



## Color image quality assessment based on sparse representation and reconstruction residual<sup>☆</sup>



Leida Li <sup>a</sup>, Wenhua Xia <sup>a</sup>, Yuming Fang <sup>b</sup>, Ke Gu <sup>c</sup>, Jinjian Wu <sup>d</sup>, Weisi Lin <sup>c</sup>, Jiansheng Qian <sup>a,\*</sup>

<sup>a</sup> School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China

<sup>b</sup> School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China

<sup>c</sup> School of Computer Engineering, Nanyang Technological University, 639798, Singapore

<sup>d</sup> School of Electronic Engineering, Xidian University, Xi'an 710071, China

### ARTICLE INFO

#### Article history:

Received 25 December 2015

Revised 8 March 2016

Accepted 7 April 2016

Available online 8 April 2016

#### Keywords:

Image quality assessment

Color distortion

Sparse representation

Reconstruction residual

### ABSTRACT

Image quality assessment (IQA) is a fundamental problem in image processing. While in practice almost all images are represented in the color format, most of the current IQA metrics are designed in gray-scale domain. Color influences the perception of image quality, especially in the case where images are subject to color distortions. With this consideration, this paper presents a novel color image quality index based on Sparse Representation and Reconstruction Residual (SRRR). An overcomplete color dictionary is first trained using natural color images. Then both reference and distorted images are represented using the color dictionary, based on which two feature maps are constructed to measure structure and color distortions in a holistic manner. With the consideration that the feature maps are insensitive to image contrast change, the reconstruction residuals are computed and used as a complementary feature. Additionally, luminance similarity is also incorporated to produce the overall quality score for color images. Experiments on public databases demonstrate that the proposed method achieves promising performance in evaluating traditional distortions, and it outperforms the existing metrics when used for quality evaluation of color-distorted images.

© 2016 Elsevier Inc. All rights reserved.

### 1. Introduction

Image quality assessment (IQA) is fundamental in image processing. Modern IQA metrics build computational models to predict image quality in a perceptually consistent way [1–3]. The current IQA models can be classified into full-reference (FR) [4–11], reduced-reference (RR) [12,13] and no-reference (NR) [14–21], depending on the amount of reference information. FR-IQA is useful for benchmarking image processing algorithms, which is the focus of this work.

Extensive FR image quality metrics have been reported in the literature. One popular FR metric is the structural similarity (SSIM) index [4], which is designed based on the assumption that human eyes judge image quality mainly according to the structure degradations. By combining image structure, contrast and luminance, SSIM can predict image quality consistently with human perception. Following SSIM, improved versions have also been proposed

to improve the performance of visual quality estimations, such as the multi-scale structural similarity (MS-SSIM) [5] and information-content-weighted structural similarity (IW-SSIM) [6]. Sheikh et al. [7] proposed the visual information fidelity (VIF) index by measuring the mutual information shared between the reference and distorted images. Larson et al. [8] proposed the most apparent distortion (MAD) model, which adopts different strategies to measure distortions in low-quality and high-quality images. Liu et al. [9] proposed to use gradient similarity (GSM) to measure both structure and contrast distortions simultaneously in images. In [10], Zhang et al. measured image distortions by combining phase congruency and gradient, and proposed the feature similarity (FSIM) model. Zhang et al. [11] proposed the sparse structural similarity (SSS) index based on global sparse dictionary. Images are first decomposed into several layers with different visual importance. The image quality score is then computed by measuring the similarity of sparse coefficients of each layer and weighted by their relative importance. Notable success has been achieved by the current IQA metrics in revealing some aspects of the influence of different types of distortions on the perceived image quality. However, most of the current IQA metrics are designed in

<sup>☆</sup> This paper has been recommended for acceptance by Zicheng Liu.

\* Corresponding author.

E-mail address: [iqaqian@gmail.com](mailto:iqaqian@gmail.com) (J. Qian).

gray-scale domain. In practice, almost all images are represented in the color format. Color also influences the perception of image quality, especially in the case where images are subject to color distortions. Currently, there are only a few studies exploring the visual quality assessment for colorful images.

In [10], the authors extended the FSIM index to a color image quality metric by incorporating color similarity simply. Specifically, the original images are first converted into YIQ color space. Then the chrominance information conveyed by I and Q are directly used to evaluate the color similarity. Chang et al. [22] presented a color IQA metric based on independent component analysis (ICA). A feature detector was first trained using color images based on ICA. Image structure features were then extracted for visual quality prediction by using the feature detector. In [23], Kolaman et al. introduced the quaternion structural similarity (QSSIM) index to evaluate the quality of color images. Color distortion is measured based on the quaternion, which is efficient for color image processing [24]. These quality metrics for color images are mainly designed simply by extending existing IQA metrics from the gray-level domain to color domain. They rarely investigate the influence of color distortions specifically in color images. Thus, the effective IQA metric for color images is much desired in the research community.

In order to measure color distortions effectively in image quality assessment, we present a color image quality metric based on Sparse Representation and Reconstruction Residual (SRRR). The underlying idea is to measure both structure and color distortions in a holistic manner using an overcomplete color dictionary, which is trained from natural color images. Dictionary-based sparse representation is employed to design the quality model, because it has been proven that dictionaries trained using natural images consist of basis vectors, which behave similarly to the simple cells in the primary visual cortex [30,31]. Therefore, it is highly related to the human visual perception [11]. Furthermore, a color dictionary is adopted, which can effectively encode the inter-channel color correlations [28], thus facilitating the evaluation of color distortions in images. We believe these characteristics of sparse representation are needed in visual quality assessment. In the proposed method, the reference and distorted images are first represented using the color dictionary, producing the sparse coefficients. Then two feature maps are generated, and the similarity between them is employed to measure the structure and color distortions. Since the feature map features are not sensitive to contrast change in images, the reconstruction residuals are also computed and used as a complementary feature. Finally, luminance similarity is incorporated to generate the overall quality score. The experimental results on public image quality databases demonstrate that the proposed metric is effective in evaluating both color distortions and traditional distortions in images, and it outperforms the state-of-the-arts.

## 2. Proposed quality model

Fig. 1 shows the diagram of the proposed quality model. The proposed method consists of three major components, namely sparse feature similarity, reconstruction residual similarity and luminance similarity. The first two features are both extracted based on sparse representation. The sparse representation-based feature maps can effectively capture the structure and color distortions in a holistic manner, while the reconstruction residual maps can capture the contrast changes in images. Feature map-induced weighting is used to adapt to the characteristics of the Human Visual System (HVS). Finally, luminance similarity is incorporated to generate the overall quality score.

### 2.1. Sparse representation

Sparse representation has been proved promising in object recognition [25–27]. It is generally believed that sparse representation can capture the underlying structures in images, based on which slightly higher level features can be extracted. The principle of sparse representation is to represent a signal as a linear combination of basis vectors in an overcomplete dictionary. With such an overcomplete dictionary  $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^K \in \mathbb{R}^{n \times K}$  ( $n < K$ ), where  $n$  denotes the dimension of the basis vector and  $K$  is the number of basis vectors, a given signal  $\mathbf{y}$  can be represented as:

$$\mathbf{y} = \mathbf{D}\mathbf{x} = \sum_{i=1}^K x_i \mathbf{d}_i, \quad \text{s.t. } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon, \quad (1)$$

where  $\mathbf{x}$  is the representation vector,  $\mathbf{d}_i$  is the  $i$ th basis vector,  $\|\cdot\|_2$  is the  $\ell_2$  norm, and  $\epsilon$  is the representation error.

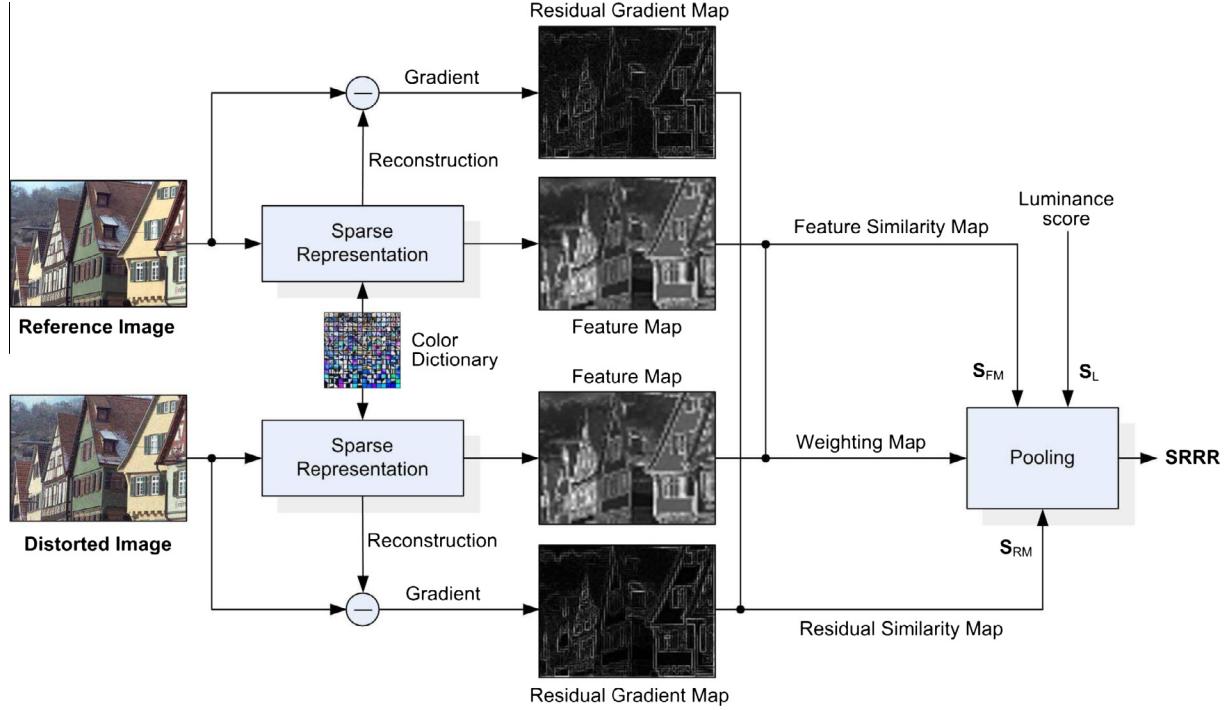
In the representation of a signal, we always hope to approximate it using as few basis vectors as possible. So the sparse representation of a signal  $\mathbf{y}$  can be achieved by:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \quad \text{s.t. } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon, \quad (2)$$

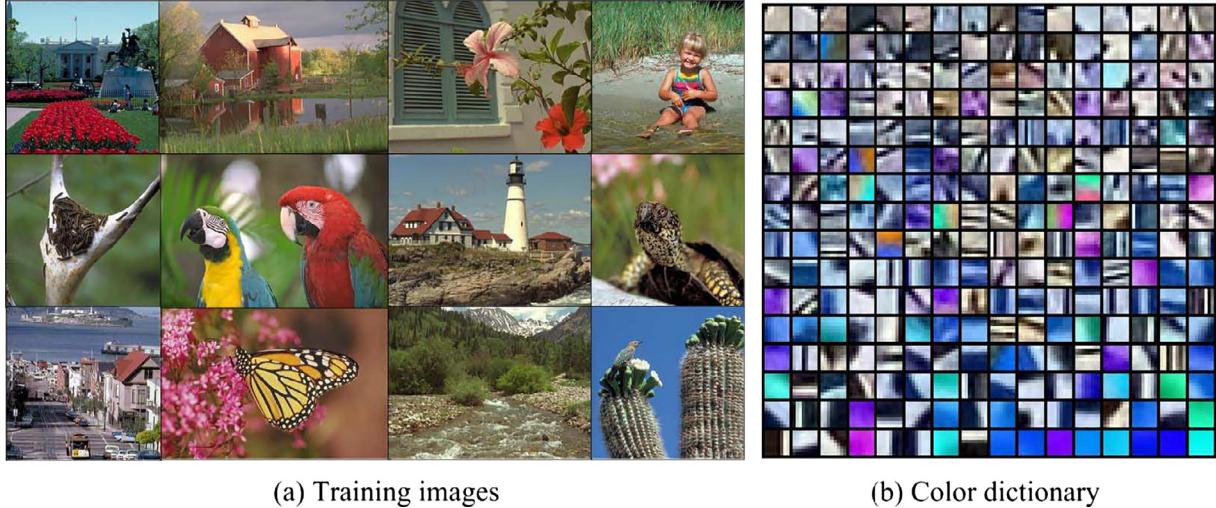
where  $\|\cdot\|_0$  denotes the  $\ell_0$  norm, which counts the number of non-zero elements in the representation vector  $\mathbf{x}$ .

In sparse representation, the overcomplete dictionary is usually trained using a large number of natural image patches. In this work, we aim to evaluate the quality of images in face of both color and traditional distortions, so a color dictionary is employed. It has been shown that with a color dictionary, the structure and color of an image can be represented in a holistic manner [28]. In this paper, we employ the K-SVD algorithm [29] to train the dictionary. Fig. 2 shows an example of dictionary learning using twelve natural images, together with the produced overcomplete color dictionary. In this example, 10,000 local patches of size  $8 \times 8 \times 3$  are first selected randomly from the training images. In order to highlight structures, the mean values of the patches are subtracted. Then each color patch is rearranged into a  $192 \times 1$  column vector and used for training. Fig. 2(b) shows the produced color dictionary of size  $192 \times 256$ . Note that each  $192 \times 1$  basis vector is rearranged into a  $8 \times 8 \times 3$  patch for visualization. Further details on the K-SVD dictionary learning algorithm can be found in [29]. It is worth mentioning that we utilize the RGB color space in this paper, because it is the most commonly used color space in digital cameras and computer monitors. Another reason for using RGB color space is that any color space conversion changes the structure of noise [28], which is one of the most important distortion types we are to evaluate.

It is observed from Fig. 2(b) that most of the basis vectors in the dictionary are edge patterns. These patterns have been shown similar to the receptive fields of neurons in the primary visual cortex [30]. Furthermore, the correlation between different color channels can be captured with these basis vectors [28], so that color distortions can be effectively represented based on the color dictionary. Images captured in natural scenes are sparse and redundant in nature, and sparse representation naturally mimics the characteristics of the human visual system by extracting sparse structures from images [11]. With these considerations, sparse representation is expected to produce better perceptual features for IQA, boosting metric performance. In this paper, the Orthogonal Matching Pursuit (OMP) algorithm is employed to obtain the sparse coefficients [32].



**Fig. 1.** Diagram of the proposed quality model.



**Fig. 2.** Training images and the color dictionary (Dict. I:  $192 \times 256$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 2.2. Sparse feature similarity

The first feature in the proposed method is generated based on sparse representation of image patches. Specifically, for an image in the RGB color space, it is first partitioned into non-overlapping local patches of size  $8 \times 8 \times 3$ . Then each patch is converted into a  $192 \times 1$  column vector. Then the overcomplete color dictionary is used to decompose the patches, producing the representation vectors, i.e., sparse coefficients. Given a reference image patch  $\mathbf{y}^r$  and the corresponding distorted image patch  $\mathbf{y}^d$ , they are first represented using the color dictionary  $\mathbf{D}$  as follows:

$$\mathbf{y}^r \approx \mathbf{D}\mathbf{x}^r, \quad (3)$$

$$\mathbf{y}^d \approx \mathbf{D}\mathbf{x}^d, \quad (4)$$

where  $\mathbf{x}^r$  and  $\mathbf{x}^d$  are the representation vectors of  $\mathbf{y}^r$  and  $\mathbf{y}^d$ , respectively. In this paper, the representation vectors of all patches in the reference and distorted images are denoted by  $\{\mathbf{x}_i^r\}_{i=1}^N$  and  $\{\mathbf{x}_i^d\}_{i=1}^N$ , where  $N$  is the total number of patches.

With the representation vectors, we propose to generate two feature maps. To be specific, the root inner product for a pair of patches are first computed as follows:

$$e_i^r = \sqrt{\langle \mathbf{x}_i^r \cdot \mathbf{x}_i^r \rangle}, \quad (5)$$

$$e_i^d = \sqrt{\langle \mathbf{x}_i^d \cdot \mathbf{x}_i^d \rangle}, \quad (6)$$

where  $\langle \cdot \rangle$  denotes the inner product. Based on  $\{e_i^r\}_{i=1}^N$  and  $\{e_i^d\}_{i=1}^N$ , two feature maps are generated, which are denoted by  $\mathbf{F}^r$  and  $\mathbf{F}^d$ ,

respectively. It should be noted that the size of the feature maps  $\mathbf{F}^r$  and  $\mathbf{F}^d$  is smaller than that of the reference and distorted images. Specifically, since we use  $8 \times 8$  patches in this work, the size of  $\mathbf{F}^r$  and  $\mathbf{F}^d$  is  $W/8 \times H/8$ , where  $W \times H$  is the size of the reference/distorted image. In this paper, we resize the feature maps to the same size of reference/distorted images for further processing.

**Fig. 3** shows some color images with different types of distortions, including Gaussian blur, JPEG compression, spatially correlated noise, change of color saturation and color quantization with dither. The corresponding feature maps of these images are shown **Fig. 4**. It is observed from the figure that the feature maps are sensitive to the distortions. This indicates that they can capture the structure and color changes among images with different distortion types.

The proposed sparse feature similarity is defined as follows:

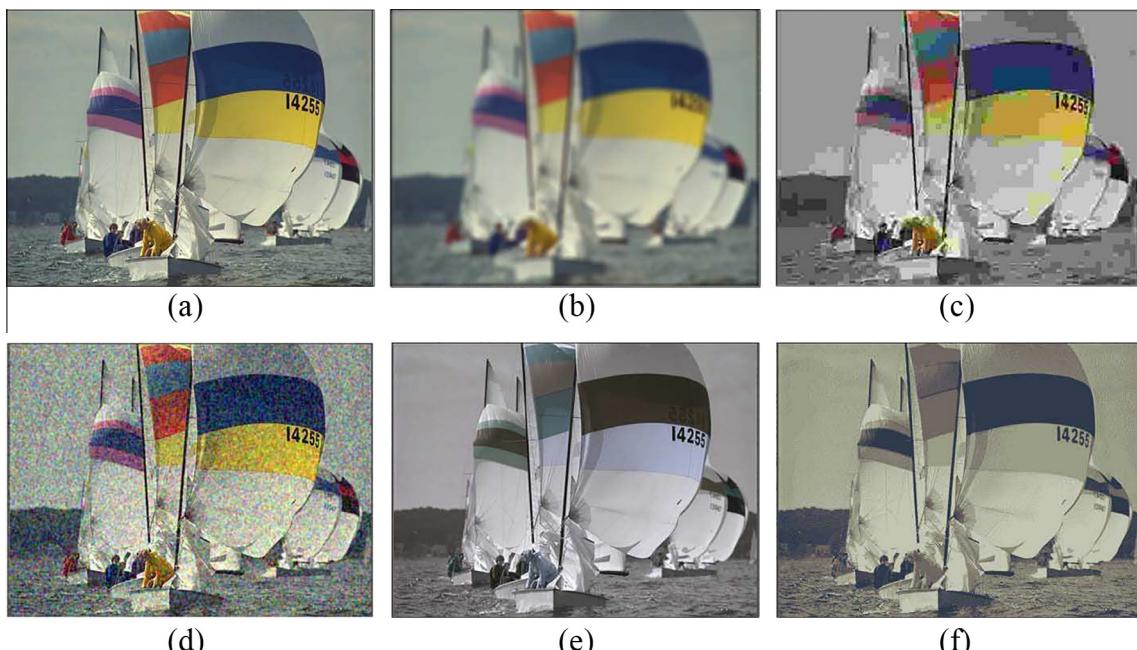
$$\mathbf{S}_{\text{FM}}(i,j) = \frac{2\mathbf{F}^r(i,j) \cdot \mathbf{F}^d(i,j) + c_1}{[\mathbf{F}^r(i,j)]^2 + [\mathbf{F}^d(i,j)]^2 + c_1}, \quad (7)$$

where  $\mathbf{S}_{\text{FM}}$  denotes the similarity map of the feature maps, and  $c_1$  is a constant used to ensure numerical stability.  $\mathbf{S}_{\text{FM}}$  can be used to measure the local structure and color similarities between the reference and distorted images, which is the first perceptual feature in the proposed method.

### 2.3. Reconstruction residual similarity

Sparse representation is characterized by the capacity to capture the underlying structure in images [26], which is beneficial for image quality assessment. However, we find that the sparse representation-based feature maps are not sensitive to contrast changes, which also have great impact on image quality. Therefore, a complementary feature is needed to measure the contrast changes in images. In this paper, we achieve this goal using the reconstruction residual.

In implementation, the representation vectors and the color dictionary are first used to reconstruct the original images:



**Fig. 3.** Example images with different types of distortions in the TID2013 database [33]. (a) Reference image, (b) Gaussian blur, (c) JPEG compression, (d) spatially correlated noise, (e) change of color saturation, (f) color quantization with dither. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\mathbf{R}_i^r = \mathbf{Dx}_i^r = \sum_{i=1}^K \mathbf{d}_i x_i^r, \quad (8)$$

$$\mathbf{R}_i^d = \mathbf{Dx}_i^d = \sum_{i=1}^K \mathbf{d}_i x_i^d, \quad (9)$$

where  $\mathbf{R}_i^r$  and  $\mathbf{R}_i^d$  denote the  $i$ th reconstructed patches of the reference and distorted images, respectively. Then the reconstruction residual images are obtained as follows:

$$\mathbf{RR}_i^r = |\mathbf{y}_i^r - \mathbf{R}_i^r|, \quad (10)$$

$$\mathbf{RR}_i^d = |\mathbf{y}_i^d - \mathbf{R}_i^d|, \quad (11)$$

where  $\mathbf{y}_i^r$  and  $\mathbf{y}_i^d$  denote original reference and distorted image patches.

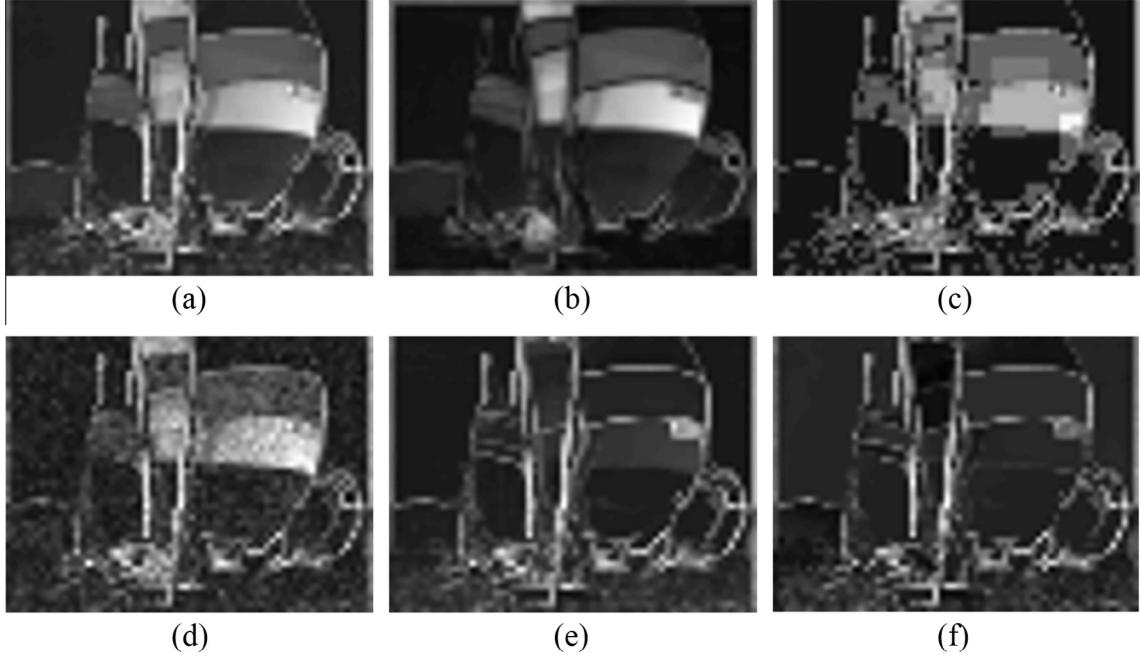
In order to highlight structure features in the residual images, their gradients are further computed using the Scharr gradient operator [10], which are denoted by  $\mathbf{G}^r$  and  $\mathbf{G}^d$ , respectively. **Fig. 5** shows a reference image and its two distorted versions with contrast changes. Their feature maps and residual gradient maps are simultaneously shown for comparison.

It is observed from **Fig. 5** that the three feature maps are quite similar, so the contrast distortions cannot be correctly represented. By contrast, the residual gradient maps can effectively capture the difference in contrast that cannot be captured by the feature maps. Based on this observation, the gradient maps of the reconstruction residual images are used as complementary features to capture the contrast changes in this paper.

The second feature of the proposed method, which we call reconstruction residual similarity, is defined as follows:

$$\mathbf{S}_{\text{RR}}(i,j) = \frac{2\mathbf{G}^r(i,j) \cdot \mathbf{G}^d(i,j) + c_2}{[\mathbf{G}^r(i,j)]^2 + [\mathbf{G}^d(i,j)]^2 + c_2}, \quad (12)$$

where  $c_2$  is also a constant.



**Fig. 4.** Feature maps of images shown in Fig. 3.

#### 2.4. Luminance similarity

Luminance has great impact on the perceived quality, so it is also considered in the proposed quality model. Similar to [22], luminance similarity is computed in a block manner. To be specific, for each patch in RGB color space, the mean values of the three respective color channels are first computed. Then their mean is further calculated and used to represent the luminance level of the patch. In implementation, the absolute luminance differences between the reference and distorted patches are denoted by  $\mathbf{m} = \{m_i\}_{i=1}^N$ , where  $m_i = |\mu(\mathbf{x}_i^r) - \mu(\mathbf{x}_i^d)|$ , and  $\mu(\cdot)$  is the mean operation. Then the patch pairs that have sufficiently large luminance differences are determined:

$$(\mathbf{m}^r, \mathbf{m}^d) = \{(\mu(\mathbf{x}_i^r), \mu(\mathbf{x}_i^d)) \mid |\mu(\mathbf{x}_i^r) - \mu(\mathbf{x}_i^d)| \geq \text{median}(\mathbf{m})\}, \quad (13)$$

where  $\text{median}(\cdot)$  denotes the median operation. Finally, the luminance similarity score is computed as follows [22]:

$$Q_L = \frac{\sum_{i=1}^T (m_i^r - \mu(\mathbf{m}^r)) \cdot (m_i^d - \mu(\mathbf{m}^d)) + c_3}{\sqrt{\sum_{i=1}^T (m_i^r - \mu(\mathbf{m}^r))^2 \cdot \sum_{i=1}^T (m_i^d - \mu(\mathbf{m}^d))^2} + c_3}, \quad (14)$$

where  $T$  denotes the number of block pairs used, and  $c_3$  is a small constant.

#### 2.5. Sparse feature-based pooling

The feature map similarity  $\mathbf{S}_{\text{FM}}$  and residual gradient map similarity  $\mathbf{S}_{\text{RR}}$  are two maps, which measure the local distortions between the reference and distorted images. In order to generate an overall quality score, a content-based pooling method is designed to adapt to the characteristics of the HVS. Similar to the FSIM index [10], a weighting map  $\mathbf{W}$  is first computed based on the two feature maps as follows:

$$\mathbf{W}(i,j) = \max\{\mathbf{F}^r(i,j), \mathbf{F}^d(i,j)\}, \quad (15)$$

where  $\max\{\cdot\}$  denotes the pixel-wise maximum. Then the weighting map is used to pool the sparse feature similarity  $\mathbf{S}_{\text{FM}}$  and reconstruction residual similarity  $\mathbf{S}_{\text{RR}}$ , producing two scores:

$$Q_{\text{FM}} = \frac{\sum_{i=1}^W \sum_{j=1}^H \mathbf{S}_{\text{FM}}(i,j) \cdot \mathbf{W}(i,j)}{\sum_{i=1}^W \sum_{j=1}^H \mathbf{W}(i,j)}, \quad (16)$$

$$Q_{\text{RR}} = \frac{\sum_{i=1}^W \sum_{j=1}^H \mathbf{S}_{\text{RR}}(i,j) \cdot \mathbf{W}(i,j)}{\sum_{i=1}^W \sum_{j=1}^H \mathbf{W}(i,j)}, \quad (17)$$

where  $W \times H$  denotes the resolution of the image. The final quality score  $Q$  is defined as a linear combination of  $Q_{\text{FM}}$ ,  $Q_{\text{RR}}$  and  $Q_L$ :

$$Q = a \cdot Q_{\text{FM}} + b \cdot Q_{\text{RR}} + c \cdot Q_L, \quad (18)$$

where  $a, b, c$  are parameters used for adjusting the relative importance of the three components, satisfying  $a + b + c = 1$ . In this work, they are experimentally set to  $a = 0.3, b = 0.45, c = 0.25$ .

### 3. Experimental results

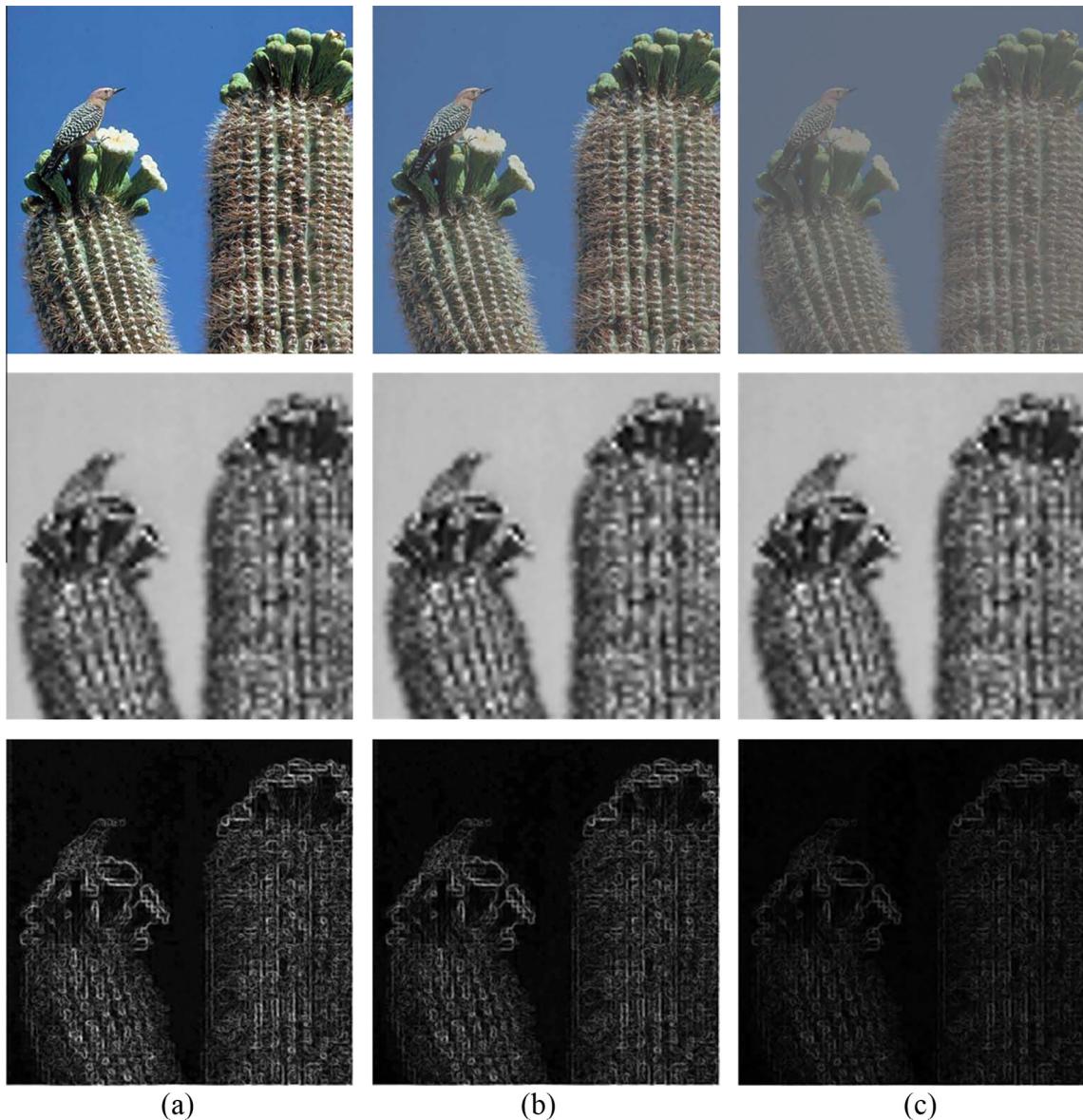
#### 3.1. Evaluation protocols

The performance of the proposed method is evaluated on four popular image quality databases, including TID2013 [33], TID2008 [34], CSIQ [8] and LIVE [4,35]. These databases contain both the reference and distorted images, together with the subjective evaluation scores. The subjective scores are measured by Mean Opinion Score (MOS) or Difference MOS (DMOS), which are used as ground truth. Table 1 summarizes the database information.

Three criteria are adopted for performance evaluation, including Pearson linear correlation coefficient (PLCC), root mean squared error (RMSE), and Spearman rank order correlation coefficient (SRCC). PLCC and RMSE are used to evaluate the prediction accuracy, while SRCC is used to measure the prediction monotonicity. A good quality metric is expected to produce high PLCC and SRCC values, and small RMSE value. In order to compute these performance values, a five-parameter logistic fitting is conducted between the subjective and objective scores:

$$f(x) = \tau_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\tau_2(x-\tau_3)}} \right) + \tau_4 x + \tau_5, \quad (19)$$

where  $\tau_i, i = 1, 2, \dots, 5$ , are the parameters to be fitted.



**Fig. 5.** Images with contrast changes, their feature maps and residual gradient maps. (a) Original image, (b) contrast change level 1, (c) contrast change level 2. From top to bottom: contrast-distorted images, feature maps and residual gradient maps.

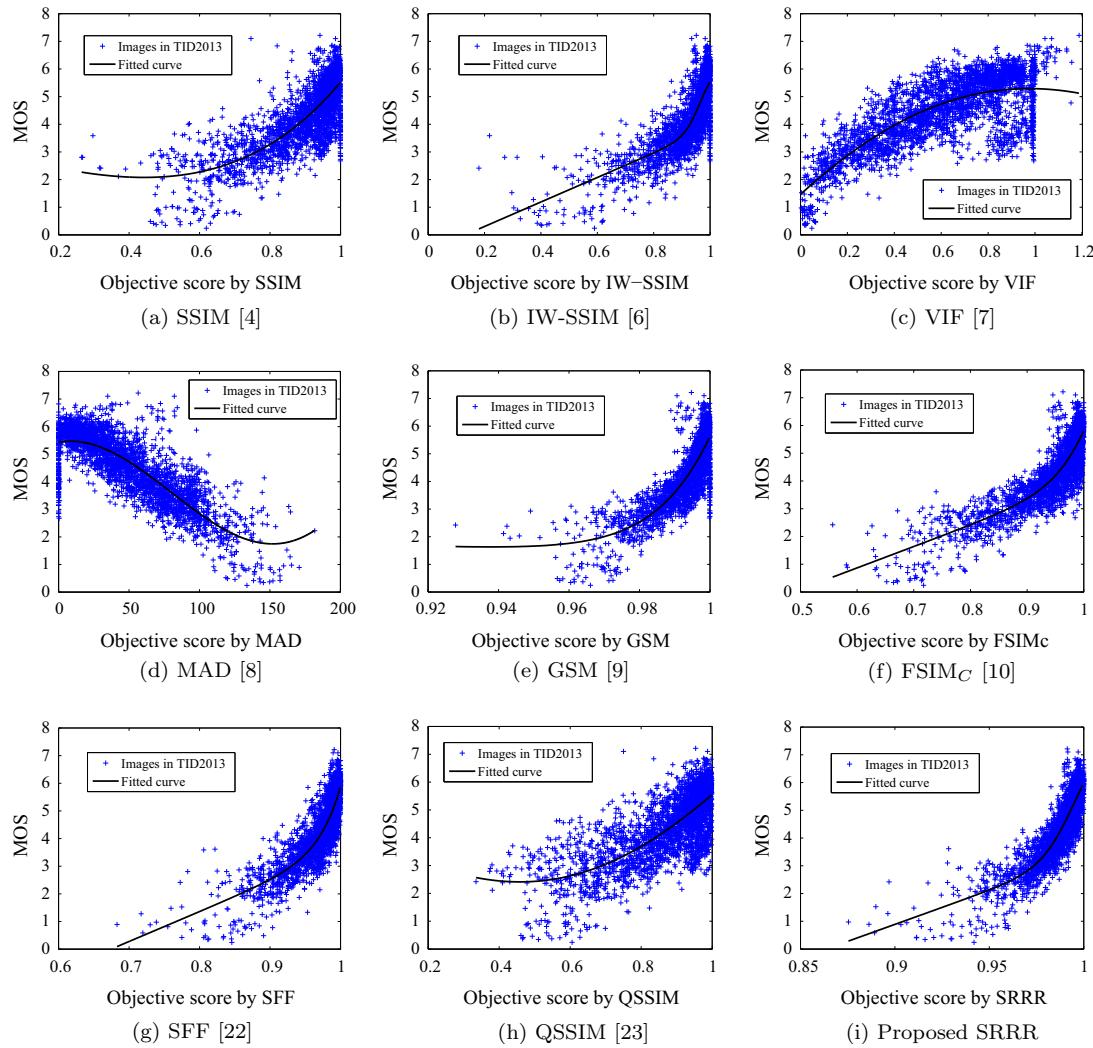
**Table 1**  
Databases for performance evaluation.

Database	Reference images	Distorted images	Distortion types	Subjects	Subjective score
TID2013	25	3000	24	971	MOS
TID2008	25	1700	17	838	MOS
CSIQ	30	866	6	35	DMOS
LIVE	29	779	5	161	DMOS

### *3.2. Performance evaluation on databases*

**Fig. 6** shows the fitting results of the proposed model and eight popular FR-IQA metrics on TID2013, which is the largest image quality database. It is observed from the figure that the proposed method produces very good fitting results, where the sample points are very close to the fitted curve. This indicates that the predicted scores of the proposed model are highly consistent with subjective scores. It is also observed that IW-SSIM [6], GSM [9],

FSIM<sub>C</sub> [10] and SFF [22] also produce very promising fitting results, which look very similar to that of the proposed metric. However, with these fitting curves we can only roughly know the performances. The reason is that there are so many (3000) images in TID2013 database, and a great number of the sample points are overlapping with each other in the fitting results (see the samples with high MOS values in Fig. 6). So it is hard to know the exact distributions of these sample points, which are closely related to the metric performances of both prediction accuracy and



**Fig. 6.** Scatter plots of subjective scores against predicted scores by different quality metrics on TID2013 database.

**Table 2**

Experimental results of SRRR and state-of-the-art metrics on four databases.

Database	Criterion	SSIM	IW-SSIM	VIF	MAD	GSM	FSIM <sub>C</sub>	SFF	QSSIM	SRRR
TID2013 (3000)	PLCC	0.790	0.832	0.772	0.827	0.846	0.877	0.871	0.746	<b>0.892</b>
	SRCC	0.742	0.778	0.668	0.808	0.795	0.851	0.851	0.716	<b>0.876</b>
	RMSE	0.761	0.688	0.788	0.698	0.660	0.596	0.610	0.826	<b>0.560</b>
TID2008 (1700)	PLCC	0.680	0.858	0.808	0.831	0.842	0.876	0.882	0.695	<b>0.883</b>
	SRCC	0.678	0.856	0.749	0.834	0.850	0.884	0.877	0.698	<b>0.889</b>
	RMSE	0.984	0.690	0.790	0.747	0.724	0.647	0.633	0.965	<b>0.631</b>
CSIQ (866)	PLCC	0.858	0.914	0.928	0.950	0.896	0.919	<b>0.964</b>	0.859	0.944
	SRCC	0.872	0.921	0.920	0.947	0.911	0.931	<b>0.963</b>	0.873	0.948
	RMSE	0.135	0.106	0.098	0.082	0.116	0.103	<b>0.070</b>	0.134	0.087
LIVE (779)	PLCC	0.921	0.952	0.960	<b>0.968</b>	0.951	0.961	0.963	0.928	0.958
	SRCC	0.923	0.957	0.964	<b>0.967</b>	0.956	0.965	0.965	0.931	0.959
	RMSE	10.63	8.347	7.614	<b>6.907</b>	8.433	7.530	7.346	10.171	7.824
Weighted Average	PLCC	0.786	0.865	0.826	0.862	0.865	0.893	0.898	0.770	<b>0.905</b>
	SRCC	0.765	0.840	0.760	0.854	0.845	0.885	0.887	0.759	<b>0.900</b>

monotonicity. As a result, the criterion values of PLCC, SRCC and RMSE should be further computed to evaluate the exact performances of the metrics.

**Table 2** summarizes the experimental results of the proposed metric and the eight popular metrics in terms of PLCC, SRCC and RMSE on the four databases. In order to know the overall performance of different metrics across the four databases, the database

size-weighted values of PLCC and SRCC are also calculated, where larger database are assigned bigger weight. In the table, the best results are marked in bold face.

It is observed from **Table 2** that in TID2013 and TID2008, the two largest databases with the most distortion types, the proposed method achieves the best performances in terms of both PLCC and SRCC. In CSIQ database, SFF [22] delivers the best performance. The

**Table 3**

Experimental results of SRRR and state-of-the-art metrics on 375 color distorted images of TID2013 database.

Criterion	SSIM	IW-SSIM	VIF	MAD	GSM	FSIM <sub>C</sub>	SFF	QSSIM	SRRR
PLCC	0.694	0.691	0.834	0.845	0.703	0.755	0.821	0.632	<b>0.871</b>
SRCC	0.235	0.234	0.810	0.828	0.242	0.629	0.782	0.419	<b>0.851</b>
RMSE	0.840	0.842	0.643	0.623	0.828	0.765	0.666	0.903	<b>0.573</b>

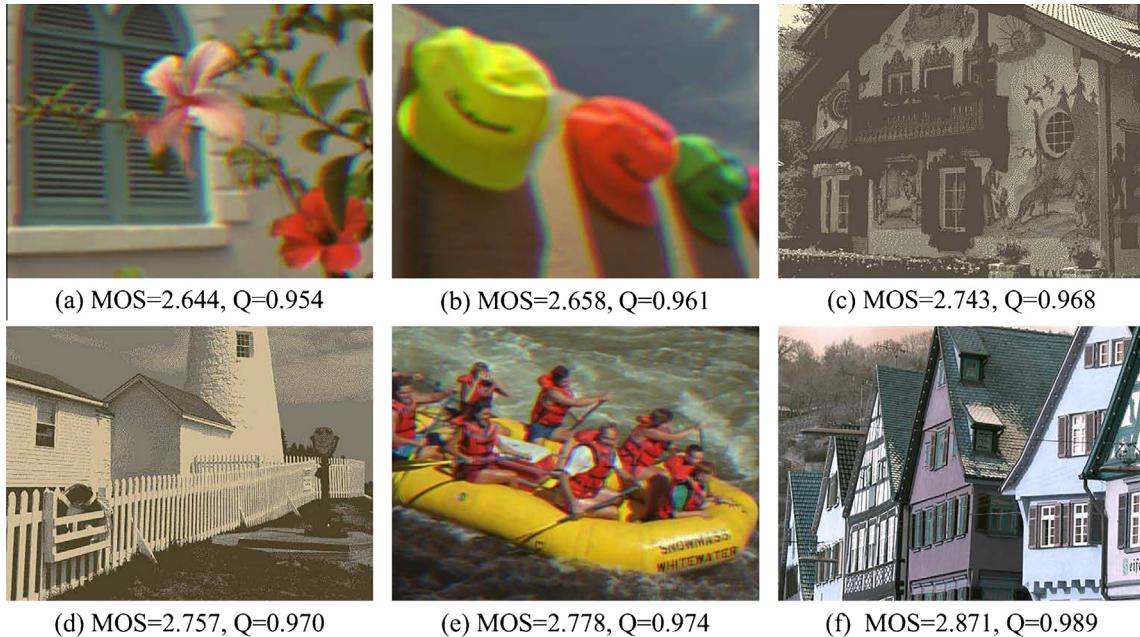


**Fig. 7.** An example image and its three color-distorted versions. (a) Reference image, (b) image color quantization with dither of level 1, (c) change of color saturation, (d) image color quantization with dither of level 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 4**

Quality scores of the distorted images shown in Fig. 7.

Image	MOS	SSIM	IW-SSIM	VIF	MAD	GSM	FSIM <sub>C</sub>	SFF	QSSIM	SRRR
Fig. 7(b)	5.351	0.975	0.986	0.760	35.092	0.998	0.987	0.994	0.962	<b>0.995</b>
Fig. 7(c)	4.684	0.999	0.999	0.992	0.000	0.999	0.998	0.996	0.994	<b>0.991</b>
Fig. 7(d)	2.895	0.852	0.892	0.380	106.501	0.990	0.888	0.941	0.828	<b>0.973</b>



**Fig. 8.** Six color-distorted images with similar subjective scores, and their objective scores predicted by the proposed SRRR metric. (a), (b) and (e) are distorted by chromatic aberrations; (c) and (d) are distorted by color quantization with dither; (f) is distorted by change of color saturation. All images are from TID2013 database [33]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

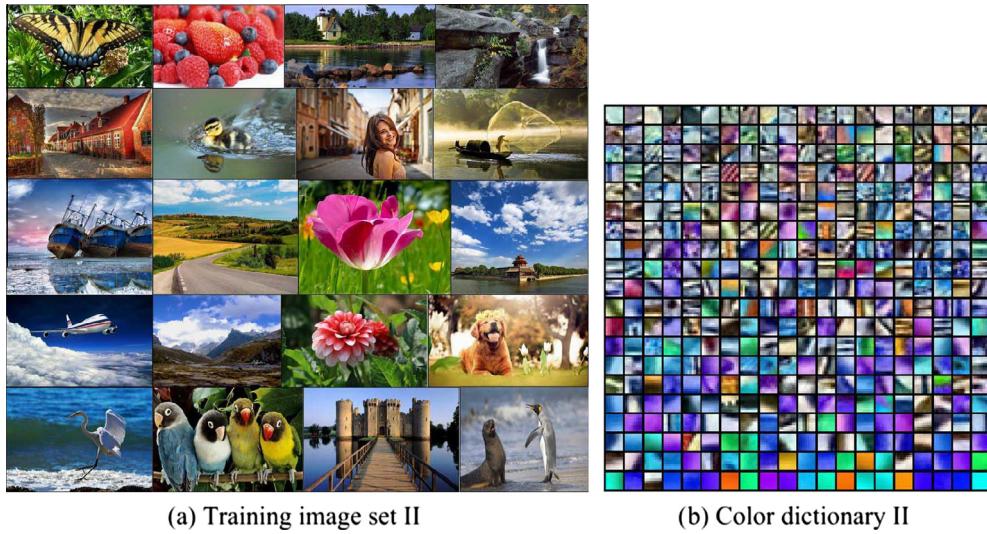
proposed method produces the second best SRCC and the third best PLCC, which are quite similar with those of MAD [8]. In LIVE, most of the quality metrics produce very good performances. The performance of the proposed method is only slightly worse than the best result. Finally, according to the weighted average values, the proposed method achieves the best overall performance across the four databases, followed by SFF [22] and FSIM<sub>C</sub> [10]. It should be noted that both SFF and FSIM<sub>C</sub> also take into consideration color

**Table 5**

Impact of sparsity degree on TID2013 database.

Sparsity	1	2	3	4	5	6	7	8
PLCC	<b>0.892</b>	0.886	0.877	0.869	0.861	0.856	0.850	0.846
SRCC	<b>0.876</b>	0.868	0.858	0.849	0.841	0.835	0.828	0.825

in their quality evaluation. This further demonstrates that color is needed for more effective image quality evaluation.



**Fig. 9.** Training image set II and the color dictionary (Dict. II:  $192 \times 400$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** Training image set III and the color dictionary (Dict. III:  $192 \times 512$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Performance evaluation on color-distorted images

The proposed method employs a color dictionary to represent the color degradations in images. The color dictionary can capture the multi-channel color distortions in a holistic manner. As a

result, the proposed method can be used to evaluate the quality of color-distorted images.

Among the four databases, only TID2013 contains images with color distortions. Specifically, there are three kinds of color distortions in TID2013, including change of color saturation, image color

**Table 6**

Performances of the proposed method on four databases when different dictionaries are used for sparse representation.

Dictionary	LIVE		CSIQ		TID2008		TID2013	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
Dict. I (192 × 256)	0.958	0.959	0.944	0.948	0.883	0.889	0.892	0.876
Dict. II (192 × 400)	0.957	0.958	0.943	0.946	0.881	0.887	0.892	0.876
Dict. III (192 × 512)	0.958	0.959	0.944	0.947	0.882	0.889	0.892	0.876

quantization with dither and chromatic aberration. For each distortion type, there are 125 images, so there are totally 375 color-distorted images. To evaluate the performance of the proposed method when used for quality evaluation of color-distorted images, we test the proposed method on these 375 images.

**Table 3** lists the experimental results of different metrics on the color-distorted images, where the best result is marked in boldface. It is obvious in the table that the proposed method achieves the best performance, and it significantly outperforms the existing metrics. This demonstrates that the proposed method is effective in evaluating the quality of color-distorted images.

**Fig. 7** shows a reference image and its three color-distorted versions. The quality of these images degrades sequentially from **Fig. 7 (b)–(d)**. **Table 4** summarizes the objective scores of the distorted images predicted by different quality metrics, where the result of the proposed method is marked in boldface. It is observed that with the decreasing MOS values, the objective scores predicted by the proposed method decrease accordingly, which indicates that the predicted scores are consistent with the subjective scores. By contrast, the compared metrics cannot correctly predict the subjective scores among the three distorted images.

**Fig. 8** further shows six color-distorted images with very similar MOS values, together with their objective scores predicted by the proposed metric. It is known from the figure that the predicted SRRR scores are also very similar, which indicates that they are consistent with the subjective scores. Furthermore, with the slight increase of the MOS values, the predicted scores also increase slightly. This indicates that the proposed method can differentiate tiny differences of color distortions.

#### 3.4. Impact of sparsity degree

Sparsity degree refers to the number of basis vectors used in sparse representation of a signal. In this part, we test the impact of sparsity degree on the performance of the proposed method. **Table 5** lists the experimental results in terms of PLCC and SRCC on TID2013 database, where the best result is marked in boldface.

It is observed from **Table 5** that the proposed method achieves the best performance when the sparsity degree equals one. With the increase of sparsity degree, the performance degrades gradually. It should be noted that this conclusion has been reached by experiments, and rigorous proof is hard to achieve. However, we may explain the potential reasons behind this result as follows. In the proposed model, two complementary features are adopted to measure the distortions in images, namely the sparse feature similarity and reconstruction residual similarity. The sparse features are employed to evaluate the underlying structure changes (usually involved in heavy low-frequency distortions) in images, while the corresponding reconstruction residuals are mainly used to capture contrast and high-frequency distortions. Therefore, setting the best sparsity degree is in fact a task to achieve an optimal trade-off between these two aspects. It has been demonstrated in Zhang et al.'s work [11] that the quality prediction performance is mainly determined by the first several basis vectors, which are most important to the perception of image quality. According to the results reported in [11], using the first basis vector (sparsity

equals one) alone contributes to more than 80% of the overall performance. Therefore, the first basis vector is most important in characterizing the structure degradations in images. This coincides with the results of the proposed model. In this paper, all experiments are conducted with sparsity degree 1. This is also beneficial in practice, because lower sparsity degree is more computationally efficient.

#### 3.5. Impact of dictionary

In the proposed method, a universal overcomplete color dictionary is used for sparse representation. Therefore, it is important to investigate the performance of the proposed metric when different dictionaries are used. To this end, we train another two overcomplete color dictionaries with bigger size (Dict. II of size 192 × 400 and Dict. III of size 192 × 512), and then test the performances of the proposed method using the new dictionaries. **Figs. 9 and 10** show the training images and the corresponding color dictionaries, which have been visualized for display. It should be noted that each of the two training image sets contains 20 natural color images, which are all collected from the Internet. Furthermore, the training images are completely different, which are also different from those used to train the aforementioned dictionary in **Fig. 2** (Dict. I of size 192 × 256).

**Table 6** summarizes the performances of the proposed method when different dictionaries are used in the experiments. From the simulation results, we have the following two observations. First, the proposed method is not dependent on training images. Although completely different images are used to train the dictionaries, the performances are quite similar. This is desirable in practical applications, meaning that we can randomly select natural images to train the dictionary and meantime achieve very satisfactory results. Second, dictionary size has very little impact on the performance of the proposed method. In our experiments, even the dictionary size is doubled from 192 × 256 to 192 × 512, the performances almost keep unchanged. With the consideration that a small-size dictionary is computationally more efficient, we use Dict. I of size 192 × 256 in this work.

## 4. Conclusion

We have presented a novel sparse representation-based color image quality assessment method in this paper. The proposed metric evaluates the quality of a color image by combining sparse representation-based feature maps and reconstruction residuals. The feature map is used to capture the underlying structure and color degradations, while the reconstruction residual is used to capture the contrast changes in images. By combining them with the luminance similarity, an overall quality score is generated to predict the visual quality of color images. We have evaluated the performance of the proposed method on public databases. The experimental results have demonstrated that the proposed method produces the state-of-the-art performance when evaluating traditional distortions and it is advantageous over the existing metrics when used for quality assessment of color-distorted images.

## Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities (2015XKMS032).

## References

- [1] W.S. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: a survey, *J. Vis. Commun. Image Represent.* 22 (4) (2011) 297–312.
- [2] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, *ISRN Signal Process.* (2013) 1–53 Article ID: 905685.
- [3] L. Ma, C.W. Deng, K.N. Ngan, W.S. Lin, Recent advances and challenges of visual signal quality assessment, *China Commun.* 20 (5) (2013) 62–78.
- [4] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [5] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: Proc. IEEE Asilomar Conference on Signals, Systems and Computers, 2003, pp. 1398–1402.
- [6] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, *IEEE Trans. Image Process.* 20 (5) (2011) 1185–1198.
- [7] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (2006) 430–444.
- [8] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, *J. Electron. Imag.* 19 (1) (2010) 1–21.
- [9] A.M. Liu, W.S. Lin, M. Narwaria, Image quality assessment based on gradient similarity, *IEEE Trans. Image Process.* 21 (4) (2012) 1500–1512.
- [10] L. Zhang, L. Zhang, X.Q. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, *IEEE Trans. Image Process.* 20 (8) (2011) 2378–2386.
- [11] X. Zhang, S.Q. Wang, K. Gu, T.T. Jiang, S.W. Ma, W. Gao, Sparse structural similarity for objective image quality assessment, in: Proc. IEEE Int. Conf. Systems, Man, and Cybernetics (SMC), 2015, pp. 1561–1566.
- [12] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, *IEEE Trans. Image Process.* 21 (8) (2012) 3378–3389.
- [13] J.J. Wu, W.S. Lin, G.M. Shi, A.M. Liu, Reduced-reference image quality assessment with visual information fidelity, *IEEE Trans. Multimedia* 15 (7) (2013) 1700–1705.
- [14] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1098–1105.
- [15] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (12) (2012) 4695–4708.
- [16] A. Mittal, R. Soundararajan, A.C. Bovik, Making a completely blind image quality analyzer, *IEEE Signal Process. Lett.* 20 (3) (2013) 209–212.
- [17] Y.M. Fang, K.D. Ma, Z. Wang, W.S. Lin, Z.J. Fang, G.T. Zhai, No-reference quality assessment of contrast-distorted images based on natural scene Statistics, *IEEE Signal Process. Lett.* 22 (7) (2015) 838–842.
- [18] L.D. Li, W.S. Lin, X.S. Wang, G.B. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, *IEEE Trans. Cybernet.* 46 (1) (2016) 39–50.
- [19] L.D. Li, Y. Zhou, J.J. Wu, W.S. Lin, H.L. Li, GridSAR: grid strength and regularity for robust evaluation of blocking artifacts in JPEG images, *J. Vis. Commun. Image Represent.* 30 (2015) 153–163.
- [20] C.W. Tang, X.K. Yang, G.T. Zhai, Image quality/distortion metric based on  $\alpha$ -stable model similarity in wavelet domain, *J. Vis. Commun. Image Represent.* 25 (7) (2014) 1746–1757.
- [21] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, *IEEE Trans. Image Process.* 24 (8) (2015) 2579–2591.
- [22] H.W. Chang, H. Yang, Y. Gan, M.H. Wang, Sparse feature fidelity for perceptual image quality assessment, *IEEE Trans. Image Process.* 22 (10) (2013) 4007–4018.
- [23] A. Kolaman, O. Yadid-Pecht, Quaternion structural similarity: a new quality index for color images, *IEEE Trans. Image Process.* 21 (4) (2012) 1526–1536.
- [24] B.J. Chen, H.Z. Shu, G. Coatrieux, G. Chen, X.M. Sun, J.L. Coatrieux, Color image analysis by quaternion-type moments, *J. Math. Imag. Vis.* 51 (1) (2015) 124–144.
- [25] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [26] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.* 15 (12) (2006) 3736–3745.
- [27] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, S.C. Yan, Sparse representation for computer vision and pattern recognition, *Proc. IEEE* 98 (6) (2010) 1031–1044.
- [28] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, *IEEE Trans. Image Process.* 17 (1) (2008) 53–69.
- [29] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.* 54 (11) (2006) 4311–4322.
- [30] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* 381 (13) (1996) 607–609.
- [31] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, *Vis. Res.* 37 (23) (1997) 3311–3325.
- [32] Y.C. Pati, R. Rezaifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers, 1993, pp. 40–44.
- [33] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo, Color image database TID2013: peculiarities and preliminary results, in: Proc. Eur. Workshop Vis. Inf. Process., 2013, pp. 106–111.
- [34] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008-a database for evaluation of full-reference visual quality assessment metrics, *Adv. Modern Radioelectron.* 10 (4) (2009) 30–45.
- [35] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Trans. Image Process.* 15 (11) (2006) 3440–3451.