 CMC, vol.56, no.3, pp.501-515, 2018

Perceptual Gradient Similarity Deviation for Full Reference Image Quality Assessment

Manyu Jin[1], Tao Wang[1], Zexuan Ji[1, *] and Xiaobo Shen[2]

Abstract: Perceptual image quality assessment (IQA) is one of the most indispensable yet challenging problems in image processing and computer vision. It is quite necessary to develop automatic and efficient approaches that can accurately predict perceptual image quality consistently with human subjective evaluation. To further improve the prediction accuracy for the distortion of color images, in this paper, we propose a novel effective and efficient IQA model, called perceptual gradient similarity deviation (PGSD). Based on the gradient magnitude similarity, we proposed a gradient direction selection method to automatically determine the pixel-wise perceptual gradient. The luminance and chrominance channels are both took into account to characterize the quality degradation caused by intensity and color distortions. Finally, a multi-scale strategy is utilized and pooled with different weights to incorporate image details at different resolutions. Experimental results on LIVE, CSIQ and TID2013 databases demonstrate the superior performances of the proposed algorithm.

Keywords: Image quality assessment, full reference, perceptual gradient similarity, multi-scale, standard deviation pooling.

1 Introduction

Objective image quality assessment (IQA) is a significant basis to measure the performance of image acquisition, image transmission and image processing algorithms. Most existing methods focus on extracting image features related to human visual system (HVS) to establish evaluation models. Based on whether there is a reference image, IQA techniques are generally classified into three categories, named full reference (FR), reduced reference (RR) and no reference (NR). FR-IQA considers that the reference image is undistorted, and measures the quality by calculating the differences between a reference image and a distorted image. By the in-depth study of imaging theory, image processing techniques such as Hermite spectral collocation, extended Hamiltonian algorithm, seeded region growing, collaborative representation [Zhang, Sun, Ji et al. (2016)] and Hidden Markov model [Zheng, Jeon, Sun et al. (2017)] have been widely used in computer vision. However, the critical problem of IQA is the extraction of

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China.

[2] School of Computer Science and Engineering, Nanyang Technology University, 639798, Singapore.

[*] Corresponding Author: Zexuan Ji. Email: jizexuan@njust.edu.cn.

quality-related features consistent with human eye perception. The most classical objective IQA algorithms are Peak Signal-to-Noise Ratio (PSNR) [Avcibas, Sankur and Sayood (2002)] and Mean-Squared Error (MSE). However, both methods are not well associated with the perceptual quality without considering the structural characteristics. To overcome this limitation, structural similarity index (SSIM) [Wang, Bovik, Sheikh et al (2004)] utilizes correlations between pixels and the concept of structure information to measure the quality score. Inspired by SSIM, a variety of improved algorithms have been proposed, including gradient based structural similarity (GSSIM) [Chen, Yang and Xie (2006)], multi-scale structural similarity Index (MS-SSIM) [Wang, Simoncelli and Bovik (2004)], and information content weighted SSIM (IW-SSIM) [Wang and Li (2011)]. These approaches are consistent with human subjective perception to a certain degree.

Moreover, many models are also developed based on the properties of human vision. Based on the perspective of information theory, Sheikh et al. [Sheikh, Bovik and De (2005)] pointed out that the natural images have statistical characteristics, and proposed information fidelity criterion (IFC) and visual information fidelity (VIF) [Sheikh and Bovik (2006)]. Both methods utilize the mutual information between the original image and the distorted image. Chandler et al. [Chandler and Hemami (2007)] proposed the visual signal-to-noise ratio (VSNR) metric to quantify the visual fidelity of natural images based on near-threshold and supra-threshold properties of human vision. Most apparent distortion (MAD) [Larson and Chandler (2010)] assumed that the HVS uses two distinct strategies when evaluating high-quality images and low-quality images. Based on the hypothesis that the visual saliency map is closely related to image quality, Zhang et al. [Zhang and Li (2013)] proposed spectral residual based similarity (SR-SIM) method. Then they further proposed visual saliency induced index (VSI) method [Zhang, Shen and Li (2014)]. VSI utilizes the visual saliency map as both feature and weighting function to reflect the importance of local regions.

Neuropsychological studies have shown that the human visual system is sensitive to the structural distortions of edge details in the image. Most distortions have a tendency to change the gradient values. The image gradient reflects the most significant part of the image brightness changes, which is often used to extract the edge and other structures. Gradient information has been employed for FR-IQA in many different ways. GSSIM improves SSIM by replacing the contrast comparison with the gradient based contrast comparison. Gradient magnitude similarity (GSM) [Liu, Lin and Narwaria (2012)] also uses such information to capture structural and contrast changes. Zhang et al. [Zhang, Zhang, Mou et al. (2011)] constructed feature similarity (FSIM/FSIMc for color images) to further improve the performance. FSIM calculates gradient similarity and phase similarity respectively, regarding the gradient as an independent feature and pooling with a phase congruency weighted average. Because of high computational complexity of phase congruency features, Xue et al. [Xue, Zhang, Mou et al. (2014)] proposed an effective metric called gradient magnitude similarity deviation (GMSD), which first calculates the gradient magnitude similarity and then uses a standard deviation pooling strategy to get the evaluation score. GMSD proves that utilization of image gradient magnitude can yield highly accurate quality prediction.

Most related methods model the FR-IQA strategy in gray space. However, color is the essential element to describe the content of images, where RGB model is the most common color model in computer vision. However, RGB model cannot separate the luminance and chrominance, which is not consistent with human subjective perception of color similarity. The other well-known color spaces include HSV, YIQ, Lab and so on. HSV (hue, saturation, value) model is a pyramid color space which is more closely associated with the way human vision perceives color-making attributes. YIQ model contains three components: luminance value Y and two chrominance values I and Q. The relationship between YIQ and RGB color space is a linear transformation, and YIQ can be adapted for the change of lightness intensity. Like the YIQ, the Lab model is comprised of one luminance and two chrominance channels. Extends from FSIM, FSIMc uses YIQ color space to achieve color image quality assessment. It computes phase congruency and gradient similarity in luminance component combined with the chromatic similarities in I and Q channels. CSV [Temel and Alregib (2016)] uses the CIEDE2000 color difference formulation to quantify low-level color degradation and calculates the Earth mover's distance between color name probability vectors to measure significant color degradation.

Generally, the perceptual quality is influenced by numerous factors, including display resolution, chrominance information and viewing distance. A natural image might have objects and structures that are relevant at different scales, and the human eye is readily able to identify and process the information presented by it. MS-SSIM proposes a multi-scale structural similarity method, which is more stable than the single-scale SSIM model. Multi-scale contrast similarity deviation (MCSD) [Wang, Zhang, Jia et al. (2016)] explores the contrast features by resorting to the multi-scale representation.

As the above discussions, the gradient only computed on the luminance channel of image, would make GMSD not work quite well for the color distortion. In this paper, we propose perceptual gradient similarity deviation (PGSD) to further improve the prediction accuracy for the distortion of color images. Inspired by GMSD, we propose a gradient direction selection method to automatically determine the pixel-wise perceptual gradient. Both the luminance and chrominance channels are taken into consideration when characterize the quality degradation caused by intensity and color distortions. Finally, a multi-scale strategy is utilized and pooled with distinct weights to incorporate image details at different resolutions and obtain the final score.

2 Perceptual gradient similarity deviation

2.1 Selection of gradient direction

Research has shown that the human visual system is sensitive to the edge of image. The gradient of the image can reflect the detail contrast and texture change, and is closely related to the perceptual quality. In this paper, we improve the existing gradient calculation method by automatically selecting the gradient directions.

Similar with GMSD, we first adapt the Prewitt filters as gradient operators to get the gradient magnitude. The operators of horizontal (x) and vertical (y) are defined as follows:

$$h_x = \frac{1}{3}\begin{bmatrix}1 & 0 & -1\\ 1 & 0 & -1\\ 1 & 0 & -1\end{bmatrix} h_y = \frac{1}{3}\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix} \tag{1}$$

These operators constitute an orthogonal coordinate system. The gradient magnitude (GM) is defined as the root mean square of direction gradients along these two orthogonal directions. GM maps of the reference (*r*) and the distorted (*d*) image are computed as follows:

$$m_{r1}(i) = \sqrt{(r \otimes h_x)^2(i) + \left(r \otimes h_y\right)^2(i)} \tag{2}$$

$$m_{d1}(i) = \sqrt{(d \otimes h_x)^2(i) + \left(d \otimes h_y\right)^2(i)} \tag{3}$$

where symbol $\otimes$ denotes the convolution operation, m_{r1} and m_{d1} are the gradient magnitude images extracted from the reference and distorted images respectively, and $i = (i_1, i_2)$ represents the pixel coordinates.

In order to find the direction of maximum gradient change more accurately, we introduce another set of filters, which are defined as follows:

$$h_c = \frac{1}{3}\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & -1\\ 0 & -1 & -1\end{bmatrix} h_m = \frac{1}{3}\begin{bmatrix}0 & 1 & 1\\ -1 & 0 & 1\\ -1 & -1 & 0\end{bmatrix} \tag{4}$$

These two operators are also orthogonal. But different from the traditional ones, they are more efficient in capturing the sloped edges. The horizontal and vertical operators can only capture one component of the sloped edges, which can be, however, captured by one of the diagonal operators. Then the corresponding GM maps can be calculated with convolution operation like Eqs. (2) and (3):

$$m_{r2}(i) = \sqrt{(r \otimes h_c)^2(i) + (r \otimes h_m)^2(i)} \tag{5}$$

$$m_{d2}(i) = \sqrt{(d \otimes h_c)^2(i) + (d \otimes h_m)^2(i)} \tag{6}$$

For each image, we can obtain two GM images (m_{r1}, m_{r2} or m_{d1}, m_{d2}). For the reference image, two GM images are compared pixel-by-pixel and the larger values are selected to construct the final GM map:

$$G_r(i) = max(m_{r1}(i), m_{r2}(i)) \tag{7}$$

The purpose of this is to determine the gradient direction that is closer to the maximum rate of change. To ensure that the gradient values being compared come from the same coordinate system, for the distorted image, GM map is constructed based on the reference one, which can be defined as

$$G_d(i) = \begin{cases} m_{d1}(i), & m_{r1}(i) \geq m_{r2}(i) \\ m_{d2}(i), & m_{r1}(i) < m_{r2}(i) \end{cases} \tag{8}$$

Therefore, we can obtain the final GM maps for both reference and distorted images. We select the gradient direction of each pixel on the maximum changing rate based on the reference image, which makes the comparison of gradient more accurately.

2.2 Pooling luminance and chrominance channels

Many researches and experiments show that the visual information enters the visual cortex through different neural channels and is then processed by different neurons. The color feature is another type of information which reflects the image content except the brightness features. Calculating the gradient similarity only on gray scale cannot guarantee an accurate evaluation for color distortions, such as change of color saturation and chromatic aberrations. Therefore, the utilization of multiple channels including luminance and chrominance channels for IQA model can extract more perceptual features that are more consistent with human perception. Hence, we transform the RGB color images into an opponent color space [Geusebroek, Boomgaard, Smeulders et al. (2001)]:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{9}$$

where L represents the luminance information and M and N contain the chrominance information. The conversion weights are optimized for the HVS.

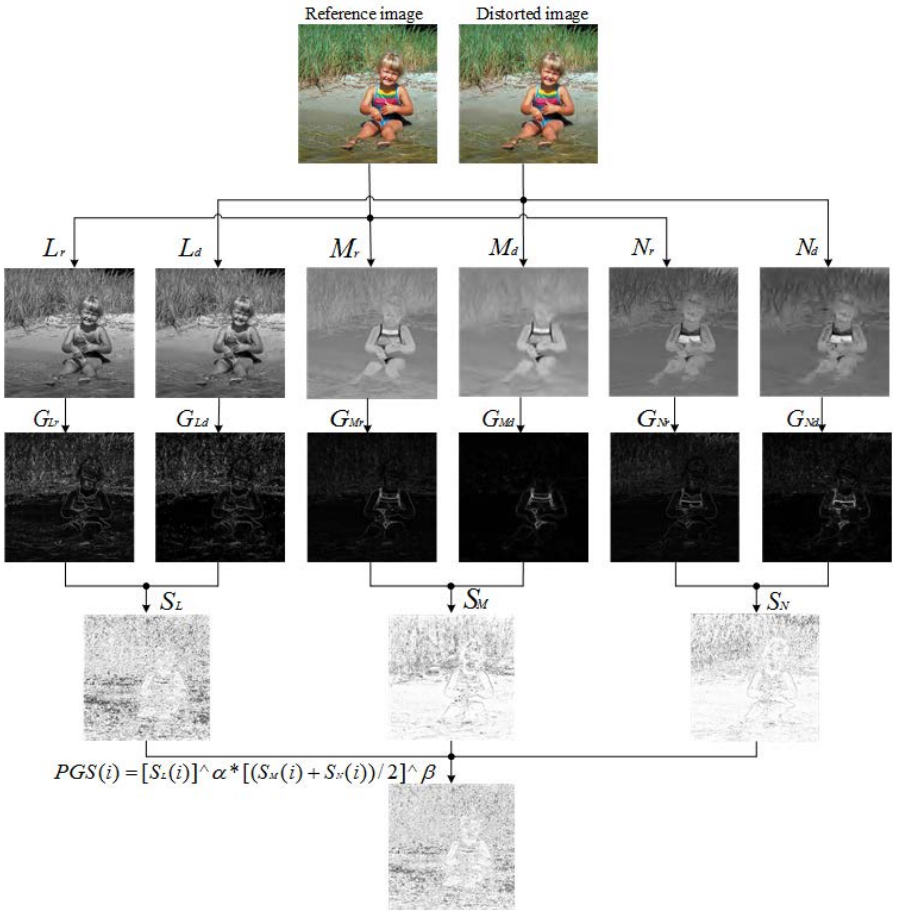

Figure 1: Illustration of the computational process of the proposed PGS map

Then the gradient magnitude maps are respectively calculated on these three channels by using the direction selection method mentioned in Section 2.1. For each image, we get three GM maps: G_L, G_M and G_N. The similarities calculated between two GM maps of each channel are defined as:

$$S_L(i) = \frac{2G_{Lr}(i)G_{Ld}(i)+c_1}{G_{Lr}^2(i)+G_{Ld}^2(i)+c_1} \tag{10}$$

$$S_M(i) = \frac{2G_{Mr}(i)G_{Md}(i)+c_2}{G_{Mr}^2(i)+G_{Md}^2(i)+c_2} \tag{11}$$

$$S_N(i) = \frac{2G_{Nr}(i)G_{Nd}(i)+c_2}{G_{Nr}^2(i)+G_{Nd}^2(i)+c_2} \tag{12}$$

where c_1 and c_2 are the positive constants. Then, we combine S_M and S_N to obtain the chrominance similarity measure, which is denoted by S_C:

$$S_C(i) = \frac{S_M(i)+S_N(i)}{2} \tag{13}$$

The similarities with respect to luminance and chrominance are described as follows:

$$PGS(i) = [S_L(i)]^\alpha \cdot [S_C(i)]^\beta \tag{14}$$

where α and β are two positive parameters to adjust the relative importance of luminance and chrominance. The procedures to calculate the PGS map are illustrated in Fig. 1.

Finally, we apply with standard deviation pooling and take the result as the IQA index called perceptual gradient similarity deviation (PGSD):

$$PGSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(PGS(i) - PGSM)^2} \tag{15}$$

where *N* is the total number of pixels in the image. And the perceptual gradient similarity means (PGSM) is the average of the PGS map and is defined as:

$$PGSM = \frac{1}{N}\sum_{i=1}^{N} PGS(i) \tag{16}$$

The value of PGSD reflects the range of distortion severities in an image, where a higher score means a larger distortion range and a lower image perceptual quality, and vice versa.

Fig. 2 shows two examples by comparing PGSD with GMSD. GMSD computes the gradient magnitude only on the gray scale. Both testing images contain declines of color saturation but with different levels. The subject scores (MOS) are 4.31707 and 3.60978, respectively. A higher MOS means better image quality. Corresponding PGSD scores are 0.0131 and 0.0499 which are consistent with the trend of MOS. However, there is no obvious change between GMSD scores, which means that GMSD method can hardly detect the change of perceptual quality. The proposed PGSD is able to consider the variation of local quality and evaluate the color distortion. Consequently, it can obtain highly relevant results to subjective image quality.

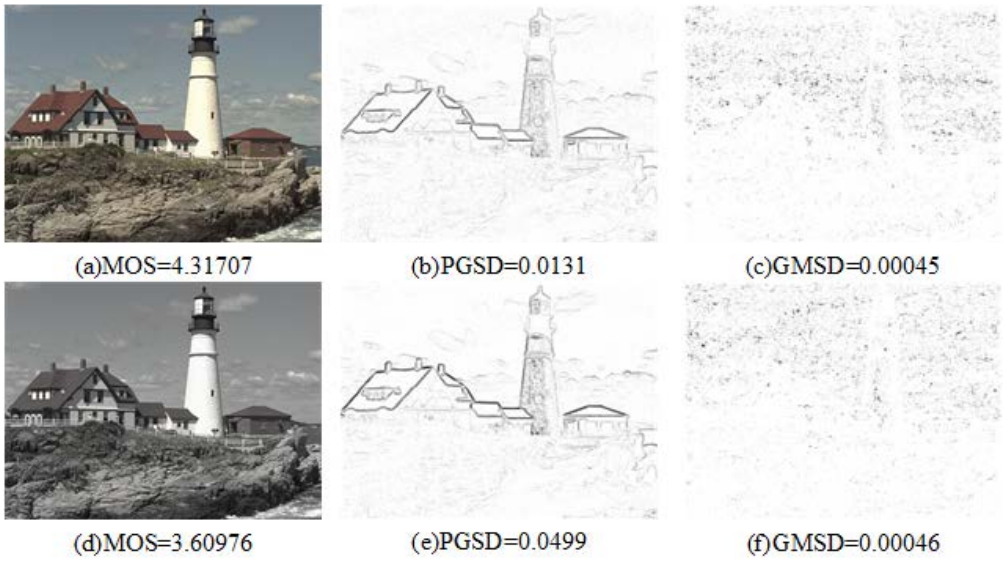

Figure 2: Examples of GMSD and PGSD for change of color saturation evaluation. (a) and (d) are two distorted images with different levels in TID2013 database. (b) and (e) are the PGS maps calculated by proposed PGSD. (c) and (f) are the GMS map calculated by GMSD model. The subjective quality scores (MOS), the PGSD indexes and GMSD indexes are listed under the corresponding images

2.3 Multi-scale pooling strategy

Typical multi-scale methods include wavelet transform, curvelet transform, and pyramid decomposition. In image quality assessment, the perceptual characteristics of the image are closely related to the observed distance and the sampling density. Therefore, scale information also has an effect on the perceptual quality.

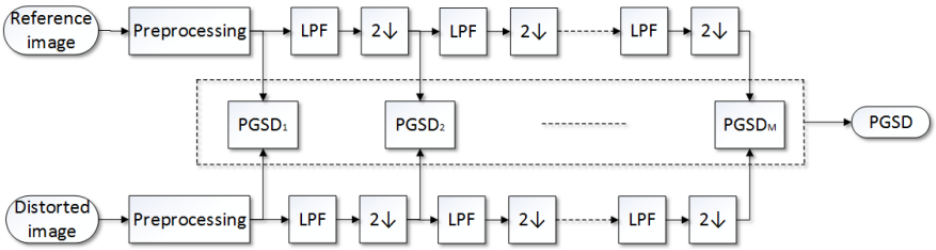

Figure 3: Multi-scale framework

In this paper, we combine multi-scale strategy by utilizing low-pass filter and down-sampling in three channels of a color image. The flow chart is presented in Fig. 3. In image preprocessing, the color image is transformed into the opponent color space as mentioned in Section 2.2.

On the basis of the previous scale level, the luminance and two chrominance images are first filtered by a low-pass filter, and then down-sampled by a factor of 2. The original

image is indexed as Scale 1 in this paper, and the largest scale is indexed as Scale M. The overall evaluation score PGSD is obtained by multiplying different weights for different scales, which are expressed as:

$$PGSD = \sum_{i=1}^{M} \omega_i * PGSD_i \quad (17)$$

where PGSD_i and ω_i respectively represent the single-scale IQA score and the weight at the *i-th* scale and $\sum_{i=1}^{M} \omega_i = 1$.

3 Experimental results

In this section, we tested the performances of the proposed PGSD on three databases: LIVE database [Sheikh, Moorthy, Wang et al. (2004)], CSIQ database [Larson and Chandler (2009)], and TID 2013 database [Ponomarenko, Ieremeiev, Lukin et al. (2013)]. Tab. 1 shows the main information of these three databases.

Table 1: Description of testing databases

Database	Reference images	Distorted images	Distorted types	MOS range	Observers
LIVE	29	779	5	0-120	161
CSIQ	30	866	6	0-1	35
TID2013	25	3000	24	0-8	971

Four common indices were used to evaluate the prediction accuracy and consistency, including the Spearman Rank-Order Correlation Coefficient (SROCC), Kendall Rank-Order Correlation Coefficient (KROCC), Pearson Linear Correlation Coefficient (PLCC) and Root Mean Squared Error (RMSE). The values of SROCC, KROCC and PLCC all range from 0 to 1, with a higher value representing a more accurate evaluation. For RMSE, a lower value indicates that the corresponding algorithm can produce more accurate estimation. Generally, a logistic regression function is utilized to provide a nonlinear mapping between the objective scores and the subjective mean opinion scores (MOS), which is defined as:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{exp(\beta_2 (x - \beta_3))} \right) + \beta_4 x + \beta_5 \quad (18)$$

where β_i, i=1,2,…,5 are regression model parameters, and x is the predicted image quality. After the regression, the above four indices can be calculated for the performance evaluations.

Unless otherwise specified, parameters involved in the proposed PGSD were set as follows. We set the constants c_1=170 and c_2=180 in Eqs. (10) and (11), respectively. The factors α and β in Eq. (14) were set as $\alpha = 0.6$ and $\beta = 0.4$, respectively. The multi-scale weights in Eq. (17) were set as ω=[0.1333, 0.3448, 0.2856, 0.2363].

The proposed PGSD was compared with eight state-of-the-art FR-IQA techniques, including SSIM, MS-SSIM, IW-SSIM, GMS, FSIMc, GMSD, CSV and MCSD. Both FSIMc and CSV can evaluate color images. It should be noted that the parameters for all the comparison algorithms were set as the suggestion values from the corresponding references.

Table 2: Performance evaluation results in LIVE, CSIQ and TID2013 databases

Database	Method	SROCC	KROCC	PLCC	RMSE
LIVE	SSIM	0.9479	0.7963	0.9449	8.9455
	IFC	0.9259	0.7579	0.9268	10.264
	MS-SSIM	0.9513	0.8045	0.9489	8.6188
	IW-SSIM	0.9567	0.8175	0.9522	8.3473
	GSM	0.9561	0.8150	0.9512	8.4327
	FSIMc	**0.9645**	**0.8363**	**0.9613**	**7.5296**
	GMSD	**0.9603**	0.8268	**0.9603**	7.6214
	CSV	**0.9610**	**0.8380**	0.9492	**7.2762**
	PGSD	0.9575	**0.8318**	**0.9572**	**6.6872**
CSIQ	SSIM	0.8755	0.6900	0.8612	0.1334
	IFC	0.7671	0.5897	0.8384	0.1434
	MS-SSIM	0.9133	0.7393	0.8991	0.1149
	IW-SSIM	0.9213	0.7529	0.9144	0.1063
	GSM	0.9108	0.7374	0.8964	0.1164
	FSIMc	0.9310	**0.7690**	0.9192	0.1034
	GMSD	**0.9570**	**0.8122**	**0.9541**	**0.0786**
	CSV	**0.9328**	0.7661	**0.9398**	**0.0897**
	PGSD	**0.9572**	**0.8122**	**0.9564**	**0.0767**
TID2013	SSIM	0.7417	0.5588	0.7895	0.7608
	IFC	0.5389	0.3939	0.5538	1.0322
	MS-SSIM	0.7859	0.6047	0.8329	0.6861
	IW-SSIM	0.7779	0.5977	0.8319	0.6880
	GSM	0.7946	0.6255	0.8464	0.6603
	FSIMc	**0.8510**	**0.6665**	**0.8769**	**0.5959**
	GMSD	0.8038	0.6334	0.8594	0.6339
	CSV	**0.8456**	**0.6544**	**0.8623**	**0.6279**
	PGSD	**0.8565**	**0.6685**	**0.8611**	**0.6303**

In the first experiment, all the four performance indices were examined in three testing databases. The results are presented in Tab. 2, in which the top three performances of each indicator are highlighted in boldface. The best ranking models are PGSD (10 times), CSV (7 times), FSIMc (7 times), MCSD (7 times) and GMSD (6 times). For the CSIQ and TID2013 datasets, all PGSD's indicators reach the top three. FSIMc, GMSD, MCSD and CSV can only obtain satisfactory results on one specific database. We also showed the weighted average of the four indices over the three databases in Tab. 3. We can observe that PGSD has the best performance in SROCC, KROCC and RMSE. The overall performance comparison shows that our improvements are effective.

Table 3: Weighted average results of LIVE, CSIQ and TID2013 databases

	SSIM	MS-SSIM	IW-SSIM	GSM	FSIMc	GMSD	CSV	MCSD	PGSD
SROCC	0.8019	0.8379	0.8351	0.8438	0.8853	0.8590	0.8816	0.8627	**0.8926**
KROCC	0.6238	0.6639	0.6642	0.6787	0.7146	0.6997	0.7066	0.7053	**0.7234**
PLCC	0.8294	0.8651	0.8678	0.8736	**0.8992**	0.8943	0.8916	0.8917	0.8952
RMSE	2.0462	1.9387	1.8919	1.8906	1.6920	1.7276	1.6667	1.6132	**1.5654**

Generally, a good IQA model should also have the ability to accurately predict image quality for each specific type of distortions. Tab. 4 lists the SROCC scores of each type in three databases in which the top three performances are highlighted in boldface. For the TID2013 database, the proposed PGSD has better performance for color distortion than the conventional GMSD, which means that the color channel decomposition is effective. We can observe that the proposed PGSD obtain 27 times ranking in top three models for all 35 distortion types, followed by MCSD and GMSD with 24 times and 18 times, respectively.

For the color distortions of contrast change and change of color saturation, the SROCC values of GMSD are 0.3235 and 0.2948, respectively. The corresponding indexes of PGSD are 0.6343 and 0.7785, both of which are ranked in the top three. All the SROCC values of PGSD are above 0.6, which indicates that PGSD is almost valid for all distortion types.

Fig. 4 shows the scatter plots of predicted quality scores against subjective DMOS scores for all the comparison IQA models on TID2013 database. The curves were obtained by a nonlinear fitting mentioned in Eq. (18). The scatter distribution of MS-SSIM is more centralized than SSIM, which means that the multi-scale strategy is effective. In the scatter distributions of GMSD and MCSD, some points concentrated in the straight line with zero predicted scores, which indicates that the corresponding distortions are accurately evaluated. Comparatively, the scatter distributions of PGSD are more concentrated than others, which mean that the subjective and objective are more consistent.

Table 4: SROCC performance comparison on each individual distortion

	Type	SSIM	MS-SSIM	IW-SSIM	GSM	FSIMc	GMSD	CSV	MCSD	PGSD
LIVE	JP2K	0.9614	0.9627	0.9649	0.9700	0.9724	0.9711	**0.9819**	**0.9825**	**0.9820**
	JPEG	0.9764	**0.9815**	**0.9808**	0.9778	**0.9840**	0.9782	0.9625	0.9613	0.9619
	AGWN	0.9694	0.9733	0.9667	0.9774	0.9716	0.9737	**0.9798**	**0.9889**	**0.9894**
	GB	0.9517	0.9542	0.9720	0.9518	0.9708	0.9567	**0.9832**	**0.9728**	**0.9752**
	FF	0.9556	0.9471	0.9442	0.9402	0.9499	0.9416	**0.9707**	**0.9723**	**0.9681**
CSIQ	AGWN	0.8974	0.9471	0.9380	0.9440	0.9359	**0.9676**	0.9556	**0.9674**	**0.9673**
	JPEG	0.9546	0.9634	0.9662	0.9632	**0.9664**	0.9651	0.9646	**0.9670**	**0.9695**
	JP2K	0.9606	0.9683	0.9683	0.9648	0.9704	0.9717	**0.9794**	**0.9746**	**0.9790**
	AGPN	0.8922	0.9331	0.9059	0.9387	0.9370	**0.9502**	**0.9629**	0.9479	**0.9513**
	GB	0.9609	0.9711	**0.9782**	0.9589	0.9729	0.9712	**0.9753**	0.9747	**0.9772**
	GCD	0.7922	**0.9526**	**0.9539**	0.9354	0.9438	0.9037	0.9462	**0.9509**	0.9488
TID2013	AGWN	0.8671	0.8646	0.8438	0.9064	0.9101	**0.9462**	0.9223	**0.9451**	**0.9521**
	ACN	0.7726	0.7730	0.7515	0.8175	0.8537	**0.8684**	0.8525	**0.8671**	**0.8688**
	SCN	0.8515	0.8544	0.8167	0.9158	0.8900	**0.9350**	0.9297	**0.9404**	**0.9463**
	MN	0.7767	**0.8073**	**0.8020**	0.7293	**0.8094**	0.7075	0.7977	0.7234	0.7561
	HFN	0.8634	0.8604	0.8553	0.8869	0.9040	0.9162	**0.9168**	**0.9166**	**0.9184**
	IN	0.7503	0.7629	0.7281	0.7965	**0.8251**	0.7637	**0.8426**	0.7769	**0.8124**
	QN	0.8657	0.8706	0.8468	0.8841	0.8807	**0.9049**	0.8791	**0.9103**	**0.8958**
	GB	0.9668	**0.9673**	**0.9701**	**0.9689**	0.8551	0.9113	0.9374	0.9095	0.9054
	ID	0.9254	0.9268	0.9152	0.9432	0.9330	**0.9525**	0.9430	**0.9537**	**0.9566**
	JPEG	0.9200	0.9265	0.9187	0.9284	0.9339	**0.9507**	0.9416	**0.9431**	**0.9468**
	JP2K	0.9468	0.9504	0.9506	0.9602	0.9589	**0.9657**	**0.9670**	0.9626	**0.9650**
	JGTE	0.8493	0.8475	0.8388	**0.8512**	**0.8610**	0.8403	0.8079	0.7466	**0.8694**
	J2TE	0.8828	0.8889	0.8656	**0.9182**	0.8919	**0.9136**	0.8927	**0.9230**	0.9092
	NEPN	0.7821	0.7968	0.8011	0.8130	0.7937	**0.8140**	0.8073	**0.8309**	**0.8306**
	LBWD	0.5720	0.4801	0.3717	**0.6418**	0.5532	**0.6625**	0.1829	**0.6718**	0.6164
	MS	0.7752	**0.7906**	**0.7833**	**0.7875**	0.7487	0.7351	0.6614	0.5961	0.6442
	CTC	0.3775	0.4634	0.4593	**0.4857**	0.4679	0.3235	0.2214	**0.6889**	**0.6320**
	CCS	0.4141	0.4099	0.4196	0.3578	**0.8359**	0.2948	**0.8076**	0.2715	**0.7772**
	MGN	0.7803	0.7786	0.7728	0.8348	0.8569	**0.8886**	0.8649	**0.8857**	**0.8968**
	CN	0.8566	0.8528	0.8762	0.9124	0.9135	**0.9298**	0.9138	**0.9333**	**0.9385**
	LCNI	0.9057	0.9068	0.9037	0.9563	0.9485	**0.9629**	0.9559	**0.9697**	**0.9721**
	ICQD	0.8542	0.8555	0.8401	0.8973	0.8815	**0.9102**	0.9087	**0.9172**	**0.9144**
	CA	0.8775	**0.8784**	0.8682	**0.8823**	**0.8925**	0.8530	0.8449	0.8390	0.8589
	SSR	0.9461	0.9483	0.9474	0.9668	0.9576	0.9638	**0.9697**	**0.9678**	**0.9676**

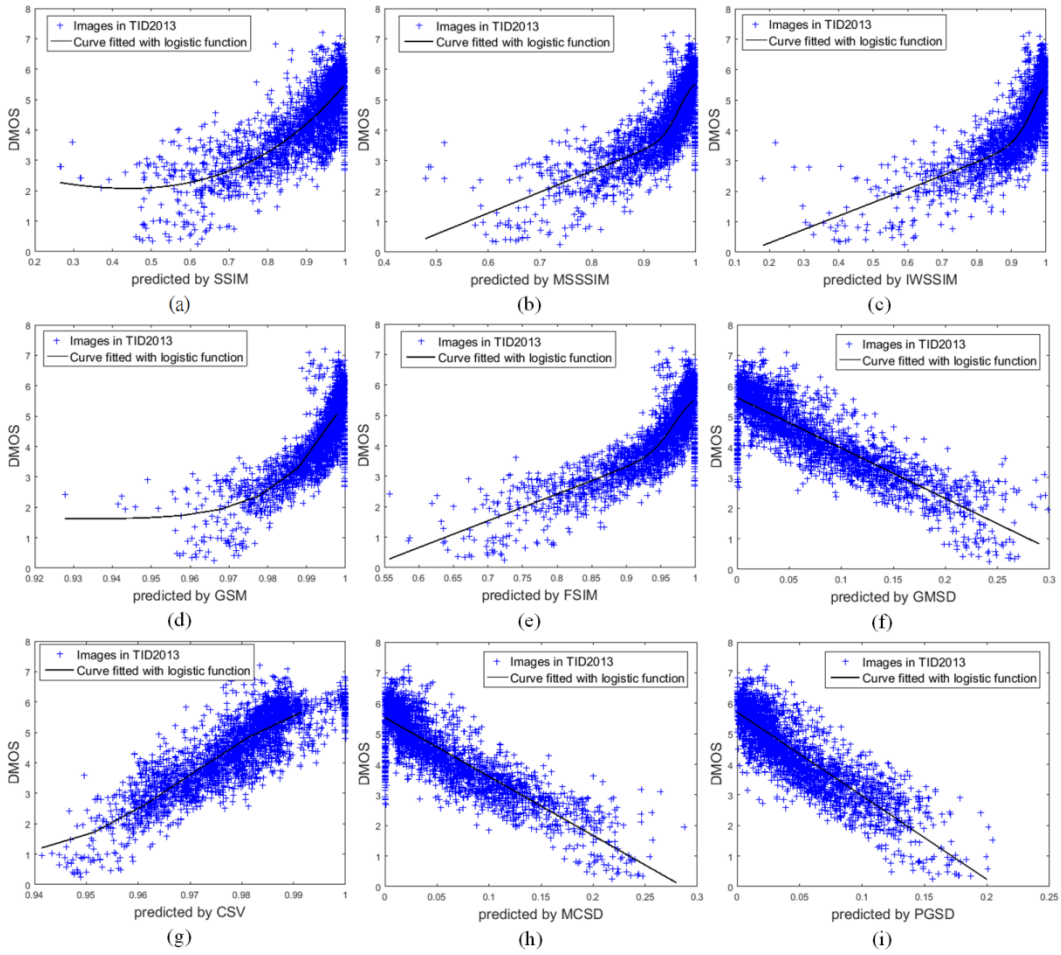

Figure 4: Scatter plots of subjective DMOS against predicted quality scores by IQA models on the TID2013 database. (a) SSIM, (b) MS-SSIM, (c) IW-SSIM, (d) GSM, (e) FSIMc, (f) GMSD, (g) CSV, (h) MCSD and (i) PGSD

In addition, we evaluated the performances using statistical significance tests to make statistically meaningful conclusions. After nonlinear regression, we compared the prediction residuals of each two models by applying the left-tailed *F*-test at a significance level of 0.05. A value of H=1 indicates the first model (represented by the row in Fig. 5) is superior to the second one (represented by the column of Fig. 5) in IQA performance. A value of H=0 means that the first model is not significantly better than the second one. Fig. 5(a)-5(c) show the significance test results on the LIVE, CSIQ and TID2013 databases, respectively. We can find that the PGSD is significantly better than most models on the CSIQ database. For the LIVE database, PGSD is significantly better than all the others except for CSV. In general, considering 0.05 as the level of significance, this evaluation apparently shows that PGSD performs steadily better than most comparison methods.

Except for accuracy, a good IQA model should also have high efficiency. To analyze the complexities of all the comparison IQA models, Tab. 5 lists the running time on the image with size 512×512, where the order is the length of execution time. All algorithms were implemented using the Matlab R2016a platform and were tested on a PC (Intel Core i3-2120 CPU, 3.30 GHz, 6 GB RAM, and 64-bit Windows 8). The execution time is the average value of 30 repetitions for each model. From Tab. 5 we can observe that the proposed PGSD is the fastest among the models that can process color image, which is 1.7 times faster than FSIMc, and 4.6 times faster than CSV. Therefore, we can conclude that the proposed method outperforms state-of-the-art methods in both terms of prediction accuracy and efficiency.

LIVE	SSIM	MSSSIM	IWSSIM	GSM	FSIMc	GMSD	CSV	MCSD	PGSD
SSIM	0	0	0	0	0	0	0	0	0
MSSSIM	0	0	0	0	0	0	0	0	0
IWSSIM	1	0	0	0	0	0	0	0	0
GSM	1	0	0	0	0	0	0	0	0
FSIMc	1	1	1	1	0	0	0	0	0
GMSD	1	1	1	1	0	0	0	0	0
CSV	1	1	1	1	1	1	0	1	1
MCSD	1	1	0	1	0	0	0	0	0
PGSD	1	1	1	1	1	1	0	0	0

(a)

CSIQ	SSIM	MSSSIM	IWSSIM	GSM	FSIMc	GMSD	CSV	MCSD	PGSD
SSIM	0	1	0	0	0	0	0	0	0
MSSSIM	1	0	0	0	0	0	0	0	0
IWSSIM	1	1	0	1	0	0	0	0	0
GSM	1	0	0	0	0	0	0	0	0
FSIMc	1	1	0	1	0	0	0	0	0
GMSD	1	1	1	1	1	0	1	0	1
CSV	1	1	1	1	1	0	0	0	0
MCSD	1	1	1	1	1	0	1	0	1
PGSD	1	1	1	1	1	0	0	0	0

(b)

TID 2013	SSIM	MSSSIM	IWSSIM	GSM	FSIMc	GMSD	CSV	MCSD	PGSD
SSIM	0	0	0	0	0	0	0	0	0
MSSSIM	1	0	0	0	0	0	0	0	0
IWSSIM	1	0	0	0	0	0	0	0	0
GSM	1	1	1	0	0	0	0	0	0
FSIMc	1	1	1	1	0	1	0	0	0
GMSD	1	1	1	1	0	0	0	0	0
CSV	1	1	1	1	0	0	0	0	0
MCSD	1	1	1	1	0	0	0	0	0
PGSD	1	1	1	1	0	0	0	0	0

(c)

Figure 5: The results of statistical significance tests for all the comparison IQA models on (a) LIVE, (b) CSIQ and (c) TID2013 databases. The value 1 indicates that the model in the row is significantly better than the model in the column, while the value 0 indicates that two comparison methods have no significant difference

Table 5: Average execution time of all comparison IQA models

models	Running time (s)
GMSD	0.0061
MCSD	0.0124
GSM	0.0189
SSIM	0.0437
MS-SSIM	0.0519
PGSD	0.1906
FSIMc	0.3493
IW-SSIM	0.4309
CSV	0.8722

4 Conclusion

In this paper, we have proposed a FR-IQA model called perceptual gradient similarity deviation (PGSD) which considers the perceptual quality related to HVS, including the texture edge, chrominance information and viewing distance. To fully take all the

direction changes into account, a gradient direction selection method has been proposed to automatically determine the pixel-wise perceptual gradient. Then both the luminance and chrominance channels have been taken into account to characterize the quality degradation caused by intensity and color distortions. Finally, we have presented multi-scale pooling strategy which is more accurate and stable than single-scale assessment. The experimental results demonstrate that the proposed PGSD outperforms state-of-the-art methods in terms of prediction accuracy and efficiency. Future work will be devoted to further reduce the complexity of the proposed algorithm and consider more visual features, such as saliency induced index in Zhang et al. [Zhang, Shen and Li (2014)] to further improve the prediction accuracy for the distortion of color images.

References

Avcibas, I.; Sankur, B.; Sayood, K. (2002): Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 206-223.

Chandler, D. M.; Hemami, S. S. (2007): VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Process*, vol. 16, no. 9, pp. 2284-2298.

Chen, G. H.; Yang, C. L.; Xie, S. L. (2006): Gradient-based structural similarity for image quality assessment. *IEEE International Conference on Image Processing*, pp. 2929-2932.

Geusebroek, J. M.; Boomgaard, R. V. D.; Smeulders, A. W. M.; Geerts, H. (2001): Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338-1350.

Larson, E. C.; Chandler, D. M. (2010): Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, vol. 19, no. 1.

Larson, E. C.; Chandler, D. M. (2009): Most apparent distortion: A dual strategy for full-reference image quality assessment. *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 7242, pp. 1-17.

Liu, A.; Lin, W.; Narwaria, M. (2012): Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500-1512.

Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Jin, L. et al. (2013): Color image database TID2013: Peculiarities and preliminary results. *European Workshop on Visual Information Processing*, pp. 106-111.

Sheikh, H. R.; Bovik, A. C.; De, V. G. (2005): An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117-2128.

Sheikh, H. R.; Bovik, A. C. (2006): Image information and visual quality. *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444.

Sheikh, H. R.; Moorthy, A. K.; Wang, Z.; Cormack, L. K. (2004): Image and video quality assessment research at LIVE 2004. http://live.ece.utexas.edu/research/quality.

Temel, D.; Alregib, G. (2016): CSV: Image quality assessment based on color, structure, and visual system. *Signal Processing Image Communication*, vol. 48, pp. 92-103.

Wang, T.; Zhang, L.; Jia, H.; Li, B.; Shu, H. (2016): Multiscale contrast similarity deviation: an effective and efficient index for perceptual image quality assessment. *Signal Processing: Image Communication*, vol. 45, pp. 1-9.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simocelli, E. P. (2004): Image quality assessment: From error measurement to structural similarity. *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612.

Wang, Z.; Li, Q. (2011): Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185.

Wang, Z.; Simoncelli, E. P.; Bovik, A. C. (2004): Multiscale structural similarity for image quality assessment. *Conference on Signals, Systems & Computers*, vol. 2, pp. 1398-1402.

Xue, W.; Zhang, L.; Mou, X.; Bovik, A. C. (2014): Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Process*, vol. 23, no. 2, pp. 684-695.

Zhang, G.; Sun, H.; Ji, Z.; Sun, Q. (2016): Label propagation based on collaborative representation for face recognition. *Neurocomputing*, vol. 171, pp. 1193-1204.

Zhang, L.; Li, H. (2013): SR-SIM: A fast and high performance IQA index based on spectral residual. *IEEE International Conference on Image Processing*, pp. 1473-1476.

Zhang, L.; Shen, Y.; Li, H. (2014): VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270-4281.

Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. (2011): FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386.

Zheng, Y.; Jeon, B.; Sun, L.; Zhang, J.; Zhang, H. (2017): Student's t-hidden markov model for unsupervised learning using localized feature selection. *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1.