Journal of Electronic Imaging

JElectronicImaging.org

Parts-based stereoscopic image assessment by learning binocular manifold color visual properties

Haiyong Xu
Mei Yu
Ting Luo
Yun Zhang
Gangyi Jiang

Haiyong Xu, Mei Yu, Ting Luo, Yun Zhang, Gangyi Jiang, "Parts-based stereoscopic image assessment by learning binocular manifold color visual properties," *J. Electron. Imaging* **25**(6), 061611 (2016), doi: 10.1117/1.JEI.25.6.061611.

Parts-based stereoscopic image assessment by learning binocular manifold color visual properties

Haiyong Xu,[a,b] Mei Yu,[a] Ting Luo,[a,b] Yun Zhang,[c] and Gangyi Jiang[a,*]
[a]Ningbo University, Faculty of Information Science and Engineering, Ningbo 315211, China
[b]Ningbo University, College of Science and Technology, Ningbo 315212, China
[c]Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology, Shenzhen 518055, China

Abstract. Existing stereoscopic image quality assessment (SIQA) methods are mostly based on the luminance information, in which color information is not sufficiently considered. Actually, color is part of the important factors that affect human visual perception, and nonnegative matrix factorization (NMF) and manifold learning are in line with human visual perception. We propose an SIQA method based on learning binocular manifold color visual properties. To be more specific, in the training phase, a feature detector is created based on NMF with manifold regularization by considering color information, which not only allows parts-based manifold representation of an image, but also manifests localized color visual properties. In the quality estimation phase, visually important regions are selected by considering different human visual attention, and feature vectors are extracted by using the feature detector. Then the feature similarity index is calculated and the parts-based manifold color feature energy (PMCFE) for each view is defined based on the color feature vectors. The final quality score is obtained by considering a binocular combination based on PMCFE. The experimental results on LIVE I and LIVE Π 3-D IQA databases demonstrate that the proposed method can achieve much higher consistency with subjective evaluations than the state-of-the-art SIQA methods. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.25.6.061611]

Keywords: stereoscopic image quality assessment; color visual property; nonnegative matrix factorization; manifold learning; binocular combination.

Paper 16365SS received Apr. 30, 2016; accepted for publication Oct. 3, 2016; published online Oct. 26, 2016.

1 Introduction

Owing to a boom in three-dimensional (3-D) imaging technologies and related applications, perceptual quality assessment of 3-D visual signals plays an important role in the 3-D imaging field. Obviously, a variety of 3-D visual distortions can be caused during 3-D visual signal creation, processing, compression, network transmission, and display. Consequently, 3-D visual signals may be unsatisfactory in terms of the end-user's 3-D quality of experience. Therefore, a stereoscopic image as an important representation form of 3-D visual signals and its quality assessment has important practical significance.[1,2]

Two-dimensional (2-D) image quality assessment (IQA) methods have been widely studied. The most popular IQA metrics use a mathematical statistic to describe pixel distortions, such as mean squared error (MSE) and peak signal to noise ratio (PSNR). However, these metrics are not friendly for the human visual experience. To improve that, Wang et al.[3] proposed the structure similarity index (SSIM) to measure the structure loss of image. Sheikh and Bovik[4] proposed visual information fidelity (VIF) based on natural scene statistics. Zhang et al.[5] proposed a feature similarity index (FSIM) by using phase congruency and the image gradient magnitude. Compared with the traditional 2-D IQA, stereoscopic image quality assessment (SIQA) methods are required to account for left-right 2-D image quality, depth perception, and the human visual mechanism. However, studies of SIQA remain limited due to the lack of the understanding of human visual perception. Generally, SIQA methods can be categorized as: (1) 2-D metric-based SIQA,[6,7] (2) binocular perception-based SIQA,[8–10] and (3) simulated receptive field-based SIQA.[11,12]

For 2-D metrics-based SIQA methods, state-of-the-art 2-D metrics were used to estimate the quality of a stereoscopic image through weighting of the left and right views' quality.[13] However, it has been proven that stereoscopic image quality cannot be expressed simply as the average of its left and right views' quality.[14] To improve the accuracy of SIQA, depth/disparity information should be taken into consideration. Benoit et al.[6] presented a linear combination for disparity distortion and 2-D IQA on both views of the stereoscopic image. You et al.[7] applied different 2-D IQA metrics on a single distorted view and integrated the disparity information into SIQA. All of those methods were used to study the depth information as an independent factor to perceive the quality of stereoscopic images. However, it is not effective to evaluate the depth perception quality of stereoscopic image since stimuli regarding perceived depth are different from those for 2-D-IQA.

To make SIQA be consistent with human visual perception, more visual properties should be taken into consideration in SIQA. Bensalma and Larabi[8] proposed an SIQA method that measures the difference of binocular energy between the reference and distorted stereopairs, and thus considered the potential influence of binocularity on perceived 3-D image quality. Chen et al.[9] addressed binocular rivalry issues by modeling the binocular suppression behaviors, which produced the state-of-the-art SIQA method. Shao et al.[10] classified a stereoscopic image into noncorresponding, binocular fusion, and binocular suppression classes.

*Address all correspondence to: Gangyi Jiang, E-mail: gyjiang@nbu.edu.cn

Each region is evaluated independently by considering its visual properties, and all effects are finally integrated into an overall quality score. Lin and Wu[15] incorporated the binocular integration behaviors (the binocular combination and the binocular frequency integration) into the existing 2-D-IQA models as the basis for measuring the quality of stereoscopic 3-D images.

Recently, many simulated receptive field-based SIQA methods have also been proposed to learn the properties of visual perception. Zhang et al.[16] applied independent subspace analysis to simulate simple cells and complex cells in the primary visual cortex (V1). Chang et al.[17] proposed sparse features fidelity to simulate simple cells in V1 via independent component analysis. Guha et al.[18] proposed a sparse representation-based quality index. However, there are only a few works about simulated receptive field-based SIQA methods. Shao et al.[11] proposed a full-reference SIQA metric by using multiscales sparse coding to learn binocular receptive field properties to be more in line with human visual perception. Li et al.[12] proposed an SIQA method based on joint structure-texture sparse coding. However, these two methods only consider the luminance information of the stereoscopic image, and the color information is lost. Actually, color information is one of the important factors that affect human visual perception. In V1, the cells are sensitive to color information (11%), brightness (60%), and color-luminance (29%). To overcome the shortcoming of these methods, we consider the color information to derive the color visual property and use nonnegative matrix factorization (NMF) with manifold regularization to simulate parts-based sparse coding and manifold perception, so that manifold color visual properties are achieved. In addition, we denote parts-based manifold color feature energy (PMCFE) as binocular combination behaviors for evaluating the quality of stereoscopic images. Since we integrate the binocular combination behaviors into the proposed metric, the proposed method is applicable to both symmetric and asymmetric distorted stereoscopic images. Note that since the proposed method does not consider depth information of stereoscopic images and binocular rivalry, the performance of the proposed method for asymmetrically distorted stereoscopic images is slightly behind Chen's method[9] which considers depth information to account for binocular rivalry.

In this paper, we propose an SIQA method by learning binocular parts-based manifold color visual properties. The main contributions of this paper are as follows:

1. In the training phase, a feature detector is created from training database based on NMF with manifold regularization by considering color information. The purpose of the feature detector is to capture parts-based manifold color properties of the image.
2. In the quality estimation phase, we consider the visual importance and compare the difference of the feature vectors to calculate the FSIM.
3. We use the estimated feature vectors to define PMCFE to get the binocular combination.

This paper is organized as follows: the relevant backgrounds are summarized in Sec. 2. The proposed SIQA method is described in Sec. 3. The experimental results are shown and discussed in Sec. 4. Finally, conclusions are made in Sec. 5.

2 Backgrounds

2.1 *Color and Stereopsis*

Apart from the enrichment of the human visual experience, more information is in a colored world than in a black-and-white world. Some studies[19,20] provided evidence that objects in colored scenes are more easily detected and more easily identified than are objects in black-and-white scenes.

For stereopsis, there is an evidence for stereoscopic perception mechanisms being sensitive to chromaticity.[21] Obviously, a prerequisite for successful stereopsis is correct matching of the corresponding parts of each view. However, a number of studies demonstrated that the chromatic difference could indeed reduce the number of false matches even if the objects of the scenes have similar luminance.

2.2 *Parts-Based Coding and Manifold Perception*

The studies of the human visual perception show that specific neurons are responsible for specific objects, and these objects are present in the human brain as part of their form.[22,23] The NMF method[24] was, therefore, proposed to learn the parts of objects such as human faces and text documents. The aim of the NMF method is to find two nonnegative matrices (the basis matrix and the encoding matrix), and the product of these matrices is the best approximation to the original matrix. In addition, the number of bases in the basis matrix is equal to the number of objects that represent the specific neurons in charge of a particular component, and the number of bases is usually small, which shows that NMF is sparse.

According to the visual perception phenomenon, it has been shown that manifolds are fundamental to perception and the visual perception of the human nervous system has the ability to capture the nonlinear manifold structure.[25] Population activity is typically described by a collection of neural firing rates, and so can be represented by a point in an abstract space with dimensionality equal to the number of neurons. The firing rate of each neuron in a population can be written as a smooth function of a small number of variables. This implies that the population activity is constrained to lie on a low-dimensional manifold. There is a great amount of information redundancy in the digital image, which needs to be processed by the dimension reduction technology. At the same time, it is expected that the essential structure of an image can be maintained. The manifold learning is used to find the low-dimensional manifold hidden in the high-dimensional data set through the nonlinear geometric variation, and it can reflect the intrinsic structure of the original high-dimensional data. For IQA, the distorted image in the manifold subspace will be in accordance with the type of change and the size of the intensity by using manifold learning.

2.3 *Binocular Combination*

Human binocular vision is a complex visual process. Light images on the two retinas are combined to form a single "cyclopean" perceptual image, in contrast to binocular rivalry, which occurs when the two eyes have incompatible inputs and only one eye's stimulus is perceived. Recently, Ding and Sperling[26] proposed the gain-control theory model for the binocular combination as

$$f_C = \left(\frac{1+E_L}{1+E_L+E_R}\right)\cdot I^L + \left(\frac{1+E_R}{1+E_L+E_R}\right)\cdot I^R, \quad (1)$$

where f_C is the perceived cyclopean image, E_L or E_R is the sum of energy over all the frequency channels for the left view or the right view, respectively, and I^L or I^R is the image presented to the left view or right view, respectively.

The gain-control model can be used to describe the binocular combination and explains the cyclopean perception.

Based on the gain-control model, Lin and Wu[15] proposed an SIQA model which integrates the binocular integration behaviors into the existing 2-D objective metrics for evaluating the quality of stereo images. They denoted the frequency-integrated metrics (FI-metrics) as follows:

$$\text{FI-metric} = \sum\{g_i^L D(V_i^L, V_I^{L\prime}) + g_i^R D(V_i^R, V_I^{R\prime})\}, \quad (2)$$

where $D(\cdot)$ represents 2-D-IQA metric, and g_i^L and g_i^R are the gains of the i'th channel for the left view and right view, respectively.

3 Proposed Parts-Based SIQA Method by Learning Binocular Manifold Color Visual Properties

In this paper, by considering binocular manifold color visual properties by using NMF with manifold regularization and binocular combination from RGB color channels, we propose an SIQA method as shown in Fig. 1. The proposed method is divided into two phases: training and quality estimation. We first consider color information, parts-based coding, and manifold perception. Using NMF with manifold regularization, the feature detector, D, is derived from the training database. Then visually important regions (VIR) are selected based on different human visual attention, and left and right views' manifold color visual features are extracted by using the feature detector to calculate manifold color visual FSIM and define PMCFE. Finally, the quality score of the stereoscopic image is derived by incorporating binocular combination by using PMCFE in the quality estimation phase. Therefore, how to capture parts-based manifold color visual features for different distortions and how to simulate binocular integration behaviors are the keys to the success of the proposed method.

3.1 *Training Phase*

3.1.1 *Selection of the training database*

To construct the training database, we randomly select nine original natural images from the Berkeley image segmentation database,[27] which includes different textures and different scenes, as shown in Fig. 2. Since the intrinsic motivation of the proposed method is to measure the similarities between the reference and distorted images based on NMF with manifold regularization, we only use the nondistorted images to construct the parts-based manifold color visual feature detector, D.

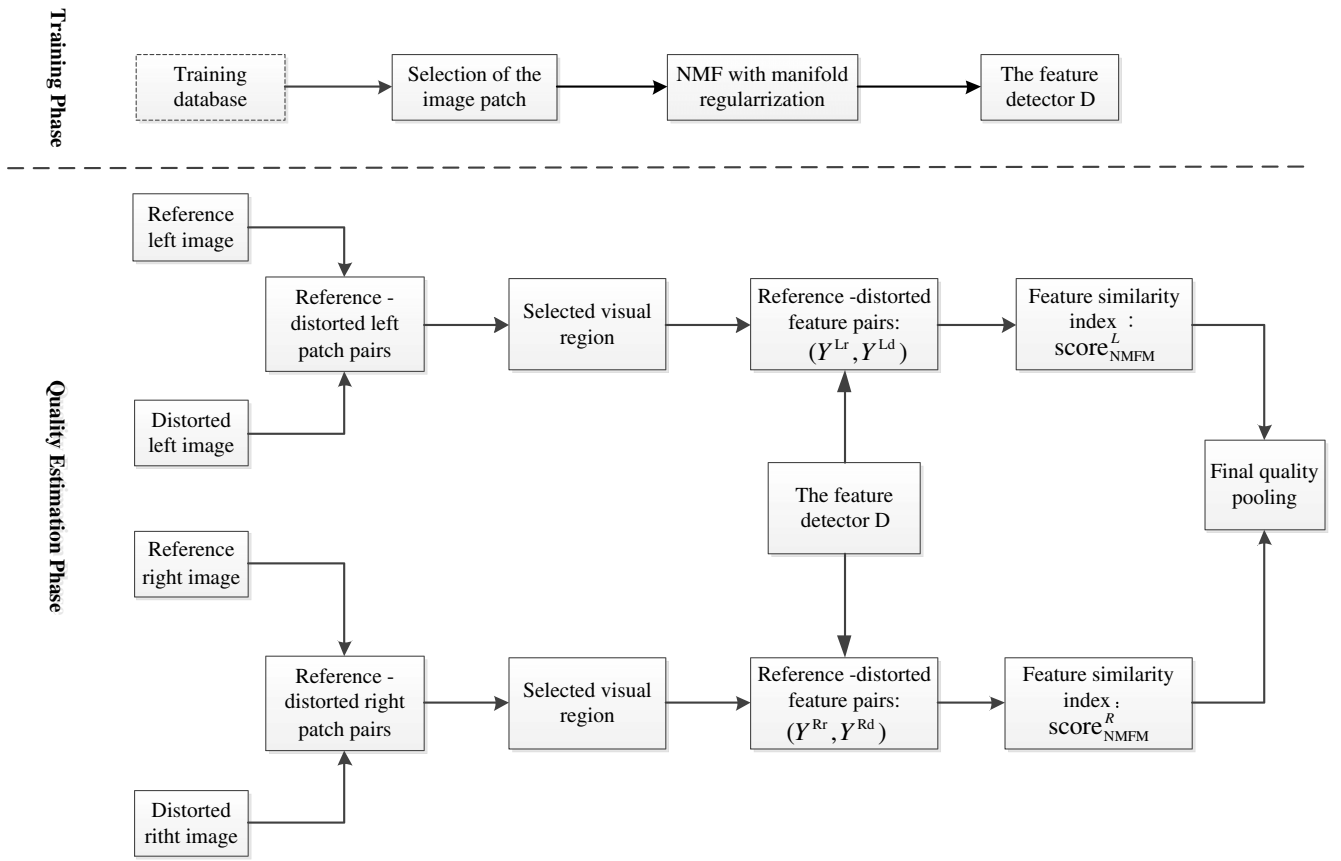

Fig. 1 The framework of the proposed parts-based SIQA method by learning binocular manifold color visual properties.

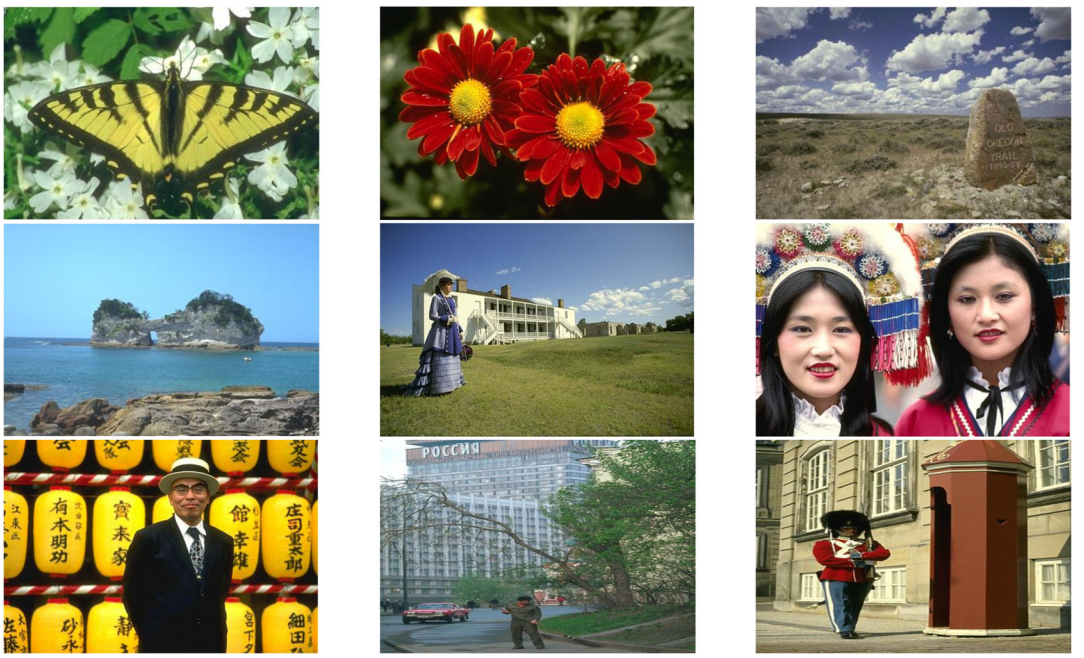

Fig. 2 Selected images for the training phase of the proposed method.

3.1.2 *Construction of the feature detector*

After constructing the training database, n nonoverlapping image patches with a size of $m \times m$ are randomly taken from the training images. In the implementation, each patch is vectorized into a column vector by scanning the values in the patch row-by-row and channel-by-channel. Since a color image has three RGB channels, the length K of the vector is $K = m \times m \times 3$. Thus, all the sample vectors form a sample matrix, $X = [x_{ij}] = [X_1, X_2, \cdots, X_n]$, where each path $X_j \in R^{K\times 1}$ contains K pixels.

Since X is nonnegative and NMF can learn a parts-based representation, the NMF method can be used to decompose the matrix X into a nonnegative feature basis matrix and a coding matrix. Specifically, for the nonnegative matrix X, NMF aims to find two nonnegative matrices, $U = [u_{ik}] \in R^{K\times r}$ and $S = [s_{jk}] \in R^{n\times r}$ such that

$$X = US^T, \tag{3}$$

where $r > 0$ is the number of the basis vector in the sample matrix. In practice, we have $r < K$ and $r < n$.

Thus, NMF is essentially used to find a compressed approximation of the sample matrix. Each sample X_j can be represented as

$$X_j \approx \sum_{k=1}^{r} U_k s_{jk}, \tag{4}$$

where U_k is the k'th column vector of U. Thus, each sample X_j is approximated by a linear combination of the columns of U, weighted by the components of S. Therefore, the matrix U can be regarded as comprising a basis that is optimized for the linear approximation of the sample matrix.

Generally, in order to find two nonnegative matrices U and S, there are two commonly used cost functions. The first one is the square of the "Frobenius norm" of two matrices' difference:

$$E_1 = \|X - US^T\|_F^2 = \sum_{i,j}\left(x_{ij} - \sum_{k=1}^{r} u_{ik}s_{jk}\right)^2. \tag{5}$$

The second one is the divergence between two matrices

$$E_2 = D(X||US^T) = \sum_{i,j}\left(x_{ij}\log\frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij}\right), \tag{6}$$

where $Y = [y_{ij}] = US^T$. The cost function of Eq. (5) is the square of the Euclidean distance between two matrices and the cost function of Eq. (6) is referred to as the divergence instead of distance between X and Y.

By using the nonnegative constraints, NMF can be used to learn a parts-based representation of image. However, it fails to discover the intrinsic geometrical and discriminating structure of the original data space, which is important for designing the IQA metric for different types and levels of distorted images. Here, a natural assumption could be that if two original data points X_j and X_l are close, then the corresponding low-dimension data points Z_j and Z_l are also close to each other. Therefore, the manifold learning theory is embedded in the NMF framework.

The low-dimensional representation of X_j with respect to the feature basis is $Z_j = [s_{j1}, \cdots, s_{jr}]^T$. The weight matrix $W = [W_{jl}]$ for measuring the closeness of two points X_j and X_l is defined as follows:

$$W_{jl} = \begin{cases} e^{-\frac{\|X_j - X_l\|^2}{\varepsilon}} & \text{if } X_l \text{ is among the } k \text{ nearst neighbors of } X_j \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $\varepsilon > 0$ is a parameter. Here, we set $\varepsilon = 1$.

Therefore, we can use the following two manifold regulation terms to measure the smoothness of the low-dimensional representation

$$\Theta_2 = \frac{1}{2}\sum_{j,l=1}^{n}[D(Z_j||Z_l) + D(Z_l||Z_j)]W_{jl} = \frac{1}{2}\sum_{j,l=1}^{n}\sum_{k=1}^{r}\left(s_{jk}\log\frac{s_{jk}}{s_{lk}} + s_{lk}\log\frac{s_{lk}}{s_{jk}}\right)W_{jl}, \tag{8}$$

and

$$\Theta_1 = \frac{1}{2}\sum_{j,l=1}^{n}\|Z_j - Z_l\|^2 W_{jl} = \sum_{j=1}^{n} Z_j^T Z_j D_{jj} - \sum_{j,l=1}^{n} Z_j^T Z_l W_{jl} = Tr(S^T DS) - Tr(S^T WS) = Tr(S^T LS), \tag{9}$$

where $Tr(\cdot)$ denotes the trace of a matrix, D is a diagonal matrix, $D_{jj} = \sum W_{jl}$, and $L = D - W$ is the graph Laplacian matrix.

Similar to NMF, two objective functions according Eqs. (8) and (9) for obtaining two nonnegative matrices U and S are defined as follows:[28]

$$\mathrm{O}_1 = \|X - US^T\|^2 + \lambda Tr(S^T LS), \tag{10}$$

or

$$\mathrm{O}_2 = \sum_{i,j}\left(x_{ij}\log\frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij}\right) + \frac{\lambda}{2}\sum_{j,l=1}^{n}\sum_{k=1}^{r}\left(s_{jk}\log\frac{s_{jk}}{s_{lk}} + s_{lk}\log\frac{s_{lk}}{s_{jk}}\right)W_{jl}. \tag{11}$$

Here, Eq. (11) is used as the objective function and to minimize this function to obtain the following two updating rules:

$$u_{ik} \leftarrow u_{ik}\frac{\sum_j x_{ij}s_{jk}/\sum_k x_{ik}s_{jk}}{\sum_j s_{jk}}, \tag{12}$$

$$S_k \leftarrow \left(\sum_i u_{ik} I + \lambda L\right)^{-1}\begin{bmatrix} s_{1\,k}\sum_i\left(x_{i1}u_{ik}/\sum_k u_{ik}s_{1\,k}\right) \\ s_{2\,k}\sum_i\left(x_{i2}u_{ik}/\sum_k u_{ik}s_{2\,k}\right) \\ \vdots \\ s_{nk}\sum_i\left(x_{\mathrm{in}}u_{ik}/\sum_k u_{ik}s_{nk}\right)\end{bmatrix}, \tag{13}$$

where I is an identity matrix, and S_k is the k'th column of S.

Thus, the two nonnegative matrices U and S are calculated, and the feature basis matrix U is needed.

After the above processing, each patch T_i can be represented as a linear combination of the feature basis matrix U, i.e.,

$$T_i = UF_i, \tag{14}$$

where F_i is a feature vector after dimensionality reduction, and its dimension is r. Thus, using the generalized inverse matrix $D = (U^T U)^{-1}U^T$, the feature vector can be obtained as follows:

$$F_i = DT_i, \tag{15}$$

where D is the feature detector.

3.2 *Feature Similarity and Quality Estimation Phase*

3.2.1 *Selection of reference-distorted patch pairs*

It is well known that not every pixel in an image receives the same level of visual attention. For IQA, the quality of the image is mainly concentrated in the VIR. Here, we use the visual saliency detector to detect the VIR in an image.

In the following section, the left and right views of a stereoscopic image are processed in the same way, and the left view is taken as an example to describe the algorithm processing. The visual saliency detection algorithm[29] is used to obtain the saliency maps $M_{\mathrm{r}}^{\mathrm{L}}$ and $M_{\mathrm{d}}^{\mathrm{L}}$ on the left view $I_{\mathrm{r}}^{\mathrm{L}}$ of a reference stereoscopic image and its corresponding distorted left view $I_{\mathrm{d}}^{\mathrm{L}}$, respectively.

To obtain VIR of the left view, $I_{\mathrm{r}}^{\mathrm{L}}$, $I_{\mathrm{d}}^{\mathrm{L}}$, $M_{\mathrm{r}}^{\mathrm{L}}$, and $M_{\mathrm{d}}^{\mathrm{L}}$ are segmented into nonoverlapping patches with the same size of 8×8 pixels, and these patches are vectorized and arranged in columns of the matrices X^{Lr}, X^{Ld}, S^{Lr}, and S^{Ld}, respectively.

In this paper, to measure the difference of saliency between the left view of the reference stereoscopic image and the distorted stereoscopic image, the term e_j is defined as

$$e_j = \frac{1}{N}\sum_{i=1}^{N}|S_{ij}^{\mathrm{Lr}} - S_{ij}^{\mathrm{Ld}}|, \qquad j = 1, \cdots, M, \tag{16}$$

where S_{ij}^{Lr} is the element of the i'th row and the j'th column of S^{Lr}, M is the number of the image patches, and N denotes the number of pixels in each patch. Here, $N = 8 \times 8 = 64$.

Then, sorting all of e_j from large to small, e_i^* is obtained as

$$e_1^* \geq e_2^* \geq \cdots \geq e_M^*. \tag{17}$$

Let $t_1 = \lambda_1 \cdot M$, where $\lambda_1 \in (0,1]$ denotes the ratio coefficient of the selected reference-distorted patch pairs and t_1 denotes the number of selected reference-distorted patch pairs. Thus, e_i^* of the former t_1 corresponding to $(Y^{\mathrm{Lr}}, Y^{\mathrm{Ld}})$ are selected as the VIR, i.e.,

$$(Y^{\mathrm{Lr}}, Y^{\mathrm{Ld}}) = \{(X_j^{\mathrm{Lr}}, X_j^{\mathrm{Lr}})|j \in \mathrm{label}\{e_i^* \text{ of the former } t_1\}\}. \tag{18}$$

After the above processing, the final visually important left reference-distorted patch pairs are selected as the VIR. Similarity, visually important right reference-distorted patch pairs $(Y^{\mathrm{Rr}}, Y^{\mathrm{Rd}})$ are also selected.

3.2.2 *Feature extraction*

After obtaining VIR of image, the parts-based manifold color feature vectors, a_i^{L} and b_i^{L}, can be extracted by using the feature detector D as follows:

$$a_i^{\mathrm{L}} = D \times Y_i^{\mathrm{Lr}}, \tag{19}$$

and

$$b_i^{\mathrm{L}} = D \times Y_i^{\mathrm{Ld}}. \tag{20}$$

Thus, the feature vectors a_i^{L} and b_i^{L} form two matrices, A^{L} and B^{L}. Since the size of D is $r \times m$, the length of all a_i^{L}, b_i^{L}, a_i^{R}, and b_i^{R} are r, and the size of all A^{L}, B^{L}, A^{R}, and B^{R} are $r \times t_1$.

3.2.3 *Feature similarity index*

In order to quantify the perceptual quality of the image, we compare the feature vector matrices A^{L} and B^{L}. Therefore, FSIM of the left distorted image among the feature vectors is defined as

$$\mathrm{Score}_{\mathrm{NMFM}}^{\mathrm{L}} = 1 - \frac{1}{r \cdot t_1} \sum_{i=1}^{r} \sum_{j=1}^{t_1} \frac{(a_{ij}^L - b_{ij}^L)^2 + C}{(a_{ij}^L)^2 + (b_{ij}^L)^2 + C}, \tag{21}$$

where t_1 denotes the number of the retained feature vectors in an image, and a_{ij}^{L} and b_{ij}^{L} denote the values of the i'th row and the j'th column of U^{L} and V^{L}, respectively. C is a constant to avoid the denominator being zero in this paper, $C = 0.08$.

In HVS, the visual response for each view of the stereoscopic image will not be the same. To simulate the visual properties from the binocular combination, we define PMCFE.

Definition 1: Given an image I, its PMCFE is defined as the average feature energy of the VIR in image I, i.e.,

$$\mathrm{PMCFE}(I) = \frac{1}{t_1} \sum_{j=1}^{t_1} \|A_j\|_2^2. \tag{22}$$

According to Eq. (22), we can calculate PMCFE for each view and define the weights for the left and right views of stereoscopic images by, respectively,

$$\omega^{\mathrm{L}} = \frac{\mathrm{PMCFE}(I^{\mathrm{L}})}{\mathrm{PMCFE}(I^{\mathrm{L}}) + \mathrm{PMCFE}(I^{\mathrm{R}})}, \tag{23}$$

and

$$\omega^{\mathrm{R}} = \frac{\mathrm{PMCFE}(I^{\mathrm{R}})}{\mathrm{PMCFE}(I^{\mathrm{L}}) + \mathrm{PMCFE}(I^{\mathrm{R}})}. \tag{24}$$

According to Eqs. (23) and (24), the final FSIM of the stereoscopic image, i.e., SIQA score, can be derived as follows:

$$Q_{\mathrm{NMFC}} = \omega^{\mathrm{L}} \cdot \mathrm{Score}_{\mathrm{NMFM}}^{\mathrm{L}} + \omega^{\mathrm{R}} \cdot \mathrm{Score}_{\mathrm{NMFM}}^{\mathrm{R}}. \tag{25}$$

4 Experimental Results and Analyses

4.1 *Databases and Performance Measures*

1. The LIVE 3-D phase I database[30] consists of 365 symmetrically distorted stereoscopic images generated from 20 reference stereoscopic images by corrupting them with five different distortion categories: JPEG 2000 (JP2K) and the JPEG compression standards, additive white Gaussian noise (WN), Gaussian blur (Blur), and a fast-fading (FF) model based on the Rayleigh fading channel.
2. The LIVE 3-D phase Π database[31] consists of 120 symmetrically distorted stereoscopic images and 240 asymmetrically distorted stereoscopic images generated from eight reference stereoscopic images. It includes the same distortion categories as phase I.

In this paper, three indices that measure the consistency between the results of the proposed method and DMOS are used: the Spearman rank order correlation coefficient (SRCC) and the Pearson linear correlation coefficient (PLCC), which measure the prediction monotonicity, and the root mean squared error (RMSE), which measures the prediction accuracy. A perfect match between the objective and subjective scores will give SRCC = PLCC = 1 and RMSE = 0. For the nonlinear regression, the four-parameter logistic function is defined as follows:[32]

$$\mathrm{DMOS}_P = \frac{\beta_1 - \beta_2}{1 + \exp\left(-\frac{x - \beta_3}{|\beta_4|}\right)} + \beta_2, \tag{26}$$

where β_1, β_2, β_3, and β_4 are the parameters of the regression model. Note that the nonlinear regression is applied to LIVE I and LIVE Π 3-D IQA databases independently since LIVE I database only consists of symmetrically distorted images and LIVE Π database consists of both symmetrically and asymmetrically distorted stereoscopic images.

4.2 *Overall Assessment Performance*

Here, we compare the proposed method with some state-of-the art SIQA metrics on the two benchmark databases in terms of SRCC, PLCC, and RMSE. Particularly, these metrics can be divided into two groups: (1) luminance information plus energy response-based information (e.g., FI-PSNR, FI-SSIM, FI-VIF, Bensalma's method,[8] and Shao's method[11]) and (2) luminance plus disparity based information (e.g., Benoit's method,[6] You's method,[7] and Chen's method[9]). FI-PSNR, FI-SSIM, FI-VIF, and Bensalma's method are based on 2-D-IQA. Shao's method uses multi-scales sparse representation and sparse energy response. Benoit's method, You's method, and Chen's method use disparity information. The performance of the three methods is highly dependent on the stereoscopic matching algorithm.

Table 1 lists the performance evaluation results of the proposed method and eight other SIQA methods on the two test databases. The best results across the nine SIQA methods for each database are highlighted in boldface. From Table 1, it is found that Chen's method and Shao's method are reasonably good for the two databases. A possible reason is that "cyclopean" perceptual map based (Chen's method) and sparse representation based (Shao's method) methods are highly in line with human visual perception. However, these methods only consider the luminance information of the image. Since the proposed method considers the color information and uses the NMF with manifold regularization to learn manifold color visual properties, the proposed method can achieve much higher results than the other SIQA methods. Actually, color is one of the important factors that affect human visual perception, and parts-based representation and manifold are fundamental to perception.

Table 1 Performance of the proposed method and the other eight methods in terms of SRCC, PLCC, and RMSE on the two databases (cases in bold denote best performance).

		FI-PSNR	FI-SSIM	FI-VIF	Bensalma[8]	Benoit[6]	You[7]	Chen[9]	Shao[11]	Proposed
LIVE I	SRCC	0.8599	0.8606	0.9188	0.8747	0.8901	0.9247	0.9157	0.9251	**0.9310**
	PLCC	0.8645	0.8699	0.9222	0.8874	0.8899	0.9303	0.9167	0.9350	**0.9381**
	RMSE	8.2424	8.0874	6.3423	7.5585	7.4786	6.0161	6.5503	5.8155	**5.6789**
LIVE II	SRCC	0.6375	0.6795	0.7213	0.7513	0.7475	0.7206	**0.9013**	0.8494	0.8879
	PLCC	0.6584	0.6844	0.7234	0.7699	0.7642	0.7744	**0.9065**	0.8628	0.8979
	RMSE	8.4956	8.2295	7.7936	7.2035	7.2806	7.1413	**4.7663**	5.7058	4.9680
Average	SRCC	0.7487	0.7701	0.8201	0.8130	0.8188	0.8227	0.9085	0.8873	**0.9095**
	PLCC	0.7615	0.7772	0.8228	0.8287	0.8271	0.8524	0.9116	0.8989	**0.9180**
	RMSE	8.3690	8.1585	7.0680	7.3810	7.3796	6.5787	5.6583	5.7607	**5.3235**

In addition, from the table, an observation is that Chen's method can achieve the best results on the LIVE 3-D phase II database while the proposed method gets close to the performance of Chen's method. A possible explanation for this situation is that Chen's method considered binocular rivalry and used depth information to construct perceived cyclopean image while the proposed method only considers binocular combination behaviors and does not use depth information. Accounting for binocular rivalry can greatly improve the performance of SIQA methods on asymmetric distorted stereoscopic images. However, Chen's method depends highly on accurate ground truth depth values.

In order to provide a visual illustration for the performance of the proposed method, the scatter plots of predicted quality scores against subjective quality scores on the LIVE 3-D phase I database and the LIVE 3-D phase II database are shown in Fig. 3. From the figure, the proposed method's points are close to each other, which means the proposed method correlates well with subjective ratings.

4.3 *Performance on Individual Distortion Types*

In this section, we comprehensively compare the proposed method with the other SIQA methods on each type of distortion. PLCC and SRCC results are listed in Tables 2 and 3, respectively, where the top three metrics have been highlighted in bold. From the tables, we can see that the proposed method is among the top 14 in terms of PLCC and SRCC, followed by Chen's method (13 times) and Shao's method (11 times). In addition, the proposed method achieves perfect results for JPEG, JP2K, and FF distortions. A possible explanation is that the proposed method is more sensitive to these three distortions. However, the proposed method is noted as very prominent for WN distortion because the localized features cannot reflect the changes of image quality

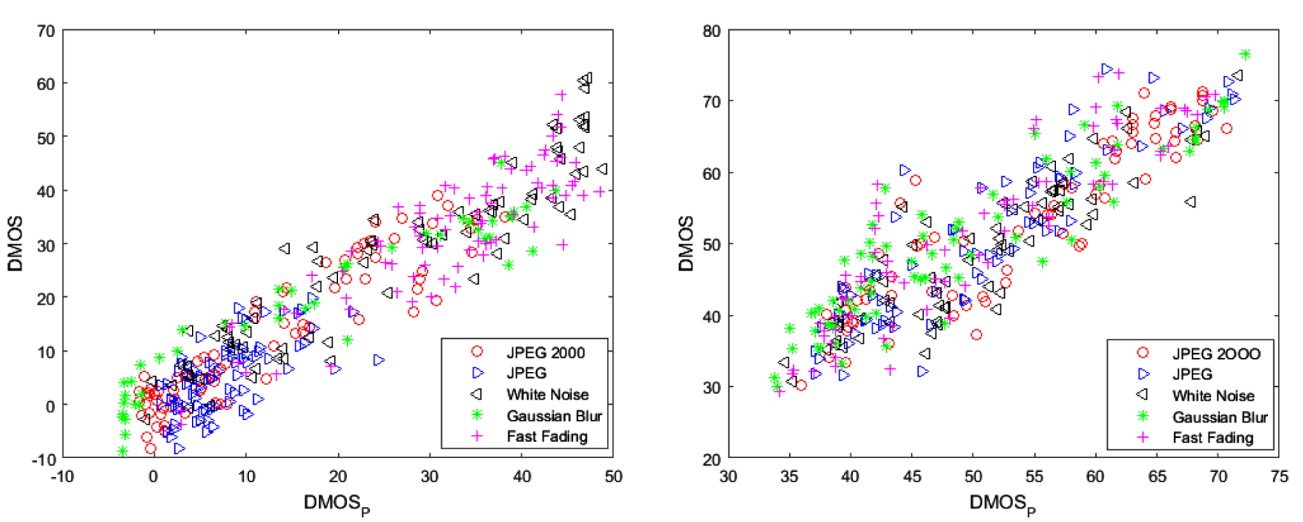

Fig. 3 Scatter plots of predicted quality scores against the subjective scores of the proposed method on the LIVE 3-D phase I database and the LIVE 3-D phase II database. (a) LIVE 3-D phase I database. (b) LIVE 3-D phase II database.

Table 2 Performance comparisons of the nine methods on each individual distortion types in terms of PLCC.

	Criteria	FI-PSNR	FI-SSIM	FI-VIF	Bensalma[8]	Benoit[6]	You[7]	Chen[9]	Shao[11]	Proposed
LIVE I	JPEG	0.2866	0.2741	**0.6545**	0.3803	0.5766	0.6333	**0.6344**	0.5200	**0.6547**
	JP2K	0.8381	0.8210	**0.9421**	0.8389	0.8859	**0.9410**	0.9164	0.9213	**0.9357**
	WN	0.9280	0.9250	0.9310	0.9147	0.9354	0.9351	**0.9436**	**0.9448**	**0.9373**
	Gblur	0.9475	0.9080	**0.9573**	0.9369	0.9217	0.9545	0.9417	**0.9592**	**0.9633**
	FF	0.7086	0.7297	0.7572	0.7339	0.7477	**0.8589**	0.7580	**0.8594**	**0.8756**
LIVE II	JPEG	0.6124	0.5486	**0.8906**	**0.8577**	0.5328	0.6741	0.8422	0.7472	**0.9038**
	JP2K	0.7457	0.7191	**0.9164**	0.6667	0.6467	0.7320	**0.8426**	0.7823	**0.9218**
	WN	0.9150	0.9139	0.8981	**0.9436**	0.8610	0.5464	**0.9602**	**0.9464**	0.9045
	Gblur	0.7083	0.7250	0.8993	0.9077	0.8814	**0.9763**	**0.9650**	**0.9580**	0.8974
	FF	0.7025	0.7342	0.7574	**0.9097**	0.8472	0.8561	**0.9097**	0.9046	**0.9183**

for this distortion. Overall, the proposed method not only predicts the image quality consistently across different types of distortions but also has an impressive consistency with human perception.

4.4 *Influence of the Parameter Selection*

In this section, we conduct experiments on the LIVE 3-D phase I database to explore the impact of our parameters: the number of sample n, patch size m, r, and λ_1. Without losing generality, PLCC values are used for analyzing the impacts of the parameters.

We first analyze the effects of different setting combinations of the number of sample n and patch size m since these two parameters affect the formation of the feature detector D in the training phase. In our experiment, the following parameter candidates are considered: $n \in \{13{,}500(1500 \times 9), 18{,}000(2000 \times 9), 22{,}500(2500 \times 9)\}$, and $m \in \{5,6,7,8,9,10,11\}$. Performance effects of n and m are shown in Fig. 4. In Fig. 4, the highest value of PLCC is obtained when n equals to 18000 and m equals to 8. From the figure, an interesting observation is that no further increase of PLCC is observed when the value of patch size m is continuously increased. A possible explanation for this situation is that too larger or too small a patch size m will reduce the image quality of evaluation. As is clearly seen from the figure, we set $n = 18{,}000$ and $m = 8$ in this paper.

We further analyze the sensitivity to the value of r, while the parameter λ_1 is fixed to be $\lambda_1 = 0.4$. First, we set 17 integer values of r from 8 to 24 with the step of 1. Figure 5 shows

Table 3 Performance comparisons of the nine methods on each individual distortion types in terms of SRCC.

	Criteria	FI-PSNR	FI-SSIM	FI-VIF	Bensalma[8]	Benoit[6]	You[7]	Chen[9]	Shao[11]	Proposed
LIVE I	JPEG	0.2070	0.2047	**0.6002**	0.3283	0.4983	**0.6008**	0.5582	0.4951	**0.6368**
	JP2K	0.8388	0.8222	**0.9125**	0.8170	0.8730	**0.9051**	**0.8956**	0.8945	0.8920
	WN	0.9284	0.9282	0.9335	0.9055	0.9369	**0.9403**	**0.9481**	**0.9405**	0.9393
	Gblur	**0.9345**	0.8788	0.9329	0.9157	0.8802	0.9300	0.9261	**0.9403**	**0.9445**
	FF	0.6581	0.6866	0.7497	0.6500	0.6242	**0.8030**	0.6879	**0.7963**	**0.8324**
LIVE II	JPEG	0.6129	0.5641	**0.8768**	0.8461	0.5078	0.5229	**0.8396**	0.7330	**0.8893**
	JP2K	0.7193	0.7003	**0.9212**	0.8038	0.6325	0.7309	**0.8334**	0.7845	**0.9027**
	WN	0.9073	0.9091	0.9341	**0.9386**	0.8569	0.4820	**0.9554**	**0.9651**	0.8861
	Gblur	0.7112	0.7387	0.8868	0.8838	0.8545	**0.9227**	**0.9096**	**0.9204**	0.8914
	FF	0.7012	0.7350	0.7586	0.8743	0.8319	0.8392	**0.8890**	**0.8905**	**0.8996**

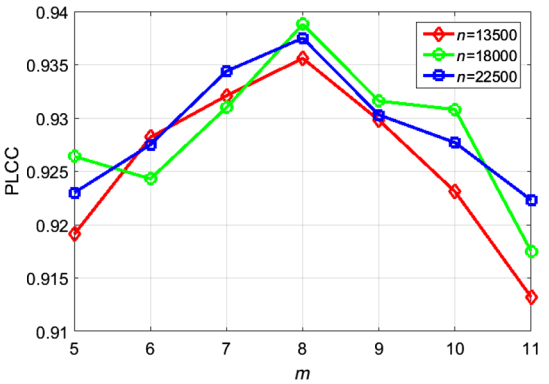

Fig. 4 Effects of different setting combinations of the number of sample n and patch size m.

performance impact of r and λ_1. Figure 5(a) shows the plot of PLCC as a function of the parameter r. As shown in Fig. 5, the best result is obtained on the LIVE 3-D phase I database when $r = 9$. Although the value of PLCC is oscillatory, the values are stable between 0.92 and 0.93. It shows that the proposed method can obtain good quality evaluations results for different parameters r. Here, we set $r = 9$.

Similarly, we analyze the inference of the parameter λ_1, while the parameter is fixed to be $r = 9$. We set 9 values of the parameter λ_1 from 0.1 to 1 with an increment of 0.1. The performance results are shown in Fig. 5(b). The parameter λ_1 affects the number of the VIR. The whole image will give $\lambda_1 = 1$. As shown in Fig. 5(b), the best value of PLCC is obtained when $\lambda_1 = 0.4$. This shows that the VIR occupy a certain proportion of the whole image. Therefore, we set $\lambda_1 = 0.4$ in this paper.

4.5 *Impact of Color Information*

Here, we discuss the impact of color information for SIQA. Color information used in the proposed method is composed of RGB three channels. It is necessary to verify the impact of color information and to test the performances when only luminance information is used or independent channel information is used. We denote these schemes by Pro-A (only luminance information is used), Pro-B (only intensity information of R-channel is used), Pro-C (only intensity information of G-channel is used), Pro-D (only intensity information of G-channel is used), and Pro-E (color information of the proposed method is used). We further compare the performances of the five schemes with different information channels. The comparison results of all these schemes on two databases are presented in Table 4. From the table, we can see that the best performance is obtained when the color information of RGB channels is jointly used and other schemes give similar results. A possible explanation for the comparison results is that the prediction accuracy and the prediction monotonicity of SIQA are improved when color information of the RGB channels is used. Certainly, satisfactory results are also obtained when only luminance information is used or independent channel information is used. Overall, color information of the RGB channels is able to improve the performance of SIQA.

Table 4 Performance comparisons for each proposed scheme.

		Pro-A	Pro-B	Pro-C	Pro-D	Pro-E
LIVE I	SRCC	0.9229	0.9226	0.9186	0.9177	**0.9310**
	PLCC	0.9263	0.9278	0.9240	0.9208	**0.9381**
	RMSE	6.1790	6.1170	6.2698	6.3944	**5.6789**
LIVE II	SRCC	0.8797	0.8714	0.8640	0.8720	**0.8879**
	PLCC	0.8887	0.8865	0.8793	0.8850	**0.8979**
	RMSE	5.1741	5.2234	5.3766	5.2551	**4.9680**

5 Conclusions

In this work, we propose a parts-based SIQA method by learning manifold color visual properties. First, color information and parts-based manifold perception are considered. Then, a color feature detector is learned from the training images by using NMF with manifold regularization. Furthermore, the VIR are selected and the local manifold

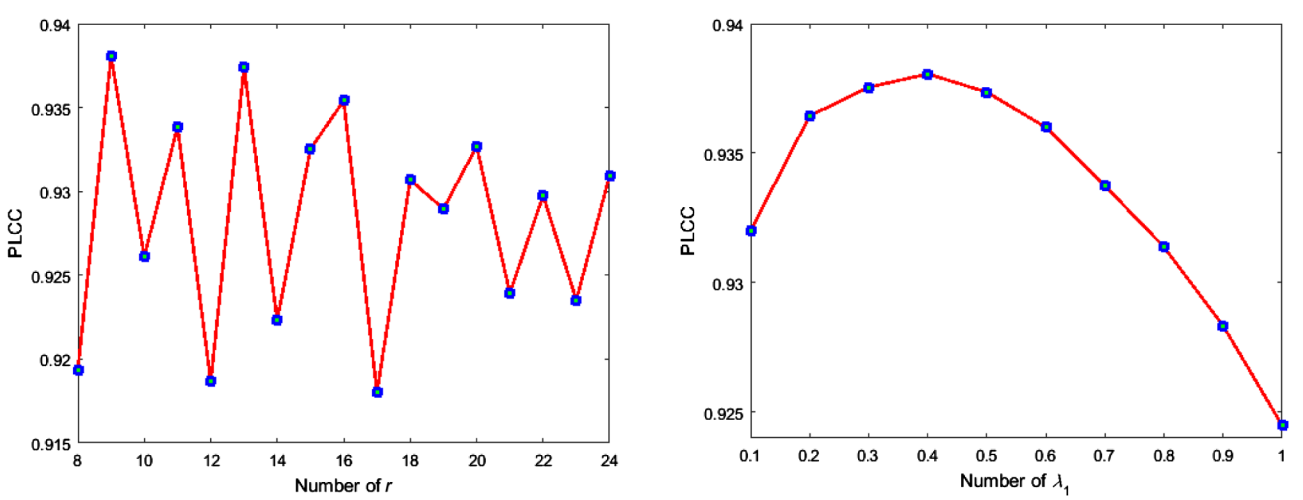

Fig. 5 Performance impact of r and λ_1. (a) Results influenced by r. (b) Results influenced by λ_1.

color visual feature of the selected VIR is extracted by using the feature detector. Finally, the quality score of the stereoscopic image is obtained by incorporating the binocular combination. Moreover, the experimental results of the proposed method are consistent with subjective quality assessment. In the future, we will extend the proposed method to measure the quality assessment of stereoscopic video.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant Nos. U1301257, 61271270, 61671258, and 61471348, the National High-tech R&D Program of China under Grant No. 2015AA015901, the Natural Science Foundation of Zhejiang Province under Grant No. LY15F010005, and the Natural Science Foundation of Ningbo under Grant No. 2016A610071. It is also sponsored by K.C. Wong Magna Fund in Ningbo University.

References

1. Y. Song et al., "Stereoscopic image quality assessment using disparity-compensated view filtering," *J. Electron. Imaging* **25**(2), 023001 (2016).
2. F. Shao et al., "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Trans. Broadcast.* **61**(2), 154–165 (2015).
3. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
4. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
5. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
6. A. Benoit et al., "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.* **2008**, 659024 (2009).
7. J. You et al., "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proc. Int. Workshop Video Processing Quality Metrics Consumer Electronics*, Scottsdale, Arizona (2010).
8. R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimension. Syst. Signal Process.* **24**(2), 281–316 (2013).
9. M.-J. Chen et al., "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process. Image Commun.* **28**(9), 1143–1155 (2013).
10. F. Shao et al., "Perceptual full reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.* **22**(5), 1940–1953 (2013).
11. F. Shao et al., "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.* **24**(10), 2971–2983 (2015).
12. K. Li et al., "Joint structure-texture sparse coding for quality prediction of stereoscopic images," *Electron. Lett.* **51**(24), 1994–1995 (2015).
13. P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE* **6803**, 680305 (2008).
14. D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery," *J. Exp. Psychol. Appl.* **7**, 143–153 (2001).
15. Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.* **23**(4), 1527–1542 (2014).
16. F. Zhang et al., "Exploring V1 by modeling the perceptual quality of images," *J. Vis.* **14**(1), 26–26 (2014).
17. H. W. Chang et al., "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.* **22**(10), 4007–4018 (2013).
18. T. Guha, E. Nezhadarya, and R. K. Ward, "Sparse representation-based image quality assessment," *Signal Process. Image Commun.* **29**(10), 1138–1148 (2014).
19. N. J. Dominy and P. W. Lucas, "Ecological importance of trichromatic vision to primates," *Nature* **410**(6826), 363–366 (2001).
20. S. K. Shevell and F. A. A. Kingdom, "Color in complex scenes," *Annu. Rev. Psychol.* **59**, 143–166 (2008).
21. D. R. Simmons and F. A. A. Kingdom, "On the independence of chromatic and achromatic stereopsis mechanisms," *Vision Res.* **37**(10), 1271–1280 (1997).
22. S. E. Palmer, "Hierarchical structure in perceptual representation," *Cognitive Psychol.* **9**(4), 441–474 (1977).
23. E. Wachsmuth, M. W. Oram, and D. I. Perrett, "Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque," *Cereb. Cortex* **4**(5), 509–522 (1994).
24. D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**(6755), 788–791 (1999).
25. H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science* **290**(5500), 2268–2269 (2000).
26. J. Ding and G. Sperling, "A gain-control theory of binocular combination," *Proc. Natl. Acad. Sci. U. S. A.* **103**(4), 1141–1146 (2006).
27. D. Martin et al., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Eighth IEEE Int. Conf. on Computer Vision*, Vol. 2, pp. 416–423 (2001).
28. D. Cai et al., "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1548–1560 (2011).
29. L. Zhang, Z. Gu, and H. Li, "SDSP: a novel saliency detection method by combining simple priors," in *IEEE Int. Conf. Image Processing*, pp. 171–175 (2013).
30. A. K. Moorthy et al., "Subjective evaluation of stereoscopic image quality," *Signal Process. Image Commun.* **28**(8), 870–883 (2013).
31. M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.* **22**(9), 3379–3391 (2013).
32. S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.* **24**(4), 591–602 (2014).

Haiyong Xu is a PhD student at Ningbo University. He received his BS degrees in mathematics and applied mathematics from Jilin University in 2003 and his MS degree in applied mathematics from Sun Yat-sen University in 2005. His current research interests include image and video quality assessment and image processing.

Mei Yu received her MS degree from Hangzhou Institute of Electronics Engineering, China, in 1993, and PhD degree from Ajou University, Republic of Korea, in 2000. She is now a professor in the Faculty of Information Science and Engineering, Ningbo University. Her research interests mainly include visual perception and image/video coding.

Ting Luo received his BSc degree in computer science from Ningbo University, Ningbo, China, in 2003, and his MSc degree in business information technology from Middlesex University, London, UK in 2004. His research interests include multimedia security, image processing, data hiding and pattern recognition.

Yun Zhang received his BS and MS degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the PhD in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China, in 2010. His current research interests include multiview video coding, video object segmentation, and content based video processing.

Gangyi Jiang received his MS degree from Hangzhou University, China, in 1992, and his PhD degree from Ajou University, Republic of Korea, in 2000. He is now a professor in the Faculty of Information Science and Engineering, Ningbo University. His research interests mainly include multiview video coding and 3-D image/video quality assessment.