

Sparse Feature Fidelity for Perceptual Image Quality Assessment

Hua-Wen Chang, Hua Yang, Yong Gan, and Ming-Hui Wang

Abstract—The prediction of an image quality metric (IQM) should be consistent with subjective human evaluation. As the human visual system (HVS) is critical to visual perception, modeling of the HVS is regarded as the most suitable way to achieve perceptual quality predictions. Sparse coding that is equivalent to independent component analysis (ICA) can provide a very good description of the receptive fields of simple cells in the primary visual cortex, which is the most important part of the HVS. With this inspiration, a quality metric called sparse feature fidelity (SFF) is proposed for full-reference image quality assessment (IQA) on the basis of transformation of images into sparse representations in the primary visual cortex. The proposed method is based on the sparse features that are acquired by a feature detector, which is trained on samples of natural images by an ICA algorithm. In addition, two strategies are designed to simulate the properties of the visual perception: 1) visual attention and 2) visual threshold. The computation of SFF has two stages: training and fidelity computation, in addition, the fidelity computation consists of two components: feature similarity and luminance correlation. The feature similarity measures the structure differences between the two images, whereas the luminance correlation evaluates brightness distortions. SFF also reflects the chromatic properties of the HVS, and it is very effective for color IQA. The experimental results on five image databases show that SFF has a better performance in matching subjective ratings compared with the leading IQMs.

Index Terms—Image quality assessment, full-reference, sparse coding, independent component analysis.

I. INTRODUCTION

IMAGE quality assessment (IQA) is essential to image applications and imaging services. The increasing demand for image processing technologies has pushed the need for

Manuscript received July 11, 2012; revised January 24, 2013 and May 8, 2013; accepted May 21, 2013. Date of publication June 6, 2013; date of current version August 30, 2013. This work was supported in part by the NSF in China under Grant 61071162. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hsueh-Ming Hang.

H.-W. Chang and Y. Gan are with the College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China (e-mail: changhuawen@gmail.com; ganyong@zzuli.edu.cn).

H. Yang is with the College of Computer Science, Sichuan University, Chengdu 610064, China, and also with the School of Computer Science, China West Normal University, Nanchong 637002, China (e-mail: hyang.yh@gmail.com).

M.-H. Wang is with the College of Computer Science, Sichuan University, Chengdu 610064, China (e-mail: wangminghui@scu.edu.cn).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The material contains the experimental results on five image databases. It includes a .mat file and a subdirectory containing scatter plots. The total size of this material is 3.5 MB. Contact changhuawen@gmail.com for further questions about this work.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2013.2266579

accurate image quality metrics (IQMs) to the forefront. Therefore, a great deal of interest and research has been devoted to the design and development of models for IQA. Over the past two decades, image and video quality assessment methods have been extensively studied, and many criteria have been proposed [1]–[6].

IQA methods fall into two categories: subjective assessment by humans and objective assessment by algorithms designed to mimic human observers. Subjective assessment can provide accurate results for any given evaluation, but it is a costly and time-consuming process. Therefore, there has been an increasing need to develop objective IQMs that can predict image quality automatically. According to the availability of a reference image, IQMs can be classified into three types: full reference (FR) metrics, reduced reference (RR) metrics, and no reference (NR) metrics [5]. This paper addresses the topic of FR IQA, where a “distortion-free” image is assumed to be known as the reference image.

As images will ultimately be viewed by human beings, it is desirable to have IQMs that can predict the perceived visual quality as measured with human subjects. Perceptual quality metrics aim to emulate the integral mechanisms of the human visual system (HVS) to correlate well with visual perception of quality. To this end, many researchers have developed sophisticated IQA models to achieve perceptual consistency in quality prediction by modeling physiological response properties of the HVS. Among all these methods, structural similarity (SSIM) index [5] is quite attractive owing to its simplicity and excellent performance relative to old methods such as the peak signal-to-noise ratio (PSNR). It is based on the hypothesis that the HVS is highly adapted for extracting structural information in images. There are also several similar structural information based IQMs, including universal image quality index (UQI) [7], complex wavelet structural similarity (CW-SSIM) index [8], content-partitioned structural similarity index [9], and information content weighted SSIM (IW-SSIM) [10]. Besides the structural approaches, there are some interesting IQMs based on other properties of the HVS. The research in [11] presented a two stage metric called visual signal-to-noise ratio (VSNR). It first determines whether the distortions are visible by a visual threshold, and then quantifies the visible distortions that are beyond the threshold in the second stage. A recent research in [12] reconfirmed the importance of divisive normalization masking models for IQA. Visual information fidelity (VIF) [13] is an alternative approach to IQA, which is based on natural scene statistics. It is also a leading IQM in terms of correlation with human perception [14].

A recent study suggested that gradient information is crucial for visual perception and can be used for IQA [15].

An ideal image quality method should be able to simulate the properties of visual perception [16], [17]. As an important part of the HVS, the visual cortex is responsible for most of our conscious perception of the visual world [18], [19], so an ideal visual model for perceptual quality assessment should be closely related to the neural response properties of the visual cortex. Sparse coding [20], [21], which is equivalent to independent component analysis (ICA) [22]–[25], can offer exact quantitative predictions that often turn out to be in line with the measurements from the visual cortex [21]. Moreover, natural scenes contain sparse structures, while sparse coding just provides a strategy for extracting these intrinsic structures in images [26]–[28]. In sparse coding, a given image may usually be represented using a set of basis vectors so that only a small number of basis vectors are activated and the rest are activated weakly or not at all at the same time [20]. Similarly, Bell and Sejnowski [23] obtained similar results by a technique called ICA. In fact, sparse coding and ICA actually use quite similar approaches, and their respective algorithms can be related mathematically [23], [29]. Therefore, we can make use of ICA models to accomplish the sparse coding process.

Inspired by above facts, we propose to use ICA models to simulate the visual processing at the level of visual cortex. Then, for each input image (patch), the output information of this model is named as sparse feature. The sparse feature can be seemed as the response of neurons in visual cortex which is closely associated with visual perception [29]. Meanwhile, the difference between the responses of the reference and distorted images can reflect the perceived difference in quality by humans. So measuring the feature fidelity to the reference image can be an effective way for evaluating the quality of a given image. Our previous research in [30], [31] has achieved some success in this way. In [30], ICA is used for training a feature detector on samples of each reference image. Because the samples come from each corresponding reference image, the training process should be involved in every evaluation, which takes a lot of time. However, our further research found that different source images for the ICA training will lead to similar results of quality assessment. This indicates that the difference of training samples has only a little effect on the quality evaluation, as confirmed in the experiment of Sec. V H. Therefore, the feature detector can be trained only once, and then be used for evaluating all the test images.

In this paper, a novel IQM called sparse feature fidelity (SFF) is proposed for FR IQA. The SFF relates the quality of an image with the fidelity to the reference image in the form of sparse features that are extracted by a feature detector which is trained on samples of natural images by ICA. After dividing a test image into image patches, the mean value of each patch will be subtracted. Then all the patch vectors with zero mean will be used for the computation of feature similarity, while all the mean values will be used to compute luminance correlation. The component of feature similarity measures the structure difference between two images, while the luminance correlation evaluates brightness distortions. Finally, the SFF

index is obtained by combining the two components into a quality score. Unlike most of the other IQA methods, SFF is based on all the RGB color components, and it can detect the color distortion in a perceptual way. Experimental results show that SFF index has high correlation with subjective quality evaluation.

Unlike our previous work, the feature detector of this paper is a universal one, once it has been obtained, it can be used for evaluating the quality of images without the time-consuming training process. Moreover, the proposed fidelity metric considers not only structure distortions but also luminance changes, which makes this metric more accurate and applicable to a wider range of distortions. Besides that, to improve the efficiency and accuracy, we only use the reference-distorted patch pairs with low quality for the computation of feature similarity and luminance correlation by a selection strategy which can be explained as a simple model of visual attention mechanism. In addition, a visual threshold is designed based on the features of the corresponding reference image to enhance the accuracy of quality evaluation.

The rest of this paper is organized as follows. Section II provides some background work about the HVS and methods for modeling of the HVS. Section III explains the sparse feature and feature detector. Section IV presents the computation procedure of the SFF index. Section V first describes the experimental method for validation of IQMs, and then shows results and comparisons. Finally, we conclude this paper in Section VI.

II. SIMULATION OF THE HVS

Images will fall upon the retina when we are viewing the natural world. Then the lateral geniculate nucleus (LGN) transmits visual signals from the retina to cortex and controls how much of the information is allowed to pass [32]. From the LGN, the signals are sent to the most important part, the primary visual cortex (V1), where most of the visual processing is performed.

A number of studies have proposed that the retina and the LGN are dedicated to whitening the input signals [18], [33], [34]. Moreover, the principle of redundancy reduction is successful in accounting for response properties of neurons in the retina and LGN [35]. Therefore, we use whitening and dimension reduction to simulate the functions of the retina and LGN. The act of whitening is a special processing, which is to attenuate the low frequencies and boost the high frequencies to yield a roughly flat power spectrum across all spatial frequencies [18].

The spatial receptive fields of simple cells in the cortex can be characterized as being localized, oriented, and band-pass [36]. In [20], Olshausen and Field proposed that a coding strategy that maximizes sparseness is sufficient to account for those three response properties of cortical simple cells. In fact, similar results can be obtained when ICA is applied to natural images [23]. Therefore, we use ICA to simulate the visual processing of cortical simple cells in the visual cortex. Another advantage of using ICA models is that ICA can generate a complete dictionary for image representations, which means we don't have to find the sparsest solutions for images.

III. SPARSE FEATURE AND FEATURE DETECTOR

Let's start with the basic assumption that an image (patch) vector, \mathbf{x} , can be represented by a linear superposition of some basis vectors in \mathbf{F} : $\mathbf{x} = \mathbf{F}\mathbf{s}$. The coefficient vector, \mathbf{s} , is a random vector that is different from image to image. Inverting this linear system, we obtain $\mathbf{s} = \mathbf{W}\mathbf{x}$ with \mathbf{W} being the (pseudo) inverse of \mathbf{F} . Generally, the goal of sparse coding is to find a weighting matrix (\mathbf{W}) that can transform every image vector, \mathbf{x} , into a sparse coefficient vector, \mathbf{s} , by maximizing the sparseness or super-gaussianity of \mathbf{s} [20].

In neurophysiological modeling, the row elements of \mathbf{W} are related to the receptive fields of neurons in the primary visual cortex [20]. Each weighting vector of \mathbf{W} can be interpreted as a model of the simple cell receptive field, while the coefficient vector, \mathbf{s} , typically represents simple cell responses (outputs). Moreover, \mathbf{W} can also be considered as a feature detector [29], since the values of the coefficient vector have a sparse distribution [20], \mathbf{s} is named as sparse feature in this paper.

IV. SPARSE FEATURE FIDELITY

The calculation process of the proposed method is divided into two stages: training and fidelity computation. Firstly, the feature detector, \mathbf{W} , is received from a training stage. Then \mathbf{W} is used to extract sparse features from reference and distorted image patches. Finally, SFF quality index is calculated on the basis of the sparse features in the second stage.

A. Training of the Feature Detector

Firstly, thousands of samples (image patches) are randomly taken from nine natural images with no distortion. Then these sample vectors are whitened by principal component analysis (PCA). Finally, \mathbf{W} will be trained on the whitened data, \mathbf{Z} , by a sparse coding algorithm. In our study, we choose FastICA [37] to implement this training process.

1) *Preprocessing*: At the beginning of the training process, 18000 image patches of size 8×8 are randomly taken from nine natural images with no distortion. In practical calculations, each patch should be vectorized into a column vector by scanning the numerical values in the patch row-by-row and channel-by-channel. Since a color image has three channels, the length of the vector is $8 \times 8 \times 3 = 192$. The statistical analysis of the vectors is not influenced by the choice of this transformation [29]. Then each vector is centred by subtracting the mean pixel value of each patch. Thus, all the sample vectors form a matrix \mathbf{X} .

2) *Whitening and Dimension Reduction*: This step serves as a model of the LGN because it could explain the center-surround structure of the receptive fields of ganglion cells in the retina, as well as those in the LGN [33]. We use PCA to reduce the dimension of the sample vectors so that the maximum amount of information is preserved and the redundant information is discarded. Then PCA can also be used for whitening of the sample data.

PCA can be done by eigenvalue decomposition of a data covariance matrix. The covariance matrix of the sample data

can be obtained by

$$\mathbf{U} = \frac{1}{S} (\mathbf{X} \times \mathbf{X}^T) \quad (1)$$

where S denotes the amount of samples. After eigenvalue decomposition, let $\mathbf{D} = \text{diag}(d_1, \dots, d_M)$ and $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_M)$ respectively denote the M largest eigenvalues and corresponding eigenvectors for the covariance matrix, \mathbf{U} . In our study, only the first 8 principal components of the samples are retained for training. That means the dimension of each whitened vector will be reduced from 192 to $M = 8$. Then the whitening matrix, \mathbf{V} , is given by

$$\mathbf{V} = \mathbf{D}^{-1/2} \times \mathbf{E}^T \quad (2)$$

where $\mathbf{D}^{-1/2} = \text{diag}(1/\sqrt{d_1}, \dots, 1/\sqrt{d_M})$. Finally, sample data, \mathbf{X} , can be whitened into \mathbf{Z} by the following multiplication

$$\mathbf{Z} = \mathbf{V} \times \mathbf{X} \quad (3)$$

3) *Training by FastICA*: The feature detector is learned from the whitened data, \mathbf{Z} , by a sparse coding algorithm. In fact, maximization of sparseness is very closely related to ICA since sparseness is a form of super-Gaussianity or kurtosis [25]. So a fixed-point ICA algorithm called FastICA [37] is used to generate the feature detector. For whitened data \mathbf{Z} , finding one maximally non-Gaussian direction by FastICA algorithm has the following basic form of iteration:

$$\begin{aligned} \mathbf{w}_{(p)}^w = E \left\{ \mathbf{Z} g \left(\left(\mathbf{w}_{(p-1)}^w \right)^T \mathbf{Z} \right) \right\} \\ - E \left\{ g' \left(\left(\mathbf{w}_{(p-1)}^w \right)^T \mathbf{Z} \right) \right\} \mathbf{w}_{(p-1)}^w \end{aligned} \quad (4)$$

where \mathbf{w}^w denotes the vector of the feature detector in the whitened space, p denotes the iteration step, $E(\bullet)$ is the expectation operator, $g'(\bullet)$ denotes the derivative of the function, $g(\bullet)$, and $G(\bullet)$ is the derivative of the function, $G(\bullet)$. Here we chose the following functions:

$$G(u) = \frac{1}{\eta} \log \cosh(\eta u) \quad (5)$$

$$g(u) = \tanh(\eta u) \quad (6)$$

where $1 \leq \eta \leq 2$, while in this paper, η is set to $\eta = 1$.

Starting from a random orthogonal matrix, each iteration of the algorithm consists of updating each row, $(\mathbf{w}^w)^T$, of \mathbf{W}^w by (4), followed by orthogonalization of the matrix through:

$$\mathbf{W}^w \leftarrow \left(\mathbf{W}^w \mathbf{W}^{wT} \right)^{-1/2} \mathbf{W}^w \quad (7)$$

The same procedure will repeat until it reaches the convergence. The convergence criterion is as follow.

$$\left\| \left\| \mathbf{W}_{(p)}^w \times \left(\mathbf{W}_{(p-1)}^w \right)^T \right\| - \mathbf{I} \right\|_F < 10^{-8} \cdot M \quad (8)$$

where $\|\bullet\|_F$ denotes the Frobenius norm, \mathbf{I} is an identity matrix, M is the number of independent components.

After the learning process, the feature detector should be transformed from the whitened space to the original space by

$$\mathbf{W} = \mathbf{W}^w \times \mathbf{V} \quad (9)$$

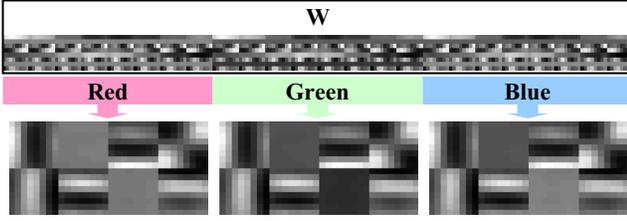


Fig. 1. An illustration for the feature detector, \mathbf{W} . \mathbf{W} has three sections, corresponding to three color channels. Each section forms 8 blocks that are similar to the spatial receptive fields.

Fig. 1 shows the retained feature detector, \mathbf{W} , where all the entries of this matrix are quantified into gray scale values. The feature detector is in the form of an 8×192 matrix which can transform an image vector of length 192 into a sparse feature vector of length 8. Each row element of \mathbf{W} consists of three sections that can form three 8×8 patches corresponding to the RGB channels. Thus all the eight row elements form three sets of 8×8 patches, each set has 8 patches. These patches are similar to receptive fields with different directions. So this feature detector serves as a model of neurons in visual cortex which transforms image patches into visual responses.

B. Procedure of Fidelity Computation

As shown in Fig. 2, the fidelity computation of SFF index consists of two components: feature similarity and luminance correlation. Before the computation, reference and distorted images should be first divided into non-overlapping patches of size 8×8 by a sliding window that scans sequentially on a regular grid. Since the images are divided in the same way, each reference patch and its distorted counterpart form a patch pair which will then be considered as a pair of vectors. Then in the computation of feature similarity, the mean value of each image patch should be removed because it does not contain any interesting information about contrast and structure. However, luminance changes can also cause perceptible distortions although they are not as annoying as contrast or structure changes [15]. Therefore, the mean values will then be employed to evaluate luminance changes in the computation of luminance correlation. Finally, the SFF quality index will be obtained by combining the feature similarity and luminance correlation together.

C. Feature Similarity

By removing the mean value of each patch, all the pixel values in an image patch form a column vector with zero mean. Then all the column vectors from each reference and the corresponding distorted image form two matrices, \mathbf{X}^{ref} and \mathbf{X}^{dis} . Finally, the similarity computation will be completed by the following 4 steps as shown in Fig. 2.

1) *Step 1: Selection of reference-distorted patch pairs.* Supra-threshold distortions can be a strong attractor of visual attention, and as a result, have a severe impact on the perceived quality [38]. This is why the HVS is more sensitive to poor quality regions in images than the good ones. Because the regions with low quality are critical to quality evaluation, a

reasonable approach to improve the prediction performance is to utilize the image patches with large differences [39].

Some researchers proposed several weighting strategies that put more weight on the poor quality regions [39]. But weighting strategies will increase the complexity of a metric. In order to keep the proposed metric simple, we propose to only use the reference-distorted patch pairs with large differences for the similarity computation. In this research, the difference between a pair of vectors is measured by the mean absolute error (MAE). Let \mathbf{x}^{ref} and \mathbf{x}^{dis} denote a reference and the corresponding distorted patch vectors, respectively, then the MAE between \mathbf{x}^{ref} and \mathbf{x}^{dis} is defined as

$$MAE(\mathbf{x}^{ref}, \mathbf{x}^{dis}) = \frac{1}{n} \sum_{i=1}^n |x_i^{ref} - x_i^{dis}| \quad (10)$$

where n denotes the number of pixels in each image vector. Finally, all the MAE values between the reference and distorted patches form a vector which is denoted by \mathbf{d} , the element of which is $d_i = MAE(\mathbf{x}_i^{ref}, \mathbf{x}_i^{dis})$.

In order to select a set of useful patch pairs for quality assessment, a threshold TH_x is designed on the basis of the median value of \mathbf{d} , and then it will be used to select reference-distorted patch pairs. If the MAE value of a patch pair is not less than the value of TH_x , this patch pair will be selected, then let \mathbf{y}^{ref} and \mathbf{y}^{dis} denote the retained reference and distorted image vectors, respectively. Finally, all the retained vectors form two matrices, \mathbf{Y}^{ref} and \mathbf{Y}^{dis} , as expressed in (11).

$$(\mathbf{Y}^{ref}, \mathbf{Y}^{dis}) = \{(\mathbf{y}_i^{ref}, \mathbf{y}_i^{dis}) \mid MAE(\mathbf{y}_i^{ref}, \mathbf{y}_i^{dis}) \geq TH_x\} \quad (11)$$

$$TH_x = T_x \cdot median(\mathbf{d}) \quad (12)$$

where $median(\bullet)$ represents the median value of a vector, T_x is an adjustment for TH_x , its default value is 1.

2) *Step 2: Feature Extraction.* After the selection step, the sparse feature vectors, \mathbf{a}_i and \mathbf{b}_i , can be extracted by a multiplication operation.

$$\mathbf{a}_i = \mathbf{W} \times \mathbf{y}_i^{ref}, \quad \mathbf{b}_i = \mathbf{W} \times \mathbf{y}_i^{dis} \quad (13)$$

Since the size of \mathbf{W} is 8×192 , the length of \mathbf{a}_i and \mathbf{b}_i is $M = 8$. For simplicity, we use a vector pair, $(\mathbf{a}_i, \mathbf{b}_i)$, to represent the features of a reference image patch together with its distorted counterpart. Moreover, all the feature vectors of \mathbf{Y}^{ref} and \mathbf{Y}^{dis} form two matrices, \mathbf{A} and \mathbf{B} , respectively.

$$(\mathbf{A}, \mathbf{B}) = \{(\mathbf{a}_i, \mathbf{b}_i) \mid i = 1, \dots, N\} \quad (14)$$

where N is the number of the selected patches in an image, \mathbf{a}_i and \mathbf{b}_i denote the column vector in \mathbf{A} and \mathbf{B} , respectively.

3) *Step 3: Thresholding.* As the feature detector serves as a model of neurons in the visual cortex, the sparse feature vectors can be seemed as visual responses to image patches. Different regions in an image will lead to different responses since they have different visual importance to the whole image [39]. Certain patches in an image may be visually more important than others, and these patches are essential to perceived quality of an image. In order to remove the perceptually unimportant signals, a visual threshold for each test image is computed based on the feature vectors that are

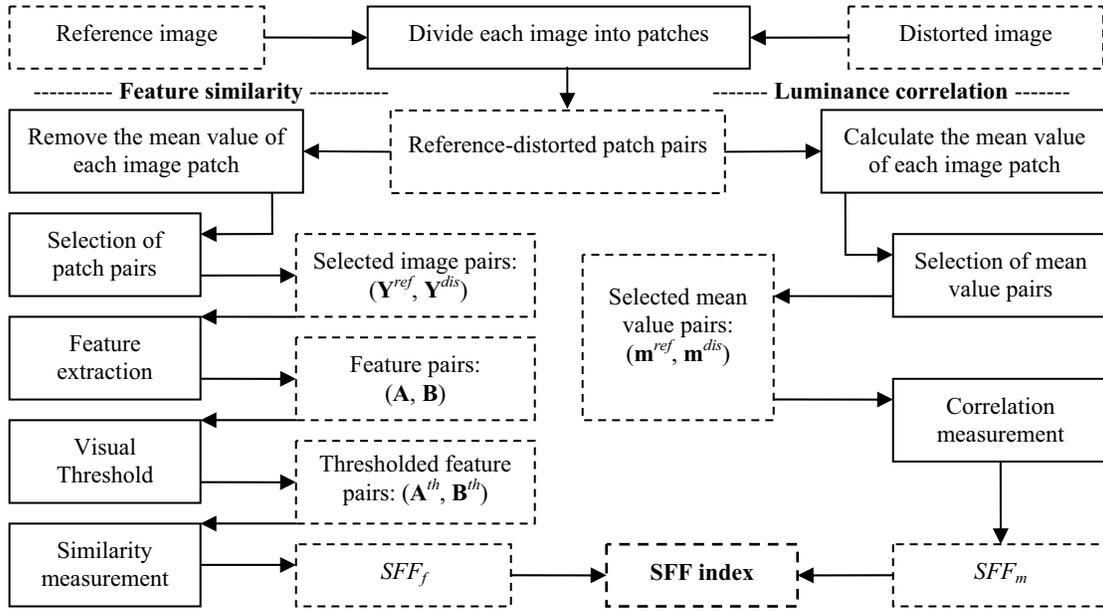


Fig. 2. Block diagram for the fidelity computation of SFF. SFF has two components, SFF_f is the result of feature similarity, and SFF_m is the result of luminance correlation.

extracted from the reference image. The threshold value is given by

$$VT = \frac{T_v}{N} \sum_{i=1}^N VR(\mathbf{a}_i) \quad (15)$$

where T_v is a parameter for adjustment of the threshold, according to our experiment (as shown in Sec. V G), its proper value is 0.4. In addition, VR represents the intensity of the visual response to an image patch, which is obtained by

$$VR(\mathbf{a}) = \sum_{j=1}^M a_j^2 \quad (16)$$

where a_j is the j -th value of \mathbf{a} .

For each reference feature vector \mathbf{a}_i , if $VR(\mathbf{a}_i) > VT$, then \mathbf{a}_i and its corresponding distorted vector, \mathbf{b}_i , will be used to compute the fidelity index. This means the patches that have strong visual responses will be used for fidelity computation. After this thresholding step, all the retained sparse feature (or response) vectors for each reference and distorted image form two matrices, \mathbf{A}^{th} and \mathbf{B}^{th} .

$$\left(\mathbf{A}^{th}, \mathbf{B}^{th}\right) = \{(\mathbf{a}_k, \mathbf{b}_k) | VR(\mathbf{a}_k) > VT, k \in \{1, \dots, N\}\} \quad (17)$$

where $(\mathbf{a}_k, \mathbf{b}_k)$ denotes the feature pairs that above the threshold value.

4) *Step 4: Similarity Measurement.* Finally, the feature similarity component of SFF, SFF_f , is given by:

$$SFF_f = \frac{1}{K \cdot M} \sum_{i=1}^K \sum_{j=1}^M \frac{2A_{ij}^{th} B_{ij}^{th} + C}{\left(A_{ij}^{th}\right)^2 + \left(B_{ij}^{th}\right)^2 + C} \quad (18)$$

where K denotes the number of the retained feature vectors in an image, A_{ij}^{th} and B_{ij}^{th} denote the values of the i -th column and j -th row in \mathbf{A}^{th} and \mathbf{B}^{th} , respectively. C is a small positive constant to stabilize the result, and its proper value is 0.08.

D. Luminance Correlation

The computation of luminance correlation is based on the mean value of every image patch. The mean values of a reference-distorted patch pair form a mean value pair. It is not necessary to use all the mean value pairs for measuring the luminance distortion, since some pairs with small differences will not affect or even moderate the perception of the regions with large luminance changes. So we only consider the mean value pairs with large differences, and then calculate the correlation between two sets of selected mean values. The computation consists of the following two steps.

1) *Step 1: Selection of mean value pairs.* The luminance difference of a patch pair is measured by the absolute error between the mean values of the patches. The absolute errors of all the mean value pairs can form a vector, \mathbf{h} , the element of which is $h_i = \left| \mu(\mathbf{x}_i^{ref}) - \mu(\mathbf{x}_i^{dis}) \right|$. Then the selection strategy is as follows.

$$\left(\mathbf{m}^{ref}, \mathbf{m}^{dis}\right) = \left\{ \left(\mu(\mathbf{x}_i^{ref}), \mu(\mathbf{x}_i^{dis}) \right) \mid \left| \mu(\mathbf{x}_i^{ref}) - \mu(\mathbf{x}_i^{dis}) \right| \geq TH_m \right\} \quad (19)$$

$$TH_m = T_m \cdot \text{median}(\mathbf{h}) \quad (20)$$

where \mathbf{m}^{ref} and \mathbf{m}^{dis} denote the selected mean values of the reference and distorted image patches, respectively, $\mu(\bullet)$ represents the mean value of a image vector, T_m is designed to adjust the value of TH_m , its default value is 1.

2) *Step 2: Correlation Measurement.* After the selection step, the component of luminance correlation, SFF_m , which is calculated on the basis of \mathbf{m}^{ref} and \mathbf{m}^{dis} , is defined as

$$SFF_m = \frac{\sum_{i=1}^K ((m_i^{ref} - \mu(\mathbf{m}^{ref})) \cdot (m_i^{dis} - \mu(\mathbf{m}^{dis}))) + C_m}{\sqrt{\sum_{i=1}^K (m_i^{ref} - \mu(\mathbf{m}^{ref}))^2 \cdot \sum_{i=1}^K (m_i^{dis} - \mu(\mathbf{m}^{dis}))^2 + C_m}} \quad (21)$$

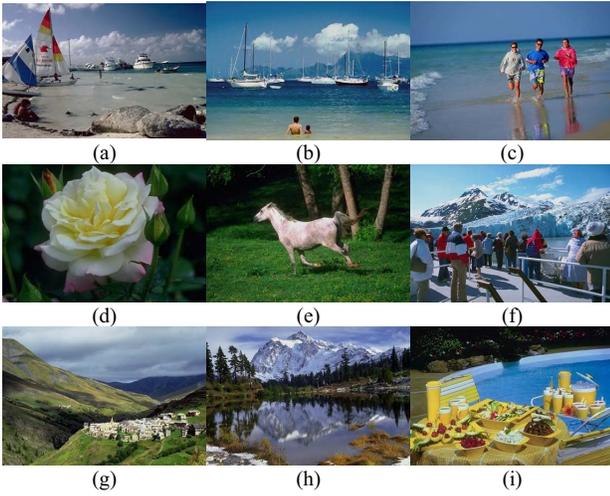


Fig. 3. Source images for the training process of SFF (D1). (a)–(i) are sample images taken from SIMPLiCity test set.

where $C_m > 0$ is a small constant to avoid dividing by zero (e. g., $C_m = 0.001$), m_i^{ref} and m_i^{dis} denote the i -th elements of \mathbf{m}^{ref} and \mathbf{m}^{dis} , respectively.

E. Sparse Feature Fidelity

Finally, the SFF index is calculated by combining SFF_m and SFF_f into a quality score.

$$SFF = \lambda \cdot SFF_m + (1 - \lambda) \cdot SFF_f \quad (22)$$

where $0 < \lambda < 1$ is a parameter for adjusting the relative importance of the two components, its proper value is 0.8.

V. EXPERIMENTS AND RESULTS

In this section, we shall present experimental results and comparisons on five color image databases. The SFF index was compared with seven state-of-the-art IQMs including: PSNR, SSIM [5], VSNR [11], IFC [40], pixel domain VIF (VIF-p) [13], IW-SSIM [10], and FSIM [41]. These IQMs were applied using their default implementations. In addition, nine source images for the training stage of SFF are shown in Fig. 3. They were randomly selected from the SIMPLiCity test set which contains 1000 images that are extracted from the COREL database [42].

To provide quantitative performance evaluations of the eight objective IQMs, we followed the performance evaluation procedures employed by the video quality experts group (VQEG) Phase-I/II validation methods [43], [44]. According to VQEG's suggestion, a nonlinear mapping function should be involved to remove the nonlinearity of objective scores and to facilitate the comparison of the IQMs in a common analysis space. In this experiment, we chose a five-parameter logistic function for nonlinear mapping [13].

$$Q(o) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(o - \beta_3))} \right) + \beta_4 o + \beta_5 \quad (23)$$

where o denotes the raw objective score of IQMs, and Q denotes the mapped score. The five fitting parameters, i.e.,

TABLE I
CHARACTERISTICS OF THE FIVE COLOR IMAGE DATABASES

Database	Distorted Images	Reference Images	Distortion Types
CSIQ	866	30	6
IVC	185	10	4
LIVE	779	29	5
TID	1700	25	17
TOY	168	14	2

$\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 , are determined by minimizing the sum of squared differences between the mapped objective scores, $Q(o)$, and the subjective ratings.

A. Databases for Validation

So far, there are five databases of ordinary color images for IQA [45], all of which were used in our validation and comparisons, i.e., LIVE [46], CSIQ [47], [48], IVC [49], TID2008 (TID) [50], and Toyama-MICT (TOY) [51]. Each of the image databases consists of hundreds of distorted images contaminated by a variety of distortion types. Each of these distorted images is provided with a subjective score, e.g., mean opinion score (MOS) or differential mean opinion score (DMOS). Main characteristics of the five databases are summarized in Table I.

In our experiment, only the distorted images in the five databases are employed (i.e., reference images are excluded). Then the performance validation is conducted by comparing these subjective ratings with the objective evaluations of IQMs.

B. Performance Metrics

According to VQEG's reports [43], [44], the performance of an objective IQM should be evaluated with respect to three aspects of its ability to estimate subjective assessment of image quality: *prediction accuracy* stands for the ability to predict the subjective quality ratings with low error; *prediction monotonicity* stands for the degree to which the metric's predictions agree with the relative magnitudes of subjective quality ratings, in other words, the objective quality scores should be monotonic in their relationship to the subjective scores; *prediction consistency* stands for the degree to which an IQM maintains prediction accuracy over the range of test images.

To provide a complete evaluation of each IQM, five metrics were employed to measure the performance of IQMs, including the Pearson linear correlation coefficient (PLCC) (after nonlinear mapping), the Spearman rank-order correlation coefficient (SRCC), the Kendall rank-order correlation coefficient (KRCC), the root mean squared error (RMSE) (after nonlinear mapping), and the outlier ratio (OR) (after nonlinear mapping). Among all these performance metrics, PLCC and RMSE are adopted to evaluate the prediction accuracy; SRCC and KRCC operate only on the rank of the data points and ignore the relative distance between data points, they are employed to assess the prediction monotonicity; OR is used as a measurement of an IQM's prediction consistency. A better objective IQM should have higher PLCC, SRCC, and KRCC, while lower RMSE and OR values.

TABLE II
PERFORMANCE COMPARISON OF 8 IQMs ON FIVE DATABASES

Database	Metric	PSNR	SSIM	VSNR	IFC	VIF-p	FSIM	IW-SSIM	SFF
CSIQ	PLCC	0.8000	0.8612	0.8002	0.8366	0.9388	0.9120	0.9144	0.9643
	SRCC	0.8057	0.8755	0.8106	0.7671	0.9280	0.9242	0.9213	0.9627
	KRCC	0.6078	0.6900	0.6247	0.5897	0.7658	0.7561	0.7529	0.8281
	RMSE	0.1575	0.1334	0.1575	0.1438	0.0905	0.1077	0.1063	0.0695
	OR	0.1674	0.1339	0.1386	0.1513	0.1028	0.1224	0.1189	0.0831
IVC	PLCC	0.7196	0.9119	0.8032	0.9093	0.8963	0.9376	0.9231	0.9324
	SRCC	0.6884	0.9018	0.7983	0.8993	0.8877	0.9262	0.9125	0.9249
	KRCC	0.5218	0.7223	0.6036	0.7202	0.6968	0.7564	0.7339	0.7553
	RMSE	0.8460	0.4999	0.7258	0.5069	0.5403	0.4236	0.4686	0.4404
	OR	0.1514	0.0703	0.1243	0.1351	0.1405	0.0703	0.0811	0.0811
LIVE	PLCC	0.8723	0.9449	0.9231	0.9268	0.9464	0.9597	0.9522	0.9632
	SRCC	0.8756	0.9479	0.9274	0.9259	0.9448	0.9634	0.9567	0.9649
	KRCC	0.6865	0.7963	0.7616	0.7579	0.8044	0.8337	0.8175	0.8365
	RMSE	13.3600	8.9460	10.5060	10.2642	8.8249	7.6780	8.3472	7.3460
	OR	0.1335	0.0873	0.1065	0.1091	0.0680	0.0719	0.0706	0.0655
TID	PLCC	0.5734	0.7732	0.6820	0.7340	0.8711	0.8738	0.8579	0.8817
	SRCC	0.5531	0.7749	0.7046	0.5675	0.8506	0.8805	0.8559	0.8767
	KRCC	0.4027	0.5768	0.5340	0.4236	0.6660	0.6946	0.6636	0.6882
	RMSE	1.0994	0.8511	0.9815	0.9113	0.6590	0.6525	0.6895	0.6333
	OR	0.1553	0.1471	0.1318	0.1535	0.1124	0.1176	0.1224	0.1212
TOY	PLCC	0.6428	0.8887	0.8680	0.8403	0.8892	0.9078	0.9248	0.9030
	SRCC	0.6132	0.8794	0.8614	0.8354	0.8846	0.9059	0.9202	0.8992
	KRCC	0.4443	0.6939	0.6762	0.6370	0.7027	0.7302	0.7537	0.7217
	RMSE	0.9587	0.5738	0.6213	0.6784	0.5727	0.5249	0.4761	0.5378
	OR	0.1012	0.0952	0.1250	0.1012	0.0952	0.0952	0.0714	0.1012
Average result on above five databases	PLCC	0.7216	0.8760	0.8159	0.8494	0.9083	0.9182	0.9145	0.9289
	SRCC	0.7072	0.8759	0.8205	0.7990	0.8991	0.9200	0.9133	0.9257
	KRCC	0.5326	0.6959	0.6399	0.6256	0.7271	0.7542	0.7442	0.7660
	RMSE	3.2843	2.2008	2.5971	2.5009	2.1375	1.8773	2.0175	1.8054
	OR	0.1418	0.1068	0.1205	0.1300	0.1038	0.0955	0.0926	0.0904
Weighted average result	PLCC	0.6999	0.8422	0.7751	0.8122	0.9049	0.9056	0.8973	0.9217
	SRCC	0.6897	0.8460	0.7881	0.7185	0.8919	0.9116	0.8982	0.9188
	KRCC	0.5184	0.6621	0.6130	0.5574	0.7217	0.7430	0.7244	0.7571
	RMSE	3.4425	2.3581	2.7654	2.6710	2.2362	1.9876	2.1453	1.9013
	OR	0.1509	0.1252	0.1263	0.1403	0.1014	0.1057	0.1060	0.0976

C. Overall Performance

Table II lists the performance evaluation results of SFF and other seven IQMs on the five databases. The best results across the eight IQMs for each database are highlighted in boldface. From Table II, we can see that SFF performs the best on LIVE and CSIQ databases, and it is comparable in performance to FSIM on TID database, meanwhile, it performs better than the other IQMs except FSIM and IW-SSIM on IVC and TOY databases. Moreover, the average results over the five databases are provided at the bottom of Table II, where the average values are computed in two cases. In the first case, the performance metric scores are directly averaged across the five databases; while in the second case, different weights are given to different databases depending upon their sizes (measured as the numbers of distorted images). From the two kinds of average results, we can see that SFF performs the best on average. In general, SFF correlates much more consistently with the subjective evaluations than the other seven IQMs.

In order to provide a visual illustration for the performance comparison among the eight IQMs, scatter plots of subjective ratings versus objective scores obtained by IQMs on CSIQ database are shown in Fig. 4, where each point represents one test image. The curves shown in Fig. 4 are obtained by a nonlinear fitting according to (23). Compared with other scatter plots, SFF's points are more close to each other, which means that SFF correlates well with subjective ratings.

D. Efficiency Evaluation

To compare the computational complexity of different IQMs, we measured the average execution time required for assessing an image of size 512×512 in CSIQ database using a laptop computer with Intel B940 processor at 2.00 GHz. All the codes were implemented with Matlab. Table III shows the results of computation time. It can be seen that SFF takes more time than PSNR and SSIM, but less time than VSNR,

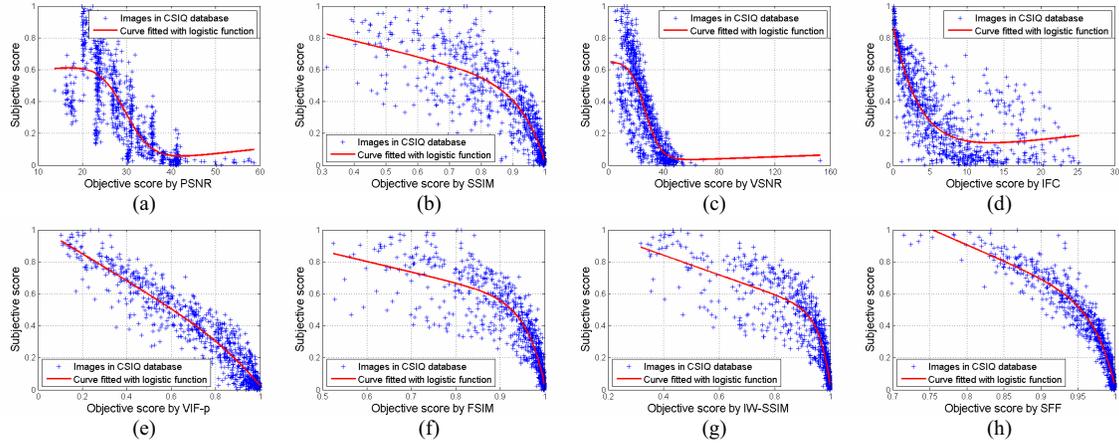


Fig. 4. Scatter plots of eight IQMs on CSIQ database. (a) PSNR, (b) SSIM, (c) VSNR, (d) IFC, (e) VIF-p, (f) FSIM, (g) IW-SSIM, and (h) SFF.

TABLE III
COMPARISON OF COMPUTATION TIME (IN SECONDS PER IMAGE)

IQM	PSNR	SSIM	VSNR	IFC	VIF-p	FSIM	IW-SSIM	SFF
Time	0.03522	0.06521	0.63642	2.47865	0.23875	0.76617	0.57353	0.18385

TABLE IV
SRCC COMPARISON OF INDIVIDUAL DISTORTION TYPES ON LIVE AND CSIQ DATABASES

Distortion	PSNR	SSIM	VSNR	IFC	VIF-p	FSIM	IW-SSIM	SFF
LIVE database								
JP2K	0.8954	0.9614	0.9551	0.9113	0.9710	0.9717	0.9649	0.9672
JPEG	0.8809	0.9764	0.9657	0.9468	0.9799	0.9834	0.9808	0.9786
AWGN	0.9854	0.9694	0.9785	0.9382	0.9855	0.9652	0.9667	0.9859
GB	0.7823	0.9517	0.9413	0.9584	0.9675	0.9708	0.9720	0.9752
FF	0.8907	0.9556	0.9027	0.9629	0.8664	0.9499	0.9442	0.9529
CSIQ database								
AWGN	0.9363	0.8974	0.9241	0.8429	0.8869	0.9262	0.9380	0.9469
JPEG	0.8879	0.9543	0.9033	0.9411	0.9585	0.9652	0.9660	0.9641
JP2K	0.9361	0.9605	0.9479	0.9251	0.9632	0.9684	0.9682	0.9762
PGN	0.9338	0.8924	0.9080	0.8261	0.9526	0.9233	0.9057	0.9549
GB	0.9291	0.9608	0.9445	0.9525	0.9644	0.9728	0.9781	0.9751
GCD	0.8623	0.7925	0.8695	0.4870	0.8894	0.9420	0.9540	0.9536

IFC, VIF-p, FSIM, and IW-SSIM, which means that SFF has relatively low computational complexity.

E. Performance on Individual Distortion Types

In this testing, we examined the performance of the competing methods on different image distortion types. This experiment was conducted on LIVE and CSIQ databases because these two databases contain the most commonly encountered distortion types. LIVE contains five types of distortions: JPEG compression (JPEG), JPEG-2000 compression (JP2K), additive white Gaussian noise (AWGN), Gaussian blur (GB), and fast fading Rayleigh channel noise (FF), while CSIQ contains six types of distortions: AWGN, JPEG, JP2K, pink Gaussian noise (PGN), GB, and global contrast decrements (GCD).

In order to simplify this experiment, only the SRCC values are listed. SRCC is chosen because it is suitable for measuring

a small number of data points and its value will not be affected by an unsuccessful nonlinear mapping.

The experimental results are summarized in Table IV. For each distortion type, the best (highest) value across the eight IQM results is highlighted in boldface. We can see that SFF performs the best for AWGN distortion on both two databases. Besides that, SFF's performance is more stable across different distortion types, e.g., the SRCC values of SFF are almost all above 0.95.

In order to discover the relationships among the results of different distortions, Fig. 5 shows the fitting curves of different distortion types for the six best IQMs (i.e., SSIM, VSNR, VIF-p, FSIM, IW-SSIM, and SFF) in our experiment. Ideally, for an IQM, these curves should lie on top of each other. If this were the case, then this IQM could stably predict subjective quality across all the distortion types [13]. From Fig. 5, we can see that all the curves of SFF are very close to each

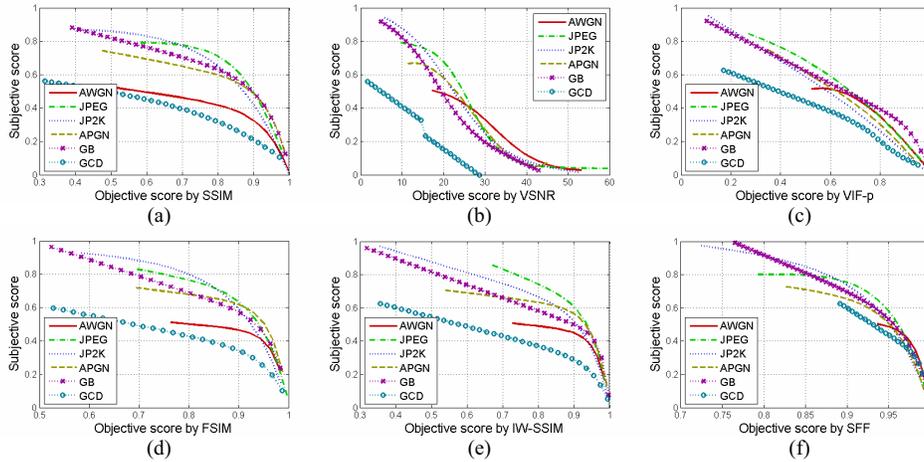


Fig. 5. Fitting curves of different distortions for six best IQMs on CSIQ. (a) SSIM, (b) VSNR, (c) VIF-p, (d) FSIM, (e) IW-SSIM, and (f) SFF.

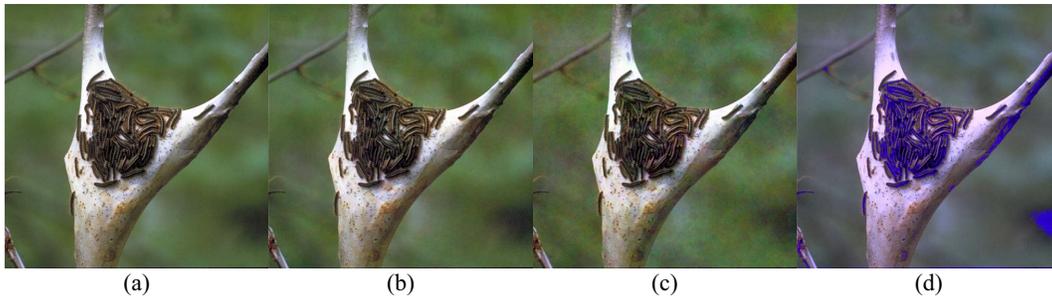


Fig. 6. Example images for the evaluation of color distortion. (a) Reference image, (b) Pink Gaussian noise of level 1, (c) Pink Gaussian noise of level 2, (d) Color distortion.

TABLE V
EVALUATION RESULTS OF THE EXAMPLE IMAGES IN FIG. 6

Image	SSIM	VSNR	VIF-p	IW-SSIM	FSIM	FSIMc	SFF
Fig. 6(b)	0.9852	36.2785	0.8487	0.9850	0.9883	0.9873	0.9922
Fig. 6(c)	0.9450	29.2942	0.6614	0.9453	0.9587	0.9559	0.9769
Fig. 6(d)	0.9970	36.6535	0.9362	0.9984	0.9987	0.9635	0.9506

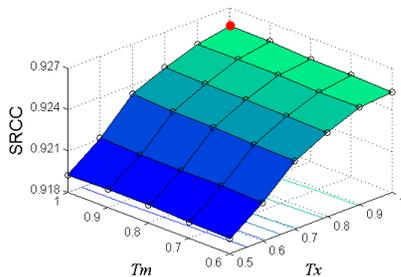


Fig. 7. Performance impact of T_x and T_m for SFF.

other, which means SFF can be stably calibrated for predicting quality for a range of distortion types.

F. Evaluation of Color Distortion

Color information is significant to visual perception and IQA. However, most IQMs are designed for gray-scale images,

and they only use the luminance component (or the average of RGB values) to make quality predictions. These IQMs can not accurately predict the perceived quality when a color-distorted image has no changes in the luminance component. Whereas, the SFF is sensitive to color distortion since it uses all the RGB color components in a perceptual way. In fact, the HVS encodes the chromatic signals conveyed by the three types of retinal cone photoreceptors in an opponent mechanism which can constitute a sparse code for natural images [52], moreover this mechanism can also be simulated by ICA [53].

Fig. 6 shows four example images. Fig. 6(a) is a reference image from CSIQ database, Fig. 6(b) and (c) are two images contaminated with different levels of pink Gaussian noise, and Fig. 6(d) is an image with color distortion (its red and luminance values remain unchanged, then green pixel values were decreased by 10%, finally its blue pixel values were computed by subtracting the red and decreased green values from the luminance values of the original image).

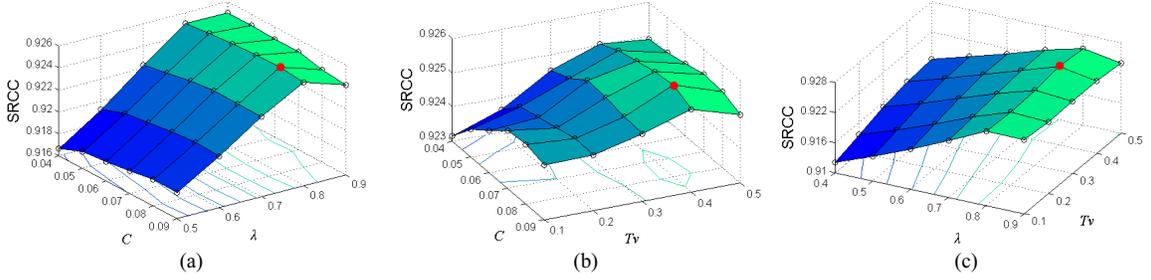


Fig. 8. Performance impact of T_v , λ , and C for SFF. (a) Results influenced by C and λ . (b) Results influenced by C and T_v . (c) Results influenced by λ and T_v .

For comparisons, SSIM, VSNR, VIF-p, FSIM, and IW-SSIM were selected to compare with SCC. Moreover, FSIMc [41], which is the chromatic version of FSIM, was also chosen for this comparison. The original quality scores evaluated by the 7 IQMs are listed in Table V. The evaluation results of SSIM, VSNR, VIF-p, IW-SSIM, FSIM, and FSIMc indicate that Fig. 6(d) is better than Fig. 6(c), or even Fig. 6(b). However, as we can see that Fig. 6(d) is worse than Fig. 6(c), and is much worse than Fig. 6(b), which is consistent with the result of SFF. Compared with other IQMs, SFF can give more accurate perceptual quality scores when evaluating color-distorted images.

G. Performance Impact of Parameters

In this section, we shall test five parameters of SFF, i.e., T_x , T_m , T_v , λ , and C to find how the parameters affect the performance of SFF. This testing was conducted on five databases, and then the average results (i.e., SRCC values) across all the databases were used for analyzing the impact of parameters.

We first analyzed the sensitivity of SFF to the values of T_x and T_m , while other parameters were fixed to be $T_v = 0.4$, $\lambda = 0.8$, and $C = 0.08$. Firstly, 5 different values of T_m (from 0.6 to 1 with increment of 0.1) and 6 different values of T_x (from 0.5 to 1 with increment of 0.1) were selected for this testing. The two sets of parameter values can provide 30 different combinations and result in 30 different results of SRCC. Fig. 7 shows the surface plot of SRCC as a function of the two parameters, each point represents an average result across the five databases. As shown in Fig. 7, SFF can give the best results when T_x as well as T_m is 1.

The other three parameters, T_v , λ , and C , are related with each other, so we analyzed the performance of all the three possible parameter combinations, i.e., (C, λ) , (C, T_v) , and (λ, T_v) . Firstly, T_x and T_m are both fixed to 1. Then 6 values of C (from 0.04 to 0.09 with increment of 0.01), 5 values of λ (from 0.5 to 0.9 with increment of 0.1), and 5 values of T_v (from 0.1 to 0.5 with increment of 0.1) were selected for this testing. The performance results of the three parameter combinations are shown in Fig. 8, from which we can see that the proper setting for SFF is $T_v = 0.4$, $\lambda = 0.8$ and $C = 0.08$.

TABLE VI
COMPARISON OF SRCC RESULTS FOR TWO TRAINING DATASETS

Data	CSIQ	IVC	LIVE	TID	TOY	Average
D1	0.9627	0.9249	0.9649	0.8767	0.8992	0.9257
D2	0.9632	0.9228	0.9634	0.8721	0.8973	0.9238

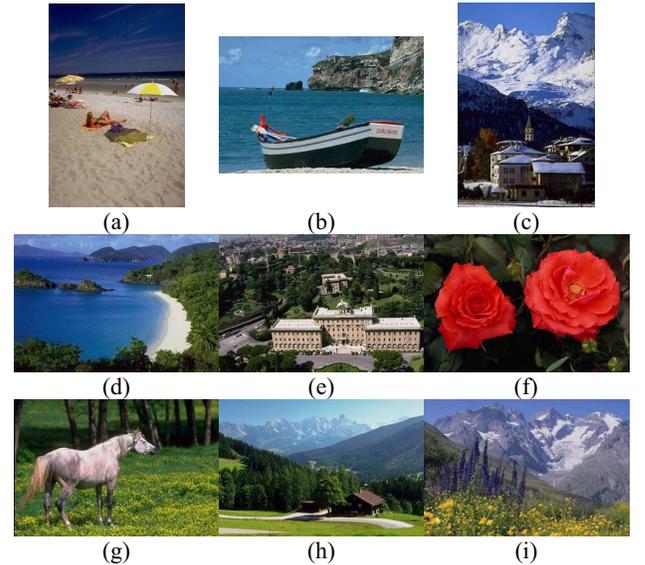


Fig. 9. Another set of source images for comparison (D2). (a)–(i) are sample images taken from SIMPLicity test set.

H. Performance Impact of Training Data

Though SFF relies on a training process, it will not cause great changes of the evaluation results if the training images are more diverse in content. In order to test the robustness of SFF, we chose another set of images for training (D2), which is shown in Fig 9. The results based on the two training sets (i.e., D1 of Fig. 3 and D2 of Fig. 9) are listed in Table VI, from which we can see that the difference between two sets of results is small, this means that the evaluation result of SFF has a certain robustness to the change of training data.

VI. CONCLUSION

In this paper, a novel FR IQM called sparse feature fidelity (SFF) was proposed for IQA. The computation of SFF is divided into two stages: training and fidelity computation.

Firstly, the universal feature detector, which will be used for extracting sparse features, is trained by FastICA on samples of natural images in the training stage. Then the fidelity computation consists of two components: feature similarity and luminance correlation. The computation of feature similarity component is based on the sparse features. Another component, luminance correlation, is computed based on the mean values of image patches. Finally, the SFF index will be achieved by combining feature similarity and luminance correlation together. To increase the accuracy and stability, two strategies, visual attention and visual threshold, were designed to remove the perceptually unimportant information from image patches. Experimental results and comparisons with other leading IQMs on five public image databases show that SFF can achieve higher consistency with the subjective evaluations than state-of-the-art IQMs.

The proposed IQM depends on a learning process which is implemented by an ICA algorithm. ICA can simulate the chromatic mechanism of the HVS (the resulting color filters resemble either blue–yellow or red–green double-opponent receptive fields with various orientations) [52], [53]. So SFF is very effective for color IQA. Moreover, SFF can also be applied to evaluating gray-scale images by training on luminance components of images.

However, a few aspects of SFF deserve further research and improvement. Firstly, MAE in the selection step is too simple to model the complex visual attention mechanism precisely, so some reliable algorithms can be considered. Secondly, in the training stage, FastICA can also be replaced with other more efficient dictionary learning algorithms for sparse representations.

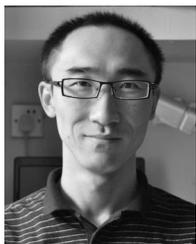
ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their valuable and insightful comments on this paper.

REFERENCES

- [1] P. Mouroulis and X. X. Cheng, "Visual instrument image-quality assessment with rotationally symmetric difference of Gaussians," *Appl. Opt.*, vol. 36, pp. 1667–1670, Mar. 1997.
- [2] A. M. van Dijk and J. B. Martens, "Subjective quality assessment of compressed images," *Signal Process.*, vol. 58, pp. 235–252, May 1997.
- [3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
- [4] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 163–172, 2004.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [6] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
- [7] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [8] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2385–2401, Nov. 2009.
- [9] C. F. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Process., Image Commun.*, vol. 25, pp. 517–526, Aug. 2010.
- [10] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
- [11] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [12] V. Laparra, J. Munoz-Mari, and J. Malo, "Divisive normalization image quality metric revisited," *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.*, vol. 27, pp. 852–864, Apr. 2010.
- [13] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [14] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
- [15] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
- [16] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.
- [17] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 469–481, 2010.
- [18] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.
- [19] B. A. Olshausen, "Principles of image representation in visual cortex," in *The Visual Neurosciences*. Cambridge, MA, USA: MIT Press, 2003, pp. 1603–1615.
- [20] B. A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, Jun. 1996.
- [21] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [22] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.
- [23] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [24] T. W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, 1999.
- [25] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 26. New York, NY, USA: Wiley, 2001.
- [26] A. Hyvärinen, "Statistical models of natural images and cortical visual representation," *Topics in Cognit. Sci.*, vol. 2, no. 2, pp. 251–264, 2010.
- [27] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*. Cambridge, MA, USA: MIT Press, 1961, pp. 217–234.
- [28] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current Opinion Neurobiol.*, vol. 14, no. 4, pp. 481–487, 2004.
- [29] A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*, vol. 39. New York, NY, USA: Springer-Verlag, 2009.
- [30] H. Chang and M. Wang, "Sparse correlation coefficient for objective image quality assessment," *Signal Process., Image Commun.*, vol. 26, pp. 577–588, Nov. 2011.
- [31] H.-W. Chang, M.-H. Wang, S.-Q. Chen, H. Yang, and Z.-J. Huang, "Sparse feature fidelity for image quality assessment," in *Proc. 21st Int. Conf. Pattern Recognit.*, 2012, pp. 1619–1622.
- [32] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. New York, NY, USA: Wiley, 2005.
- [33] J. J. Atick, "Could information theory provide an ecological theory of sensory processing?" *Netw., Comput. Neural Syst.*, vol. 3, no. 2, pp. 213–251, 1992.
- [34] Y. Dan, J. J. Atick, and R. C. Reid, "Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory," *J. Neurosci.*, vol. 16, no. 10, pp. 3351–3362, 1996.
- [35] H. Barlow, "Redundancy reduction revisited," *Netw., Comput. Neural Syst.*, vol. 12, no. 3, pp. 241–253, 2001.
- [36] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.

- [37] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [38] U. Engelke, H. Kaprykowsky, H. J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 50–59, Nov. 2011.
- [39] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.
- [40] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
- [41] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [42] J. Z. Wang, J. Li, and G. Wiederhold. (2001). *SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries* [Online]. Available: <http://wang.ist.psu.edu/~jwang/test1.tar>
- [43] VQEG. (2000, Apr.). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*, Singapore [Online]. Available: <http://www.vqeg.org>
- [44] VQEG. (2003, Aug.). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*, Singapore [Online]. Available: <http://www.vqeg.org/>
- [45] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 10, pp. 616–625, Oct. 2012.
- [46] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2006). *LIVE Image Quality Assessment Database Release 2* [Online]. Available: <http://live.ece.utexas.edu/research/quality/subjective.htm>
- [47] E. C. Larson and D. M. Chandler. (2009). *Consumer Subjective Image Quality Database* [Online]. Available: <http://vision.okstate.edu/index.php?loc=csiq>
- [48] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, pp. 011006-1–011006-21, Mar. 2010.
- [49] P. Le Callet and F. Atrousseau. (2005). *Subjective Quality Assessment IR-CCyN/IVC Database* [Online]. Available: <http://www.irccyn.ec-nantes.fr/ivcdb/>
- [50] N. Ponomarenko and K. Egiazarian. (2008). *Tampere Image Database* [Online]. Available: <http://www.ponomarenko.info/tid2008.htm>
- [51] N. Ponomarenko, et al. (2008). *TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics* [Online]. Available: <http://www.ponomarenko.info/tid2008.htm>
- [52] T. W. Lee, T. Wachtler, and T. J. Sejnowski, "Color opponency is an efficient representation of spectral properties in natural scenes," *Vis. Res.*, vol. 42, no. 17, pp. 2095–2103, 2002.
- [53] D. R. Tailor, L. H. Finkel, and G. Buchsbaum, "Color-opponent receptive fields derived from independent component analysis of natural images," *Vis. Res.*, vol. 40, no. 19, pp. 2671–2676, 2000.



Hua-Wen Chang received the Ph.D. degree in computer science and technology from Sichuan University, Chengdu, China, in 2012, and the M.S. degree in computer science from the Guilin University of Technology, Guilin, China, in 2007. He is currently an Associate Professor with the Zhengzhou University of Light Industry, Zhengzhou, China. His current research interests include image and video quality assessment, image fusion, sparse representation, independent component analysis, and wavelet analysis.



Hua Yang received the M.S. degree in computer science from China West Normal University, Nanchong, China, in 2005. Currently, she is pursuing the Ph.D. degree with the College of Computer Science, Sichuan University, Chengdu, China. She is a Lecturer with the School of Computer Science, China West Normal University. Her current research interests include image and video feature extraction and network security.



Yong Gan received the Ph.D. degree in computer science and technology from Xi'an Jiao Tong University, Xi'an, China, in 2006. He is currently the Dean of the College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. He is a Technical Committee Member for embedded systems with the China Computer Federation, the Director of the Henan Computer Federation, and the Vice Chairman of computer education with the Colleges of Henan, Henan, China. His current research interests include computer vision, emergency alert systems, and pattern recognition.



and Graphics.

Ming-Hui Wang received the Ph.D. and M.S. degrees in electronics engineering from Northwestern Polytechnical University and Xi'dian University, Xi'an, China. He is currently a Professor with the College of Computer Science, Sichuan University, Chengdu, China. He was a Post-Doctoral Research Fellow of information and communication engineering with Tsinghua University, Beijing, China. His current research interests include information fusion and computer vision. He is the Deputy Director of a technical committee with the China Society of Image