

无参考屏幕内容图像质量评价*

朱映映, 曹磊, 王旭

(深圳大学 计算机与软件学院, 广东 深圳 518060)

通讯作者: 王旭, E-mail: wangxu@szu.edu.cn



摘要: 随着多客户端交互多媒体应用的快速发展, 屏幕内容图像(screen content image, 简称 SCI)的分发和处理与日俱增. 图像质量评价课题的研究是其他许多应用的基础, 至今图像质量评价课题研究的重点是传统自然图像, 因此, 针对屏幕图像质量评价的研究就变得非常迫切和必要. 客观图像质量评价算法提出的基础建立在标准图像质量评价数据库上. 首先构建了一个大规模的屏幕内容图像质量评价数据库(immersive media laboratory screen content image quality database, 简称 IML-SCIQD). IML-SCIQD 数据库包含参考图像 25 幅以及经过 10 种失真处理的 1 250 幅失真图像. 以建立的 IML-SCIQD 数据库为基础, 考虑到屏幕内容图像的图像区域与文本区域的视觉感知差异, 在基于自然场景统计的无参考方法的启发下, 提出了针对屏幕内容图像的无参考评价算法(natural scene statistics based no reference screen content image quality assessment metric, 简称 NSNRS). NSNRS 算法首先分别计算图像区域和文本区域的质量分数, 再将这两个区域的质量分数结合起来得到整幅失真图像的质量分数. 该算法与其他 12 种经典的客观评价算法, 包括全参考算法、部分参考算法与无参考算法, 在 IML-SCIQD 数据库和 SIQAD 数据库上进行了性能测试和对比, 结果表明, 所提出的算法优于经典的无参考评价算法; 就整个数据库而言, 所提出的算法可以达到与全参考方法相当的性能.

关键词: 图像质量评价; 主观质量评价; 屏幕内容图像; 无参考; 自然场景统计
中图法分类号: TP391

中文引用格式: 朱映映, 曹磊, 王旭. 无参考屏幕内容图像质量评价. 软件学报, 2018, 29(4): 973-986. <http://www.jos.org.cn/1000-9825/5410.htm>

英文引用格式: Zhu YY, Cao L, Wang X. No reference screen content image quality assessment. Ruan Jian Xue Bao/Journal of Software, 2018, 29(4): 973-986 (in Chinese). <http://www.jos.org.cn/1000-9825/5410.htm>

No Reference Screen Content Image Quality Assessment

ZHU Ying-Ying, CAO Lei, WANG Xu

(College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China)

Abstract: With the rapid development of multi-device interactive applications, the transmission and processing of screen content image (SCI) is growing every day. Image quality assessment, which is the basis of many other research topics, has mainly focused on traditional natural images so far. Image quality assessment specifically for screen content image is therefore becoming very important and urging.

* 基金项目: 国家自然科学基金(61602314, 61602312, 61501299, 61672443); 广东省自然科学基金(2016A030313043, 2016A030310058); 深圳市科技计划基础研究项目(JCYJ20160331114551175, JCYJ20150324141711630, JCYJ20130326105637578)

Foundation item: National Natural Science Foundation of China (61602314, 61602312, 61501299, 61672443); Natural Science Foundation of Guangdong Province of China (2016A030313043, 2016A030310058); Fundamental Research Project in the Science and Technology Plan of Shenzhen (JCYJ2016 0331114551175; JCYJ20150324141711630; JCYJ20130326105637578)

本文由“多媒体大数据处理与分析”专题特约编辑赵耀教授、李波教授、华先胜研究员、文继荣教授、蒋刚毅教授、常冬霞副教授推荐.

收稿时间: 2017-04-30; 修改时间: 2017-06-26; 采用时间: 2017-10-13; jos 在线出版时间: 2017-12-01

CNKI 网络优先出版: 2017-12-04 06:46:58, <http://kns.cnki.net/kcms/detail/11.2560.TP.20171204.0646.010.html>

Considering that image quality assessment database is the basis of objective image quality assessment metrics, this paper first constructs a large scale Immersive Media Laboratory screen content image quality database (IML-SCIQD). The IML-SCIQD database contain 25 reference images and 1250 distorted images that are distorted by 10 distortions. Based on the IML-SCIQD database, the visual perception difference of pictorial region and textual region is studied. At the same time, inspired by the idea of natural scene statistics (NSS) based no reference (NR) image quality assessment metrics, a NSS based NR content image quality assessment metric (NSNRS) is proposed. The quality scores of textual region and pictorial region are first computed in the NSNRS metric. Then, the quality scores of these two regions are combined to get the quality score of the whole screen content image. For performance comparison, the proposed metric is compared with 12 state-of-the-art objective image quality assessment metrics, including full reference, reduced reference and no reference algorithms, on the IML-SCIQD database and the SIQAD database. Extensive experiments support that the proposed algorithm outperforms the existing representative no reference techniques, and that the new metric has comparable performance with those full reference metrics for the whole database.

Key words: image quality assessment; subjective image quality assessment; screen content image; no reference; natural scene statistics

1 引言

1.1 研究背景及意义

随着计算机和移动互联网技术的快速发展,多屏互动^[1]、视频游戏^[2]、远程教育等多客户端通信系统也得到了飞速的发展.在这类系统中,各类终端之间可以相互通信,以实现屏幕内容图像(终端设备绘制并显示在屏幕上的视觉内容,包括网页、邮件、地图、动画、文档和图像等)的分发和处理.与传统自然图像相比,屏幕内容图像包含的多媒体形式更多,蕴含着较多的信息,具有非常大的应用价值.例如,消费者可以在好友圈中直播购物体验,并直接分享商品的图片和文字介绍信息;汽车追尾现场,事主可直接分享手机摄像头内容给车险理赔员,远程完成理赔的认定和文件签署;远程教育中,教师可实时批阅学生的随堂练习并完成答疑等.屏幕内容图像一般同时包含图像和文本^[3].传统的自然图像和屏幕内容图像的例子如图 1 所示.其中,图 1(a)、图 1(b)所示为传统自然图像,图 1(c)和图 1(d)所示为屏幕内容图像.

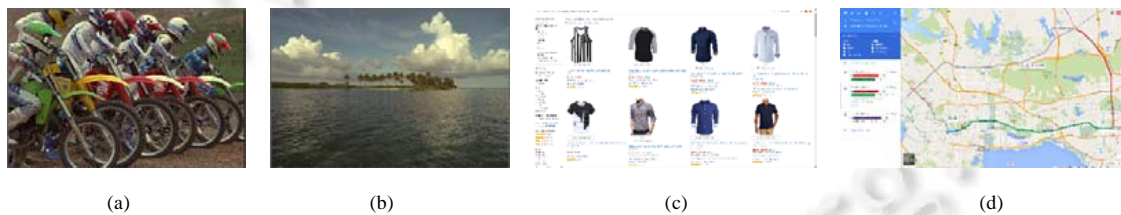


Fig.1 Examples of traditional natural images and screen content images

图 1 传统自然图像和屏幕内容图像的示例

在实时多客户端通信系统中,屏幕内容图像的分发和处理通常涉及到获取、压缩^[4]和传输等环节.然而,由于系统资源有限,各类处理环节都可能在屏幕内容图像中引入失真,影响用户体验.例如,在图像获取过程中,由于视频采集设备的缺陷,屏幕内容图像包含噪声;由于硬件设备资源的约束,屏幕内容图像的分辨率有限.在带宽约束下,屏幕内容图像的压缩处理会引入压缩失真,降低图像的视觉质量.在传输过程中,比特位错误会引起图像内容的随机丢失,引入难以预测的失真.以上失真会影响屏幕内容图像的画质,进而影响到用户体验.当前图像质量评价的研究主要针对传统自然图像,为了能够反映系统中各环节对最终用户体验的影响,进而实现系统优化的目的,需要研究和设计针对屏幕内容图像的质量评价方法.

1.2 图像质量评价方法研究现状

图像质量评价方法分为主观评价方法和客观评价方法^[55-8].主观评价方法由观察者来对图像评分,一般采用平均主观得分(mean opinion score,简称 MOS)或平均主观得分差异(differential mean opinion score,简称 DMOS)作为失真图像的质量指标.观察者是图像视觉信息的最终接收者,因此主观评价方法是最准确的方法.但

是,主观评价方法需要耗费大量人力、物力以及时间代价并且不能嵌入到实时系统,因此,它主要在建立标准数据库时使用.客观评价方法是由相应的计算模型计算得到图像的质量指标.客观评价方法在计算机视觉和图像处理等领域中扮演着重要角色.一些图像处理和传输系统的参数需要根据图像的质量来进行调节,可以将客观评价方法嵌入到系统之中实时地优化系统的参数.图像检索系统可以根据图像的质量来排列检索的结果.图像处理算法的性能可以根据图像质量来评估,因此客观评价方法可以作为图像处理算法的基准.客观评价方法也可以动态监测和调节图像处理系统输出图像的质量.

客观评价方法根据所需无失真参考图像信息量的多少进一步可以分为全参考(full reference,简称 FR)方法、部分参考(reduced reference,简称 RR)方法和无参考(no reference,简称 NR)方法.全参考方法在计算失真图像的质量指标时需要获取失真图像的整幅无失真参考图像的全部信息,通过对比计算二者的差异得到失真图像的质量指标.部分参考方法在计算失真图像的质量指标时不需要整幅无失真参考图像的全部信息,仅需从无失真参考图像中提取的某些特征信息即可.无参考方法不需要无失真参考图像的任何信息即可计算失真图像的质量指标.由于没有无失真参考图像,无参考方法是这3类方法中最困难的方法.

当前图像质量评价研究的重点是传统自然图像.针对屏幕内容图像质量评价的研究较少.传统自然图像质量评价的全参考方法经过几十年的发展已具有坚实的理论基础和许多成熟的评价算法.其中,思想最简单且应用最广泛的是均方误差算法(MSE)和峰值信噪比算法(PSNR),这两种算法直接将失真图像对应于无失真参考图像的像素差异作为失真图像的质量指标.虽然这两种方法计算简便、快捷,但是它们计算得到的结果与人的视觉感知相关性较低.Chandler等人通过对人类视觉系统的认知并建模提出了VSNR^[9]算法.Wang等人提出的SSIM^[10]算法将计算得到的无失真参考图像和失真图像的结构相似度作为失真图像的质量指标.在文献[11]中,Zhang等人提出的FSIM算法将计算得到的无失真参考图像和失真图像的特征相似度作为失真图像的质量指标.Sheikh等人通过量化失真过程中信息的丢失并研究丢失的信息和图像感知质量之间的关系而提出了VIP^[12]方法.

部分参考方法多用于通信系统中.在发送端提取无失真参考图像的部分信息,比如特征等,然后将这部分信息通过辅助信道传送到接收端或者先将这部分数据以水印的形式嵌入到发送端的无失真参考图像之后,再与无失真参考图像一起传送到接收端以节省带宽.对于在接收端收到的失真图像,提取同样的信息.最后将在发送端提取的无失真参考图像的信息和接收端提取的失真图像的信息结合起来计算失真图像的质量指标.

传统自然图像质量评价无参考方法可以分为针对如模糊、噪声和块状效应等特定失真类型的方法及通用的方法.针对特定失真类型利用到了特定失真的某种特性,它们因此也只能用于计算受到特定失真影响的失真图像的质量指标.然而,在现实应用情景中,失真绝不会仅限于某种特定的失真,因此,这类无参考方法的应用范围有限.但通用的方法可以应用于多种失真类型,并且进行简单的扩充之后,通用的方法就具有针对新的失真类型的能力.通用的方法主要分为基于自然场景统计^[13]的方法和基于学习或训练^[14]的方法.

理论研究发现自然场景具有某些统计特性,当失真存在时会改变这些统计特性,并且不同的失真以及不同的失真程度对统计特性的改变也不同.基于自然场景统计的方法将失真图像偏离自然场景统计特性程度的多少作为失真图像的质量指标.Moorthy等人提出的DIIVINE^[15]方法首先对图像进行小波变换,然后对变换得到的系数进行统计分布的拟合,并将得到的系数作为特征.Liu等人提出的SSEQ^[16]方法首先将图像划分为许多小块,然后分别在这些小块上计算特征,最后将在所有小块上计算得到的特征聚集起来进行池化得到统计特征.这类方法在计算得到特征后再利用回归方法建立起特征与失真图像质量分数之间的关系.

基于学习或训练的方法与基于自然场景统计的方法类似,但其直接提取与图像质量相关的特征,而不是自然场景统计特征,最后利用支持向量机或神经网络等机器学习技术建立起图像特征和图像质量指标之间的关系.Li等人^[17]首先提取与图像质量相关的相位一致性、熵和梯度等特征,然后利用广义回归神经网络建立起图像特征和图像质量指标之间的关系.Ye等人首先利用提取的40维的Gabor特征向量集构建码本,然后将码字直方图作为图像特征建立起与图像质量指标之间的关系,基于此而提出了CODEBOOK^[18]方法.Gao^[19]等人受到机器学习排序思想的启发,先从可明显区分的图像对入手提取差特征向量,然后利用多核学习算法建立起差特

征向量和图像质量指标之间的关系,该方法的明显优点是不需要图像的主观得分(MOS 或 DMOS).

目前,屏幕内容图像质量评价的研究还处于起步阶段,Yang^[20]等人构建了第一个大规模屏幕内容图像质量评价数据库 SIQAD.基于此数据库,他们得出屏幕内容图像的文本区域和图像区域带给人的视觉感知特性是不同的这一结论.基于此结论,他们提出了分开考虑文本区域和图像区域,针对屏幕内容图像的全参考方法 SPQA. Ni 等人^[21]基于图像梯度的局部信息提取梯度方向特征,提出了针对屏幕内容图像的 GSS 全参考方法.Gu^[22]等人利用简单的卷积操作凸显失真引起的结构失真以及失真更容易引起的显著性区域这一结论而提出了 SQMS 全参考方法.Wang 等人^[23]根据屏幕内容图像不同于自然图像的视觉感知特性提出了 RRS 部分参考方法.

综上所述可以得出结论:目前针对屏幕内容图像的研究还较少,并且提出的方法大部分都是全参考方法.在许多应用情景中很难获取无失真参考图像,因此,屏幕内容图像的无参考评价方法研究就变得非常有意义.

本文第 2 节介绍 IML-SCIQD 数据库的构建过程,主要包括参考图像和失真类型的选取、主观评价方法的确认、具体实验的进行以及最后主观数据的处理.第 3 节介绍本文提出的针对屏幕内容图像的无参考算法的相关细节.具体有图像区域和文本区域的划分、不同区域的统计特征以及提取的区域统计特征等细节.第 4 节介绍进行算法评估时所用到的数据库以及最后的评估结果.最后对本文工作进行总结并提出未来的展望.

2 IML-SCIQD 数据库的构建

2.1 标准图像质量评价数据库的重要性及概况

标准图像质量评价数据库的建立是客观图像质量评价算法研究的基础,新提出的算法需要在数据库上进行性能评估,不同的算法也需要在数据库上进行性能优劣比较.目前,图像质量评价的研究主要还是针对传统自然图像,针对屏幕内容图像的研究还处在起步阶段.因此,目前已公开发表的图像质量评价数据库大部分都是针对传统自然图像的,例如 LIVE 数据库^[24]、CSIQ 数据库^[25]、TID 2008 数据库^[26]、TID 2013 数据库^[27].SIQAD 数据库是目前唯一一个公开发表的大规模屏幕内容图像数据库^[20].

LIVE 数据库包含 29 幅参考图像和经过 5 种失真处理的 779 幅失真图像,数据库中失真图像的主观分数为 DMOS.TID 2008 数据库包含 25 幅参考图像和经过 17 种失真类型处理的 1 700 幅失真图像.数据库中失真图像的主观分数为 MOS.TID 2013 数据库是 TID 2008 数据库的扩展,它包含 25 幅参考图像和经过 24 种失真类型处理的 3 000 幅失真图像.数据库中失真图像的主观分数为 MOS.CSIQ 数据库包含 30 幅参考图像和经过 6 种失真类型处理的 866 幅失真图像.数据库中失真图像的主观分数为 DMOS.SIQAD 数据库包含 20 幅参考图像和经过 7 种失真类型处理的 980 幅失真图像.数据库中失真图像的主观评分为 DMOS.SIQAD 数据库中参考图像的来源主要有新闻网页、电子杂志、幻灯片和 PDF 文件等.在真实的应用环境中,可能出现各种各样的情况,单一数据库难以覆盖.因此,屏幕内容图像质量评价数据库的构建就显得非常有必要.

图像质量评价数据库的构建通常涉及到参考图像的选取,失真类型的确定和失真图像的生成,主观评价方法的选取、主观实验平台的搭建以及招募主观实验人员进行主观实验,最后将得到的初始主观数据进行处理得到失真图像的主观分数.

2.2 IML-SCIQD 数据库概况

IML-SCIQD 数据库包含 25 幅参考图像以及 25 幅参考图像经过 10 种失真每种失真 5 个失真等级处理的 $1250=25 \times 10 \times 5$ 幅失真图像.这 25 幅参考图像都是通过屏幕截图得到的,它们主要来自于社交网站、购物网站、生活服务网站、电子地图、电子杂志、PDF 文件、幻灯片、电影、动画和新闻网页等,因此,参考图像的视觉内容具有一定的代表性.25 幅参考图像视觉内容布局和分辨率也不尽相同.IML-SCIQD 数据库中的 10 种失真类型是:(1) JPEG 压缩(JPEG);(2) JPEG 2000 压缩(JP2K);(3) 高斯模糊(GB);(4) 运动模糊(MB);(5) 高斯白噪声(GWN);(6) 椒盐噪声(SPN);(7) 乘性噪声(MN);(8) 对比度变化(CC);(9) JPEG 2000 比特流经过快速衰落瑞丽信道时产生的比特位传输错误(FF);(10) 屏幕内容压缩(SCC).

10 种失真类型既包含了传统自然图像相关应用中经常遇到的失真,也包含了屏幕内容图像所特有的失真.

第1种~第9种失真经常出现在与自然图像相关的应用中,当然也会出现在屏幕内容图像中.例如,图像获取、图像压缩、图像传输等.这9种失真也经常被其他数据库选作失真类型.最后一种失真是为提高屏幕内容压缩效率而特别针对屏幕内容而设计的压缩算法.

2.3 主观实验

选择合适的主观实验方法是获得稳定主观实验结果的重要基础.国际电信联盟(ITU)^[28]推荐了几种不同的主观评价方法,不同的方法适用于不同的情景.这些方法可以大致分为单刺激方法和双刺激方法.对于单刺激方法,在进行主观实验时,每次屏幕上仅出现1幅图像供实验者评分.但是对于双刺激方法,在进行主观实验时,每次屏幕上会出现两幅图像,它们是同一幅图像的不同失真版本,实验者仅需判断这两幅图像是否有视觉感知差异或哪一幅图像的视觉感知更好或更差即可.但当需要评估的图像数量较大时,需要对比的图像对的数量将难以接受.为了操作方便,本文选用单刺激连续质量评估方法(SSCQE)作为主观实验方法.每次只有1幅图像显示在屏幕上,主观实验者需要根据他自己的视觉感知,在一个连续的区间内对图像进行评分.

WEST^[29]开源软件平台为从多个地点上的多个设备上收集主观数据提供了解决方案,其优点是多个主观实验者可以同时参加主观测试,只要他们参加测试的设备满足实验条件且连接在实验的本地服务器上即可.它原本是为视频数据设计的,本文对其进行简单修改之后作为本次主观实验的软件平台.主观实验的一大缺点是耗时,随着WEST软件平台的应用将大大缩短主观实验的时间.在本文的主观实验中,所有参考图像也被包含在主观测试之中,因此整个主观实验过程一共有1275幅图像需要进行评分.这1275幅图像根据参考图像被划分为5个子集,每个子集有255幅图像,也就是5幅参考图像以及与其对应的失真图像.每一位参加实验的人员必须一次完成一个子集中所有255幅图片的评分.完成一个子集的评分工作大概耗时30分钟左右.每位实验者必须完成所有5个子集的评分工作.为了防止实验者的疲劳对实验数据造成的影响,一个实验者完成的两个子集评分之间的时间间隔必须大于1小时.

本次实验招募的35位实验者均来自于深圳大学的研究生和本科生,年龄在18岁~28岁之间.他们都没有图像处理 and 图像质量评价的相关经验.在正式的实验开始之前,每一位实验者都将接受简单的培训,以此向他们介绍图像质量评价的背景以及如何对图像评分.整个主观实验在4台安装有WIN7操作系统和谷歌CHROME网络浏览器的台式机上并行进行.所有实验图像都以原始分辨率在联想LT2223wA LED显示器上加以显示.

实验者登录WEST本地服务器之后,可开始一个子集的实验,之后,子集中的255幅图像会一张接一张地出现,直到这个子集中的最后一幅图像.每当出现一幅图像,实验者观看后鼠标单击图像即可出现图2所示的评分界面.这时实验者需要根据他们自己的视觉感受对图像评分,评分区间位于1~100之间.分数越高表明图像带给实验者自己的视觉感受越好.从图2可以看出,整个评分区间又被5个参考点划分为4个小区间,这5个参考点分别代表图像质量非常差(bad)、差(poor)、一般(fair)、好(good)、非常好(excellent).实验者在给图像评分时首先根据这些参考点确定自己大概认为图像所属的分数区间,然后再微调至自己所认为的分数.评分的操作只需用鼠标拖动评分界面的滚动条即可,当滚动条定位好后,评分界面右边会有相应位置对应的质量分数.鼠标点击VOTE就可以完成对图像的评分,然后下一幅图像就会显示在屏幕上.点击VOTE对一幅图像进行评分之后不可返回进行修改.若未对图像进行评分之前可以点击BACK返回,重新对图像进行观察之后再行评分.



Fig.2 Quality judgement interface of subjective experiment

图2 主观实验评分界面

2.4 数据处理

在得到初始主观数据之后,为了使得主观分数能够更好地表示图像的视觉感知质量,初始主观数据还要根

据文献[30]中提到的方法进行统计处理以得到图像的平均主观得分差异(DMOS).令 S_{ijk} 表示主观实验人员 i 对在子集 k 中失真图像 j 的初始主观评分.第 1 步将每一个子集内的初始主观评分 S_{ijk} 转变为差异分数 d_{ijk} :

$$d_{ijk} = S_{ij_{ref}k} - S_{ijk} \quad (1)$$

这里, $S_{ij_{ref}k}$ 表示主观实验人员 i 对子集 k 内的失真图像 j 对应的参考图像的初始主观评分.接下来,我们将每个子集内的差异分数 d_{ijk} 转变为 z 分数 z_{ijk} :

$$\mu_{ik} = \frac{1}{250} \sum_{j=1}^{250} d_{ijk} \quad (2)$$

$$\sigma_{ik} = \frac{1}{250} \sum_{j=1}^{250} (d_{ijk} - \mu_{ik})^2 \quad (3)$$

$$z_{ijk} = \frac{(d_{ijk} - \mu_{ik})}{\sigma_{ik}} \quad (4)$$

每个子集内失真图像的总数是 250.最后,我们可以得到一个 z 分数矩阵,行是主观实验人员的索引,列是失真图像的索引.

并不是所有主观实验人员的评分数据都是可信的,不可信的主观人员的评分数据将会对最后得到的失真图像的主观评分对失真图像视觉感知的表示带来偏差.得到 z 分数矩阵之后,本文采用文献[30]中的方法丢弃不可信的主观实验人员的评分数据.首先,验证某位实验人员的 z 分数数据(即上述得到的 z 分数矩阵的对应的某一行)是否服从正态分布.这可以通过计算该行数据的峰度值得到,如果计算得到的峰度值位于 2~4 以内,则认为该实验人员的评分数据服从正态分布,这时,若实验人员给出的 1 250 幅失真图像中多于 5% 的评分数据位于相对于数据均值 2 倍的标准差的范围之外,则丢弃该实验人员的全部数据.反之,则保留.如果实验人员的主观评分数据的峰度值不在区间 2~4 以内,即不服从正态分布,这时,若实验人员给出的 1 250 幅失真图像多于 5% 的数据位于相对于数据均值 4.47 倍的标准差的范围之外,则丢弃该实验人员的全部数据.反之,则保留.经过上述处理,本次实验招募的 35 位实验人员中有 3 个人的数据被丢弃.为了使得最后得到的失真图像的主观分数位于区间 1~100 以内,本文对最后丢弃相应主观实验人员评分数据后的 z 分数矩阵进行简单的线性变换,使得 z 分数矩阵中的数据均位于区间 1~100 以内.数据库中失真图像最后的主观评分 DMOS 可以通过以下公式计算得到:

$$DMOS_j = \frac{1}{32} \sum_{i=1}^{32} z_{ij} \quad (5)$$

至此,整个数据库的构建工作全部完毕.图 3 所示为数据库中全部失真图像的主观分数的直方图.

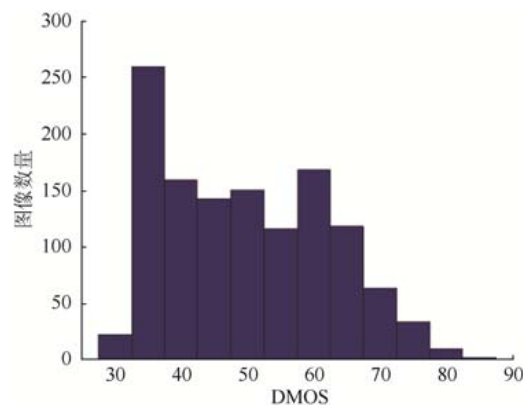


Fig.3 Histogram of quality scores of distorted images in the IML-SCIQD database

图 3 IML-SCIQD 数据库中失真图像质量分数的直方图

3 NSNRS 无参考算法

3.1 算法框架

屏幕内容图像的文本区域和图像区域带给人的视觉感受是不同的,这一点已在文献中得到证实.本文进一步发现图像区域和文本区域的统计特性也是不同的,甚至对于同一区域,当图像区域或者文本区域遭受不同失真的时候,这一区域的统计特性也是不同的.基于上述结论以及自然场景统计的无参考算法的思想,本文提出了针对屏幕内容图像的 NSNRS 无参考算法.算法框架如图 4 所示.

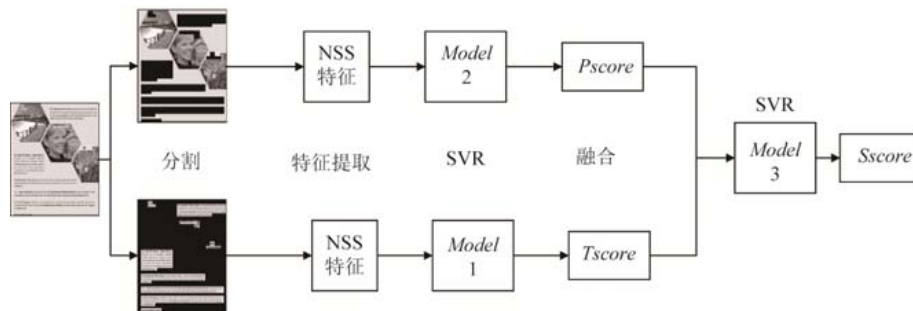


Fig.4 Framework of our proposed no reference metric

图 4 本文提出的无参考算法框架

本文提出的方法是基于学习的,因此包含训练和测试阶段.在训练阶段训练,Model 1,Model 2 和 Model 3 一共 3 个模型.Model 1 是文本区域质量分数模型,用来预测文本区域的质量分数.Model 2 是图像区域质量分数模型,用来预测图像区域质量分数.Model 3 是结合模型,用来结合文本区域和图像区域的质量分数而得到失真图像的质量分数.NSNRS 无参考算法首先将屏幕内容图像划分为图像区域和文本区域,然后在划分的两个区域上分别提取自然场景统计特征,区域的具体划分后面会加以介绍.在每一个区域上,用提取的自然场景统计特征以及对应的整幅失真图像的主观质量分数 DMOS,利用支持向量回归(support vector regression,简称 SVR)可以训练得到 Model 1 和 Model 2.反过来,分别在两个区域上提取的自然场景统计特征还要再次输入,利用它们训练的对应的模型来得到训练集中失真图像文本区域和图像区域的质量分数 T_score 和 P_score .将 T_score 和 P_score 组合成二维向量,将它们对应的整幅失真图像的主观质量分数 DMOS 作为标签,利用支持向量回归可以训练得到 Model 3.

在测试阶段,首先,屏幕内容图像同样被划分为图像区域和文本区域,然后,分别在两个区域上提取与训练阶段同样的统计特征.将提取的统计特征输入 Model 1 或 Model 2 就可以得到两个区域的质量分数.最后,将得到的两个区域的质量分数组合成二维向量输入到 Model 3,就可以得到失真图像的质量分数.

3.2 图像区域和文本区域的划分

本文基于文献[31]中的算法对屏幕内容图像进行区域的划分.为了将屏幕内容图像划分为图像区域和文本区域,首先计算屏幕内容图像的文本区域索引图 T .给定一幅屏幕内容图像,其文本区域索引图 T 是与屏幕内容图像的灰度图一样大小的矩阵,本文的工作都是在灰度级上进行的.在文本区域,索引图 T 中对应着屏幕内容图像中文本区域的值全为 1,其他地方的值全为 0.得到屏幕内容图像的文本区域索引图之后,可以非常容易地得到屏幕内容图像的图像区域索引图 P ,它仅仅是屏幕内容图像文本区域索引图的取反,也就是屏幕内容图像文本区域索引图 T 中 1 值变 0 值,0 值变 1 值就可以得到对应的屏幕内容图像的图像区域索引图 P .当得到屏幕内容图像的文本区域索引图和图像区域索引图之后,文本区域和图像区域可以通过相应的区域索引图和屏幕内容图像的点积得到.一幅屏幕内容图像和对应的文本区域索引图、图像区域索引图以及文本区域和图像区域如图 5 所示.其中,图 5(a)所示为参考图像,图 5(b)所示为文本区域索引图,图 5(c)所示为图像区域索引图,图 5(d)所

示为文本区域,图 5(e)所示为图像区域.在索引图中,黑色区域代表索引图中的 0 值,白色区域代表索引图中的 1 值.在区域图中,无论是文本区域还是图像区域,黑色区域均代表该区域的灰度值为 0.因此可知,区域图中,文本区域或图像区域,只是保留了屏幕内容图像中相应区域的灰度值而让其他区域的灰度值为 0.区域图中大量的零灰度值会对在区域图提取的特征对视觉感知的表示造成影响,在区域图上提取特征时需要排除这些零值的影响,具体细节会在后面详细加以介绍.

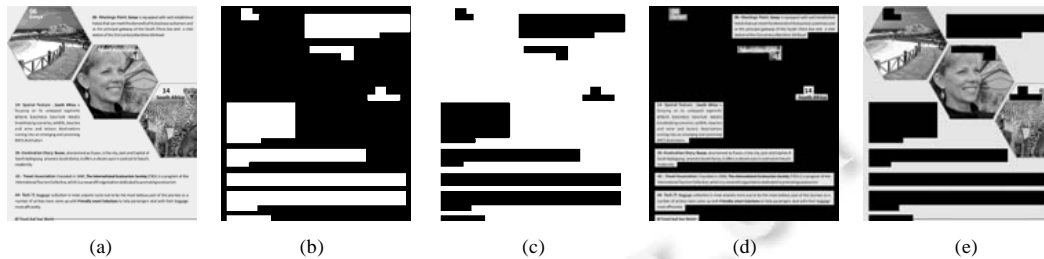


Fig.5
图 5

3.3 图像区域和文本区域的统计差异性

屏幕内容图像不同于自然图像,它一般同时包含图像和文本,然而人们对这两个区域的视觉感知是不一样的.为了探究屏幕内容图像文本区域和图像区域的统计差异性,本文选择了 IML-SCIQD 数据库中的一幅参考图像以及与其对应的 10 幅分别经过 10 种失真类型最高失真级别处理的失真图像.该参考图像如图 6 所示.这 11 幅图像首先被划分为图像区域和文本区域.由这些划分得到的总共 22 幅区域图都进行 Daubechies 9/7 小波基的小波变换.在上一小节提到区域图像中包含大量的零值,为了更好地研究区域图像的统计特性,就需要排除这些零值的影响.由于这些零值区域在变换后也会产生大量的零值和一些奇异值,因此,在变换之后,首先丢弃这些变换之后的零值和奇异值,然后再进行下一步的研究.在丢弃变换之后的零值和奇异值之后,本文分区地画出了图像区域和文本区域变换之后系数的直方图.结果如图 7 所示.

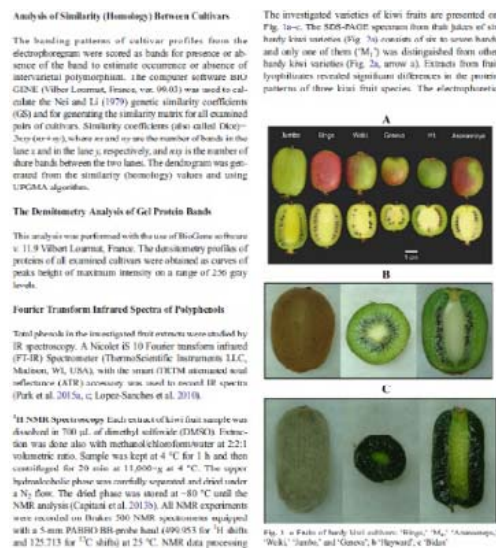


Fig.6 Reference image selected for research of statistical properties

图 6 选作统计特性研究的参考图像

从图 7 可以看出,屏幕内容图像的图像区域和文本区域的统计特性有明显的不同.文本区域的直方图相比

较图像区域的直方图要平滑一些.对于图像区域,特别是除3种噪声失真以外的7种失真,其直方图异常地尖锐,表明其大部分数据都集中于零值附近,近似服从拉普拉斯分布.对于文本区域,除了两种模糊失真的直方图较尖锐之外,其他失真的直方图都较平滑,近似服从正态分布.根据统计特性的差异,可以分开考虑屏幕内容图像的文本区域和图像区域.至于单个区域,其文本区域或图像区域当遭受不同失真的影响时,统计特性也是不同的.因此,可以在单个区域上运用基于自然场景统计方法的思想来预测单个区域的质量.

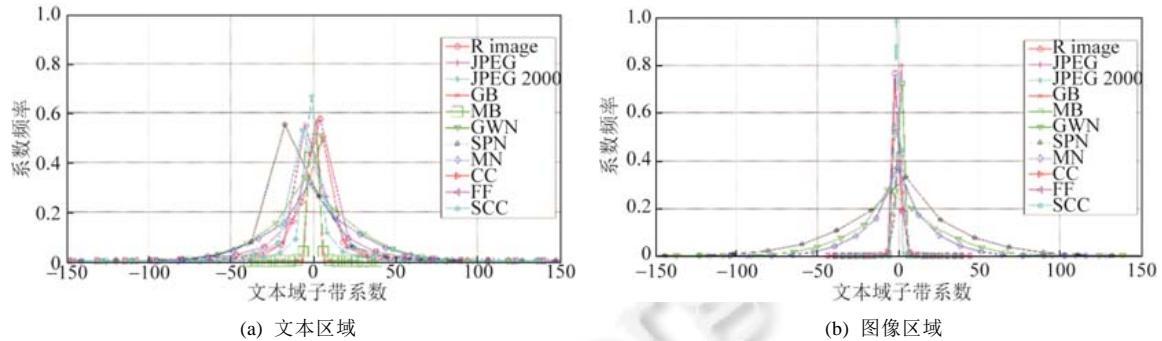


Fig.7 Histogram of transform coefficients of textual region and pictorial region of screen content image and corresponding 10 distorted ones

图7 屏幕内容图像以及与其对应的10种失真图像的文本区域和图像区域变换系数的直方图

3.4 算法中提取的自然场景统计特征

目前已有大量基于自然场景统计思想的算法被提了出来,因而在这些算法中包含了各种各样的自然场景统计特征.自然场景统计特征的设计非常耗时,而且需要足够强大的背景知识.本文没有自己设计自然场景统计特征,而是选取他人的适合于本文的自然场景统计特征.正如上面提到的划分得到的区域图像中存在大量的零灰度值,这会严重影响提取的特征对区域图像视觉感知质量的表示,因此,选取的自然场景统计特征应该可以较好地避开这些零值的影响.就如本文在探究屏幕内容图像文本区域和图像区域统计特性时所做的那样.

不论是文本区域图像还是图像区域图像,本文都提取同样的90个自然场景统计特征.这90个自然场景统计特征来自于4种经典的基于自然场景统计的无参考评价方法:SSEQ、BLIINDS-II^[32]、BIQI^[33]和BRISQUE^[34].SSEQ方法提取12个自然场景统计特征,BLIINDS-II方法提取24个自然场景统计特征,BIQI方法提取18个自然场景统计特征,BRISQUE方法提取36个自然场景统计特征.这些特征可以排除大量零灰度值的影响.

SSEQ算法先将图像分为许多小块,之后分别在每个小块上计算特征,然后将所有小块上计算得到的相同的特征聚成一个向量,最后对得到的特征向量进行池化,将池化后得到的特征作为图像最后的特征.为了摆脱零灰度值的影响,在提取SSEQ方法的特征时,在图像小块上计算完特征之后,先丢弃那些图像小块上计算得到的零值或奇异值特征,这是因为,由包含大量零值的图像小块计算得到的特征肯定也是零值或是一些奇异值.在丢弃这些图像小块上计算得到的零值或奇异值特征之后,再进行最后的池化操作,这样就可以摆脱零值的影响.BLIINDS-II方法也是先将图像分为许多小块,然后分别在这些小块上计算特征,最后将在所有小块上计算得到的相同的特征聚成一个向量进行池化,以得到最后的特征.这里摆脱上面提到的零灰度值的影响的方法与SSEQ方法一样.

BIQI方法先将图像在3个尺度和3个方向上进行小波变换,然后将变换得到的系数利用广义高斯分布进行拟合,最后将广义高斯分布拟合得到的参数作为特征.为了摆脱上面提到的零灰度值的影响,在提取BIQI方法的特征时,对于图像小波变换之后的系数先丢弃那些零值和奇异值系数,再去进行拟合,这是因为,区域图像矩阵中大量的零灰度值经过小波变换后会产生大量零值和奇异值,这样就可以摆脱零灰度值的影响.BRISQUE方法和BIQI方法类似,只不过它没有将图像先进行类似于小波变换这样的频域处理,而只是在空间域中进行了相应的处理,然后将处理得到的系数进行拟合,最后将拟合的系数作为特征.为了摆脱上面提到的零灰度值的影

响,在提取 BRISQUE 方法的特征时的处理方法与 BIQI 方法类似,这里不再赘述.

4 实验结果

4.1 测试数据库及评估标准

为了探究提出的无参考算法的性能,本文将提出的 NSNRS 无参考算法与 12 种客观评价算法在 IML-SCIQD 数据库以及 SIQAD 数据库上进行了性能对比,这 12 种客观评价算法分别是:(1) GSS;(2) SQMS;(3) RSS;(4) PSNR;(5) FSIM;(6) SSIM;(7) VIF;(8) SSEQ;(9) BIQI;(10) DIIVINE;(11) BLINDS-II;(12) BRISQUE.前两种算法是专门针对屏幕内容图像的全参考评价方法,第 3 种算法是专门针对屏幕内容图像所提出来的部分参考评价算法,第 4 种~第 7 种是经典的针对自然图像提出来的全参考算法,最后 5 种是经典的基于自然场景统计的无参考评价算法.IML-SCIQD 数据库包含参考图像 25 幅,10 种失真类型,每种失真类型又有 5 个失真等级,因此一共有 $25 \times 10 \times 5 = 1250$ 幅失真图像.SIQAD 数据库包含参考图像 20 幅,7 种失真类型,每种失真类型又有 7 个失真等级,因此一共有 $20 \times 7 \times 7 = 980$ 幅失真图像.这 7 种失真类型是:(1) 高斯噪声(GN);(2) 高斯模糊(GB);(3) 运动模糊(MB);(4) 对比度变化(CC);(5) JPEG 压缩(JPEG);(6) JPEG 2000 压缩(JPEG 2000);(7) 基于层划分的压缩(LSC).前 6 种失真经常出现在与自然图像相关的应用中,最后一种失真是为提高屏幕内容压缩效率的压缩算法.

人是图像视觉信息的最终接收者,客观评价方法性能的评估需要通过对比客观评价算法预测得到的质量分数和主观分数的相关性而得到.本文用到的评估标准有斯皮尔曼等级相关系数(the Spearman rank-order correlation coefficient,简称 SROCC)、皮尔森线性相关系数(Pearson product-moment correlation coefficient,简称 PLCC)和根均方误差(root mean square error,简称 RSME).斯皮尔曼等级相关系数衡量算法预测的单调性,其值位于 0~1 之间,值越大表示算法的性能越好.皮尔森线性相关系数衡量算法预测值与主观分数之间的线性相关程度,其值位于 0~1 之间,值越大表明算法的性能越好.根均方误差衡量算法预测值和主观分数之间的绝对差异,值越小表明算法的性能越好.

4.2 结果分析

所有参与比较的 12 种客观算法的源代码均来自于相关作者和项目的网站.因为本文所提出的算法和最后 5 种方法都是基于学习的算法,因此在数据库上进行测试时,数据库中 80%的失真图像用于训练,20%的失真图像用于测试.在 IML-SCIQD 数据库上进行测试时,20 幅参考图像对应的失真图像用于训练,剩下的 5 幅参考图像对应的失真图像用于测试.在 SIQAD 数据库上进行测试时,16 幅参考图像对应的失真图像用于训练,剩下的 4 幅参考图像对应的失真图像用于测试.由于图像内容的不同可能对算法的性能产生影响,因此,这种根据图像内容而进行的训练和测试集划分随机进行 1 000 次,最后所有 1 000 次的平均性能作为算法最后的性能.因为本文提出的算法是基于学习的,我们使用 LIBSVM 软件包实现 SVR^[35].在 Model 1 和 Model 2 的训练中,使用径向基核函数.在 Model 3 的训练中使用线性核函数.

为了消除主观评价过程中的非线性以及便于不同的图像质量评价算法之间的性能对比,MATLAB 中的非线性最小二乘回归函数 nlinfit 被用来先将客观算法的预测分数映射到 $DMOS_p$,然后再与主观分数进行相关系数的计算.映射函数是具有 5 个参数的逻辑斯蒂函数:

$$DMOS_p = \frac{p_1}{2} - \frac{p_1}{1 + \exp(p_2 \cdot (q - p_3))} + p_4 \cdot q + p_5 \quad (6)$$

p_1, p_2, p_3, p_4 和 p_5 是逻辑斯蒂函数的参数.

在 IML-SCIQD 数据库和 SIQAD 数据库上的测试结果分别见表 1 和表 2.从表 1 可以看出,本文提出的算法相比 5 种经典的基于自然场景统计的无参考算法,在整个数据库上针对 SROCC、PLCC 和 RSME 这 3 个评估标准都有一定程度的提升.对于针对某种特定失真的情况,在 CC、FF、SCC 这 3 种失真上,本文提出的算法在性能上有明显的提升.对于其他失真,本文提出的算法也有可比的性能.然而,对于全参考方法,本文提出的算法在性能上还有一定的差距,特别是在针对特定失真的情况下.因为本文提出的方法是无参考方法,无法像全参考

方法那样获取参考图像.但是考虑到整个数据库,本文提出的算法依然可以达到与全参考方法相当的性能,至于在 SIQAD 数据库上的表现,从表 2 可以看出,与在 IML-SCIQD 数据库上的结果类似,本文提出的算法在总体上均要优于其他的无参考方法.至于全参考方法的情况,本文提出的算法性能总体上还是与在 IML-SCIQD 数据库上的情况一样,而且在某些情况下,本文提出的算法的性能甚至超越了全参考方法.比如对于整个数据库而言,本文提出的算法的性能超过了 PSNR 和 FSIM 算法.

Table 1 Test result on the IML-SCIQD database

表 1 IML-SCIQD 数据库上的测试结果

	Distortion	GSS ¹	SQMS ¹	RRS ²	PSNR ³	FSIM ³	SSIM ³	VIF ³	SSEQ ⁴	BIQI ⁴	DIIVINE ⁴	BLIINDS-II ⁴	BRISQUE ⁴	NSNRS
SROCC	JPEG	0.821	0.888	0.788	0.851	0.863	0.899 1	0.906 3	0.2533	0.243 8	0.110 5	0.331 7	0.155	0.461 8
	JP2K	0.867 7	0.946 3	0.887 2	0.917 4	0.943 5	0.922 2	0.945 7	0.3758	0.137 2	0.225 1	0.589 8	0.385 2	0.636 9
	GB	0.932 5	0.918 8	0.952 7	0.892	0.887	0.938 4	0.925 3	0.1377	0.744 4	0.607 4	0.202	0.697	0.624 2
	MB	0.886 5	0.864 8	0.944 9	0.778 3	0.785 9	0.919 8	0.847 5	0.1348	0.621 6	0.574 3	0.163 6	0.623 2	0.576 5
	GWN	0.847 2	0.916 9	0.922 8	0.882 8	0.849 6	0.904	0.918 6	0.8197	0.907 7	0.810 8	0.694 2	0.886 3	0.884 8
	SPN	0.278 6	0.862 3	0.862 3	0.778 5	0.737 3	0.842 4	0.847 9	0.7693	0.826 7	0.755 9	0.659 7	0.816	0.868
	MN	0.796	0.862	0.889 4	0.804 3	0.827 9	0.824	0.872 5	0.7184	0.856 6	0.729	0.629 4	0.796 9	0.835 5
	CC	0.219 3	0.493 7	0.438	0.606 7	0.554 2	0.257 9	0.485 3	0.2006	0.514 5	0.361 5	0.166 7	0.189 5	0.398 7
	FF	0.680 7	0.860 5	0.763 9	0.839 8	0.877 1	0.857 8	0.869 4	0.517	0.322 5	0.329	0.363 8	0.405 4	0.543 4
	SCC	0.726 5	0.817 7	0.736 7	0.797 1	0.794 2	0.792 7	0.809 1	0.13	0.155 3	0.182	0.178	0.225 8	0.266 4
	ALL	0.833 1	0.890 6	0.842 3	0.762 1	0.807 9	0.881 4	0.862 9	0.5597	0.665 6	0.630 3	0.555	0.651 6	0.731 6
PLCC	JPEG	0.837	0.902 3	0.825 6	0.719 9	0.872	0.891 5	0.941 3	0.3426	0.266 5	0.224 2	0.444 7	0.225 8	0.439 2
	JP2K	0.884 7	0.961 4	0.894 9	0.841 1	0.953 1	0.930 2	0.956 8	0.4063	0.256	0.288 6	0.594 1	0.405 1	0.645 8
	GB	0.952 5	0.932 4	0.969 9	0.777 7	0.885 7	0.957 7	0.945 2	0.2313	0.780 6	0.622 4	0.337	0.744 2	0.661 4
	MB	0.898 1	0.890 2	0.950 6	0.783 9	0.803 5	0.921	0.876 5	0.2046	0.616 6	0.565 8	0.282 9	0.655 7	0.586 5
	GWN	0.872 7	0.959	0.954 5	0.949 4	0.934 1	0.944 7	0.960 1	0.8709	0.942 2	0.902 2	0.72	0.942 2	0.903 9
	SPN	0.678 3	0.901 1	0.891 4	0.870 7	0.808 9	0.882 6	0.893 3	0.8123	0.897 6	0.832 4	0.724 4	0.895 2	0.899 2
	MN	0.88	0.940 3	0.945	0.917 7	0.923 9	0.919 4	0.948 9	0.7863	0.901 7	0.848 1	0.677 8	0.850 9	0.821 7
	CC	0.292 1	0.597 9	0.430 4	0.739 4	0.685	0.379 7	0.505 2	0.2438	0.694 7	0.500 9	0.226	0.251 7	0.615 6
	FF	0.86	0.942 6	0.914 9	0.915 8	0.951 8	0.940 1	0.940 3	0.6439	0.422 6	0.398 2	0.491 5	0.429 4	0.648 2
	SCC	0.816 8	0.897 2	0.814 9	0.873 4	0.905 7	0.868 1	0.931 4	0.2081	0.225 9	0.242 5	0.317 1	0.273 3	0.387 4
	ALL	0.863 8	0.897 5	0.856 3	0.777 6	0.815 6	0.886 1	0.871 9	0.5921	0.455 1	0.567 1	0.556 3	0.602 4	0.710 3
RMSE	JPEG	4.164 9	3.281 1	4.295 4	5.282 7	3.725 8	3.448 7	2.570 4	6.8723	7.080 6	7.211 4	6.558 5	7.272 4	6.552 3
	JP2K	5.754 8	3.396 6	5.507 5	6.676 9	3.737 1	4.530 5	3.587 4	10.768	11.68 7	11.384	9.193 1	10.751	8.743 5
	GB	3.254	3.861	2.603 4	6.715 4	4.959 6	3.074 4	3.487 7	10.218	6.053 8	7.285 6	9.735 2	6.604 4	7.462 9
	MB	3.655 6	3.786 3	2.581 2	5.161	4.947 9	3.238	4.000 7	7.9817	6.082 4	6.066 6	7.754 1	5.767 3	6.368 6
	GWN	5.242 9	3.042 6	3.203 8	3.373 5	3.834 5	3.520 9	3.001 7	4.8147	3.323 1	4.381 9	6.814 8	3.173 3	3.589 1
	SPN	5.877 6	3.469 1	3.625 2	3.939 6	4.703 7	3.760 2	3.596	4.2604	3.360 9	4.164 6	5.145 5	3.393	3.337 3
	MN	4.725 7	3.385 6	3.254 6	3.954 1	3.807 4	3.914 4	3.140 1	5.7079	3.480 8	5.047 5	6.833 8	4.502 8	4.267 6
	CC	6.281 2	5.264 4	5.928 4	4.421 5	4.785	6.075 8	5.667 8	6.1562	4.192 4	4.976 2	6.247 3	6.143 9	4.703 6
	FF	5.875 1	3.845 2	4.647 1	4.625 1	3.532 8	3.923 5	3.917 3	8.1352	9.814 9	9.957 1	9.301 6	9.665 9	7.926 6
	SCC	4.493 8	3.440 2	4.514 7	3.793 6	3.301 6	3.865 8	2.835	7.4015	7.414	7.255 6	7.050 3	7.208 1	6.870 8
	ALL	6.170 8	5.401 7	6.325 7	7.699 9	7.086 3	5.675 8	5.996 7	9.2651	9.589	9.186 3	9.489	8.963	8.045 7

说明:注 1 为针对屏幕内容图像的全参考评价算法;注 2 为针对屏幕内容图像所提出来的部分参考评价算法;注 3 为经典的针对自然图像提出来的全参考算法;注 4 为经典的基于自然场景统计的无参考评价算法

Table 2 Test result on the SIQAD database

表 2 SIQAD 数据库上的测试结果

	Distortion	GSS ¹	SQMS ¹	RRS ²	PSNR ³	FSIM ³	SSIM ³	VIF ³	SSEQ ⁴	BIQI ⁴	DIIVINE ⁴	BLIINDS-II ⁴	BRISQUE ⁴	NSNRS
SROCC	GN	0.852	0.886	0.866 4	0.879	0.737 3	0.869 4	0.888 8	0.816 1	0.870 9	0.878 7	0.780 2	0.887 5	0.855 4
	GB	0.904 9	0.914 9	0.871 5	0.857 3	0.728 6	0.892 1	0.905 9	0.809 6	0.812 9	0.862 6	0.703 8	0.871 5	0.892 5
	MB	0.839	0.869 5	0.843 4	0.713	0.664 1	0.804 1	0.849 2	0.774 1	0.661 9	0.840 7	0.501 8	0.865 3	0.825 3
	CC	0.449 7	0.694 9	0.529 1	0.682 8	0.717 5	0.640 5	0.643 3	0.215 4	0.414 4	0.386 8	0.333 8	0.313 5	0.461 8
	JPEG	0.795 9	0.789 3	0.760 5	0.756 9	0.587 9	0.757 6	0.792 4	0.321	0.327 3	0.292 2	0.219 8	0.220 8	0.466 4
	JP2K	0.813 7	0.819 4	0.661 7	0.774 6	0.636 3	0.760 3	0.813 1	0.269 2	0.443 4	0.491 5	0.498 4	0.301 8	0.584 4
	LSC	0.814 2	0.829 3	0.711 6	0.793	0.597 9	0.737 1	0.846 3	0.193 8	0.230 2	0.217 4	0.185 9	0.184 4	0.412 6
	ALL	0.835 6	0.880 3	0.765 5	0.560 4	0.527 9	0.756 6	0.806 5	0.569 6	0.639 5	0.655 8	0.526	0.662 2	0.737 4
PLCC	GN	0.879 9	0.900 4	0.879 8	0.905 3	0.742 8	0.880 6	0.901 1	0.827 2	0.884 6	0.904 4	0.727 2	0.904 5	0.871 7
	GB	0.908 5	0.912 6	0.881	0.860 3	0.720 6	0.901 4	0.910 2	0.826 9	0.84	0.869 6	0.695 8	0.890 9	0.910 4
	MB	0.839 9	0.867 3	0.846 5	0.704 4	0.687 4	0.806	0.849	0.786 5	0.682 9	0.841 2	0.515 5	0.857 1	0.839 6
	CC	0.61	0.802 7	0.681 2	0.740 1	0.750 7	0.743 5	0.707 6	0.307 7	0.533 2	0.522 2	0.445 2	0.541 7	0.554 4

Table 2 Test result on the SIQAD database (Continued)
表 2 SIQAD 数据库上的测试结果(续)

	Distortion	GSS ¹	SQMS ¹	RRS ²	PSNR ³	FSIM ³	SSIM ³	VIF ³	SSEQ ⁴	BIQI ⁴	DIIVINE ⁴	BLIINDS-II ⁴	BRISQUE ⁴	NSNRS
PLCC	JPEG	0.793 5	0.785 7	0.763 8	0.754 5	0.556 6	0.748 7	0.798 6	0.414 9	0.323 3	0.337 8	0.319 8	0.288 7	0.492
	JP2K	0.820 8	0.826 3	0.680 7	0.789 3	0.667 5	0.774 9	0.820 5	0.363 7	0.49	0.515 2	0.512 1	0.350 9	0.595
	LSC	0.806 8	0.812 5	0.711	0.780 5	0.596 4	0.730 7	0.838 5	0.337 8	0.279	0.287 6	0.258 6	0.286 5	0.442 4
	ALL	0.851 5	0.887 2	0.801 4	0.586 9	0.538 9	0.756 1	0.819 8	0.595	0.492 1	0.520 5	0.534 6	0.359 1	0.784 8
RMSE	GN	7.085 8	6.490 6	7.089 5	6.337 2	9.986	7.067 9	6.467 3	7.507	6.350 9	6.150 4	9.006 5	6.013 7	6.957 7
	GB	6.341	6.204 1	7.179 2	7.737 6	10.523	6.570 1	6.285 9	7.958 9	7.617 2	6.935 2	9.934 5	6.479 2	5.931 6
	MB	7.054 7	6.472 2	6.921 2	9.228 7	9.443 2	7.696 7	6.870 4	7.541 6	8.815 2	6.515 8	10.353	6.198 2	6.676 3
	CC	9.966 6	7.501 3	9.208 4	8.459 1	8.319	8.411 6	8.887 6	11.486	9.817 4	9.586 9	10.65	9.966	9.694 2
	JPEG	5.717 7	5.812 4	6.065 4	6.166 5	7.807 2	6.229 5	5.655 1	8.134 3	8.388 3	8.412 7	8.443 9	8.603 7	7.572 3
	JP2K	5.935 5	5.854	7.614 1	6.381 9	7.740 4	6.569 1	5.941 2	8.988 2	8.202 3	8.154 1	7.983 5	9.094	7.422 6
	LSC	5.039 7	4.973 3	5.999 2	5.333 6	6.848 6	5.825 3	4.649 7	7.716	7.852 8	7.854 2	7.980 3	7.871 2	7.266 2
	ALL	7.504 5	6.603 9	8.562	11.589	12.058	9.367 6	8.196 9	10.888	11.287	10.871	11.538	11.717	8.694 1

说明:注 1 为针对屏幕内容图像的全参考评价算法;注 2 为针对屏幕内容图像所提出来的部分参考评价算法;注 3 为经典的针对自然图像提出来的全参考算法;注 4 为经典的基于自然场景统计的无参考评价算法

5 总结与展望

目前屏幕内容图像质量评价的研究还处于起步阶段,SIQAD 数据库是目前仅有的一个大规模的屏幕内容图像质量评价数据库,单一数据库难以覆盖实际应用中出现的各种可能的情况.图像质量评价数据库是客观图像质量评价算法研究的基础,本文因此首先构建了一个大规模的具有一定代表性的屏幕内容图像质量评价数据库 IML-SCIQD.在参考图像的视觉内容、视觉内容分布以及失真类型种类等属性上,IML-SCIQD 数据库都具有一定的代表性.IML-SCIQD 数据库中所有失真图像的质量分数覆盖了分数区间的大部分区域,这说明,IML-SCIQD 数据库中的失真图像都考虑到了各种可能的失真等级的情况.通过对失真图像主观质量分数和失真参数之间关系的分析,本文了解到图像的视觉质量与图像的视觉内容之间具有一定的关系.在进行屏幕内容图像相关课题,如针对屏幕内容图像的压缩算法的研究时需要将其考虑进来.

基于在构建 IML-SCIQD 数据库过程中的主观经验,本文进一步证实了屏幕内容图像的文本区域和图像区域带给人们的视觉感知特性的不同,并且发现屏幕内容图像的文本区域和图像区域的统计特性也是不同的.单个区域的统计特性在不同失真情况下所受到的影响也随之变化.与 12 种客观评价算法的性能对比表明,本文提出的算法与 5 种无参考评价算法相比在性能上有一定提升.相比 7 种全参考算法,本文提出的算法的性能虽然还有一定差距,但仍然具有相当的性能,并且在个别情况之下的性能甚至超越了全参考方法.

本文利用基于自然场景统计的无参考方法的思想来评估单个区域的视觉质量,以后的工作中可以考虑其他的方法.将屏幕内容图像分区域分析是考虑到了各自区域不同的视觉感知特性.未来的工作中可以从屏幕内容图像整体的视觉感知特性出发.深度学习方法目前已被广泛应用于计算机视觉、语音识别、自然语言处理、音频识别与生物信息学等领域,并获取了令人满意的效果.未来的工作中运用深度学习的方法来进行屏幕内容图像质量评价的研究也是一个不错的方向.

References:

- [1] Lu Y, Li SP, Shen HF. Virtualized screen: A third element for cloud-mobile convergence. *IEEE Multimedia*, 2011,18(2):4-11. [doi: 10.1109/MMUL.2011.33]
- [2] Shen HF, Lu Y, Wu F, Li SP. A high-performance remote computing platform. In: *Proc. of the IEEE Int'l Conf. on Pervasive Computing and Communications*. 2009. 1-6. [doi: 10.1109/PERCOM.2009.4912855]
- [3] Yang H, Fang YM, Yuan Y, Lin WS. Subjective quality evaluation of compressed digital compound images. *Journal of Visual Communication and Image Representation*, 2015,26:105-114. [doi: 10.1016/j.jvcir.2014.11.001]
- [4] Zhu WJ, Ding WP, Xu JZ, Shi YH, Yin BC. Screen content coding based on HEVC framework. *IEEE Trans. on Multimedia*, 2014,16(5):1316-1326. [doi: 10.1109/TMM.2014.2315782]

- [5] Jiang GY, Huang DJ, Wang X, Yu M. Overview on image quality assessment methods. *Journal of Electronics & Information Technology*, 2010,32(1):219–226 (in Chinese with English abstract). <http://jeit.ie.ac.cn/CN/abstract/abstract13977.shtml> [doi: 10.3724/SP.J.1146.2009.00091]
- [6] Wang ZM. Review of no-reference image quality assessment. *Acta Automatica Sinica*, 2015,6:1062–1079 (in Chinese with English abstract). <http://www.aas.net.cn/CN/abstract/abstract18682.shtml> [doi: 10.16383/j.aas.2015.c140404]
- [7] Vipin K, Bhurchandi KM. No-Reference image quality assessment algorithms: A survey. *Optik-Int'l Journal for Light and Electron Optics*, 2015,126(11):1090–1097. [doi: 10.1016/j.ijleo.2015.02.093]
- [8] Manap RA, Shao L. Non-Distortion-Specific no-reference image quality assessment: A survey. *Information Sciences*, 2015,301:141–160. [doi: 10.1016/j.ins.2014.12.055]
- [9] Chandler DM, Hemami SS. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. on Image Processing*, 2007,16(9):2284–2298. [doi: 10.1109/TIP.2007.901820]
- [10] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. on Image Processing*, 2004,13(4):600–612. [doi: 10.1109/TIP.2003.819861]
- [11] Zhang L, Zhang L, Mou XQ, Zhang D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. on Image Processing*, 2011,20(8):2378–2386. [doi: 10.1109/TIP.2011.2109730]
- [12] Sheikh HR, Bovik AC. Image information and visual quality. *IEEE Trans. on Image Processing*, 2006,15(2):430–444. [doi: 10.1109/TIP.2005.859378]
- [13] Moorthy AK, Bovik AC. Statistics of natural image distortions. In: *Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. 2010. 962–965. [doi: 10.1109/ICASSP.2010.5495298]
- [14] Xu L, Lin WS, Jay Kuo C-C. *Visual Quality Assessment by Machine Learning*. Singapore: Springer-Verlag, 2015. 1–142. [doi: 10.1007/978-981-287-468-9]
- [15] Moorthy AK, Bovik AC. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. on Image Processing*, 2011,20(12):3350–3364. [doi: 10.1109/TIP.2011.2147325]
- [16] Liu LX, Liu B, Huang H, Bovik AC. No-Reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 2014,29(8):856–863. [doi: 10.1016/j.image.2014.06.006]
- [17] Li CF, Bovik AC, Wu XJ. Blind image quality assessment using a general regression neural network. *IEEE Trans. on Neural Networks*, 2011,22(5):793–799. [doi: 10.1109/TNN.2011.2120620]
- [18] Ye P, Doermann D. No-Reference image quality assessment using visual codebooks. *IEEE Trans. on Image Processing*, 2012,21(7):3129–3138. [doi: 10.1109/TIP.2012.2190086]
- [19] Xue WF, Zhang L, Mou XQ. Learning without human scores for blind image quality assessment. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2013. 995–1002. [doi: 10.1109/CVPR.2013.133]
- [20] Yang H, Fang YM, Lin WS. Perceptual quality assessment of screen content images. *IEEE Trans. on Image Processing*, 2015, 21(11):4408–4421. [doi: 10.1109/TIP.2015.2465145]
- [21] Ni ZK, Ma L, Zeng HQ, Cai CH, Ma K-K. Gradient direction for screen content image quality assessment. *IEEE Signal Processing Letters*, 2016,23(10):1394–1398. [doi: 10.1109/LSP.2016.2599294]
- [22] Gu K, Wang SQ, Yang H, Lin WS, Zhai GT, Yang XK, Zhang WJ. Saliency-Guided quality assessment of screen content images. *IEEE Trans. on Multimedia*, 2016,18(6):1098–1110. [doi: 10.1109/TMM.2016.2547343]
- [23] Wang SQ, Gu K, Zhang XF, Lin WS, Ma SW, Gao W. Reduced-Reference quality assessment of screen content images. *IEEE Trans. on Circuits and Systems for Video Technology*, 2016,PP(99):1–1. [doi: 10.1109/TCSVT.2016.2602764]
- [24] Sheikh HR, Wang Z, Cormack L, Bovik AC. LIVE image quality assessment database release 2. 2005. <http://live.ece.utexas.edu/research/quality>
- [25] Larson EC, Chandler DM. Categorical image quality (CSIQ) database. 2009. <http://vision.okstate.edu/csiq>
- [26] Ponomarenko N, Egiazarian K. Tampere image database 2008 TID2008. 2008. <http://www.ponomarenko.info/tid2008.html>
- [27] Ponomarenko N, Jeremeiev O, Tampere V. Image database (TID2013). 2013. <http://www.ponomarenko.info/tid2013.html>

- [28] ITU, International Telecommunication Union. BT.500: Methodology for the subjective assessment of the quality of television pictures. 2002. <https://www.itu.int/rec/R-REC-BT.500>
- [29] NTIA/ITS. Web-Enabled subjective test (WEST) software. 2014. <http://www.its.bldrdoc.gov/resources/video-quality-research/web-enabled-subjective-test-west.aspx>
- [30] Seshadrinathan K, Soundararajan R, Bovik AC, Cormack LK. Study of subjective and objective quality assessment of video. IEEE Trans. on Image Processing, 2010,19(6):1427–1441. [doi: 10.1109/TIP.2010.2042111]
- [31] Yang H, Lin WS, Deng CW. Image activity measure (IAM) for screen image segmentation. In: Proc. of the IEEE Int'l Conf. on Image Processing. 2012. 1569–1572. <https://sites.google.com/site/yanghuan689/>
- [32] Moorthy AC, Bovik AC. Blind image quality assessment: From scene statistics to perceptual quality. IEEE Trans. on Image Processing, 2011,20(12):3350–3364. [doi: 10.1109/TIP.2011.2147325]
- [33] Moorthy AK, Bovik AC. A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters, 2010,17(5):513–516. [doi: 10.1109/LSP.2010.2043888]
- [34] Mittal A, Moorthy AK, Bovik AC. No-Reference image quality assessment in the spatial domain. IEEE Trans. on Image Processing, 2012,21(12):4695–4708. [doi: 10.1109/TIP.2012.2214050]
- [35] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001. <http://www.csie.ntu.edu.tw/~cjlin/index.html>

附中文参考文献:

- [5] 蒋刚毅,黄大江,王旭,郁梅.图像质量评价方法研究进展.电子与信息学报,2010,32(1):219–226. <http://jeit.ie.ac.cn/CN/abstract/abstract13977.shtml> [doi: 10.3724/SP.J.1146.2009.00091]
- [6] 王志明.无参考图像质量评价综述.自动化学报,2015,6:1062–1079. <http://www.aas.net.cn/CN/abstract/abstract18682.shtml> [doi: 10.16383/j.aas.2015.c140404]



朱映映(1976—),女,山东莒南人,博士,研究员,主要研究领域为多媒体信息处理,多媒体信息安全.



王旭(1986—),男,博士,研究员,CCF 专业会员,主要研究领域为视频编码压缩,图像,视频质量评价.



曹磊(1989—),男,硕士,主要研究领域为图像质量评价.