

基于随机森林的空域-频域联合特征全参考彩色图像质量评价方法

杨小琴 刘国军 郭建慧 马文涛

宁夏大学数学统计学院 银川 750021 (yxq258351@163.com)



摘 要 文中旨在设计一种可以自动评估图像质量,并达到与人类视觉系统相一致的客观评价算法。针对大多数传统的全参 考图像质量评价方法只在空域中分析图像,并且在池策略上存在不足,文中提出了一种基于随机森林的空域-频域联合特征全 参考彩色图像质量评价方法。该方法首先在空域上提取色度和梯度特征,刻画图像的颜色信息和空间结构信息;在频域上提取 log-Gabor 滤波器组响应后的纹理细节信息以及空间频率特征,将二者作为联合特征;然后利用随机森林学习特征向量与主观 意见得分之间的映射关系,预测客观质量得分。在 TID2013,TID2008 和 CSIQ 3 个标准数据库上的实验结果表明,所提方法的 综合评价性能优于目前主流的全参考评价算法,尤其是在 TID2013 数据库上其皮尔逊线性相关系数值达到了 0.9397。 关键词:全参考图像质量评价;空域;频域;随机森林 中图法分类号 TP391

Full Reference Color Image Quality Assessment Method Based on Spatial and Frequency Domain Joint Features with Random Forest

YANG Xiao-qin, LIU Guo-jun, GUO Jian-hui and MA Wen-tao School of Mathematics and Statistics, Ningxia University, Yinchuan 750021, China

Abstract This paper is to design an objective evaluation algorithm that automatically evaluates image quality and is consistent with the human visual system. In view of the fact that most traditional full reference image quality assessment methods only analyze images in the spatial domain, and have shortcomings in pooling strategies, this paper proposes a random forest based spatial-frequency domain joint feature full reference color image quality evaluation method. Firstly, this method extracts the chroma and gradient features in the spatial domain, which are used to characterize the color information and spatial structure information of images. The texture detail information of the response of the log-Gabor filter bank and spatial frequency features are extracted in the frequency domain, which are used to be joint features. Then, random forest is implemented for learning the mapping relationship between the feature vector and the subjective opinion score to predict the objective quality score. Experiments conducted on three standard databases, i. e. TID2013, TID2008, and CSIQ show that the comprehensive evaluation performance by our method is better than the state-of-the-art full reference assessment algorithms, especially on TID2013 database, the Pearson linear correlation coefficient value can reach 0, 9397.

Keywords Image quality assessment, Spatial domain, Frequency domain, Random forest

1 引言

数字图像在信息技术发展和日常生活服务中发挥着至关 重要的作用,然而图像在获取、压缩、处理、传输、显示等过程 中难免会出现一定程度的失真,因此需要建立有效的客观图 像质量评价(Image Quality Assessment, IQA)模型。根据对 原始图像或参考图像的参考程度,客观 IQA 分为全参考型 (Full-Reference, FR)、部分参考型(Reduced-Reference, RR) 和无参考型(No-Reference, NR)^[1]。FR IQA 方法由于可获 取的参考图像丰富,评价结果更加准确、可靠,成为了 IQA 领 域的研究热点。

大多数传统的全参考 IQA 方法基于图像结构的角度设计算法。例如,SSIM^[2]算法通过图像的亮度特征、结构特征和对比度特征来表征图像感知质量。后续很多 IQA 方法在

到稿日期:2020-07-16 返修日期:2020-09-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目: 宁夏自然科学基金(2018AAC03014); 国家自然科学基金(61461043, 51769026); 宁夏回族自治区重点研发项目(2019BEG03056); 宁夏大学研究生创新项目(GIP2019011)

This work was supported by the Natural Science Foundation of Ningxia(2018AAC03014), National Natural Science Foundation of China (61461043,51769026), Key Research and Development Projects of Ningxia(2019BEG03056) and Graduate Innovation Project of Ningxia University(GIP2019011).

通信作者:刘国军(liugj@nxu.edu.cn)

空域中提取图像的结构信息,如FSIM^[3],PSIM^[4]和GMSD^[5] 算法都是在空域中提取梯度特征,描述图像的结构变化。 Yang 等^[6]提出了 Riesz 变换和基于视觉对比度敏感度的特 征相似度(Riesz transform and Visual contrast sensitivitybased feature SIMilarity,RVSIM)度量,比传统 IQA 模型具 有更好的预测能力。文献[7]提出了一种改进的客观 FR IQA 方法,即多尺度多方向可见性差异(Multiscale and Multidirectional Visibility Differences,MMVD)算法。MMVD考虑了 离散不可分 shearlet 变换域内的多尺度、多方向可见性差异, 仿真了人眼视觉系统的多通道信息处理结构。

此外,在 IQA 池化步骤中,对各个特征进行加权处理会 影响预测的图像质量得分。由于图像的不同特征及不同区域 对人类视觉系统(Human Visual System, HVS)的影响程度不 一样,SSIM 对亮度、对比度和结构特征采用相同的权重,导 致图像质量得分的预测准确性较低。因此 FSIM 算法将相位 一致性作为权重函数,加权处理相位特征和梯度特征。 VSI^[8]算法将视觉显著性函数作为加权函数,处理视觉显著 性特征和梯度特征。GMSD 仅利用标准偏差来处理梯度特 征,具有较快的计算速度和较高的预测准确性。

上述大多算法在度量局部图像质量方面有较好的效果, 但只能在部分图像数据库中获得较佳的评价性能,鲁棒性较 差。主要原因包括:1)上述算法大多只是在空域中分析图像, 然而频域中也具有重要的视觉感知信息;2)这些方法在池化 过程中采取的池化策略一般有算术平均、加权平均、调和平均 等,但这些池化方法缺少理论基础和可靠的计算模型支持,往 往会导致图像的预测质量得分较低。然而,基于机器学习的 池化策略可以弥补这方面的不足。

常用的机器学习方法有支持向量回归(Support Vector Regression, SVR)^[9]、极限学习机(Extreme Learning Machine, ELM)^[10]、随机森林(Random Forest, RF)^[11]和神经网络(Neural Network, NN)^[12]。它们被广泛地应用于客观 IQA 算法。 Narwaria 等^[13]利用奇异值分解方法将奇异向量作为特征来 量化图像的结构信息,然后采用 SVR 进行图像质量预测。 Liu 等^[14]利用机器学习将多个 FR IQA 进行融合,获得了较 好的评价性能。Pei 等^[15]提出了一种基于视网膜过程学习的 FR IQA 方法,从几个高斯差异频带中提取特征,利用 RF 将 提取的特征向量映射到与其对应的主观得分。其实验结果表 明,RF比SVR和BP神经网络有着更高的预测精度以及更 快的计算速度。Gao 等^[16]从深度 NN 架构如何影响 IQA 性 能的角度出发,提出了一种新的 FR IQA 框架,即深度相似 性。该框架利用 VGGnet 提取图像特征,然后计算相似度,给 出最终评分。Liu 等^[17]对文献 [16] 所提出的框架进行了改 进,提出了一种基于视觉显著性和梯度特征的深度 IQA 模 型。然而,深度学习网络涉及大量的训练数据和超参数,使得 理论分析变得极其困难且大多数网络都被极度耗时的训练过 程所困扰。

客观 IQA 的基本目标是设计出能精确和自动感知图像 质量的模型,并且该模型的评价性能应尽可能接近人的主观 感受。基于此,研究者将 HVS 与 IQA 方法相结合,以提高算 法的评价性能。然而,由于 HVS 过于复杂,人们对视觉感知 原理的认知不足,很难建立能够模拟 HVS 视觉信号处理过 程的数学模型。因此,提取 HVS 体现出来的视觉特征成为了 目前 IQA 的研究重点。其中 HVS 对空间频率响应特性、纹 理信息等具有高度敏感性。

基于以上分析,本文通过在空域和频域中提取互补特征 进行图像质量预测,提出了一种基于随机森林的空域-频域联 合特征(Spatial and Frequency Domain Joint Features-Random Forest,SFDJF-RF)全参考 IQA 方法。首先在空域中分别对 参考图像和失真图像提取颜色和梯度特征图,在频域中提取 互补特征,即多尺度多方向 log-Gabor 滤波器组提取纹理特 征图以及空间频率特征图;其次对每对特征图进行相似性度 量,以此来量化特征差异,从而表征图像的失真程度;然后利 用 RF 建立特征差异与图像平均意见得分(Mean Opinion Score,MOS)值之间的映射关系。本文在 TID2008,TID2013, CSIQ 3 个标准数据库上进行实验,检验本文方法的评价 性能。

2 特征提取

SFDJF-RF方法的流程图如图 1 所示。首先将参考图像 和失真图像由原始的 RGB 空间转换到 LMN 空间;其次对亮 度通道提取梯度特征、纹理特征和空间频率特征,对色度通道 *M*和*N*提取颜色特征,其中纹理特征由 log-Gabor 滤波器组 提取,并且将图像离散余弦变换的子块系数用于提取空间频 率特征;然后为了量化参考图像和失真图像的特征差异,对以 上提取的特征进行相似性度量;最后采用 RF 进行特征融合, 预测最后的客观质量得分。



图1 身伝航性图 Fig. 1 Algorithm flow chart

2.1 空域特征提取

(1)颜色特征

彩色图像一般包含亮度信息和颜色信息,并且它们对图 像质量都有着显著影响。在进行 IQA 时,选取恰当的颜色空 间,对设计与 HVS 相吻合的彩色 IQA 方法有着重要作用。 原始 RGB 空间不能将亮度信息和色度信息进行分离,因而也 不能分别对亮度通道和色度通道进行处理。因此,为了将亮 度和色度信息进行分离,我们选取 IQA 中常用的一种颜色空 间 LMN。原始 RGB 空间转换为 LMN 空间的具体计算如下:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$
(1)

其中,L是亮度通道,M和N是色度通道。

众所周知,对于彩色 IQA 而言,颜色特征是重要的低层特征,大多 IQA 方法只是在亮度通道上对图像的视觉感知进行数学建模,却忽略了颜色失真对图像质量也有很大的影响。因此本文在 *M* 和 *N* 色度通道中提取颜色特征来刻画图像颜色失真程度。

(2)梯度特征

图像边缘是一种重要的视觉信息,梯度特征可以充分刻 画图像边缘结构信息和对比度变化信息,因此在 IQA 中应用 广泛。比较常见的图像边缘检测滤波器有 Sobel, Prewitt, Scharr等,我们选取 Sobel 滤波器计算图像亮度通道的梯度。 利用 Sobel 滤波器分别对参考图像和失真图像进行卷积,参 考图像与失真图像的梯度幅值计算公式如下:

$$G_{R}(i) = \sqrt{(h_{x} * R)^{2}(i) + (h_{y} * R)^{2}(i)}$$
(2)

$$G_D(i) = \sqrt{(h_x * D)^2(i) + (h_y * D)^2(i)}$$
(3)

其中, $G_R(i)$ 和 $G_D(i)$ 分别为参考图像和失真图像在位置索引 *i*处的梯度幅值;符号"*"为卷积运算;R和D分别表示参考 图像和失真图像; h_x 和 h_y 分别为 Sobel 滤波器水平和垂直方 向模板,定义如下:

$$h_x = \frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, h_y = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$
(4)

2.2 频域特征提取

(1)多尺度多方向纹理特征

HVS比较复杂,除了边缘结构信息外,还有很多信息可 以被 HVS捕获。现有的很多全参考 IQA 方法忽略了图像纹 理细节信息。因此,本文利用 log-Gabor来提取多尺度多方向 的纹理细节特征,作为在空域中所提取特征的补充。

作为传统的图像频域表示方法,傅里叶变换只是刻画图像的频率统计特性,不能有效刻画图像的纹理信息。因此,文献[18]在频域上对图像进行高斯加窗,设计了一种 Gabor 滤波器。Gabor 函数的二维形式如下:

$$G(x,y) = \frac{1}{d} \exp\left(-\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \cos(\omega x + \varphi)$$
(5)

其中,(x,y)为图像的像素点; ω 为频率; φ 为相位偏移; σ_x 和 σ_y 分别为高斯函数水平与垂直方向上的标准差;常数 d 是为 了确保 $\Sigma G(x,y)=1$ 。

然而,Gabor 滤波器的频率响应过于集中在低频区域,导 致高频区域大量细节信息丢失。基于此,文献[19]提出了 log-Gabor 滤波器。二维 log-Gabor 滤波器的频率响应为:

$$LG(f,\theta) = \exp\left(-\frac{(\log(f/f_0))^2}{2(\log(\sigma_f/f_0))^2} - \frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right)$$
(6)

其中, f_0 为滤波器的中心频率; θ_0 为滤波器方向; σ_f 定义径向带宽; σ_θ 定义角度带宽。

由式(6)可以看出,一个具体的 log-Gabor 滤波器由方向 和尺度两个参数确定,这说明它所提取的特征不仅是多方向 的,而且是多尺度的,这与 HVS 多通道、多分辨率特性具有 很好的相关性。本文采用 4 个尺度和 4 个方向(0°,45°,90°, 135°)的 log-Gabor 滤波器组对图像进行特征提取。一幅 m× *n* 的图像经过 log-Gabor 滤波器滤波之后可得到一个 $m \times n$ 的复矩阵,且该图像的能量为:

$$E_{m,n}(x,y) = \sqrt{D_{m,n}^{r}(x,y)^{2} + D_{m,n}^{i}(x,y)^{2}}$$
(7)

其中, $D_{m,n}^{r}(x,y)$ 和 $D_{m,n}^{i}(x,y)$ 分别为复矩阵的实部和虚部。

式(7)表示的能量图反映了图像的纹理细节信息。图 2 给出了取自 TID2008 数据库的一幅参考图像,以及该图像经 过4个尺度4个方向的 log-Gabor 滤波器组响应后的能量图。 由图中可以看出,更大尺度的滤波器可以更好地消除噪声,同 时也会丢失更多的细节信息。

(2)空间频率特征

除了上述提出的多尺度多方向的纹理特征外,空间对比 敏感度函数(Contrast Sensitivity Function,CSF)也是 HVS 的 重要视觉特征,对图像不同的失真有着不同的视觉敏感度。 因此,我们对图像亮度通道 L 进行离散余弦域(Discrete Cosine Transform,DCT),使用图像 DCT 后的 3 个子带特征来 模拟 HVS 中的 CSF,即在 $n \times n$ 的 DCT 子块 B 中获取图像 的高频、中频和低频能量。

首先对 DCT 子块 B 的系数进行归一化处理,归一化后 DCT 系数幅值计算如下:

$$p(\omega) = (\varepsilon + |c(\omega^{\lambda})|)/Z$$
(8)

其中, $p(\omega)$ 是在 ω 处归一化 DCT 系数的幅值; $c(\omega)$ 是 ω 的 DCT 系数; ε 是一个正常数,以避免分母为 0; λ 为调整参数; ω 为第(u,v) 个 DCT 系数的空间频率值,u 是水平方向频率,v是垂直方向频率,且 $\omega = \delta \cdot \sqrt{u^2 + v^2}, \delta$ 为非负常数;Z 为归 一化因子,计算式如下:

$$Z = \sum_{\omega \in B, \omega \neq 0} (\varepsilon + |c(\omega)^{\lambda}|)$$
(9)

因此,一幅图像高频、中频和低频能量的计算如下:

$$\phi_{\rm H} = \sum_{CP} P(\omega) \tag{10}$$

$$\phi_{\rm M} = \sum_{\omega \in R_{\rm M}} P(\omega) \tag{11}$$

$$L = \sum_{\omega \in \mathcal{P}} P(\omega) \tag{12}$$

其中,R_H,R_M,R_L分别为 DCT 子块 B 的高频、中频、低频区域。



图 2 一幅参考图像经过 log-Gabor 滤波器组滤波后在 4 个尺度 4 个方向的纹理信息

Fig. 2 Texture information of reference image after filtering by log-Gabor filter bank in 4 scales and 4 directions

2.3 特征相似性度量

图像的失真会导致提取的特征发生变化,因此,本节通过 计算原始图像和失真图像的特征相似性来度量图像失真程度。对于不同的特征,运用不同的相似性计算方法。将参考 图像和失真图像之间的相似性度量分为4个部分:色度、梯度 幅值、纹理和基于 CSF 的空间频率。

首先进行色度相似性度量,参考图像 R 和失真图像 D 在 M 通道的颜色相似性为:

$$S_{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{2M_{R}(i) * M_{D}(i) + C_{1}}{M_{R}^{2}(i) + M_{D}^{2}(i) + C_{1}}$$
(13)

其中, M_R 和 M_D 分别为参考图像R和失真图像D在M通道的颜色信息;N为图像中像素的总数; C_1 是稳定分式的正常数,避免分母为 0。

参考图像 R 和失真图像 D 在 N 通道上的颜色相似性为:

$$S_{N} = \frac{1}{N} \sum_{i=1}^{N} \frac{2N_{R}(i) * N_{D}(i) + C_{1}}{N_{R}^{2}(i) + N_{D}^{2}(i) + C_{1}}$$
(14)

其中, N_R , N_D 分别为参考图像R和失真图像D在N通道的颜色信息。

因此,参考图像 R 和失真图像 D 的色度相似性计算如下:

$$S_c =$$

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(\frac{2M_{R}(i) * M_{D}(i) + C_{1}}{M_{R}^{2}(i) + M_{D}^{2}(i) + C_{1}} \cdot \frac{2N_{R}(i) * N_{D}(i) + C_{1}}{N_{R}^{2}(i) + N_{D}^{2}(i) + C_{1}}\right)} (15)$$

参考图像 R 和失真图像 D 的梯度幅值之间的相似性采 用相似性标准差来进行度量:

$$G(i) = \frac{2G_R(i) * G_D(i) + C_2}{G_R^2(i) + G_D^2(i) + C_2}$$
(16)

$$S_{G} = \sqrt{\frac{\sum_{i=1}^{N} (G(i) - \bar{G})^{2}}{N}}$$
(17)

其中,C₂为正常数,以保证分式稳定,避免分母为0;N为像 素的总数。

log-Gabor 滤波器组提取的纹理特征之间的相似性采用 相关系数度量:

$$CC_{s}(E_{R},E_{D}) = \frac{\sum_{i=1}^{N} (E_{R}(i) - \bar{E}_{R}) \cdot (E_{D}(i) - \bar{E}_{D})}{\sum_{i=1}^{N} (E_{R}(i) - \bar{E}_{R})^{2} \cdot \sum_{i=1}^{N} (E_{D}(i) - \bar{E}_{D})^{2}} (18)$$

其中, E_R , E_D 分别为参考图像R和失真图像D在第s个尺度 上 log-Gabor 滤波器的响应能量。

基于 CSF 的空间频率特征相似性,继续采用相似性标准 差来度量。图像高频区域的能量相似性度量如下:

$$\phi_{\rm H} = \frac{2\phi_{R_{\rm H}} * \phi_{D_{\rm H}} + C_3}{\phi_{R_{\rm H}}^2 + \phi_{D_{\rm H}}^2 + C_3}$$
(19)
$$S_{H} = \sqrt{\frac{\sum_{\omega \in R_{\rm H}} (\phi_{\rm H}(\omega) - \bar{\phi}_{\rm H})^2}{N}}$$
(20)

其中, $\phi_{R_{\text{H}}}$ 和 $\phi_{D_{\text{H}}}$ 分别为参考图像 R 和失真图像 D 在高频区域的能量; C_3 为正常数。

同样,参考图像 R 和失真图像 D 在中频区域、低频区域 的能量相似性计算如下:

$$\phi_{\rm M} = \frac{2\phi_{R_{\rm M}} * \phi_{D_{\rm M}} + C_4}{\phi_{R_{\rm M}}^2 + \phi_{D_{\rm M}}^2 + C_4} \tag{21}$$

$$\phi_{\rm L} = \frac{2\phi_{R_{\rm L}} * \phi_{D_{\rm L}} + C_5}{\phi_{R_{\rm L}}^2 + \phi_{D_{\rm L}}^2 + C_5} \tag{22}$$

$$S_{\rm M} = \sqrt{\frac{\sum\limits_{\omega \in R_{\rm M}} (\phi_{\rm M}(\omega) - \bar{\phi}_{\rm M})^2}{N}}$$
(23)

$$S_{\rm L} = \sqrt{\frac{\sum\limits_{\omega \in R_{\rm L}} (\phi_{\rm L}(\omega) - \bar{\phi}_{\rm L})^2}{N}}$$
(24)

其中, ϕ_{R_M} 和 ϕ_{D_M} 分别为参考图像 R 和失真图像 D 在中频区域的能量; ϕ_{R_L} 和 ϕ_{D_L} 参考图像 R 和失真图像 D 在低频区域的能量; C_4 和 C_5 为正常数,以避免分母为 0。

3 随机森林回归

通过对参考图像和失真图像提取特征,以及对特征的相 似性度量,一幅图像可以获得一个9维的特征向量,如下所 示:

 $f = [S_{c}, S_{G}, CC_{1}, CC_{2}, CC_{3}, CC_{4}, SD_{H}, SD_{M}, SD_{L}]$ (25) 其中, S_{c} 为色度相似性; S_{G} 为梯度幅值相似性; $CC_{1}, CC_{2}, CC_{3}, CC_{4}$ 为4个尺度的纹理相似性; SD_{H}, SD_{M}, SD_{L} 分别为 高、中、低频区域的空间频率相似性。

接下来需要在9维的特征向量和图像的 MOS 值之间建 立回归映射关系。对于回归问题,机器学习方法能有效解决。 SVR 对数据格式要求严格,虽然其计算速度较快,但准确性 不高。NN 参数较多,模型较为复杂。ELM 虽然学习速度快, 但易出现过拟合现象,且泛化性较差。RF 由很多决策树构 成,不仅理论简单,而且耗时短,准确性较高。因此,本文选取 RF 来学习特征向量和 MOS 值之间的回归映射关系。

假设训练集 $D = \{(x_1, y_1), \dots, (x_n, y_n)\},$ 每个样本具有 k 维特征,训练具有 M 棵决策树的 RF 算法的步骤如算法 1 所示。

算法1 RF 算法

- 输入:训练集 D,特征维数为 k,随机选取特征数目为 d,决策树学习算法 f
- 输出:集成模型 F(x)
- 步骤1 for m=1 to M do
- 步骤 2 采用 Bootstrap 抽样,从训练集 D 中抽取大小为 n 的数据集 D_m。
- 步骤 3 从 k 维特征中随机选择 d 维特征,在基于步骤 2 获得的样本 数据集 D_m中随机选择 d 维特征,通过决策树模型学习得到 一棵决策树 f_m(x)。

步骤4 回归问题,输出集成模型: $F(\mathbf{x}) = \frac{1}{M} (\sum_{n=1}^{M} f_n(\mathbf{x})).$

4 实验结果与分析

为了验证本文方法 SFDJF-RF 的性能,在 TID2013^[20], TID2008^[21],CSIQ^[22]3个标准数据库上进行实验,采用斯皮 尔曼等级相关系数(Spearman Rank Order Correlation Coefficient,SROCC)、肯德尔等级相关系数(Kendall Rank Order Correlation Coefficient,KROCC)、皮尔逊相关系数 (Pearson Linear Correlation Coefficient,PLCC)和均方根误差 (Root Mean Square Error,RMSE)来验证 SFDJF-RF 的评价 性能。数据库参数信息详见表 1,包括参考图像的数量、失 真图像的数量、失真类型以及失真强度等。本文将所提方 法与目前主流的 11 种经典 FR IQA 方法进行了比较,包 括 SSIM^[2]、FSIMc^[3]、GMSD^[5]、VSI^[6]、DOG-SSIMc^[13]、SC- QI^[23]、GWL-SSIM^[24]、PSIM^[4]、MMVD^[7]、DeepSim^[16]和文献[17]的方法。

表 1 验证 IQA 性能的数据库

Table 1 Databases to verify IQA performance

数据库	CSIQ	TID2008	TID2013
参考图像	30	25	25
失真图像	866	1 700	3 0 0 0
失真类型	6	17	24
失真强度	$4 \sim 5$	4	5
观测者人数	35	838	971

在图像学习过程中,将每个数据库中的图像随机分为训 练集和测试集,分别随机选取图像的50%训练和50%测试, 60%训练和40%测试,以及80%训练和20%测试进行了 实验。结果表明,随机选取80%图像训练和20%图像测 试的评价性能最好。为了公平比较,重复训练-测试过程

Та

1000次,并取1000次迭代之后的中值为最终结果,以消除性能偏差。

4.1 单个数据库的性能表现

首先,将本文方法 SFDJF-RF 与 11 种 FR IQA 方法在整 个数据库上进行性能比较。表 2 列出了 12 种方法在 3 个数 据库上的 4 个评价指标值,用黑色粗体标出了性能表现最好 的两种算法。由表 2 可以看出,SFDJF-RF 在 3 个数据库中的 性能都优于其他 11 种 FR IQA 算法,具有最好的预测准确性 和单调性,位于第二的算法是 DOG-SSIMc,并且在 TID2008 数据库上与两个基于深度学习的方法(DeepSim、文献[17]) 相比,SFDJF-RF 具有更好的预测精度。此外,表 2 还列出了 4 个评价指标按照数据库中图像数量加权后的平均值。可以 看出,SFDJF-RF 方法的加权平均值均高于其他方法。因此, SFDJF-RF 的综合评价性能更好。

表 2 12 种算法在 3 个数据库上的性能比较

ble 2 Performance	comparison	of 12	algorithms	on 3	databases
-------------------	------------	-------	------------	------	-----------

数据库	评价 指标	GWL- SSIM	SSIM	FSIMc	GMSD	VSI	PSIM	SC-QI	DOG- SSIMc	MMVD	DeepSim	文献[17]	SFDJF-RF
	SROCC	0.7957	0.7417	0.8015	0.8044	0.8965	0.8925	0.9052	0.9145	0.9106	_	-	0.9316
TID	KROCC	0.6274	0.5588	0.6289	0.6339	0.7183	0.7161	0.7352	0.7542	0.7426	—	_	0.7720
2013	PLCC	0.8529	0.7895	0.8589	0.8590	0.9000	0.9080	0.9071	0.9279	0.9171	—	_	0.9397
	RMSE	0.6472	0.7608	0.6349	0.6346	0.5404	0.5193	0.5219	0.4606	0.5157	—	_	0.4256
	SROCC	0.8747	0.7749	0.8805	0.8907	0.8979	0.9120	0.9051	0.9235	0.9347	0.9199	0.9401	0.9444
TID	KROCC	0.6954	0.5768	0.6946	0.7092	0.7123	0.7395	0.7249	0.7721	0.7547	-	-	0.8017
2008	PLCC	0.8696	0.7732	0.8738	0.8788	0.8762	0.9078	0.8899	0.9270	0.9217	0.9030	0.9250	0.9485
	RMSE	0.6627	0.8511	0.6525	0.6404	0.6466	0.5628	0.6120	0.5036	0.4279	0.7790	0.7030	0.4277
	SROCC	0.9617	0.8756	0.9242	0.9570	0.9423	0.9621	0.9434	0.9488	0.9351	_	-	0.9731
CELO	KROCC	0.8210	0.6907	0.7567	0.8129	0.7857	0.8273	0.7870	0.8128	0.7674	—	_	0.8584
CSIQ	PLCC	0.9441	0.8613	0.9120	0.9541	0.9279	0.9642	0.9268	0.9445	0.9365	—	_	0.9773
	RMSE	0.0866	0.1334	0.1077	0.0786	0.0979	0.0696	0.0986	0.0857	0.0845	-	-	0.0577
	SROCC	0.8432	0.7705	0.8430	0.8523	0.9032	0.9082	0.9104	0.9220	0.9245	_	-	0.9413
加权	KROCC	0.6753	0.5826	0.6670	0.6820	0.7257	0.7387	0.7391	0.7678	0.7498	-	-	0.7931
平均	PLCC	0.8707	0.7943	0.8709	0.8783	0.8964	0.9157	0.9044	0.9299	0.9212	—	_	0.9476
	RMSE	0.5749	0.7026	0.5678	0.5595	0.5126	0.4710	0.4918	0.4224	0.4290	—	_	0.3756

其次,为了说明 SFDJF-RF 与人眼主观得分之间的一致 性,绘制了 SFDJF-RF 在 TID2013, TID2008, CSIQ 3 个数据 库上评估结果的散点图,如图3所示。该算法均匀分布在曲 线两侧,这说明其客观评价得分与主观得分有较高的一致性。





4.2 不同失真类型的性能表现

4.1 节比较了本文方法在 3 个数据库上的整体性能表 现,而算法在各种失真类型上的性能表现同样值得关注。一 个好的客观质量评价方法应该对各类不同失真类型也具有很 好的评价性能,而不是只对特定或者某几种失真类型有效。 因此,本节对 TID2008 数据库中的 17 种不同失真类型进 行了实验。TID2008 数据库中各种失真类型在 8 种算法中 的 SROCC 值比较结果如表 3 所列,其中性能表现最好的 两种方法用黑色粗体标出。由表 3 可以看出,SFDJF-RF 标黑 11 次,SC-QI 标黑 8 次,VSI 标黑 5 次,GWL-SSIM 标 黑 5 次。

SFDJF-RF 在大多数失真类型上都有较好的表现,在, mean shift 和 jpg-trans-error 这两种失真类型上表现一般。综 上所述,本文算法在整体上仍比其他算法更具优势。

表 3 8 种鼻法对 TID2008 数据库中 17 种矢具类型的 SROCC 比较	表 3	3 种算法对 TID20	8数据库中17	7 种失真类型的	SROCC 比较	
--	-----	--------------	---------	----------	----------	--

Table 3 SROCC comparison of 17 distortion types on TID2008 database by 8 methods

失真类型	GWL-SSIM	SSIM	FSIMc	IW-SSIM ^[25]	GMSD	VSI	SC-QI	SFDJF-RF
awgn	0.8666	0.8107	0.8566	0.7869	0.9175	0.9229	0.9393	0.9450
awgn-color	0.8311	0.8029	0.8527	0.7920	0.8975	0.9118	0.9074	0.9280
spatial corr-noise	0.8912	0.8144	0.8483	0.7714	0.9132	0.9296	0.9412	0.9424
masked noise	0.7657	0.7795	0.8021	0.8087	0.7087	0.7734	0.8345	0.8723
high-fre-noise	0.9068	0.8729	0.9093	0.8662	0.9188	0.9253	0.9196	0.9368
impulse noise	0.7110	0.6732	0.7452	0.6465	0.6606	0.8298	0.8605	0.9201
quanti-noise	0.8548	0.8531	0.8564	0.8177	0.8897	0.8731	0.9027	0.8997
blur	0.9456	0.9544	0.9472	0.9636	0.8967	0.9529	0.9621	0.9265
denoising	0.9734	0.9530	0.9603	0.9473	0.9753	0.9693	0.9588	0.9703
jpg-comp	0.9374	0.9252	0.9279	0.9184	0.9517	0.9616	0.9513	0.9524
jpe2k-comp	0.9818	0.9625	0.9773	0.9738	0.9795	0.9848	0.9637	0.9778
jpg-trans-error	0.8650	0.8678	0.8708	0.8588	0.8617	0.9160	0.8864	0.8778
jpg2k-trans-error	0.8986	0.8577	0.8544	0.8203	0.8826	0.8942	0.8996	0.8904
pattern-noise	0.7883	0.7107	0.7491	0.7724	0.7603	0.7699	0.7256	0.8442
block-distortion	0.8346	0.8462	0.8492	0.7623	0.8968	0.6295	0.8622	0.8996
mean shift	0.6499	0.7231	0.6720	0.7067	0.6501	0.6714	0.7047	0.6812
contrast	0.6624	0.5246	0.6481	0.6301	0.4647	0.6557	0.6169	0.7386

4.3 感知特征分析

利用 RF 方法进行特征融合时,每种特征对最终预测的 质量得分的贡献不同。为了度量每个特征独自的评价性能以 及它们之间的相互影响,我们比较了每个感知特征对最终评 价结果的贡献。SFDJF-RF 提取了色度、梯度幅值、基于 log-Gabor 的纹理信息以及空间频率 4 个特征。为了比较每个特 征对最终质量评价结果的影响,我们通过组合这 4 种特征来 衡量每个特征的重要性。图 4(a)和图 4(b)分别为使用 4 种 特征组合并利用 RF 融合得到的 SROCC 和 PLCC 值。如图 4 所示,3 条不同的虚线分别表示使用 1 种特征、2 种特征以及 3 种特征预测得到的结果,黑色实线表示同时使用这 4 种特 征得到的结果。可以看出,SROCC 和 PLCC 这两个相关系数 都是在同时使用 4 种特征建立回归模型的情形下达到了最高 值。因此,我们提取的 4 种特征对最终的预测结果都是有用 的,具有互补的效果。





此外,本文对仅使用空域或频域特征时 SFDJF-RF 方法 的表现进行了对比实验,表 4 列出了在 3 个数据库上分别使 用空域特征、频域特征以及两组特征时的 SROCC 值和 PLCC 值。由表 4 可以看出,单独使用空域特征或频域特征时,得到 的评价结果之间相差不大。当两组特征共同作用时,得到的 评价性能远远优于仅使用一组特征。因此,空域和频域特征 对于算法的提出都是必要的,它们之间具有互补性。

表 4 空域和频域特征对预测精度的影响比较

Table 4 Comparison of influence of spatial and frequency domain

features on prediction accuracy

数据库	评价指标	空域特征	频域特征	空域 十频域特征
TID2012	SROCC	0.8735	0.8579	0.9316
11D2013	PLCC	0.8930	0.8813	0.9397
TID2008	SROCC	0.8958	0.8925	0.9444
	PLCC	0.8813	0.9046	0.9485
CSIQ	SROCC	0.9601	0.9537	0.9731
	PLCC	0.9612	0.9603	0.9773

4.4 不同回归方法的性能比较

选取不同回归方法得到的评价结果也会不同,因此,为了 选择合适的回归方法对特征进行融合,本节对 RF,SVR 和 ELM 3 种回归方法进行了比较。表 5 列出了在 3 个数据库上 分别利用 RF,SVR 和 ELM 建立模型得到的 SROCC,PLCC, KROCC 和 RMSE 值。从各个数据库上的 4 个评价指标值可 以看出,采用 RF 进行回归得到的评价指标值均高于 SVR 和 ELM。从加权平均值也可以看出,由 RF 得到的预测性能明 显优于 SVR 和 ELM。出现上述结果的原因在于 RF 由很多 的决策树构成,不仅理论简单,而且耗时较短,准确性较高。 因此,本文选取 RF 为回归工具进行图像特征融合。

表 5 3种回归方法的性能比较

Table 5 Performance comparison of three regression methods

数据库	回归方法	SROCC	PLCC	KROCC	RMSE
	RF	0.9316	0.9397	0.7720	0.4256
TID2013	SVR	0.9108	0.9291	0.7427	0.4579
	ELM	0.8935	0.9083	0.7224	0.5179
	RF	0.9444	0.9485	0.8017	0.4277
TID2008	SVR	0.9356	0.9372	0.7840	0.4697
	ELM	0.9197	0.9218	0.7608	0.5211
	RF	0.9731	0.9773	0.8584	0.0577
CSIQ	SVR	0.9701	0.9705	0.8472	0.0652
	ELM	0.9662	0.9650	0.8396	0.0710
	RF	0.9413	0.9476	0.7927	0.3756
加权平均	SVR	0.9267	0.9373	0.7700	0.4075
	ELM	0.9116	0.9203	0.7505	0.4574

结束语 针对目前大多数 FR IQA 方法只是在空域上或 其变换域上提取图像特征,以及池化策略上存在的局限性,本 文提出了一种彩色的全参考方法:SFDJF-RF。该方法在空域 和频域上分别提取了互补的4个特征:色度、梯度、纹理能量 和空间频率特征,以刻画图像的感知质量。此外,使用 RF 建 立回归模型,预测最终图像感知质量得分。最后,本文在3个 公开的标准数据库上进行了具体实验。单个数据库性能比较 的实验结果表明 SFDJF-RF 的综合性能优于目前一些主流的 FR IQA 算法,并与人眼主观评价具有较高的一致性。不同 失真类型性能比较的结果表明,SFDJF-RF 的鲁棒性较好。 感知特征分析的实验结果表明,提取的特征能够很好地反映 图像感知质量,具有互补作用。不同回归方法的性能比较证 明,利用 RF 进行特征融合后的预测性能优于 SVR 和 ELM。 由于对不同空间互补特征的提取对 FR IQA 重要且有效,因 而本文算法具有较好的泛化性和鲁棒性,相比深度学习方法, 其具有较好的解释性和实时性。

参考文献

- [1] WANG Z,BOVIK A C,WANG Z, et al. Modern Image Quality Assessment[M]. Morgan & Claypool,2006.
- [2] WANG Z, BOVIK A C, SHEIKHH R, et al. Image quality assessment: from error visiblity to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4):600-612.
- [3] ZHANG L,ZHANG L, MOU X Q, et al. FSIM: a feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378-2386.
- [4] GU K,LI L,LU H, et al. A fast reliable image quality predictor by fusing micro and macro-structures[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 3903-3912.
- [5] XUE W F,ZHANG L,MOU X Q,et al. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index [J]. IEEE Transactions on Image Processing, 2014, 23(2):684-695.
- [6] YANG G,LI D,LU F, et al. RVSIM: a feature similarity method for full-reference image quality assessment[J]. EURASIP Journal on Image & Video Processing, 2018, 2018(1):6.
- [7] DONG W.BIE H.LU L, et al. Image quality assessment by considering multiscale and multidirectional visibility differences in shearlet domain[J]. IEEE Access, 2019, 7:78715-78728.
- [8] ZHANG L, SHEN Y, LI H Y. VSI: A visual saliency-induced index for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2014, 23(10):4270-4281.
- [9] SMOLA A J, GUNTUKU S C. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14:199-222.
- [10] BREIMAN L. Random forests[J]. Machine Learing, 2001, 45(1):5-32.
- [11] HUANG G B,ZHU Q Y,SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1):489-501.
- [12] HANSEN L K, SALAMON P. Neural network ensembles [J].
 IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(10):993-1001.
- [13] NARWARIA M,LIN W. 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.

- [14] LIU T J,LIN W,KUO C C J. Image quality assessment using multi-method fusion [J]. IEEE Transactions on Image Processing, 2013, 22(5):1793-1807.
- [15] PEI S C, CHEN L H. Image quality assessment using human visual DOG model fused with random forest[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3282-3292.
- [16] GAO F.WANG Y.LI P.et al. DeepSim:Deep similarity for image quality assessment[J]. Neurocomputing, 2017, 257(27): 104-114.
- [17] LIU H.YANG H.PAN Z K.et al. A learning based image quality assessment model assisted with visual saliency and gradient features[C]//2019 IEEE 4th International Conference on Signal and Image Processing.
- [18] GROSS M H.KOCH R. Visualization of multidimensional shape and texture features in laser range data using complex-valued Gabor wavelets [J]. IEEE Transactions on Visualization and Computer Graphics, 1995, 1(1):44-59.
- [19] FIELD D J. Relations between the statistics of natural images and the response properties of cortical cells[J]. Journal of the Optical Society of America. A. Optics and Image Science, 1987, 4(12):2379-2394.
- [20] PONOMARENKO N, JIN L, LEREMEIEV O, et al. Image database TID2013: Peculiarities, results and perspectives[J]. Signal Processing: Image Communication, 2015,1(30):55-77.
- [21] PONOMARENKO N,LUKIN V,ZELENSKY A.et al. TID2008: a database for evaluation of full-reference visual quality assessment metrics[J]. Advances of Modern Radio electronics,2009,1(10):30-45.
- [22] LARSON E C, CHANDLER D M. Most apparent distortion:full reference image quality assessment and the role of strategy[J]. Journal of Electronic Imaging,2010,19(1):1-21.
- [23] WANG Z.LI Q. Information content weighting for perceptual image quality assessment[J]. IEEE Transaction on Image Processing, 2011, 20(5):1185-1198.
- [24] BAE S H.KIM M. A novel image quality assessment with globally and locally consilient visual quality perception[J]. IEEE Transactions on Image Processing, 2016, 25(5):2392-2406.
- [25] GAO L X,LIU G J,ZHANG X D. Fast lifting IQA algorithm of the structure similarity index [J]. Journal of Computer-Aided Design and Computer Graphics, 2017, 29(7):1314-1323.



YANG Xiao-qin, born in 1996, postgraduate. Her main research interests include image quality assessment and machine learning.



LIU Guo-jun, born in 1978, Ph.D, professor, Ph.D supervisor, master tutor. His main research interests include wavelet and partial differential equations for image processing, image quality assessment, and machine learning.