

DOI:10.19322/j.cnki.issn.1006-4710.2019.02.013

# 基于视觉感知高度相关的图像质量评价

于淼淼<sup>1,2</sup>, 郑元林<sup>1,2</sup>, 廖开阳<sup>1,2</sup>, 唐棣森<sup>1,2</sup>

(1. 西安理工大学 印刷包装与数字媒体学院, 陕西 西安 710048;

2. 西安理工大学 陕西省印刷包装工程重点实验室, 陕西 西安 710048)

**摘要:** 评估并监控图像质量是数字图像处理技术的基础工作。客观图像质量评价(IQA)旨在通过计算机开发与人眼视觉感知密切相关的算法。本文充分模拟人眼视觉系统(HVS)和大脑机制,提出了一种新的基于机器学习的全参考型图像质量评价模型CGDR。该模型融合了图像的色度特征、梯度特征、对比敏感度函数(CSF)特征以及 Gauss 差分(DOG)频带特征。其中,改进后的梯度算法不仅包含更丰富的相邻信息和多方向边缘信息,同时强调了参考图像和失真图像的边缘相关性。在三个基准图像数据库上的实验结果表明,CGDR 的预测性能优于八种主流方法,跨数据库测试体现出其强大的鲁棒性,预测结果能够与人眼主观感知保持高度一致性。

**关键词:** 图像质量评价; 人眼视觉系统; 梯度特征; Gauss 差分模型

**中图分类号:** TS8

**文献标志码:** A

**文章编号:** 1006-4710(2019)02-0224-10

## Image quality evaluation based on high correlation of visual perception

YU Miaomiao<sup>1,2</sup>, ZHENG Yuanlin<sup>1,2</sup>, LIAO Kaiyang<sup>1,2</sup>, TANG Zhisen<sup>1,2</sup>

(1. School of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China; 2. Key Lab of Printing and Packaging Engineering of Shaanxi Province, Xi'an University of Technology Xi'an 710048, China)

**Abstract:** Evaluating and monitoring image quality is the basic work of digital image processing technology. The objective image quality assessment (IQA) aims to develop algorithms closely related to human visual perception. This paper fully simulates the human visual system (HVS) and brain mechanism, with a new machine learning-based full reference image quality assessment (FR-IQA) model CGDR proposed. This model combines the chrominance features, gradient features, contrast sensitivity function (CSF) features and difference of Gaussian (DOG) band features of the image. Among them, the improved gradient algorithm not only contains richer adjacent information and multi-directional edge information, but also emphasizes the edge correlation between reference image and its distorted version. The experimental results from the three benchmark image databases show that the proposed method has a better prediction performance than the current eight mainstream approaches. The cross-database validation shows its strong robustness, which is highly consistent with the human subjective perception.

**Key words:** image quality assessment; human visual system; gradient feature; difference of Gaussian model

数字信息时代拉近了人与人之间的距离,影响和改变着人们的生活方式。随着多媒体通信系统的日益发展,图像质量评价(IQA)已然成为业界研究者关注的焦点,对于图像处理、图像压缩编码、视频编码技术等意义重大。IQA旨在量化人眼对图

像质量的视觉感知,包括主观评价法和客观评价法。前者是评价者通过观察待评估的失真图像从而给出主观感知得分,是最符合人眼视觉系统(Human Visual System, HVS)视觉感知的评价方法,但是由于该方法需要耗时耗力耗财的人工完成,因此无

**收稿日期:** 2018-09-10

**基金项目:** 国家自然科学基金资助项目(61671376);陕西省自然科学基金资助项目(2016JM6022)

**作者简介:** 于淼淼,女,硕士生,研究方向为图像质量评价。E-mail:731705019@qq.com

**通讯作者:** 郑元林,男,副教授,硕导,研究方向为图像质量评价、印刷色彩复制理论及应用。E-mail:zhengyuanlin@xaut.edu.cn

法用于大规模实时系统。基于此,客观评价法应运而生,其基本思想是通过计算机建立数学模型从而模拟人眼对图像的视觉判断,目前被广泛用于实时监控设备、图像处理算法中的参数优化等。

依据对参考图像的依赖程度,客观图像质量评价方法可分为全参考型(FR)、部分参考型(RR)以及无参考型(NR)。近几年,FR和NR受到广泛关注,也是应用最多的两类方法,并且已有研究<sup>[1]</sup>将FR-IQA应用到了NR-IQA中,旨在借助前者预测准确率高的特性来增强NR-IQA算法性能。本文主要研究FR-IQA。

现有的FR-IQA依据算法设计原理的不同大致分为两种类型,一类是传统的基于计算引导的方法,另一类是基于机器学习的方法。早期的FR-IQA,如峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和均方根误差(Mean Squared Error, MSE),主要通过计算参考图像及其失真版本在像素水平上的差异来评价失真图像质量,但这种方法假定图像像素之间是独立存在的,这显然违背了HVS在观察图像时的视觉感知机制。随后,SSIM(Structural Similarity Index Milarity)<sup>[2]</sup>作为里程碑式的方法,极大推动了IQA的研究进展,对于之后的很多研究成果意义重大。由于图像的最终接收者是人,一个性能优越的IQA方法必然是通过模拟人眼视觉特性来实现的,因此从SSIM提出后,便掀起了通过考虑人眼感知特性来设计评价算法的热潮,即通过从不同角度模拟HVS的不同功能来提取低水平图像特征。MS-SSIM<sup>[3]</sup>利用HVS的多尺度特性对SSIM进行改进,评价准确性也得到了进一步提升。FSIM<sup>[4]</sup>提取图像的相位一致性特征以及梯度特征,并将相位一致性特征作为加权函数来强调其在全局图像中的重要性。VSI<sup>[5]</sup>在提取梯度特征的同时也考虑到了图像的视觉显著性(VS)特征,并将其作为加权函数,VS特征描述了一幅图像中能够引起人眼注意的局部区域,能够与主观感知有高度的相关性。VSM<sup>[6]</sup>在VSI的基础上,将梯度及VS特征的提取方法进行改进,表现出更高的预测结果。大部分方法均在空间域中提取图像特征,Balanov等<sup>[7]</sup>尝试提取图像的变换域特征,并提出了一种基于DCT子带相似性的IQA算法。HVS的不同功能决定了其具有可同时处理空域和变换域图像特征的特性,因此SC-QI<sup>[8]</sup>将色度特征、对比敏感度函数(CSF)及结构对比度指标(SCI)进行融合,同时提取空域和变换域图像特征,充分模拟了HVS对局部图像特征和各种失真类型的感知机制。

近年来,随着各领域对机器学习的研究进一步深入,基于机器学习的IQA方法变得越来越广泛,也开发出了很多较为成熟的回归工具。Pei等<sup>[9]</sup>通过提取Gauss频带特征建立基于随机森林(Random Forest, RF)的回归模型,所提出的Gauss差分(DOG)模型可与现有其他方法相结合,极大提高了模型的预测性能。Ding等<sup>[10]</sup>和Narwaria等<sup>[11]</sup>分别通过基于支持向量回归(Support Vector Regression, SVR)及奇异值分解(Singular Value Decomposition, SVD)的方法建立质量回归模型。

研究表明,基于机器学习的IQA方法在图像质量预测精度以及鲁棒性等方面均显示出绝对优势,并且一个性能优越的基于机器学习的IQA算法主要表现在两个方面:一是通过模拟HVS对视觉场景的视觉感知来提取图像特征,包括空域特征和变换域特征;二是通过模拟大脑机制建立质量回归模型。因此,本文提出了一种新的基于视觉感知高度相关的机器学习算法CGDR。

## 1 算法设计

首先,为充分模拟HVS观察视觉场景时的感知机制,在色度通道中提取参考图像与失真图像的颜色信息,在亮度通道中提取梯度特征、CSF特征以及5级DOG频带特征;然后计算两幅图像间的色度相似性、梯度相似性、对比敏感度相似性以及各级频带特征相似性。为降低特征空间维度、控制运算成本,池化策略分别提取色度相似性图、梯度相似性图以及对比敏感度相似性图的均值、标准偏差及熵;最后,通过随机森林RF进行多特征融合并训练回归模型,得到质量预测得分。CGDR的总体算法框架图见图1。

### 1.1 色度相似性特征

就彩色图像而言,人眼对其颜色空间的感知特性是质量评估的重要组成部分<sup>[12]</sup>。色度特征是其中一个重要特征之一,它直接影响了彩色图像的视觉效果。当图像颜色发生饱和和失真时,提取图像的颜色特征就显得尤为重要<sup>[13]</sup>。从另一个角度讲,一幅彩色图像是由色度与亮度组成的,而大部分图像特征主要集中在亮度通道上,因此,提取色度特征的结果必然是将图像的亮度通道与色度通道进行分离。HVS在观察真实场景时也总是将亮度和色度分开处理的,为了模拟HVS的这种功能,本文首先将图像由RGB颜色空间转化为YIQ三个通道,实现亮度信息Y与两个色度信息I、Q的分离。色彩转换方式具体如下:

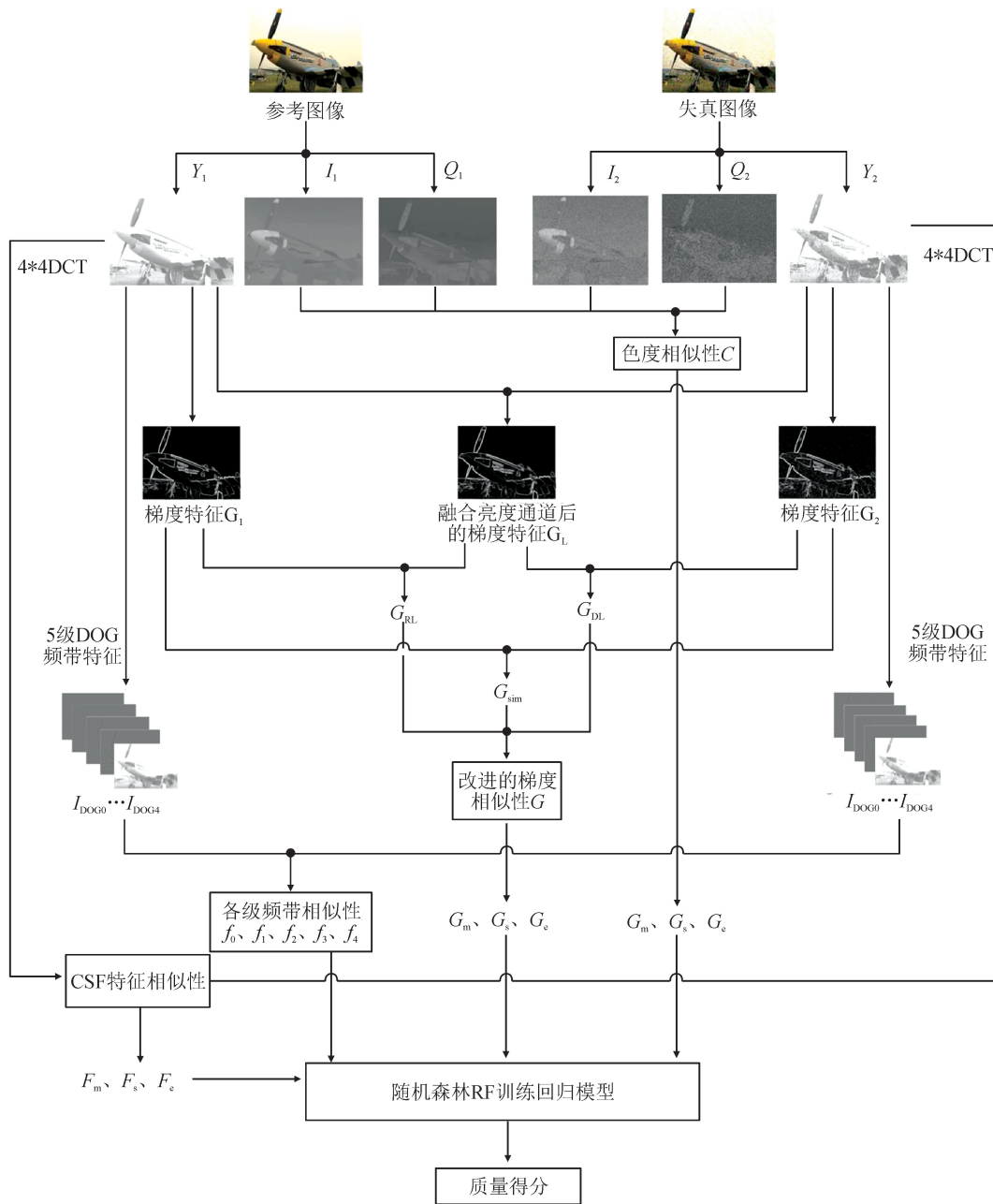


图 1 提出算法 CGDR 计算流程

Fig. 1 Computational process of the proposed IQA index CGDR

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

自 SSIM 结构相似性指标被提出后,之后的大部分 IQA 算法都采用一种特定的形式来计算特征之间的相似度,这种特定形式可表示为:  $(2ab + c)/(a^2 + b^2 + c)$ ,其中  $a, b$  表征待比较的两个物理量,  $c$  为常量。这种计算方法具有很强的掩模效应,当物理量(如:亮度和对比度)增强的同时,这些物理量之间的感知差异会变小,即物理量之间的相似度会增大,这是符合 HVS 的视觉特性的。基于此,本文同样采用这种特定形式来计算两幅图像之间的色度相似性,将参考图像(失真

图像)的  $I$  和  $Q$  通道分别定义为  $I_1(I_2)$  和  $Q_1(Q_2)$ ,则参考图像及其失真图像的色度相似性计算为:

$$C(X) = \frac{2I_1(X) \cdot I_2(X) + C_1}{I_1^2(X) + I_2^2(X) + C_1} \cdot \frac{2Q_1(X) \cdot Q_2(X) + C_2}{Q_1^2(X) + Q_2^2(X) + C_2} \quad (2)$$

式中:  $C_1, C_2$  均为正常量,根据经验取  $C_1 = C_2 = 130$  用于防止分母为零而造成  $C$  的不稳定。  $X$  为  $I, Q$  通道中的像素点。

### 1.2 改进的梯度相似性特征

对于一幅空间图像而言,边缘附近像素点的灰度变化是最剧烈的,在边缘附近的失真要比在纹理

或者平滑区域的失真更容易引起人眼的注意。梯度作为描述图像边缘特征最常用的指标,可以有效地描述由于结构和对比度失真而引起的图像变化。传统的梯度计算方法主要包括 Sobel 算子、Schar 算子、Prewitt 算子等,它们的共同点是采用包含水平和垂直两个方向的模板来计算图像梯度,并且模板的最大尺寸为  $3 \times 3$ ,这会存在两个问题:首先,只考虑水平和垂直两个方向的梯度特征是不准确的,因为图像的边缘分布是非常丰富的,很难用两个方向来描述所有边缘的变化情况;其次,模板太小则无法包含足够多的相邻信息。随后,很多方法<sup>[6, 14]</sup>在此基础上进行了改进,采用包含四个方向且尺寸更大的梯度模板对图像做卷积,并选择四个方向梯度分

量的最大值作为最终提取的图像梯度特征(记为最大分量法)。虽然使用这种方法的效果要比传统方法好很多,但是选择四个方向上的最大值并不是最佳的选择,因为另外三个方向的信息对于梯度的形成同样至关重要。本文后面会对最大分量法与我们所提出的梯度方法进行比较。

基于以上研究,本文采用一种新的计算方法来提取图像的梯度特征,在使用多方向大尺寸卷积模板的基础上,通过联合四个方向上的梯度信息共同捕获梯度特征,尽可能多的保留边缘细节。图 2 所示为四方向高通滤波模板,模板大小为  $5 \times 5$ ,包含  $0^\circ$ 、 $90^\circ$ 、 $45^\circ$ 、 $135^\circ$  四个方向。

0	0	0	0	0	0	1	0	-1	0	0	0	1	0	0	0	0	1	0	0
1	3	8	3	1	0	3	0	-3	0	0	0	8	3	0	0	0	3	8	0
0	0	0	0	0	0	8	0	-8	0	1	3	0	-3	-1	-1	-3	0	3	1
-1	-3	-8	-3	-1	0	3	0	-3	0	0	0	-3	-8	0	0	-8	-3	0	0
0	0	0	0	0	0	1	0	-1	0	0	0	-1	0	0	0	0	-1	0	0
(a) $M_x$					(b) $M_y$					(c) $M_u$					(d) $M_v$				

图 2 四方向高通滤波模板 (a) $0^\circ$ ; (b) $90^\circ$ ; (c) $45^\circ$ ; (d) $135^\circ$

Fig. 2 High-pass filter templates in four directions (a)  $0^\circ$ ; (b)  $90^\circ$ ; (c)  $45^\circ$ ; (d)  $135^\circ$

具体计算方法如下:

$$G_k(X, Y) = \frac{1}{16} \sum_{i=1}^5 \sum_{j=1}^5 I(X-3+i, Y-3+j) \times M_k(i, j) \quad (3)$$

式中:  $G_k(X, Y)$  (其中  $k$  分别代表  $x, y, u, v$ ) 表示分别使用  $M_k(k=x, y, u, v)$  模板所得到的四个梯度分量。 $(X, Y)$  表示梯度分量中各个像素点的横纵坐标值。

然后,定义  $G_1(X, Y)$  为  $0^\circ$  和  $90^\circ$  两个方向的梯度分量,定义  $G_2(X, Y)$  为  $45^\circ$  和  $135^\circ$  两个方向的梯度分量,并计算如下:

$$\begin{cases} G_1(X, Y) = \sqrt{G_x^2(X, Y) + G_y^2(X, Y)} \\ G_2(X, Y) = \sqrt{G_u^2(X, Y) + G_v^2(X, Y)} \end{cases} \quad (4)$$

最终,将一幅图像的全局梯度特征计算为两个梯度分量的加和,即:

$$GMaP(X, Y) = G_1(X, Y) + G_2(X, Y) \quad (5)$$

图 3 是对使用最大分量法与所提出的梯度分量加和法进行的比较。所使用的图片来源于 TID2013 数据库,图 3(a) 为原始图像,图 3(b) 和图 3(c) 为图 3(a) 对应的两种失真类型的失真图像,图 3(d) ~ (f) 分别为图 3(a) ~ (c) 使用最大分量法得到的梯度图,图 3(g) ~ (i) 分别为图 3(a) ~ (c) 使用梯度分量加和法得到的梯度图。从图中可以明显的观察到,

无论哪种失真类型,使用所提出的方法计算得到的梯度特征要比使用最大分量法捕获更丰富的边缘信息,从而有效地反映出了图像在边缘处的特征变化。

通过采用这种新的梯度计算方法,分别得到参考图像及其失真版本的梯度特征,记为  $GMaP_1(X, Y)$  和  $GMaP_2(X, Y)$ 。与计算色度相似性的原理类似,这里使用相同的方法计算参考图像及其失真版本之间的梯度相似性,过程如下:

$$G_{sim}(X, Y) = \frac{2GMaP_1(X, Y) \cdot GMaP_2(X, Y) + C_3}{GMaP_1^2(X, Y) + GMaP_2^2(X, Y) + C_3} \quad (6)$$

式中:  $C_3$  为一个正常量,根据经验取  $C_3 = 386$  用于防止分母为零而造成  $G_{sim}$  的不稳定。

当失真图像产生多余边缘时,尤其是当增加的多余边缘颜色与参考图像对应像素点的颜色差异较大时,那么独立的计算参考图像和失真图像各自的梯度特征是有一定缺陷的,因为这种方法会误认为该边缘是参考图像经失真后保留下来的正确的边缘,造成误判。为了弥补此类计算缺陷,将参考图像与失真图像的边缘相关性融合到所提出的梯度相似性的计算过程中。具体做法是,首先将参考图像与失真图像的亮度通道进行融合,见式(7),再通过式(3)~(5),计算融合亮度通道后的图像梯度特征,记为  $G_L$ 。与 SSIM 中计算相似性类似,接下来通过式

(8)~(9)分别计算  $G_L$  与参考图像的梯度特征  $GMap_1$  以及失真图像的梯度特征  $GMap_2$  之间的相似性,分别用  $G_{RL}$  和  $G_{DL}$  表示。其中,  $C_4$  为正常量,根据经验取  $C_4 = 55$  用于稳定式(8)~(9),防止分母为零。

$$L = (L_1 + L_2)/2 \quad (7)$$

$$G_{RL}(X) = \frac{2GMap_1(X) \cdot G_L(X) + C_4}{GMap_1^2(X) + G_L^2(X) + C_4} \quad (8)$$

$$G_{DL}(X) = \frac{2GMap_2(X) \cdot G_L(X) + C_4}{GMap_2^2(X) + G_L^2(X) + C_4} \quad (9)$$

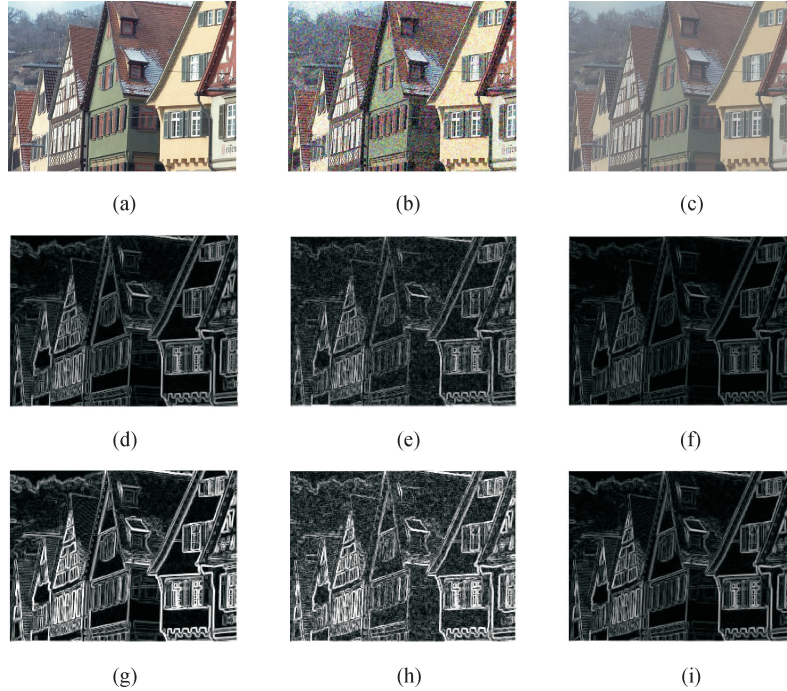


图 3 两种梯度计算方法的比较

Fig. 3 Comparison of two gradient calculation methods

最后,改进后的梯度相似性特征见式(10),与只采用梯度分量加和法得到的梯度相似性特征  $G_{sim}$  相比,该公式增加的部分  $(G_{DL} - G_{RL})$  将突出强调参考图像经失真后所丢失的边缘多于其失真版本所增加的边缘,降低对平滑边缘处的强调。

$$G = G_{sim} + [G_{DL} - G_{RL}] \quad (10)$$

### 1.3 DOG 频带特征

由于梯度特征只强调了图像中的少部分信息,而图像大部分的能量主要集中在低频区域,因此,只在空间域中提取特征不足以准确描述整幅图像。本文采用 Gauss 差分(DOG)模型,通过提取不同频带特征来模拟 HVS 可同时处理空域和频域特征的特性。

DOG 响应用于描述具有不同标准偏差的两个 Gauss 响应之间的差值,见式(11),其中  $G_{\sigma_1}(X)$  和  $G_{\sigma_2}(X)$  分别代表标准偏差为  $\sigma_1$  和  $\sigma_2$  的两个 Gauss 函数。一幅图像的 DOG 响应见式(12),即将一幅空间图像转化成带通表示。

$$DOG_{\sigma_1, \sigma_2}(X) = G_{\sigma_1}(X) - G_{\sigma_2}(X) \quad (11)$$

$$I_{DOG_{\sigma_1, \sigma_2}}(X) = I \circ (G_{\sigma_1}(X) - G_{\sigma_2}(X)) \quad (12)$$

通过式(12),可将一幅图像分解为多个频带表示:

$$I_{DOG_i} = \begin{cases} I \circ (\delta - G_{\sigma_1}) = I_{G_0} - I_{G_1}, & i = 0 \\ I \circ (G_{\sigma_i} - G_{\sigma_{i+1}}) = I_{G_i} - I_{G_{i+1}}, & i = 1, \dots, N-2 \\ I \circ G_{\sigma_i} = I_{G_{N-1}}, & i = N-1 \end{cases} \quad (13)$$

式中:  $N$  代表图像分解级数,  $\sigma_i = k^{i-1} (i = 1, \dots, N-1)$ ,  $\circ$  是 Hardamard 乘积。

通过式(13),经过  $N$  级分解后的图像  $I$  可表示为:

$$I = I_{G_0} = I_{DOG_0} + I_{G_1} = I_{DOG_0} + I_{DOG_1} + I_{G_2} = I_{DOG_0} + I_{DOG_1} + I_{DOG_2} + \dots + I_{G_{N-1}} = \sum_{i=0}^{N-1} I_{DOG_i} \quad (14)$$

利用上述方法将参考图像  $I^1$  与失真图像  $I^2$  分解为 5 级( $N=5$ )子带,并计算两幅图像在各级频带间的相似性,即:

$$f_i = SSIM(I_{DOG_i}^1, I_{DOG_i}^2) \quad (15)$$

式中:  $SSIM(\cdot)$  表示相似性计算指标,  $i = 0, 1, 2, 3, 4$ 。

图 4 所示为图像 5 级( $N=5$ )DOG 分解示例。图 4(a)为 TID2013 数据库中其中一幅原始图像的  $L$  通道图,图 4(b)~(f)分别为分解后的 5 个频带图。

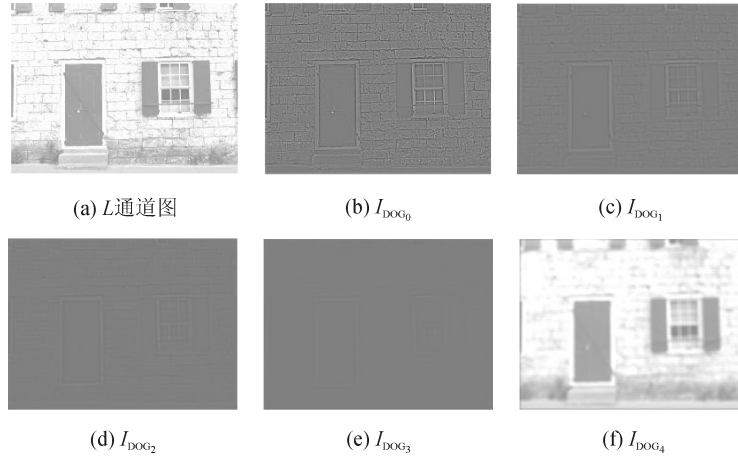


图4 图像5级 DOG 分解图

Fig. 4 5-level DOG decomposition of an image

#### 1.4 CSF 对比敏感度相似特征

CSF 特征有效反映了 HVS 与空间频率之间的关系。实验表明,将 DOG 频带特征以及 CSF 两种特征结合使用有助于增强模型对不同空间频率的分辨能力,因此本文使用 CSF 与 DOG 频带特征共同反映人眼视觉系统的频率响应特性。这里使用 SC-QI 算法中提出的方法来提取与 CSF 相关的图像特征,首先将参考图像与失真图像从空间域转换为 DCT 系数,然后将  $csf_{L_1}(csf_{L_2})$ ,  $csf_{M_1}(csf_{M_2})$  和  $csf_{H_1}(csf_{H_2})$  分别定义为参考图像(失真图像)在低频(LF)、中频(MF)和高频(HF)区域的分量。对于一幅参考图像,其对比度能量值计算如下:

$$csf_K = \sum_{(u,v) \in R_K} p(u,v) \quad (16)$$

式中:  $K = L_1, M_1, H_1$ , 分别对应参考图像  $R_K$  的 LF, MF 和 HF 三个区域,  $p(u,v)$  表示在  $(u,v)$  点处 DCT 系数的归一化值。同理,对失真图像用同样方法计算这三个指标,计算结果记为:  $csf_{L_2}$ 、 $csf_{M_2}$  以及  $csf_{H_2}$ 。参考图像与失真图像在 LF, MF 和 HF 区域的相似性值分别计算如下:

$$csfLMap(\mathbf{X}) = \frac{2csf_{L_1}(\mathbf{X}) \cdot csf_{L_2}(\mathbf{X}) + C_5}{csf_{L_1}^2(\mathbf{X}) + csf_{L_2}^2(\mathbf{X}) + C_5} \quad (17)$$

$$csfMMap(\mathbf{X}) = \frac{2csf_{M_1}(\mathbf{X}) \cdot csf_{M_2}(\mathbf{X}) + C_6}{csf_{M_1}^2(\mathbf{X}) + csf_{M_2}^2(\mathbf{X}) + C_6} \quad (18)$$

$$csfHMap(\mathbf{X}) = \frac{2csf_{H_1}(\mathbf{X}) \cdot csf_{H_2}(\mathbf{X}) + C_7}{csf_{H_1}^2(\mathbf{X}) + csf_{H_2}^2(\mathbf{X}) + C_7} \quad (19)$$

式中:  $C_5, C_6, C_7$  为正常量,根据 SC-QI 算法,取  $C_5 = 20.001111, C_6 = 1.7, C_7 = 0.0063$  用于防止分母为零而造成  $csfLMap$ 、 $csfMMap$  及  $csfHMap$  的不稳定。 $\mathbf{X}$  为 LF, MF 和 HF 三个区域中像素点的坐标值。最终,参考

图像与失真图像的对比敏感度相似性特征表示为:

$$F = csfLMap(\mathbf{X}) \cdot csfMMap(\mathbf{X}) \cdot csfHMap(\mathbf{X}) \quad (20)$$

#### 1.5 池化策略与多特征融合

经上述特征提取过程,可得到参考图像与失真图像之间的色度相似性特征、梯度相似性特征、DOG 频带特征以及对比敏感度相似性特征。池化策略的目的是将提取的图像特征映射到低维空间中,通过低维空间保留下来的特征要最大程度地代表全局图像。本文提取色度相似性图的均值、标准偏差及熵,分别用  $C_m, C_s$  及  $C_e$  表示,提取梯度相似性图的均值、标准偏差及熵,分别用  $G_m, G_s$  及  $G_e$  表示,提取对比敏感度相似性图的均值、标准偏差及熵,分别用  $F_m, F_s$  及  $F_e$  表示。DOG 频带特征是一个 5 维的特征向量,记为  $[f_0 f_1 f_2 f_3 f_4]$ 。接下来,将 3 维色度相似性特征、3 维梯度相似性特征、3 维对比敏感度相似性特征以及 5 维频带特征进行多特征融合,融合后的特征空间用一个 14 维特征向量表示,即:

$$V = [C_m C_s C_e G_m G_s G_e f_0 f_1 f_2 f_3 f_4 F_m F_s F_e] \quad (21)$$

#### 1.6 回归策略

与传统基于计算引导的方法不同,一旦得到代表每幅失真图像质量的特征向量与主观分数 MOS 后,基于机器学习的方法可自动进行特征学习并通过回归工具建立回归模型。在训练阶段,提取训练集中所有图像的 14 维特征向量,与主观分数一同输入到回归工具中,建立回归模型。测试阶段,将测试集中提取的图像特征向量输入到训练好的 IQA 回归模型中,最终映射为图像质量预测得分。

目前已经开发出了很多回归工具,其中 RF 和 SVR 两种工具使用最为广泛,也显示出卓越的学习

能力和回归效果。Pei 等<sup>[9]</sup>利用这两种回归技术训练了六个 DOG 模型,实验结果表明,RF 的预测性能要优于 SVR,尤其是当考虑到图像色度特征时结果最为显著。因此,本文采用随机森林 RF 训练质量回归模型,并设置参数(ntree, mtry)=(500, 2)。

## 2 实验结论与分析

### 2.1 数据库与评价指标

本实验在应用最广泛的四个图像数据库上进

行,包括 TID2013<sup>[15]</sup>、TID2008<sup>[16]</sup>、LIVE<sup>[17]</sup> 以及 CSIQ<sup>[18]</sup>,表 1 比较了这四个图像数据库的部分信息。其中,TID2013 数据库是四个数据库中规模最大的,是 TID2008 数据库的拓展版本,共有 25 幅原始图像、24 种失真类型,每种失真类型有 5 个失真等级,该数据库总共包含 3 000 幅失真图像。另外,CSIQ 数据库中所包含的参考图像的数量是最丰富的,LIVE 数据库中图像的分辨率最高。

表 1 四种图像质量评价数据库各种信息比较

Tab.1 Comparison of various information in four image quality evaluation databases

类别	TID2013	TID2008	LIVE	CSIQ
参考图像数量	25	25	29	30
失真图像数量	3 000	1 700	779	866
图像分辨率	512×384	512×384	不固定	512×512
失真类型数量	24	17	5	6
失真等级	5	4	6	4/5
图像类型	彩色	彩色	彩色	彩色
主观测试方法	配对比较法	配对比较法	单一刺激连续 缩放法	多刺激绝对 测量法
观察者数量	971	838	161	35
MOS/DMOS	MOS	MOS	DMOS	DMOS

目前,用于表征图像质量评价性能的指标主要有四种:Pearson 线性相关系数(PLCC)、Spearman 秩相关系数(SRCC)、Kendall 次秩相关系数(KRCC)以及均方根误差(RMSE)。其中,SRCC 和 KRCC 均是计算秩相关系数,即评价 IQA 算法的预测单调性。PLCC 和 RMSE 用于评价算法的预测准确性,需要特别注意的是,在实际应用之前还需要通过回归分析得出客观预测值与主观得分 MOS 之间的非线性映射关系,本文采用式(17)的映射函数对 MOS 值与客观预测分数进行非线性回归。其中, $X_i$  为 IQA 算法预测的第  $i$  幅图像的质量, $\hat{X}_i$  为非线性回归值,参数  $\beta_k (k = 1, 2, 3, 4, 5)$  为固定值。

$$\hat{X}_i = f(X_i) = \beta_1 \left[ \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(X_i - \beta_3))} \right] + \beta_4 X_i + \beta_5 \quad (22)$$

### 2.2 算法综合性能评价

本文所提出的方法与八种主流的 FR-IQA 方法进行了比较,这八种方法分别是 SSIM, MS-SSIM, IW-SSIM<sup>[19]</sup>, FSIMc, GMSD<sup>[20]</sup>, VSI, SC-QI 以及 DOG-SSIM<sup>[9]</sup>,其中,前七种为基于计算引导的方法,最后一种为基于机器学习的方法。由于 TID2013 数据库是 TID2008 数据库的拓展版本,因

此,本实验仅在 TID2013、LIVE 以及 CSIQ 三个数据库中进行测试,实验所得 PLCC 和 SRCC 性能指标见表 2。除包括在各数据库中的评价结果外,表 2 右半部分还列出了每种方法在三个数据库中的平均值及加权平均值评价结果,权重因子依赖于各个数据库中失真图像数量。表 2 中,加粗部分表示排名最高的图像质量评价方法所得结果。值得一提的是,为防止过拟合现象,保证训练集和测试集之间互不重叠,本文采用  $K$ -折交叉验证的方法避免此类问题。具体做法是在 TID2013 数据库中采用 8-折交叉验证策略,在 LIVE 及 CSIQ 数据库中采用 10-折交叉验证策略,每次训练选择其中 1-折作为训练集,其余作为测试集,每次实验进行 1 000 次迭代,并取中位数作为评价指标的最终结果。

由表 2 可知,提出算法 CGDR 在各个数据库中的预测性能远远优于其他八种方法,平均结果和加权平均结果同样显示出其卓越的综合表现能力。具体来看,CGDR 算法的平均 PLCC 和 SRCC 结果分别高于排名第二位方法的 2.98% 和 2.60%,加权平均结果分别高了 3.21% 和 2.99%。

另外,图 5 显示了在 TID2013 数据库上通过九种 IQA 方法获得的预测得分与主观分数的散点图

以及通过式(22)得到的最佳拟合逻辑函数,散点图横纵坐标分别表示 IQA 算法的预测得分以及主观 MOS 值,每个蓝色加号表示数据库中的一幅图像。

从图 5 中可以看出,与其他八种方法相比,提出的 CGDR 算法的所有采样点更接近于拟合曲线,这说明与主观得分有更高的一致性。

表 2 九种 FR-IQA 方法总体性能比较

Tab. 2 Comparison of the overall performance of nine FR-IQA methods

FR-IQA 算法	TID2013		LIVE		CSIQ		均值		加权均值	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
SSIM	0.686 0	0.627 4	0.902 4	0.910 4	0.861 3	0.875 6	0.816 6	0.804 5	0.755 0	0.721 1
MS-SSIM	0.832 9	0.785 9	0.948 9	0.951 3	0.899 8	0.913 8	0.893 9	0.883 7	0.864 8	0.837 5
IW-SSIM	0.831 9	0.777 9	0.951 5	0.960 4	0.914 4	0.921 3	0.899 3	0.886 5	0.867 3	0.835 2
FSIMc	0.876 9	0.851 0	0.961 3	0.964 5	0.919 2	0.931 0	0.919 1	0.915 5	0.898 9	0.884 9
GMSD	0.869 0	0.804 4	0.960 3	0.960 3	0.954 1	0.957 0	0.927 8	0.907 2	0.900 2	0.859 0
VSI	0.900 0	0.896 5	0.943 1	0.946 4	0.927 9	0.942 3	0.923 7	0.928 4	0.912 4	0.913 4
SC-QI	0.907 1	0.905 2	0.937 3	0.948 0	0.926 8	0.943 4	0.923 7	0.932 2	0.915 8	0.919 5
DOG-SSIM	0.918 6	0.907 3	0.962 7	0.961 3	0.941 8	0.951 9	0.941 0	0.940 2	0.930 3	0.924 7
CGDR	<b>0.950 3</b>	<b>0.940 0</b>	<b>0.978 9</b>	<b>0.975 3</b>	<b>0.977 8</b>	<b>0.975 0</b>	<b>0.969 0</b>	<b>0.964 6</b>	<b>0.960 2</b>	<b>0.952 4</b>

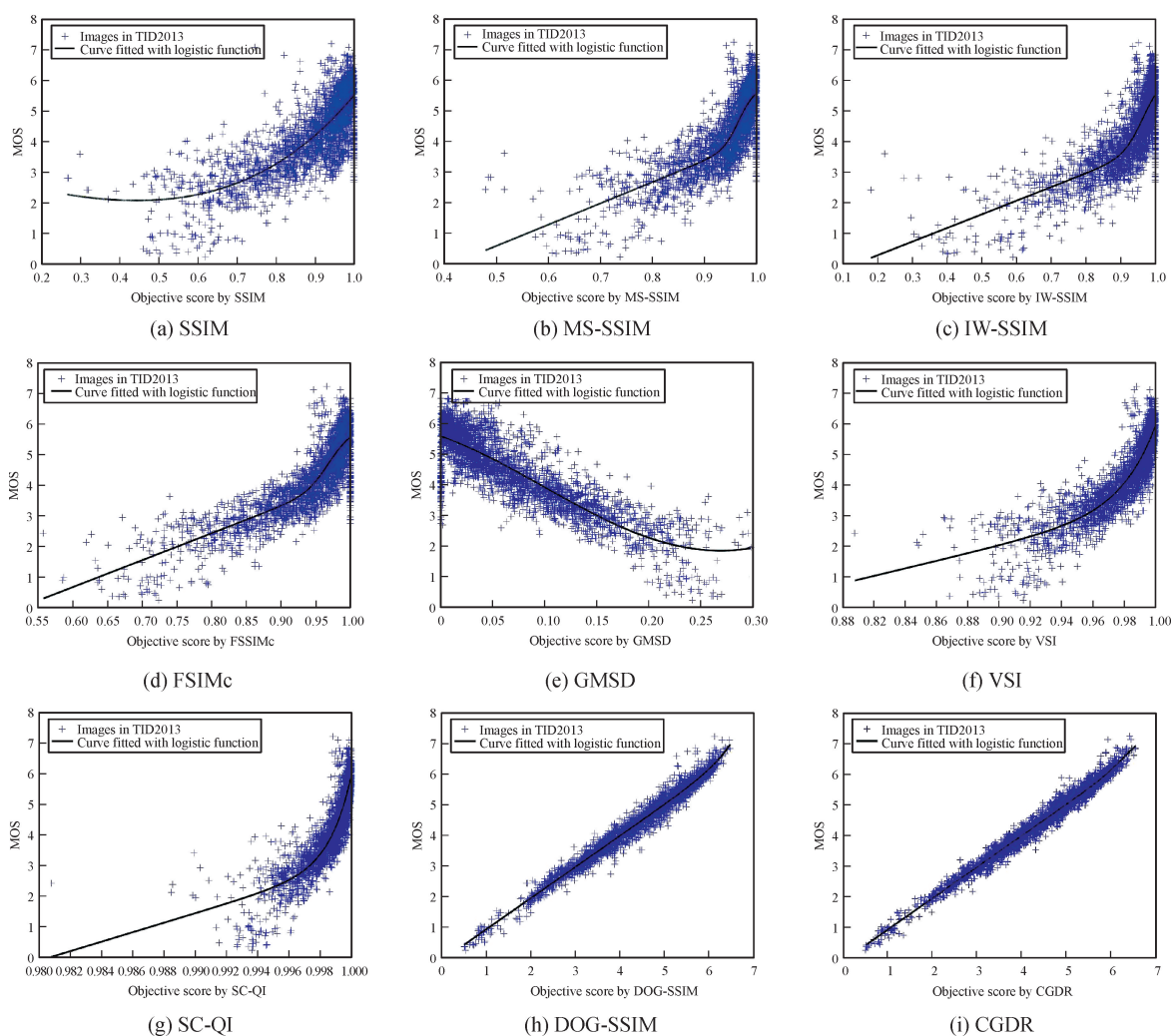


图 5 TID2013 数据库上通过 IQA 方法获得的预测得分与主观分数的散点图

Fig. 5 Scatter plots of subjective MOS and predicted scores obtained by several IQA methods on TID2013 database



### 2.3 跨数据库验证

基于机器学习的 IQA 方法的预测性能直接取决于训练数据集与测试数据集,有些方法在一个数据集上的性能非常好,但在其他数据集上却表现平平。与  $K$ -折交叉验证不同,跨数据库验证是在不同数据库中分别训练和测试 IQA 模型,这对于评估模型的整体性能方面更有说服力。

表 3 为所提出的 CGDR 算法与其他三种基于机器学习方法的跨数据验证的 PLCC 结果。其中,前两种方法分别使用 SVR 和 SVD 回归技术,后一种使用随机森林 RF 回归策略。值得一提的是,由于 TID2013 数据库与 TID2008 数据库中包含的原始图像相同,因此本实验排除在这两个数据库间进行验证。同理,LIVE 与 TID2008 和 TID2013 数据库均有重叠内容,因此,为保证实验的严谨和规范,跨数据库验证实验仅在 CSIQ/LIVE 以及 CSIQ/TID2008 数据库间进行训练和测试。表 3 中用粗体标出了每次交叉验证实验的最佳值,结果无疑证明了 CGDR 算法比其他三种方法有更强的鲁棒性。

表 3 跨数据库验证的 PLCC 结果

Tab. 3 PLCC values of cross-database verification results

训练集	CSIQ	LIV	CSIQ	TID2008
测试集	LIVE	CSIQ	TID2008	CSIQ
[10]	0.885 2	0.862 0	0.776 1	0.886 3
[11]	0.908 6	0.858 1	0.755 0	0.883 1
DOG-SSIM	0.894 7	0.892 5	0.854 7	0.912 3
CGDR	<b>0.961 2</b>	<b>0.943 9</b>	<b>0.904 6</b>	<b>0.948 8</b>

### 2.4 运算复杂性分析

将 IQA 算法运用到实时系统中时,设计者总是希望在算法的预测精度与运算复杂度之间找到很好的平衡。表 4 列出了包括提出的 CGDR 算法在内的九种 FR-IQA 方法在 TID2013 数据库中每幅图像的平均运行时间,所有方法均使用 Matlab 2016a 在配有智能英特尔酷睿 i7 四核处理器的个人计算机上进行计算,其余八种方法的代码均来自原始作者。

由表 4 可知,SSIM 的运算速度是九种方法中最快的,平均每 0.049 7 s 运行一对图像,其次是 GMSD,速度快的原因是这两种方法无论在特征提取过程还是池化阶段,其计算过程都比较简单,计算的复杂程度相对较低。CGDR 算法的平均运算时间与 MS-SSIM 大致相同,速度均落后于其他方法。CGDR 算法运行速度慢的原因可能是在特征提取的过程中,对特征空间的维度控制的不够低,每一幅

失真图像要提取 14 维特征向量,并且不仅需要在空间域中提取图像特征,还要将图像转变为 DCT 系数以及分解为 DOG 子带来反映其频率特征,所以导致其运算效率受到影响。虽然平均运算时间相对较高,但 CGDR 算法与其他八种方法相比保持了最高的预测精度。

表 4 TID2013 数据库上九种 FR-IQA 方法的平均运行时间  
Tab. 4 Average running time of nine FR-IQA methods in TID2013 database

FR-IQA 算法	平均运行时间/s
SSIM	0.049 7
MS-SSIM	0.095 3
IW-SSIM	0.561 3
FSIMc	0.176 8
GMSD	0.052 5
VSI	0.257 6
SC-QI	0.092 7
DOG-SSIM	0.067 7
CGDR	0.561 4

## 3 结 语

本文提出了一种基于视觉感知高度相关图像质量评价模型 CGDR。该模型在考虑图像色度特征的同时,使用改进后的梯度计算方法提取图像的空域梯度特征,并联合 DOG 频域特征以及 CSF 特征,从而充分模拟人眼对视觉场景的感知机制。在四个数据库上进行的实验表明,与主流 FR-IQA 方法相比,所提出的 CGDR 算法具有更高的预测精度和更强的鲁棒性,能够与人眼主观感知保持高度一致性。

### 参考文献:

- [1] KIM J, LEE S. Fully deep blind image quality predictor [J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(1): 206-220.
- [2] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [3] WANG Zhou, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment [C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, IEEE, 2003: 1398-1402.
- [4] ZHANG Lin, ZHANG Lei, MOU Xuanqin, et al. FSIM: a feature similarity index for image quality assessment [J]. IEEE Transactions on Image Processing, 2011, 10(8): 2378-

- 2386.
- [5] ZHANG Lin, SHEN Ying, LI Hongyu. VSI: a visual saliency-induced index for perceptual image quality assessment[J]. *IEEE Transactions on Image Processing*, 2014, 23(10): 4270-4281.
- [6] ZHOU Qiangqiang, LIU Xianhui, ZHANG Lin, et al. Saliency-based image quality assessment metric[C]//2016 3rd International Conference on Systems and Informatics (ICSAI), IEEE, 2016: 918-924.
- [7] BALANOV A, SCHWARTZ A, MOSHE Y, et al. Image quality assessment based on DCT subband similarity[C]//2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015: 2105-2109.
- [8] BAE S H, KIM M. A novel image quality assessment with globally and locally consistent visual quality perception [J]. *IEEE Transactions on Image Processing*, 2016, 25(5): 2392-2406.
- [9] PEI S C, CHEN Liheng. Image quality assessment using human visual DOG model fused with random forest [J]. *IEEE Transactions on Image Processing*, 2015, 24(11): 3282-3292.
- [10] DING Yong, ZHAO Yang, ZHAO Xinyu. Image quality assessment based on multi-feature extraction and synthesis with support vector regression[J]. *Signal Processing: Image Communication*, 2017: 81-92.
- [11] NARWARIA M, LIN Weisi. SVD-based quality metric for image and video using machine learning[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 347-364.
- [12] 范彩霞,陈亚军. 颜色空间的感知均匀性评价及其在图像处理中的应用[J]. *西安理工大学学报*, 2016, 32(2): 191-198.
- FAN Caixia, CHEN Yajun. Evaluation of perceptual uniform color spaces and application in image processing[J]. *Journal of Xi'an University of Technology*, 2016, 32(2):191-198.
- [13] 唐棣森,郑元林,廖开阳,等. 基于视觉感知特征融合的图像质量评价[J]. *西安理工大学学报*, 2018, 34(4): 454-459.
- TANG Zhisen, ZHENG Yuanlin, LIAO Kaiyang, et al. Image quality assessment based on visual perception feature fusion[J]. *Journal of Xi'an University of Technology*, 2018, 34(4): 454-459.
- [14] LIU Anmin, LIN Weisi, NARWARIA M. Image quality assessment based on gradient similarity[J]. *IEEE Transactions on Image Processing*, 2011, 21(4): 1500-1512.
- [15] PONOMARENKO N, IEREMEIEV O, LUKIN V, et al. Color image database TID2013: peculiarities and preliminary results[C]//European Workshop on Visual Information Processing (EUVIP), IEEE, 2013: 106-111.
- [16] PONOMARENKO N, LUKIN V, ZELENSKY A, et al. TID2008-a database for evaluation of full-reference visual quality assessment metrics [J]. *Advances of Modern Radioelectron*, 2009, 10: 30-45.
- [17] SHEIKH H R, SABIR M F, BOVIK A C, et al. A statistical evaluation of recent full reference image quality assessment algorithms[J]. *IEEE Transactions on Image Processing*, 2006, 15(11): 3440-3451.
- [18] LARSON E C, CHANDLER D M. Categorical image quality (CSIQ) database[EB/OL]. <http://vision.okstate.edu/csiq>.
- [19] WANG Zhou, LI Qiang. Information content weighting for perceptual image quality assessment [J]. *IEEE Transactions on Image Processing*, 2011, 20(5): 1185-1198.
- [20] XUE Wufeng, ZHANG Lei, MOU Xuanqin, et al. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index[J]. *IEEE Transactions on Image Processing*, 2014, 23(2): 684-695.

(责任编辑 王绪迪)