bovik A C. Modern Image Quality Assessment[J]. Synthesis Lectures on Image Video & Multimedia Processing, 2006, 2(1): 156.

[2] Yu M, Lakshman H, Girod B. A framework to evaluate omnidirectional video coding schemes[C]. ISMAR Conference Proceedings-IEEE. 2015: 31-36.

[3] Sun Y L, Lu A, Yu L. Weighted-to-spherically-uniform quality evaluation for omnidirectional video[J]. IEEE Signal Processing Letters, 2017, 24(9): 1408-1412.

[4] Zakharchenko V, Choi K P, Alshina E, et al. Omnidirectional video quality metrics and evaluation process[C]. DCC Conference Proceedings-IEEE. 2017: 472-472.

[5] Chen S J, Zhang Y X, Li Y M, et al. Spherical structural similarity index for objective omnidirectional video quality assessment[C]. ICME Conference Proceedings-IEEE. 2018: 1-6.

[6] Zhou Y F, Yu M, Ma H L, et al. Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video[C]. ICSP Conference Proceedings-IEEE. 2018: 54-57.

[7] Youvalari R G, Aminlou A, Hannuksela M M. Analysis of regional down-sampling methods for coding of omnidirectional video[C]. PCS Conference Proceedings-IEEE. 2016: 1-5.

[8] Li C, Xu M, Du X Z, et al. Bridge the gap between VQA and human behavior on omnidirectional video[C]. ACM Multimedia Conference Proceedings-American Computer Association. 2018: 932-940.

[9] Lim H, Kim H G, Ra Y M. VR IQA NET: deep virtual reality image quality assessment using adversarial learning[C]. ICASSP Conference Proceedings-IEEE. 2018: 6737-6741.

[10] Li C, Xu M, Jiang L, et al. Viewport proposal CNN for 360° video quality assessment[C]. CVPR Conference Proceedings-IEEE. 2019: 10169-10178.

[11] Xu J, Zhou W and Chen Z. Blind Omnidirectional Image Quality Assessment With Viewport Oriented Graph Convolutional Networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(5): 1724-1737.

[12] Xu M, Li C, Zhang S and Callet P L. State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression[J]. IEEE Journal of Selected Topics in Signal Processing, 14(1): 5-26.

[13] Ye Y, Boyce J. JEVT-K1004: Algorithm descriptions of projection format conversion and video quality metrics in 360Lib Version 7[C/OL]. 2018. https://jvet-experts.org/doc_end_user/current_document.p

hp?id=4118.

[14] Sitzmann V, Serrano A, Paval A, et al. Saliency in VR: How do people explore virtual environments[J]. IEEE Transactions on Visualization and Computer Graphics. 2018, 24(4): 1633-1642.

[15] Xu M, Li C, Chen Z Z, et al. Assessing visual quality of omnidirectional videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(12): 3516-3530.

[16] Cornia M, Baraldi L, Serra G, et al. A deep multi-level network for saliency prediction[C]. ICPR Conference Proceedings-IEEE. 2016: 3488-3493.

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C/OL]. ICLR Conference Proceedings-International Conference on Learning Representations, 2015.

[18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. CVPR Conference Proceedings-IEEE. 2016: 2334-2342.

[19] Kong S, Shen X H, Lin Z, et al. Photo aesthetics ranking network with attributes and content adaptation[C]. ECCV Conference Proceedings-Springer Verlag, 2016: 662-679.

[20] Jiang M, Huang S S, Duan J Y, et al. SALICON: Saliency in context[C]. CVPR Conference Proceedings-IEEE. 2015: 1072-1080.

[21] Yang S, Zhao J Z, Jiang T T, et al. An objective assessment method based on multi-level factors for panoramic videos[C]. VCIP Conference Proceedings-IEEE, 2017: 1-4.

[22] Mittal A, Moorthy A K and Bovik A C. No-Reference Image Quality Assessment in the Spatial Domain[J]. IEEE Transactions on Image Processing, 2012, 21(12): 4695-4708.

[23] 唐银凤, 黄志明, 黄荣娟, 姜佳欣,卢昕.基于特征提取和SVM分类器的纹理图像分类[J]. 计算机应用与软件, 2011, 28(06): 22-25+46.

[24] Duan H Y, Zhai G T, Min X K, et al. Perceptual quality assessment of omnidirectional images[C]. ISCAS Conference Proceedings-IEEE. 2018: 1-5.