

文章编号: 1006-4710(2014)01-0085-06

基于感兴趣区域图像眼动数据库的建立

闫苗苗^{1,2}, 郑元林^{1,2}, 戚永红^{1,2}

(1. 西安理工大学 印刷包装工程学院, 陕西 西安 710048;

2. 陕西省印刷包装工程重点实验室, 陕西 西安 710048)

摘要: 针对人眼对图像不同区域关注程度不同的现象, 本论文设计眼动实验, 由 30 名观察者自由观看 55 幅选自主流图像质量评价数据库以及印刷标准彩色图像数据库中的图像, 建立了基于图像感兴趣区域的 XROIEF 数据库。该数据库中包含原图、蜂群图、注视图、热点图及簇图各 55 幅, 且包含其相应的数据信息, 如注视顺序、各注视点位置信息、各感兴趣区域的眼动数据等, 并通过对眼动数据分析得出了人眼对不同内容图像不同的感兴趣区域。

关键词: 感兴趣区域; 眼动; 图像质量评价; 图像数据库

中图分类号: TS805.4 **文献标志码:** A

Establishment of image eye fixation database based on ROIs

YAN Miaomiao^{1,2}, ZHENG Yuanlin^{1,2}, QI Yonghong^{1,2}

(1. Faculty of printing and packaging engineering, Xi'an University of Technology, Xi'an 710048, China;

2. Key Lab of Printing and Packaging Engineering of Shaanxi Province, Xi'an 710048, China)

Abstract: For the phenomenon that human eyes concern differently with different regions of one image, this paper designs an eye fixation experiment conducted by 30 subjects freely watching 55 images chosen from main-stream database of image quality evaluation and standard colorful image database in printing engineering, on the basis of which the database XROIEF based on the eye fixation database of the region of interest(ROI) on images has been established. This database includes 55 original images, bee swarm maps, gaze plot maps, heat maps, cluster maps in each kind, including the corresponding relevant data information of the images, such as fixation index, the position information of each fixation, the eye fixation data of every ROI etc. Furthermore, from the fixation data analysis, the ROIs received from human eyes of different images under different semantics categories are obtained.

Key words: region of interest (ROD); eye fixation; image quality evaluation; image database

在这个高速发展的信息时代, 人类对信息传递准确性的要求越来越高, 图像作为传递信息较为丰富有效的载体, 已经成为我们生活中必不可少的元素。据研究, 人类对外界信息的感知约有 83% 来自于视觉感知, 而其中 70%~80% 是以图像形式获得的。相对于其他信息, 图像作为传播载体有着直观、准确、高效等优点。然而数字图像在获取、压缩、处理、传输和再现的过程中, 由于会受到不同种类的降

质^[1], 图像会发生失真, 无法达到图像再现中“所见即所得”的目的, 因此, 一幅图像是否满足我们的视觉需求并较好地完成从原图像到目标载体上的再现, 则需要一个好的图像质量评价的方法来进行评判和指导。

目前, 图像质量评价的方法主要有主观评价和客观评价两种^[2]。根据对原始图像信息的依赖程度, 客观质量评价可分为全参考、部分参考和无参考

收稿日期: 2013-09-03

基金项目: 陕西省教育厅科学技术研究计划资助项目(09JK681); 陕西省教育厅重点实验室科学研究计划资助项目(13JS082)。

作者简介: 闫苗苗, 女, 硕士生, 研究方向为印刷图像质量客观评价。E-mail: yanmiaomiao163@163.com。郑元林, 男, 副教授, 硕士, 研究方向为印刷品质量评价、印刷色彩复制理论及应用。E-mail: zhengyuanlin@xaut.edu.cn。

三类。

由于一些传统的图像处理算法针对的是整幅图像,忽略了人眼对图像中不同区域的不同感受,因此在有些应用场合算法并不是很准确。而实际上在不同的应用中,图像各部分区域具有不同的重要性,从而使得观察者对各部分感受不尽相同。对人眼影响较大的区域称为感兴趣区域(Region of Interest, ROI),而相应的其他区域则称为非感兴趣区域或背景区域(Background, BC)^[3]。在一幅图像中,感兴趣区域最能影响观察者的兴趣,也最能表现图像的内容,因此当图像失真时,人眼往往对感兴趣区域的失真更为敏感。

如图1所示,(a)为Lenat图的原图,(b)为对(a)中除过脸部的区域进行高斯模糊得到的效果,(c)为对(a)中脸部区域进行同样参数高斯模糊的效果。相比之下,人眼的感受为(c)图的失真程度较大,而(b)图的失真程度较小,因此将感兴趣区域应用到图像质量评价中去,将更能有针对性地对图像质量进行评价和分析。



图1 Lenat图的原图及其不同区域的失真图
Fig.1 Original and distortion image from different regions of the image Lenat

人类观察一幅图像中的感兴趣区域时,注意既是自上而下(Top-down),任务驱动的,也是自下而上(Bottom-up),图像数据驱动的。即在观察图像时,该注意不仅由人的高层知识所控制,用由意志控制的与任务相关的方式来扫描场景(自上而下),也由外部刺激对人注意力的吸引所影响,用基于重要性的与任务无关的形式来扫描场景(自下而上)。

而在图像数据库建立方面,目前数据信息较为丰富完善的是MIT眼动数据库和NUSEF数据库。2009年,Tilke Judd等^[4]建立了MIT眼动数据库,该实验得出人眼的注意力集中在屏幕中央25%的区域,如图2所示,人眼对图像的注视程度以图像25%的区域为中心,依次向四周递减。2010年,Subramanian Ramanathan等^[5]建立了基于眼动数据的NUSEF数据库,该数据库将图像按照语义学范畴分类。该实验得出人类和哺乳动物肖像的脸部

吸引着人眼大部分的注意力,特别是眼睛、鼻子和嘴巴的区域。包含有动作(读、写、看等)的图像,其注视点在交互物体之间有着频繁的移动,动作发生者和动作接受者之间有着很多过渡的发生。

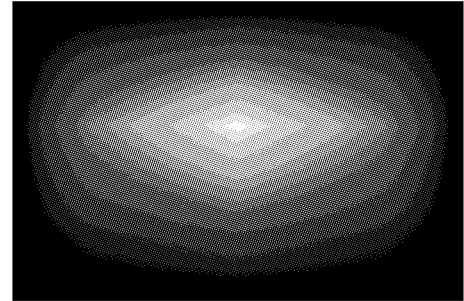


图2 注意力分布权重图
Fig.2 Weight map of attention distribution

通过学习和参照MIT眼动数据库以及NUSEF数据库,本文在图像质量评价的方向上,进一步对图像的眼动数据进行分析,建立了一个基于人眼感兴趣区域的图像数据库,为今后结合人眼感兴趣区域的图像质量评价方法提供一些借鉴和参照。

1 眼动实验

1.1 图像的选择

本论文将所建数据库命名为XROIEF数据库(XUT Region of Interest eye fixation 西安理工大学感兴趣区域眼动数据库)。在建立数据库的实验中使用了瑞典Tobii公司生产的Tobii X2-60眼动仪,被试者为30名年龄在20~35岁的本科生、研究生及教师,对55幅图像进行观察实验,并使用Tobii眼动仪对眼动数据进行记录。该数据库中的55幅图片分别选自IVC数据库^[6](9幅)、CSIQ数据库^[7](13幅)、Live数据库^[8-10](13幅),以及用于印刷的GB/T 18721-2002 CMYK标准彩色图像数据库和ISO 12640数据库(20幅)。所有图像均保证其像素在 512×512 以上,以确保用户在观看时可以完全理解图像的内容。表1所示为按图像的来源所分类整合XROIEF数据库中的所有原图像。

为确保数据库的全面性,XROIEF数据库涵盖了包含不同内容的图像。根据图像的场景语义(图像内容)划分,该数据库的组成为:人物肖像图(包括雕塑)13幅,动物肖像图6幅,风景图8幅,建筑物图3幅,静物图20幅,动作图5幅,具体如表2所示。

表1 XROIEF 图像数据库原图
Tab.1 Original images from XROIEF image database

图像数据库及相应图像信息	所选图像
IVC 数据库 (选取 9 幅图像, 均为 bmp 格式)	
CSIQ 数据库 (选取 13 幅图像, 均为 png 格式)	
Live 数据库 (选取 13 幅图像, 均为 bmp 格式)	
CMYK 标准彩色图像数据库 (用于印刷) (选取 20 幅图像, 均为 tif 格式)	

表2 数据库按图像内容的分类结果
Tab.2 The database classification based on semantics

图像场景语义	图像描述	图像数
人物肖像图(包括雕塑)	单个或多个人物的肖像(包括脸部、身体以及背景等)	13
动物肖像图	单个或多个动物的肖像(包括脸部、身体以及背景等)	6
风景图	室外风景,包括山水,或相对整幅图像较小人物或动物	8
建筑物图	图中以大篇幅的建筑物为主	3
静物图	包括飞机、船、汽车、花卉、杂物等静物图	20
动作图	有动作交互的对象或动作正在发生(读、看等动作)	5

1.2 实验过程

在眼动实验中,30 名被试者在未被告知实验意图的前提下,自由观看 55 幅随机排序的图像,以最小化外界因素对被试的心理干扰,从而确保实验的典型性和准确性。实验环境为普通的办公环境,照明及实验场地都无特殊限制。实验所用图像大小不一,分别以宽和高的像素为衡量标准,较小的有 512 * 512 像素、480 * 720 像素,较大的有 1280 * 1024 像素和 960 * 1200 像素,由于显示器需要满足所有图像按其原图大小显示,所以实验选用 Samsung SyncMaster T220P 显示器,屏幕大小为 22 in,

分辨率为 1920 * 1200,未做过色彩管理。将每幅图像观看的持续时间设定为 5 s,每两幅中间插入一幅为时 1 s 的灰色纯色图像(中心有定标符号),以降低上幅图像对人眼的影响。该灰色图像中心的定标符号意在将人眼的注视点从显示器的周边聚焦到中心,尽可能高质量的记录和提取有效的眼动数据。由于实验所用图像均不能覆盖整个显示器屏幕,因此图像周围部分以黑色显示,尽可能少的干扰人眼对图像的注视情况。为保证尽可能多并准确地采集到眼动数据,观察时,被试者与显示屏幕的距离保持在 45~90 cm 之间。实验所使用的眼动仪安装于显

示器屏幕下方,体积小,连接简单,拥有 50 cm × 36 cm 的自由头动范围,其采样率为 60 Hz。对眼睛注视点进行校正时采用五点定标,定标完成后观察者随意观看图像,其注视点和扫视点,以及注视路径和扫视路径由眼动仪的红外感受摄像头记录,并由其外接的处理器进行处理。经过 Tobii 公司所研发的 Tobii Studio 软件的计算,得到每个注视点的坐标和注视时长等有用数据,以及图像的热点图、集簇图和注视图等能直观表现人眼感兴趣区域的分析图像。

该数据库包括每幅图像的蜂群图、注视图、热点图及集簇图。对于数据库中的所有图片,经过 Tobii 眼动仪的记录和 Tobii Studio 软件的计算,得出其相应的热点图、集簇图以及注视图等,以数据库 XROIEF 中的图像 1n 为例,如图 3 所示。

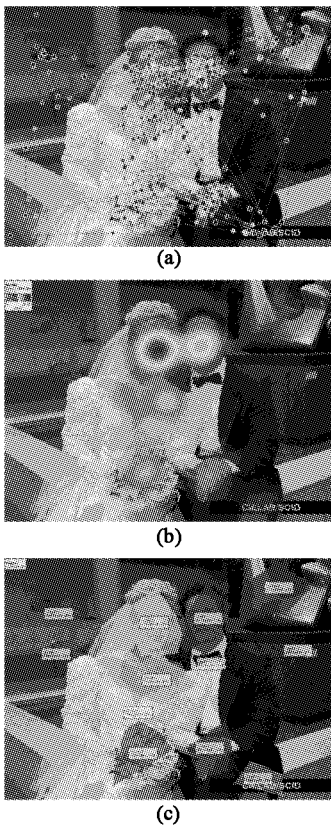


图 3 图 1n 的注视图、热点图以及集簇图
Fig. 3 The gaze plot map, heat map and cluster map of the image 1n

图 3(a)为注视图(Gaze plot map),图中不同大小的圆点记录了被试者观察该图像时的注视点位置及其注视轨迹,圆点的大小表示注视时间,圆点越大,则表明注视时间越长,反之亦然。圆点中的数字则表示被试者观察时的眼动顺序,该图为 30 名被试者观察图像时注视点的叠加。图 3(b)为该图像的热点图(Heat Map),如图所示其以热点的形式直观表现出了人眼对图像的关注区域,热点处的颜色深度表示人眼对该位置区域的关注程度,颜色越深表

示人眼在该区域的注视点越多、注视时间越长,颜色深度之间包含很多过渡层次。图 3(c)中所示是该图像的集簇图(Cluster),如图中所示,人眼对图像的关注区域以不同颜色的不规则图形所标注出,在生成集簇图时,本文根据需要,在 Tobii Studio 软件中将距离基准值(Distance Threshold)设置为 15% (默认为 50%),以划分出较为精细并符合研究需要的集簇(关注区域)。将由注视点数据生成的集簇图导入兴趣区域 AOIs(Areas of Interest)中去,可由此获得不同兴趣区域的注视点数据,如该兴趣区的注视时长、注视点个数、访问时长、访问次数及注视百分比等。其中,首次进入兴趣区用时是指被试者第一次注视到一个兴趣区所用的时间,时间计算开始于包含兴趣区的图像首次呈现时,结束于被试者的注视点出现在该兴趣区时;首次进入前的注视点个数是指被试者在首个注视点进入兴趣区之前的注视点个数,是从包含兴趣区的图像呈现开始直到被试的注视点第一次出现在兴趣区为止。首次进入用时越短,进入前的注视点个数越少,说明被试者对该兴趣区越敏感,越易被此兴趣区吸引,反之亦然。注视点持续时间指一个兴趣区中所有注视点的持续时间之和;注视点个数指的是被试者在一个兴趣区投入的注视点个数;总访问时间指访问某个兴趣区时所持续的时间总长。注视点持续、总访问时间越长,注视点个数越多,则说明被试者对该区域越感兴趣。注视百分比指注意到该兴趣区域的人数占有所有被试者人数的百分比,百分比越大,说明看到该兴趣区域的人数越多。通过此数据,可以进而进行兴趣区域中眼动数据的统计和分析。图 4 显示了图像 1n 集簇图中新娘脸部 Cluster2 和新娘身后的背景 Cluster7 所在的具体位置及大小信息。

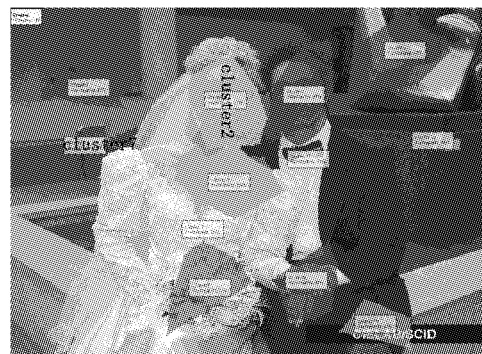


图 4 Cluster2 和 cluster7 在图 1n 的位置信息
Fig. 4 The position information of cluster2 and cluster7 in image 1n

表 3 所示为 Tobii Studio 软件对 Cluster2 和 Cluster7 这两个兴趣区域所计算而得的部分眼动数据,包括 30 人观看图片时各眼动数据的平均值

(Mean)、最大值(Max)、最小值(Min)、总值(Sum)、中值(Median)以及标准差(Stedv)。由该表数据分析可知人眼对 Cluster2 的感兴趣程度明显比 Clus-

ter7 高出很多。由此,通过对每幅图注视点数据的整合分析,可以得到人眼对该图的感兴趣区域分布情况及感兴趣程度。

表 3 不同兴趣区域眼动数据示例
Tab. 3 Example of fixation data in different ROIs

兴趣区域(Cluster)	注视数据						
	首次进入兴趣区用时/s	首次进入前注视点个数/个	注视点持续时间/s	注视点个数/个	总访问时间/s	注视百分比/%	
Cluster2	Mean	0.48	0.93	1.04	3.93	1.11	97
	Max	1.79	3	2.5	12	2.86	100
	Min	0	0	0	0	0	0
	Sum	13.91	27	31.06	118	33.28	2 900
	Median	0.29	1	0.97	4	1.12	100
	Stedv	0.48	0.84	0.71	2.79	0.76	18
Cluster7	Mean	3.91	12.25	0.02	0.13	0.02	13
	Max	4.61	14	0.2	1	0.2	100
	Min	2.85	10	0	0	0	0
	Sum	15.65	49	0.55	4	0.55	400
	Median	4.1	12.5	0	0	0	0
	Stedv	0.82	1.71	0.05	0.35	0.05	35

2 数据分析

分别对 13 幅人物图像、6 幅动物图像、8 幅风景图像、3 幅建筑图像、20 幅静物图像和 5 幅动作图像的眼动数据做统计分析,以下以人物图像为例做数据统计及分析,其它类别的图像与其分析方法相同。通过分析可得以下几点。

1) 在描述人物的图像中,人眼对人脸的感兴趣程度最高。将 13 幅人物图分为 9 幅单人图和 4 幅多人图,如表 4 所示,将 9 幅单人图像中的人脸关注区域和图像中除过人脸区域关注最多的兴趣区域

(次关注区)分别做平均,选取以下指标的平均值(对 30 个被试者的眼动数据做平均)来做数据统计。如表 4 中数据所示,人脸区域的首次进入兴趣区用时和首次进入前注视点个数均少于次关注区域,说明相较其它区域,人脸区域能够更快吸引人眼的注意力;在注视点持续时间、注视点个数及总访问时间上,人脸区域明显多于次关注区,这说明人眼对人脸区域关注最多,关注时间最长;在注视百分比上,97%的人关注到了人脸区域,而只有 71%的人关注到了次关注平均区,这说明关注人脸的人数最多,基本所有人都关注到了人脸区域。

表 4 人脸区域与次关注区域的眼动数据均值对比

Tab. 4 The comparison of fixation data in mean values between face regions and the second attention regions

区域类别	首次进入兴趣区用时/s	首次进入前注视点个数/个	注视点持续时间/s	注视点个数/个	总访问时间/s	注视百分比/%
人脸区域平均	0.534	0.968	1.687	6.079	2.114	97
次关注区平均	1.385	3.474	0.379	1.835	0.458	71

2) 当图像中包含有多个人物时,通过对各个人脸兴趣区域的数据比较,可得出结论:人脸位置对关注度的影响最大,位置靠近图像中央 25%处的人脸受关注度最高,而两边的关注度则较低,其关注程度从中央 25%处向四周递减;其次是人脸肤色的影响,肤色与其周围环境或衣服等的对比度(反差)越大,首次进入兴趣区用时就越短,首次进入前的注视点个数越少,从而被试者对该人脸就越敏感,该人脸区域受关注度就越高,而通过数据还发现,人眼在肤色为白色的人脸区域注视时间相对黄色和黑色人脸较长,注视点个数较多。

或动物的风景图像,人眼最感兴趣的是含有人物和动物的兴趣区域,其次由于位置的影响,人眼主要关注的是图像中央 25%的位置区域;再次是色彩和明度、对比度较大的位置区域。对于背景比较杂乱的风景图、静物图和建筑图,影响最大的是位置因素,图像中央 25%处到四周关注程度递减,其次则是色彩和明度、对比度较大的区域引起的人眼关注度也较高。对于有动作交互的图像(如读书的人与书),人眼最感兴趣的是交互物体(如书和人),在交互物体上会有很多注视点,其之间有很多注视路径和扫视路径。

3) 与人物图像相似,对于动物图像,人眼关注度最高,也就是最感兴趣的是动物脸的区域,尤其是眼睛、鼻子和嘴巴等区域;对于包含面积较小的人物

3 XROIEF 数据库建立

通过眼动实验,本论文建立了基于视觉感兴趣区

域的 XROIEF 数据库,该数据库包含原图像 55 幅,经过眼动实验所得的注视图、蜂群图、热点图、集簇图各 55 幅,每幅集簇图各集簇区域中的眼动数据(包括首次进入兴趣区用时、首次进入前的注视点个数、首次注视时长、注视点持续时间、注视点个数、访问时间、访问次数及注视百分比等),以及观察图像时的注视点位置等信息(图像的像素位置信息、左右眼到眼动仪的直线距

离、左右眼扫视点在屏幕上的水平和垂直坐标等),上述数据信息均以图像内容为依据,划分为人物肖像图、动物肖像图、风景图、建筑物图、静物图以及动作图六种类别,分类整理记录于 Excel 表格中,以便于以后查询和进一步研究,并有利于以后扩展和分类管理数据库。

表 5 所示为该数据库所包含的数据信息。

表 5 XROIEF 数据库的组成及其各部分所含信息
Tab. 5 Establishment of XROIEF database and the information of each part

图像类别	数量/幅	所包含的数据信息
原图	55	图像名称、图像的长度和宽度、在整个屏幕上所处的像素位置点。
蜂群图	55	图像注视点位置及关注程度的直观表现(通过注视点的密集程度来表现)。
注视图	55	每个注视点的注视顺序和注视时长、注视点在图像上的像素位置、左右眼扫视点及其平均值的像素及坐标信息、左右眼相对于屏幕位置的 3D 位置坐标、左右眼到眼动仪的直线距离、左右眼的瞳孔尺寸、眼动仪捕捉对于每个扫视点左右眼数据的有效性等。
热点图	55	人眼对图像的注视位置及关注程度的直观表现(通过兴趣区域的颜色深浅程度来表现)。
集簇图 (AOI 区域)	55	首次进入兴趣区的时间、首次进入前的注视点个数、首个注视点的注视时间、每个注视点的持续时间、注视点持续时间总和、注视点个数、每次访问兴趣区的时长、对某兴趣区的访问总时长、访问次数、注视百分比等。

4 结论及展望

本文主要建立了基于视觉感兴趣区域的 XROIEF 数据库,数据库包括原图、蜂群图、注视图、热点图和集簇图,以及其相应的数据信息。

由于 XROIEF 数据库中的图像选择是建立在图像质量评价的主流图像数据库以及用于印刷的标准图像基础之上的,因此对基于人眼感兴趣区域的图像质量评价有着很大的参考价值。数字图像在获取、压缩、处理、传输和重建的过程中,由于受到不同种类的降质,会产生压缩、噪声、模糊、对比度衰减等失真,而在印刷行业中,由于印刷环境、印刷材料及印刷设备等因素的影响,会造成印刷图像的失真,有该数据库中的眼动数据等作为参考,可以建立并进一步优化图像质量评价模型,以对基于内容的数字图像质量评价和印刷图像质量的在线评价做出一定的贡献。

致谢:论文中的眼动实验学术租赁了 Tobii 公司生产的 Tobii X2-60 眼动仪,在此特别感谢 Tobii 公司对该实验的大力支持;特别感谢西安交通大学张昀博士给予的支持和帮助!

参考文献:

- [1] 王保云. 图像质量客观评价技术研究[D]. 合肥:中国科学技术大学,2010.
Wang Baoyun. Research on digital image objective quality assessment[D]. Hefei: University of Science and Technology of China, 2010.
- [2] 刘书琴,毋立芳,宫玉,等. 图像质量评价综述[J]. 中国科技论文在线,2011,6(7):501-506.

- [3] 杨琬,吴乐华,李淑云,等. 基于感兴趣区域的图像质量评价方法[J]. 计算机应用,2008,28(5):1310-1312.
Yang Wan, Wu Lehua, Li Shuyun, et al. Method of image quality assessment based on region of interest[J]. Computer Application, 2008, 28(5): 1310-1312.
- [4] Judd Tilke, Ehinger Krista, Durand Fréd, et al. Learning to predict where humans look[C]//2009 IEEE 12th international conference on Computer Vision. IEEE, 2009: 2106-2113.
- [5] Ramanathan Subramanian, Katti Harish, Sebe Nicu, et al. An eye fixation database for saliency detection in images[C]//Computer Vision-ECCV 2010. Springer, 2010:30-43.
- [6] Ninassi Alexandre, Le Callet Patrick, Atrusseau Florent. Pseudo no reference image quality metric using perceptual data hiding[C]//Electronic Imaging 2006. International Society for Optics and Photonics, 2006: 60570G-60570G-12.
- [7] Larson Eric C, Chandler Damon M. Most apparent distortion: full-reference image quality assessment and the role of strategy[J]. Journal of Electronic Imaging, 2010, 19(1): 011006-011006-21.
- [8] Sheikh H R, Wang Z, Cormack L, et al. LIVE image quality assessment database release 2[EB/OL]. <http://live.ece.utexas.edu/research/quality>,2013-12-18.
- [9] Sheikh Hamid R, Sabir Muhammad F, Bovik Alan C. A statistical evaluation of recent full reference image quality assessment algorithms[J]. IEEE Transactions on Image Processing, 2006, 15(11): 3440-3451.
- [10] Wang Zhou, Bovik Alan C, Sheikh Hamid R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

(责任编辑 王卫勋)