J. Vis. Commun. Image R. 46 (2017) 269–279

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions ☆

Gangyi Jiang [a,*], Haiyong Xu [a,b,*], Mei Yu [a], Ting Luo [a,b], Yun Zhang [c]

[a] Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[b] College of Science and Technology, Ningbo University, Ningbo 315211, China
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 28 November 2016
Revised 7 March 2017
Accepted 27 April 2017
Available online 29 April 2017

Keywords:
Stereoscopic image quality assessment
Non-negative matrix factorization
Schmidt orthogonalization
Monocular perception
Binocular interaction

ABSTRACT

In this paper, we propose a novel stereoscopic image quality assessment (SIQA) method by learning non-negative matrix factorization (NMF)-based color visual characteristics for monocular perception and considering binocular interactions. In training phase, a feature basis matrix is learned based on NMF by considering color information and a feature detector is designed by performing Schmidt orthogonalization on the feature basis matrix. In construction of SIQA phase, for monocular perception, visual saliency regions are selected and parts-based feature similarity indexes of left and right views based on the feature vectors extracted by the feature detector are calculated. For binocular interactions, we calculate cyclopean feature similarity index by considering binocular fusion and rivalry. Finally, support vector regression is used to simulate nonlinear relationship between monocular perception and binocular interactions. Experimental results on LIVE 3D image databases and NBU 3D IQA database demonstrate that the proposed SIQA method is more consistent with human perception.

1. Introduction

Image quality assessment (IQA) is a crucial aspect in various image processing applications, such as image communication, image fusion, medical imaging, and so on [1–3]. Recently, three dimensional (3D) video/image and the related technologies have drawn great concern in academic researches and relative applications. Unfortunately, 3D visual distortion is almost unavoidable during 3D content creation, compression, network transmission and stereoscopic display. These distortions may cause 3D visual signals to be unsatisfactory in terms of the end-user 3D quality of experience [4]. Therefore, stereoscopic image quality assessment (SIQA) is essential to optimize 3D content production and processing [5–8].

Generally, SIQA in 3D video/image systems can be categorized into subjective and objective assessments [9–12]. Since human is the final receiver of the visual signal, subjective SIQA methods provide the ultimate perceptual quality evaluation of stereoscopic images. Recently, there are 8 subject-rated image databases including LIVE 3D Image Quality Database Phase I [13], LIVE 3D Image Quality Database Phase II [14], IRCCyN/IVC 3D Images Database [15], MICT 3D Image Quality Evaluation Database [16], MMSPG 3D IQA Database [17], NBU 3D IQA database [18], Tianjin University 3D IQA database [19], and Waterloo-IVC 3D Image Quality database [20]. Subjective data is important for understanding the visual perception of stereoscopic images. However, subjective quality assessment is infeasible for many applications since the subjective tests are inconvenient, time-consuming and expensive. Therefore, it is necessary to develop objective SIQA methods for evaluating the performance of 3D image processing technologies. During recent years, a number of SIQA methods have been proposed. According to the use of 3D perception, existing objective SIQA methods may be grouped into two categories: (1) the SIQA methods based on monocular IQA (MIQA) metrics [21,22,15,23], (2) the SIQA methods based on 3D perceptual properties [24–28].

For the SIQA methods based on MIQA metrics, the most straightforward way is to use state-of-the-art MIQA metrics to evaluate the quality of the left and right views and to get the quality score of the stereoscopic image through the left and right

☆ This paper has been recommended for acceptance by 'Zicheng Liu'.

* Corresponding authors at: Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (G. Jiang). College of Science and Technology, Ningbo University, Ningbo 315211, China (H. Xu).

E-mail addresses: jianggangyi@126.com (G. Jiang), xuhaiyong@nbu.edu.cn (H. Xu).

http://dx.doi.org/10.1016/j.jvcir.2017.04.010
1047-3203/

weights [29]. However, some researchers have provided evidence that the perceived quality of stereoscopic images cannot be expressed simply as the average quality of the left and right views [22]. To evaluate the stereoscopic image quality more accurately, researchers have taken into account the effect of disparity/depth information. Yang et al. [22] proposed a SIQA model based on the average PSNR of left and right views and the absolute difference with respect to the disparity map. Benoit et al. [15] presented a linear combination for disparity distortion and the measurement of 2D image quality on both views. You et al. [23] investigated different MIQA metrics on a single view and integrated the disparity information into SIQA. However, it is not effective to assess the quality of perceived depth information since stimuli regarding perceived depth are different from those for MIQA. In addition, the ground truth of the disparity map is generally not available, and the estimated disparity map is usually inaccurate.

To make SIQA better consistent with human perception, more binocular perceptual characteristics of the human visual system (HVS) should be considered. Ryu et al. [24] proposed a 3D-version of SSIM based on binocular quality perception. Chen et al. [25] proposed a "cyclopean" SIQA model accounting for binocular rivalry issues. Shao et al. [26] classified stereoscopic images into non-corresponding, binocular fusion and binocular suppression regions and all effects are finally integrated into an overall quality score. Zhou et al. [27] proposed a perceptual modulated feature similarity (PMFS) metric for SIQA by considering the monocular and binocular perception properties. Lin and Wu [28] incorporated binocular combination and binocular frequency integration into the existing MIQA metrics to measure the perceived quality of stereoscopic images. Lee and Lee [30] proposed a 3D perception-based stereoscopic image quality pooling model, which segments a stereoscopic image into binocular and monocular vision segments, and respectively evaluates them. Wang et al. [20] proposed a binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images from the single-view images. To improve SIQA method and simulate the properties of visual perception, many simulated receptive field methods have been proposed. Bensalma and Larabi [31] proposed a binocular energy quality metric which simulated the HVS by modeling the simple cells responsible for the local spatial frequency analysis and the complex cells responsible for the generation of the binocular energy. Shao et al. [32] used multi-scales sparse coding to learn binocular receptive field properties to be more in line with human visual perception, and proposed a corresponding metric. However, these methods only considered the luminance information of stereoscopic image, and the color information is lost. To overcome the shortcoming, the color information of stereoscopic image is considered and non-negative matrix factorization (NMF) is used to simulate parts-based perception of HVS. For parts-based perception of HVS, a representation of the object in the brain is segmented into separate parts and these parts and their relationships are indexed [33,34]. In mathematics, NMF is able to learn the intrinsic parts underlying the object being pictured [35]. This parts-based representation under the non-negativity constraint in NMF encourages sparsity and is conceptually similar to sparse coding. Here, NMF is used to simulate parts-based perception of HVS and learn local color visual characteristics.

In this paper, we propose a new SIQA method by learning NMF-based color visual characteristics and considering binocular interactions. The main contributions of this paper are as follows.

(1) A feature detector is derived based on NMF by considering color information from the RGB channel of the training images. The feature detector is able to extract the local color properties of the testing image.
(2) In construction of SIQA phase, we consider the visual saliency and compare the difference of the color feature vector to calculate the feature similarity index for monocular perception.
(3) Based on gain-control theory, binocular interactions are considered to simulate complex binocular perception.

This paper is organized as follows: The motivations of the proposed SIQA method are presented in Section 2. The proposed SIQA method is described in Section 3. Experimental results are discussed in Section 4. Finally, conclusions are given in Section 5.

2. Motivations

Different from MIQA methods, SIQA method should not only consider MIQA method, but also binocular mechanism. Based on [36], given left and right views of stereoscopic image, I_L and I_R, the perceived quality ζ can be modeled as the posteriori probability with the given left and right views:

$$P(\zeta|I_L,I_R)=\frac{P(I_L,I_R|\zeta)\cdot P(\zeta)}{P(I_L,I_R)} \tag{1}$$

where $P(I_L,I_R|\zeta)$ is the inverse posterior probability for the given left and right views with the perceived quality ζ, $P(\zeta)$ is the prior probability for the perceived quality ζ, and $P(I_L,I_R)$ is the prior joint probability for the given left and right views.

$P(I_L,I_R|\zeta)$ can be decomposed as

$$P(I_L,I_R|\zeta)=P(I_L|\zeta)+P(I_R|\zeta)-\alpha P(I_L\cap I_R|\zeta) \tag{2}$$

where $P(I_L\cap I_R|\zeta)$ is the binocular interaction between the left and right views with the perceived quality ζ, and $\alpha>0$ is a positive constant to balance these aspects.

Using Eq. (2), Eq. (1) is rewritten as follows:

$$\begin{aligned}P(\zeta|I_L,I_R)&=\frac{(P(I_L|\zeta)+P(I_R|\zeta)-\alpha P(I_L\cap I_R|\zeta))\cdot P(\zeta)}{P(I_L,I_R)}\\&=\frac{\left(\frac{P(I_L)\cdot P(\zeta|I_L)}{P(\zeta)}+\frac{P(I_R)\cdot P(\zeta|I_R)}{P(\zeta)}-\alpha\frac{P(I_L\cap I_R)\cdot P(\zeta|I_L\cap I_R)}{P(\zeta)}\right)\cdot P(\zeta)}{P(I_L,I_R)}\\&=\underbrace{\frac{P(I_L)}{P(I_L,I_R)}\cdot P(\zeta|I_L)+\frac{P(I_R)}{P(I_L,I_R)}\cdot P(\zeta|I_R)}_{\text{Monocular perception}}\\&\quad-\alpha\cdot\frac{P(I_L\cap I_R)}{P(I_L,I_R)}\cdot\underbrace{P(\zeta|I_L\cap I_R)}_{\text{Binocular interaction}}\end{aligned} \tag{3}$$

According to Eq. (3), the quality of monocular perception is measured by the visual quality of the left and right views, and the visual quality of binocular interactions is measured by visual quality of the corresponding cyclopean image. Thus, perceived quality of stereoscopic image can be obtained by

$$\begin{aligned}Q&=\frac{P(I_L)}{P(I_L,I_R)}\cdot Q_L+\frac{P(I_R)}{P(I_L,I_R)}\cdot Q_R-\beta\cdot Q_C\\&=\omega_L\cdot Q_L+\omega_R\cdot Q_R-\beta\cdot Q_C\end{aligned} \tag{4}$$

where $\omega_L=\frac{P(I_L)}{P(I_L,I_R)}$ and $\omega_R=\frac{P(I_R)}{P(I_L,I_R)}$ are weights of the left and right views, respectively, and $\beta>0$ is a parameter to balance monocular perception and binocular interactions.

Due to the complexity of binocular vision, the relationship between monocular perception and binocular interactions is not clear. It is difficult to quantify easily with the weight coefficient, and β in Eq. (4) needs to be artificially controlled, which leads to certain difficulties to the model. To solve this problem, we use some nonlinear tool, such as support vector regression (SVR), to better describe the complex relationship between monocular perception and binocular interactions.

2.1. Monocular perception

From a neuro-biological point of view, the goal of IQA is to simulate the perceptual mechanism of HVS. In HVS, there is a series of cells from the retina to the cerebral cortex, which is described in the receptive field (RF) model. The RF is the basic structural and functional unit of the information processing in HVS and is usually characterized to understand the behavior of visual perception. As an important property of the RF, sparse coding can capture the salient structures in line with the visual perception and the perceptual structural degradation of the distorted image is important for IQA methods [37].

Since NMF does not allow negative entries in the matrix factors and these non-negativity constraints lead to parts-based representation and sparsity, we want to use NMF to simulate properties of the RF for monocular perception. Here, the color information of the image is considered and the feature detector from the training samples is obtained by using NMF. The perceptual structural degradation of the distorted image can be evaluated by comparing the difference between the features of the reference and distorted images extracted by the feature detector. In addition, considering the effect of visual attention on human perception, the saliency detection model is used to extract the visually important region.

2.2. Binocular interactions

In addition to monocular perception, we also need to consider binocular interactions in SIQA method. When different images are presented to the left and right eyes, only a single, combined "cyclopean" is perceived. Some biological models have been proposed to describe binocular interactions. Levelt [38] proposed an eye-weighting model for describing the binocular combination, which is,

$$f_C = \omega_1 \cdot I_L + \omega_2 \cdot I_R \tag{5}$$

where ω_1 and ω_2 are the weights of the left and right views respectively, and $\omega_1 + \omega_2 = 1$. The eye-weighting model is simple but cannot explain Fencher's paradox and the cyclopean perception [39].

To consider binocular fusion and rivalry, Ding and Sperling [40] proposed the gain-control model as

$$\begin{aligned} f_C(I_L, I_R) &= \left(\frac{1+E_L}{1+E_L+E_R}\right) \cdot I_L + \left(\frac{1+E_R}{1+E_L+E_R}\right) \cdot I_R \\ &= g^L \cdot I_L + g^R \cdot I_R \end{aligned} \tag{6}$$

where E_L and E_R are the sums of energy over all the frequency channels for the left and right views, respectively. The gain-control model can be used to describe binocular fusion and rivalry and explain the cyclopean perception in SIQA.

Here, the gain control model and difference-of-Gaussian (DOG) response are used to synthesize the cyclopean image and the visual quality of the cyclopean image is employed to measure the quality of binocular interactions.

3. The proposed SIQA method

The diagram for the proposed SIQA method is shown in Fig. 1. The proposed SIQA method includes two stages: training of the feature detector (Fig. 1(a)) and construction of SIQA (Fig. 1(b)). In the training phase, color information of stereoscopic image is considered and the NMF method is used to derive the color visual feature detector D to simulate the RF. In the construction of SIQA phase, we consider monocular perception as well as binocular interactions. For monocular perception, a saliency detection model is exploited to select visual saliency regions. Then, by using the feature detector D to extract local color visual features from reference and distorted image patches of visual saliency regions, the image quality of left and right views are calculated, respectively. For binocular interaction, we consider binocular fusion and rivalry by constructing cyclopean image. Then, the quality of the cyclopean image is obtained by calculating similarity between the reference and distorted cyclopean image. Finally, the final quality score is obtained by using nonlinear pooling method to simulate complex relationship between monocular perception and binocular interactions.

3.1. Training phase

In this subsection, we assume that when the training sample is sufficient, the testing sample is similar to the training sample. In other words, we can get a feature extractor by enough training image patches, then the extractor can be used to extract features from the testing image patch. Therefore, a training database is first constructed, and then a sufficient number of training image patches are obtained.

To construct a training database, nine original images with different scenes are chosen from the Berkeley image segmentation database [41], as shown in Fig. 2. Since the essence of the proposed model is to measure the feature similarity between the reference and distorted images based on NMF, we only consider the color visual feature detector from the original images. Note that, the proposed SIQA method does not highly depend on the training database. The specific influences of the training database are discussed in the next section.

After selecting the training database, numerous overlapping image patches with the size of $m \times m$ are randomly taken from the each training image. In the implementation, 2000 patches are randomly selected for each training image (the number of image patch equals $n = 2000 \times 9 = 18,000$). Then, each patch is vectorized into a sample vector by scanning the values in the patch row-by-row and channel-by-channel. Here, since color information is considered, the length K of the vector is $K = m \times m \times 3$. Thus, all the sample vectors form a sample matrix, $X = [x_{ij}] = [X_1, X_2, \ldots, X_n]$, where each column $X_i \in R^{K\times 1}$ contains K pixels. In this paper, we set $m = 8$.

In the cerebral cortex, human has parts-based perception. In mathematics, NMF method can be used to simulate this mechanism well. Here, NMF method is used to decompose the matrix X into a non-negative basis matrix and a coding matrix [35]. Specifically, for the non-negative matrix X, NMF aims to find two non-negative matrices, $W = [w_{ik}] = [W_1, W_2, \ldots, W_r] \in R^{K\times r}$ and $S = [s_{ik}] = [S_1, S_2, \ldots, S_n] \in R^{r\times n}$ such that:

$$X = WS \tag{7}$$

where $r > 0$ is chosen to be smaller than n or K so that W and S are smaller than the original matrix X.

To find an approximate factorization $X = WS$, the Frobenius norm is used to construct the cost function,

$$E = \min_{W\geqslant 0, S\geqslant 0} \|X - WS\|_F^2 \tag{8}$$

In this paper, we use the Lee-Seung multiplicative iterative algorithm, and the specific process of the algorithm is shown in Table 1.

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

After the above processing, each patch T_i can be represented as a linear combination of a set basis function, e.g.,

$$T_i = WF_i \tag{11}$$

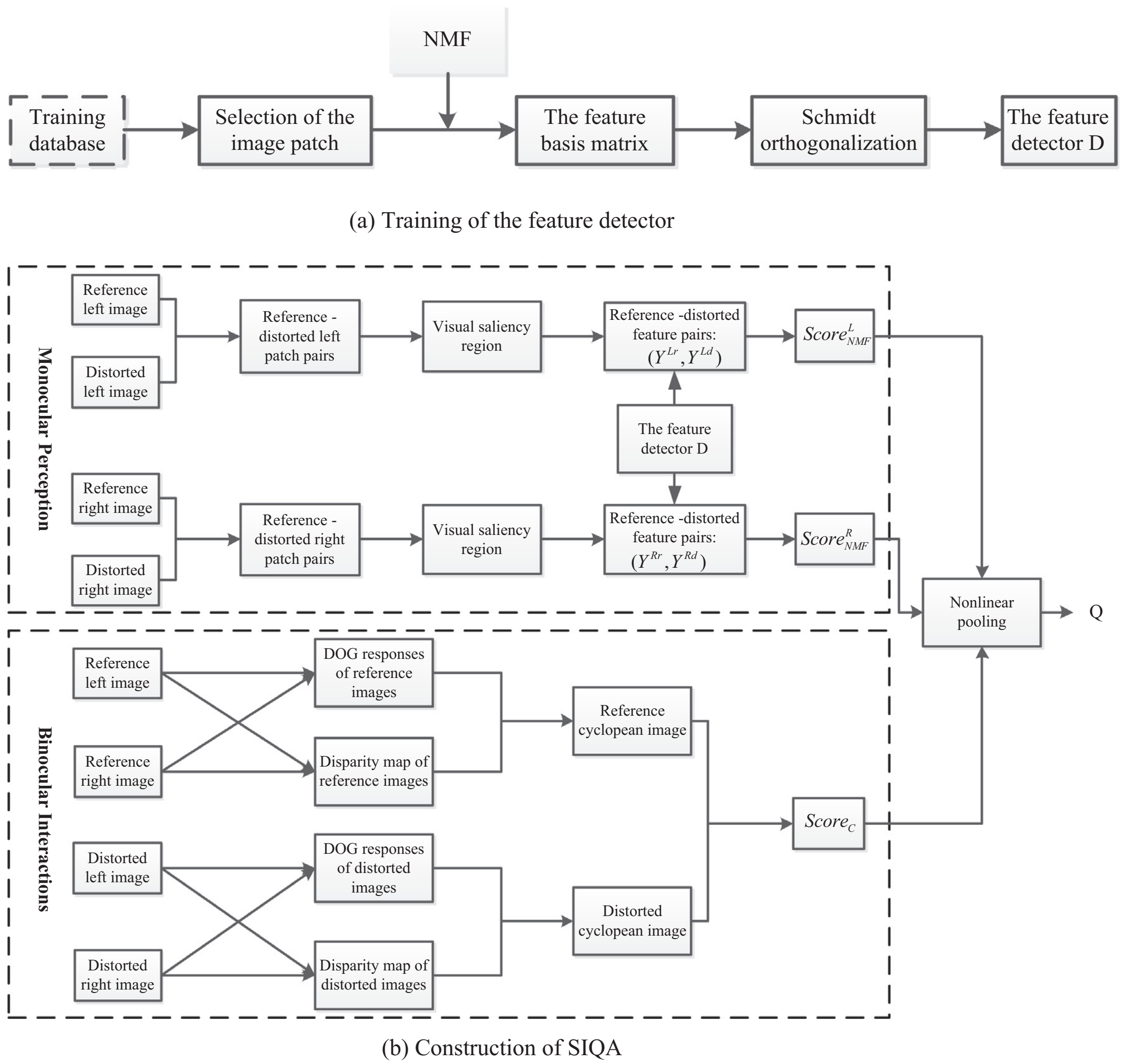

Fig. 1. Proposed SIQA method. (a) Training of the feature detector. (b) Construction of SIQA.

where F_i is a feature vector after dimensionality reduction, and its dimension is r.

Thus, using the generalized inverse matrix $W^+ = (W^TW)^{-1}W^T$, the feature vector can be obtained as follows:

$$F_i = W^+T_i \tag{12}$$

Here, Schmidt orthogonalization is used to remove correlations between the feature bases and a new orthonormal basis matrix is obtained. The specific process is as follows,

$$\begin{aligned} \alpha_1 &= W_1, \\ \alpha_2 &= W_2 - \frac{W_2^T\alpha_1}{\alpha_1^T\alpha_1}\alpha_1, \ldots, \\ \alpha_k &= W_k - \frac{W_k^T\alpha_1}{\alpha_1^T\alpha_1}\alpha_1 - \cdots - \frac{W_k^T\alpha_{k-1}}{\alpha_{k-1}^T\alpha_{k-1}}\alpha_{k-1} \end{aligned} \tag{13}$$

Thus, the new feature basis matrix $W^{orth} = [\alpha_1, \alpha_2, \ldots, \alpha_r]$ is obtained, e.g., $(W^{orth})^T W^{orth} = I$. Further, the generalized inverse matrix of W^{orth} is as follows,

$$D = ((W^{orth})^T W^{orth})^{-1}(W^{orth})^T = (W^{orth})^T \tag{14}$$

Finally, Eq. (12) can be rewritten as follows,

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

where D is the feature detector.

Fig. 3 shows an illustration of the feature detector, D, which includes three RGB sections. Each element of the feature detector is quantified into gray scale value. The feature detector $D \in R^{r\times 192}$ can transform each patch vector $T_i \in R^{192\times 1}$ into a parts-based feature vector with length r. Each row element of D consists of three sections corresponding to RGB channels, all row elements form three sets, and each set contains r patches of size 8×8. From Fig. 3, it is clear that these patches are similar to the RF with different directions. Investigating its reason NMF can learn parts-based visual properties which is similar to the RF. Therefore, the feature detector D can capture visual response of the RF.

3.2. Construction of SIQA

In this subsection, we consider monocular perception and binocular interactions to model the proposed SIQA method and use SVR to describe the nonlinear relationship between monocular perception and binocular interactions.

3.2.1. Monocular perception

3.2.1.1. Selection of visual important regions. It is well-known that the human focuses on some important regions when viewing an

Fig. 2. Selected original images for the training process of the proposed model.

Table 1
Description of the algorithm.

The Lee-Seung multiplicative iterative algorithm
Input: The matrix X and the parameter r.
Output: The basis matrix W and coding matrix S.
Algorithm:
Step 1: Initialize $w_{ia}^1 > 0$ and $s_{bj}^1 > 0$. Set $T_{iter} = 100$.
Step 2: For $k = 1, 2, \ldots, T_{iter}$, do
$w_{ia}^{k+1} = w_{ia}^k \frac{(X(S^k)^T)_{ia}}{(W^k S^k (S^k)^T)_{ia}}$ (9)
$s_{bj}^{k+1} = s_{bj}^k \frac{((W^{k+1})^T X)_{bj}}{((W^{k+1})^T W^{k+1} S^k)_{bj}}$ (10)
End
Step 3: Return the basis matrix W and coding matrix S.

image. For IQA, the visual quality of the image is directly related to the visually important regions. Here, we use the saliency detection model to select the visually important regions of an image.

In the following section, the left and right images are processed in the same way, and the left image is taken as an example to describe the algorithm processing. We use the saliency detection model [42] to obtain the saliency maps of the left-reference image I_{ref}^L and its corresponding distorted image I_{dis}^L, respectively. Let M_{ref}^L and M_{dis}^L denote the saliency maps of I_{ref}^L and I_{dis}^L, respectively. A maximum combined saliency map M^L at each point of M_{ref}^L and M_{dis}^L is defined as follows,

$$M^L = \max(M_{ref}^L, M_{dis}^L) \tag{16}$$

the significance of M^L is to obtain the maximum saliency map of the left reference image and the corresponding distorted image.

To obtain visually import regions of the left image, I_{ref}^L, I_{dis}^L and M^L are segmented into overlapping patches with the same size of 8×8, and these patches are vectorized and arranged in columns of the matrices X^{Lr}, X^{Ld} and S^L respectively.

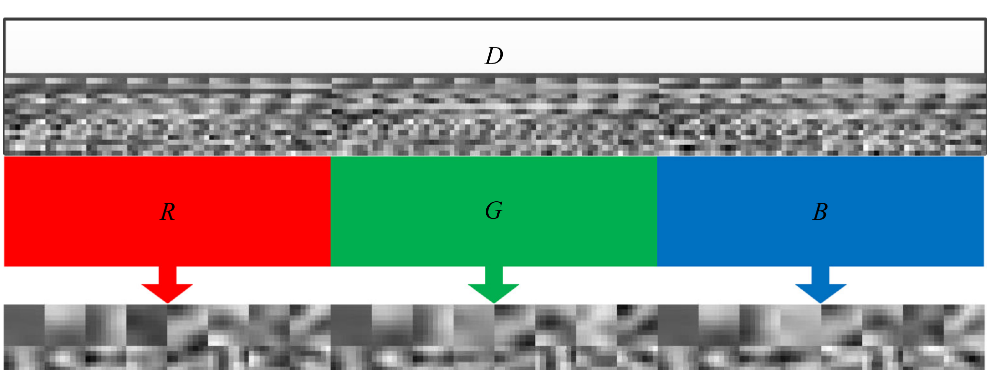

Fig. 3. An illustration for the feature detector, D, which includes three sections corresponding to RGB channels. Patches of each section are similar to the RF.

Let S_j^L denote the j-th column of the matrix S^L, i.e., the j-th patch of the maximum combined saliency map M^L. Therefore, the saliency value of the j-th patch is obtained as follows,

$$d_j = \sum_{i=1}^{N} S_{ij}^L \tag{17}$$

where S_{ij}^L is the saliency value of the i-th pixel in the j-th patch and N denotes the number of pixels in each patch. Here, $N = 8 \times 8 = 64$. d_i^* is obtained by sorting all of d_j from large to small as follows,

$$d_1^* \geqslant d_2^* \geqslant \cdots \geqslant d_k^* \tag{18}$$

where k denotes the number of image patches.

Let t_1 denote the number of selected maximum saliency patches, $t_1 = \lambda_1 \cdot k$, where $\lambda_1 \in (0, 1]$ is the ratio coefficient of the selected maximum saliency patches. Thus, the reference left image patch Y^{Lr} and the distorted left image patch Y^{Ld} corresponding to the maximum combined saliency patch d_i^* of the former t_1 are selected, i.e.,

$$(Y^{Lr}, Y^{Ld}) = \{(X_j^{Lr}, X_j^{Ld}) \big| j \in label\{d_i^* \text{ of the former } t_1\}\} \tag{19}$$

After the above processing, the final visually important left reference-distorted patch pairs are selected. Similarly, the visually important right reference-distorted patch pairs (Y^{Rr}, Y^{Rd}) can also be determined. Based on [43], we select 10% highest saliency patches, that is, we set $\lambda_1 = 0.1$ for LIVE 3D databases and NBU 3D IQA database. Of course, we can choose three different optimal λ_1 values for LIVE 3D databases and NBU 3D IQA database, respectively.

3.2.1.2. Feature extraction. After the selection processing, the feature detector D is used to extract the feature vectors of the left and right views, respectively:

$$u_i^L = D \times Y_i^{Lr} \tag{20}$$

and

$$v_i^L = D \times Y_i^{Ld} \tag{21}$$

Thus, the feature vectors $u_i^L \in R^{r\times 1}$ and $v_i^L \in R^{r\times 1}$ form two matrices, $U^L \in R^{r\times t_1}$ and $V^L \in R^{r\times t_1}$. Since $r < m$, $u_i^L \in R^{r\times 1}$ and $v_i^L \in R^{r\times 1}$ can be represented by the active neurons, and the values of these feature vectors represent the level of activity of the neurons. Thereby, the visual responses of the image patches are estimated by using the feature vectors.

3.2.1.3. Feature similarity index. To quantify the perceived quality of the image, we compare the feature vector matrices U^L and V^L. Therefore, the feature similarity $Score_{NMF}^L$ of the left distorted image among the feature vectors is defined as

$$Score_{NMF}^L = 1 - \frac{1}{r \cdot t_1} \sum_{i=1}^{r} \sum_{j=1}^{t_1} \frac{(u_{ij}^L - v_{ij}^L)^2 + C}{(u_{ij}^L)^2 + (v_{ij}^L)^2 + C} \tag{22}$$

where u_{ij}^L and v_{ij}^L denote the values of the i-th row and the j-th column in $U^L \in R^{r\times t_1}$ and $V^L \in R^{r\times t_1}$, respectively. The parameter r denotes the dimension of feature vectors $u_i^L \in R^{r\times 1}$ and $v_i^L \in R^{r\times 1}$. t_1 denotes the number of selected image patches. C is a constant to avoid the denominator being zero. Here, we set $C = 0.08$. Similarly, the feature similarity index $Score_{NMF}^R$ of the right distorted image can be obtained.

3.2.2. Binocular interactions

In this subsection, we consider the influence of binocular interactions on SIQA. We use gain control model and DOG responses to synthesize reference and distorted cyclopean images. Then, we calculate SSIM index to obtain the quality of distorted cyclopean image.

To measure binocular interactions accurately, we reproduce the cyclopean view in accordance with human perception. Generally, the localized linear model and Gabor filter bank are used to synthesize a cyclopean image [25]. Here, different from Chen's method, we adopt gain control model and DoG responses to synthesize reference and distorted cyclopean images. Specific methods are described below.

The disparity map between the left and right views is first estimated by using a stereoscopic matching method [44]. The disparity value is used to reflect the depth information of the stereoscopic image. With the disparity map, Eq. (6) of the gain control model can be rewritten by

$$\begin{aligned} f_C(I_L, I_R) = & \left(\frac{1 + E_L(x,y)}{1 + E_L(x,y) + E_R(x+d,y)}\right) \cdot I_L \\ & + \left(\frac{1 + E_R(x+d,x)}{1 + E_L(x,y) + E_R(x+d,y)}\right) \cdot I_R \end{aligned} \tag{23}$$

where d is the disparity value, and $E_L(x,y)$ and $E_R(x+d,y)$ are the sums of the total visually weighted contrast energies of the left and right views, respectively.

From Eq. (23), it is found that $E_L(x,y)$ and $E_R(x+d,y)$ must be calculated to synthesize reference and distorted cyclopean images. Here, we use DOG responses of all frequency channels to calculate $E_L(x,y)$ and $E_R(x+d,y)$. DOG decomposition I_{DOG}^L of the left view is defined as follows

$$I_{DOG}^L = (G(k\sigma) - G(\sigma)) * I_L \tag{24}$$

where $G(\cdot)$ is the Gaussian kernel function with standard deviation σ, and k is a constant.

Therefore, according to Eq. (24), the different frequency channel of I_{DOG}^L can be defined as follows

$$D_i^L = (G(k^i\sigma) - G(k^{i-1}\sigma)) * I_L, \quad i = 1, \ldots, n \tag{25}$$

Thus, the vector $D(I_L)$ which represents different frequency responses of the left image I_L is defined by

$$D(I_L) = (D_0^L, D_1^L, \ldots, D_n^L) \tag{26}$$

where $D_0^L = G(\sigma) * I_L - I_L$. Using Eq. (26), the vector $D(I_{ref}^L)$ of the reference left image can be defined. Similarly, the vector $D(I_{ref}^R)$ of the reference right image can also be defined.

According to $D(I_{ref}^L)$ and $D(I_{ref}^R)$, $E_{ref}^L(x,y)$ and $E_{ref}^R(x+d,y)$ can be calculated as

$$E_{ref}^L(x,y) = \|D(I_{ref}^L)\|_2^2 \tag{27}$$

and

$$E_{ref}^R(x+d,y) = \|D(I_{ref}^R)\|_2^2 \tag{28}$$

Then, using Eq. (23), a single cyclopean image I_{ref}^C of the reference stereoscopic image can be calculated. Similarly, a distorted cyclopean image I_{dis}^C can be obtained.

Finally, the quality value $Score_C$ of the distorted cyclopean image is obtained by using the SSIM index to measure the similarity between the reference cyclopean image I_{ref}^C and the distorted cyclopean image I_{dis}^C.

3.2.3. Nonlinear pooling

After obtaining monocular quality $Score^{L}_{NMF}, Score^{R}_{NMF}$ and quality $Score_C$ of the cyclopean image, how to measure the relationship between monocular perception and binocular interactions remains a problem. As discussed in Section 2, we use SVR to measure the nonlinear relationship between monocular perception and binocular interactions.

The overall quality score, *Q*, of stereoscopic image can be calculated by using three quality scores $\{Score^{L}_{NMF}, Score^{R}_{NMF}, Score_C\}$ and a prediction function *f*, that is,

$$Q = f(Score^{L}_{NMF}, Score^{R}_{NMF}, Score_C) \tag{29}$$

where $f : R^3 \to R$ is the quality prediction function trained in advance using SVR. Here, SVR trained the mapping function *f* which takes three quality scores $\{Score^{L}_{NMF}, Score^{R}_{NMF}, Score_C\}$ and produces output as a corresponding difference mean opinion score (DMOS). After obtaining the mapping function *f*, we use this function *f* to predict the visual quality of the stereoscopic image.

4. Experimental results and analyses

4.1. Databases and performance measures

(1) The LIVE 3D Phase I database [13] consists of 365 symmetrically distorted stereoscopic images generated from 20 reference stereoscopic pairs by corrupting them with five different distortion categories: JPEG 2000 (JP2K), the JPEG compression standards, additive white Gaussian noise (WN), Gaussian blur (Gblur) and a fast-fading (FF) model based on the Rayleigh fading channel.
(2) The LIVE 3D Phase II database [14] consists of 120 symmetrically distorted stereoscopic images and 240 asymmetrically distorted stereoscopic images generated from 8 reference stereoscopic pairs. It includes the same distortion categories as Phase I. These types of distortions are symmetrically and asymmetrically applied to the left and right reference stereoscopic images at different levels.
(3) The NBU 3D IQA database [18] consists of 312 distorted stereoscopic images generated from 12 reference stereoscopic pairs. Five types of distortions, JP2K, JPEG, WN, Gblur and H.264, are symmetrically applied to the left and right reference stereoscopic pairs at various levels.

To benchmark the performance of SIQA methods, three indexes are used: the Spearman rank order correlation coefficient (SROCC), the Pearson linear correlation coefficient (PLCC), and the root mean squared error (RMSE). A perfect matching between the objective and subjective scores will give $SROCC = PLCC = 1$ and $RMSE = 0$. For the nonlinear regression, the 4-parameter logistic function is defined as follows [45]:

$$DMOS_P = \frac{\beta_1 - \beta_2}{1 + \exp\left(-\frac{x-\beta_3}{|\beta_4|}\right)} + \beta_2 \tag{30}$$

where β_1, β_2, β_3 and β_4 are parameters of the regression model.

In the experiments, we randomly select 80% of a database content for training and the remaining 20% for testing. Specifically, 1000 randomly chosen training and testing sets are obtained, and the average PLCC, SROCC and RMSE values are regarded as the final result.

4.2. Overall assessment performance

In this subsection, to make a comprehensive analysis on the proposed SIQA method, we compare the proposed SIQA method with some existing SIQA metrics—five luminance information plus energy response based SIQA methods (FI-PSNR, FI-SSIM, FI-VIF, Bensalma's method [31] and Shao's method [32]) and three luminance plus disparity based SIQA methods (Benoit's method [15], You's method [23] and Chen's method [25])—on the overall distortions of the three benchmark databases in terms of SROCC, PLCC, and RMSE.

The values of SROCC, PLCC and RMSE on the LIVE 3D phase I, the LIVE 3D phase II and the NBU 3D IQA database are listed in Table 2. From Table 2, it can be seen that Chen's method and Shao's method are reasonable good for these three databases. The possible reason is that "cyclopean" based (Chen's method) and sparse representation based methods (Shao's method) are highly in line with human visual perception. For the proposed SIQA method, it has best performances across all these three databases. Since NMF is able to learn local color visual characteristics and both monocular perception and binocular interactions are consider, moreover, SVR is used to simulate the nonlinear relationship between monocular perception and binocular interactions, the proposed SIQA method can achieve much higher consistency with human visual perception than the other SIQA methods.

To further verify the stability of the proposed SIQA method, the variances of SROCC and PLCC are calculated and listed in Table 3. From Table 3, it can be concluded that the proposed SIQA method remains stable between trials.

4.3. Performance on individual distortion types

In this subsection, we comprehensively compare the proposed SIQA method with the eight SIQA methods to measure a SIQA method's ability degraded by specific types of distortion. Values of PLCC and SROCC are respectively listed in Tables 4 and 5 where the highest performance have been highlighted in boldface. From Tables 4 and 5, it can be seen that the proposed SIQA method and Shao's method are among the top 9 times in terms of PLCC and SROCC, followed by FI-VIF (5 times), You's method (3 times) and Chen's method (2 times). A possible explanation is that the proposed SIQA method based on NMF and Shao's method based on sparse coding have impressive consistency with human perception. However, it should be noted that the proposed SIQA method is very prominent for Gblur distortion because the localized features cannot reflect the changes of image quality for this distortion.

To further evaluate the predictive performance of the proposed SIQA method for both symmetrically and asymmetrically distorted stereoscopic images, we compare the proposed SIQA method with PMFS metric [27] on LIVE 3D Phase II database. The values of PLCC, SROCC and RMSE of each distortion types are listed in Table 6. From Table 6, it can be observed that the proposed SIQA method shows the highest performance on JPEG, JP2K and FF distortions, respectively. In addition, the proposed SIQA method is also superior to the PMFS metric in terms of overall performance. In summary, the proposed SIQA method is an accurate, monotonic, and consistent objective SIQA method.

4.4. Influence of the parameter

In this subsection, we discuss the impact of the parameter *r* which controlled the number of the feature basis in the feature detector. Particularly, the parameter *r* is similar to the number of activated cells in the RF. Without losing generality, we only conduct the performances on the LIVE 3D Phase I database to discuss the influence of the parameter *r* selection. In the experiment, we consider nine values of the parameter: $r \in \{8, 10, 12, 14, 16, 18, 20, 22, 24\}$. Performance effects of the parameter *r* are shown in Fig. 4. As shown in Fig. 4, the highest

Table 2
Performance of the proposed SIQA method and the other eight methods in terms of SROCC, PLCC and RMSE on the three databases (cases in bold denote best performance).

		FI-PSNR	FI-SSIM	FI-VIF	Bensalma and Larabi [31]	Benoit et al. [15]	You et al. [23]	Chen et al. [25]	Shao et al. [32]	Proposed
LIVE I	SROCC	0.8599	0.8606	0.9188	0.8747	0.8901	0.9247	0.9157	0.9251	**0.9336**
	PLCC	0.8645	0.8699	0.9222	0.8874	0.8899	0.9303	0.9167	0.9350	**0.9459**
	RMSE	8.2424	8.0874	6.3423	7.5585	7.4786	6.0161	6.5503	5.8155	**5.2763**
LIVE II	SROCC	0.6375	0.6795	0.7213	0.7513	0.7475	0.7206	0.9013	0.8494	**0.9030**
	PLCC	0.6584	0.6844	0.7234	0.7699	0.7642	0.7744	0.9065	0.8628	**0.9162**
	RMSE	8.4956	8.2295	7.7936	7.2035	7.2806	7.1413	4.7663	5.7058	**4.5233**
NBU	SROCC	0.8889	0.9093	0.8463	0.9381	0.8812	0.7324	0.9093	**0.9411**	0.9206
	PLCC	0.9077	0.9143	0.8455	0.9367	0.8760	0.7346	0.9083	**0.9413**	0.9330
	RMSE	7.2081	6.9565	9.1739	6.0172	8.2864	11.6556	7.1852	**5.7999**	6.3630
Average	SROCC	0.7954	0.8164	0.8288	0.8547	0.8396	0.7925	0.9087	0.9052	**0.9190**
	PLCC	0.8102	0.8228	0.830367	0.8646	0.8433	0.8131	0.9105	0.9130	**0.9317**
	RMSE	7.9820	7.7578	7.769933	6.9264	7.6818	8.2710	6.1672	5.7737	**5.3875**

Table 3
Stability of trained model in the LIVE 3D Phase I, the LIVE 3D Phase II and NBU 3D IQA database.

	Variance (SROCC)	Variance (PLCC)
The LIVE 3D Phase I	0.0144	0.0115
The LIVE 3D Phase II	0.0233	0.0188
NBU	0.0182	0.0186

value of PLCC and SROCC is obtained when the parameter r equals to 16. As can be seen from the figure, the curves are monotonically increasing and decreasing monotonically before and after the parameter r equals to 16 fore both PLCC and SROCC. Therefore, we set $r = 16$ in the proposed SIQA method.

4.5. Contributions of each part in the proposed SIQA method

Three quality scores $\{Score^{L}_{NMF}, Score^{R}_{NMF}, Score_{C}\}$ used in the proposed SIQA method is composed of two parts: visual quality of monocular perception ($Score^{L}_{NMF}$ and $Score^{R}_{NMF}$) and visual quality of binocular interactions ($Score_{C}$). It is necessary to explore how the final perceived quality is impacted by each part in the proposed SIQA method. Therefore, we designed three different methods for performance comparison, denoted by method-A ($Score^{L}_{NMF}$ and

Table 4
Performance comparison of nine metrics on each distortion types in terms of PLCC.

	Criteria	FI-PSNR	FI-SSIM	FI-VIF	Bensalma and Larabi [31]	Benoit et al. [15]	You et al. [23]	Chen et al. [25]	Shao et al. [32]	Proposed
LIVE I	JPEG	0.2866	0.2741	0.6545	0.3803	0.5766	0.6333	0.6344	0.5200	**0.7074**
	JP2K	0.8381	0.8210	**0.9421**	0.8389	0.8859	0.9410	0.9164	0.9213	0.9412
	WN	0.9280	0.9250	0.9310	0.9147	0.9354	0.9351	0.9436	0.9448	**0.9503**
	Gblur	0.9475	0.9080	0.9573	0.9369	0.9217	0.9545	0.9417	0.9592	**0.9681**
	FF	0.7086	0.7297	0.7572	0.7339	0.7477	0.8589	0.7580	0.8594	**0.8677**
LIVE II	JPEG	0.6124	0.5486	0.8906	0.8577	0.5328	0.6741	0.8422	0.7472	**0.9242**
	JP2K	0.7457	0.7191	0.9164	0.6667	0.6467	0.7320	0.8426	0.7823	**0.9419**
	WN	0.9150	0.9139	0.8981	0.9436	0.8610	0.5464	**0.9602**	0.9464	0.9174
	Gblur	0.7083	0.7250	0.8993	0.9077	0.8814	**0.9763**	0.9650	0.9580	0.9204
	FF	0.7025	0.7342	0.7574	0.9097	0.8472	0.8561	0.9097	0.9046	**0.9194**
NBU	JPEG	0.9433	0.9420	**0.9467**	0.8950	0.8835	0.7266	0.9011	0.9279	0.9200
	JP2K	0.9420	0.9441	0.9259	0.9510	0.8900	0.6811	0.9102	**0.9529**	0.9339
	WN	0.9158	0.9320	0.9551	0.9333	0.9386	0.8770	0.9498	**0.9626**	0.9532
	Gblur	0.9596	0.9578	0.9696	0.9596	0.9324	0.8258	0.9528	**0.9808**	0.9513
	H. 264	0.9640	**0.9665**	0.9672	0.9547	0.8385	0.7044	0.9337	0.9619	0.9355

Table 5
Performance comparison of nine metrics on each distortion types in terms of SROCC.

	Criteria	FI-PSNR	FI-SSIM	FI-VIF	Bensalma and Larabi [31]	Benoit et al. [15]	You et al. [23]	Chen et al. [25]	Shao et al. [32]	Proposed
LIVE I	JPEG	0.2070	0.2047	0.6002	0.3283	0.4983	0.6008	0.5582	0.4951	**0.6092**
	JP2K	0.8388	0.8222	**0.9125**	0.8170	0.8730	0.9051	0.8956	0.8945	0.8842
	WN	0.9284	0.9282	0.9335	0.9055	0.9369	0.9403	**0.9481**	0.9405	0.9246
	Gblur	0.9345	0.8788	0.9329	0.9157	0.8802	0.9300	0.9261	**0.9403**	0.9221
	FF	0.6581	0.6866	0.7497	0.6500	0.6242	**0.8030**	0.6879	0.7963	0.7957
LIVE II	JPEG	0.6129	0.5641	0.8768	0.8461	0.5078	0.5229	0.8396	0.7330	**0.8888**
	JP2K	0.7193	0.7003	**0.9212**	0.8038	0.6325	0.7309	0.8334	0.7845	0.9093
	WN	0.9073	0.9091	0.9341	0.9386	0.8569	0.4820	0.9554	**0.9651**	0.8725
	Gblur	0.7112	0.7387	0.8868	0.8838	0.8545	**0.9227**	0.9096	0.9204	0.8645
	FF	0.7012	0.7350	0.7586	0.8743	0.8319	0.8392	0.8890	**0.8905**	0.8865
NBU	JPEG	0.9390	0.9456	**0.9514**	0.9148	0.8889	0.7606	0.9133	0.9346	0.9167
	JP2K	0.9469	0.9439	0.9282	0.9508	0.8951	0.6754	0.9145	**0.9516**	0.9231
	WN	0.8604	0.9163	0.9233	0.9157	0.9278	0.8341	0.9149	**0.9453F**	0.9355
	Gblur	0.9526	0.9692	0.9737	0.9560	0.9304	0.8175	0.9473	**0.9769**	0.9361
	H. 264	**0.9555**	0.9536	0.9513	0.9379	0.8402	0.6685	0.9040	0.9553	0.9253

$Score^{R}_{NMF}$ are used with SVR procedure), method-B (only $Score_C$ is used) and method-C (the proposed SIQA method), respectively. Table 7 shows the comparison results of these three methods. From Table 7, it is easy to see that the best result is obtained when both monocular perception and binocular interactions are considered.

4.6. Performance impact of the training database

Since feature vectors in the proposed SIQA method are extracted by using the feature detector from a selected training database, it is necessary to discuss whether the performance depends on a training database. To test robustness of the proposed SIQA method, we construct two different training databases randomly selected from the IVC MIQA database [46] and the Toyama-MICT (TOY) MIQA database [47], respectively (denoted as training database 1 and training database 2 in Fig. 5) and compare the evaluation performance using these training databases to learn the feature detector D. Table 8 shows SROCC, PLCC and RMSE results with different training databases. From Table 8, we can see that, no matter which training database is used, the difference in the results is small. This means that the evaluation performance

Table 6
Performance comparison of the metrics on LIVE 3D Phase II database (the case in bold: the best performance).

		JPEG	JP2K	WN	Gblur	FF	ALL
PMFS[27]	PLCC	0.829	0.806	**0.958**	**0.971**	0.912	0.902
	SROCC	0.805	0.788	**0.945**	**0.909**	0.886	0.892
	RMSE	4.112	5.815	**3.267**	**3.290**	4.693	4.894
Proposed	PLCC	**0.924**	**0.942**	0.917	0.920	**0.919**	**0.916**
	SROCC	**0.889**	**0.909**	0.873	0.865	**0.887**	**0.903**
	RMSE	**3.518**	**4.021**	4.184	3.768	**4.681**	**4.523**

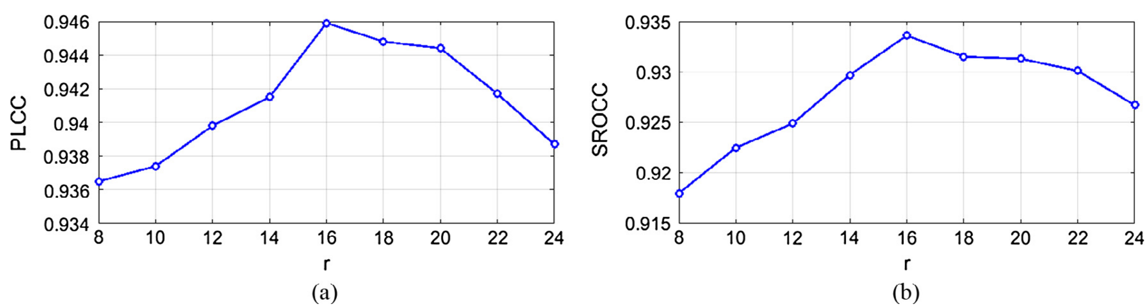

Fig. 4. Performance influence of the parameter r on the LIVE 3D Phase I database. (a) Values of PLCC influenced by r. (b) Values of SROCC influenced by r.

Table 7
Comparison of SROCC, PLCC and RMSE for different schemes (the case in bold: the best performance).

		Method-A	Method-B	Method-C
LIVE I	SROCC	0.9278	0.9180	**0.9336**
	PLCC	0.9378	0.9276	**0.9459**
	RMSE	5.5864	6.1242	**5.2763**
LIVE II	SROCC	0.8631	0.8282	**0.9030**
	PLCC	0.8706	0.8386	**0.9162**
	RMSE	5.5538	6.1483	**4.5233**
NBU	SROCC	0.8992	0.9072	**0.9206**
	PLCC	0.9057	0.9188	**0.9330**
	RMSE	7.1495	6.9413	**6.3630**

Fig. 5. Two different training databases. (a) Training database 1 randomly selected from the IVC MIQA database. (b) Training database 2 randomly selected from the TOY MIQA database.

Table 8
Comparison of SROCC, PLCC and RMSE for two training databases (the case in bold: the best performance).

		Training database 1	Training database 2	Proposed database
LIVE I	SROCC	0.9247	0.9261	**0.9336**
	PLCC	0.9382	0.9387	**0.9459**
	RMSE	5.5876	5.5792	**5.2763**
LIVE II	SROCC	0.8940	0.8902	**0.9030**
	PLCC	0.9070	0.9038	**0.9162**
	RMSE	4.6823	4.7997	**4.5233**
NBU	SROCC	0.9058	0.9082	**0.9206**
	PLCC	0.9245	0.9236	**0.9330**
	RMSE	6.4556	6.5183	**6.3630**

of the proposed SIQA method is usually stable on different databases and has robustness in terms of the change of the training database.

5. Conclusions

In this work, we propose a novel method for objective stereoscopic image quality assessment (SIQA) by learning non-negative matrix factorization (NMF)-based color visual characteristics and considering binocular interactions. Since the NMF method can reflect parts-based perception of the human visual system, the NMF method is first used to learn a feature detector by considering color information from the training images. Then, in construction of SIQA phase, visual attention is considered and the feature detector is used to extract localized color visual features for monocular perception. After obtain the quality scores of the left and right views, the cyclopean is considered for binocular interactions. The gain control model and the DOG responses of the left and right views are used to form the cyclopean image. Furthermore, quality score of the distorted cyclopean image is calculated. Finally, nonlinear pooling method is used to reflect nonlinear relationship between monocular perception and binocular integration, and the final quality score is obtained by integrating three quality scores. In the future, we will extend the proposed SIQA method to measure the quality of stereoscopic video sequences.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant U1301257, Grant 61671258, Grant 61620106012, Grant 61471348, and Grant 61501270, the National High-tech R&D Program of China (2015AA015901), and in part by Zhejiang Provincial Natural Science Foundation of China under Grant LY15F010005, Natural Science Foundation of Ningbo under Grant No. 2016A610071, and in part by the K.C. Wong Magna Fund in Ningbo University.

References

[1] J. Han, X. Ji, X. Hu, et al., Representing and retrieving video shots in human-centric brain imaging space, IEEE Trans. Image Process. 22 (7) (2013) 2723–2736.
[2] Z. Zhang, D. Tao, Slow feature analysis for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 34 (3) (2012) 436–450.
[3] M. Wang, B. Ni, X. Hua, T.S. Chua, Assistive tagging: a survey of multimedia tagging with human-computer joint exploration, ACM Comput. Surv. 44 (4) (2012) Article 25.
[4] L. Coria, D. Xu, P. Nasiopoulos, Quality of experience of stereoscopic content on displays of different sizes: a comprehensive subjective evaluation, in: Proc. Int. Conf. Consum. Electron., 2011, pp. 755–756.
[5] Y. Lv, M. Yu, G. Jiang, et al., No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network, Signal Process. Image Commun. 47 (2016) 346–357.
[6] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook, IEEE Trans. Broadcast. 61 (2) (2015) 154–165.
[7] F. Battisti, E. Bosc, M. Carli, P.L. Callet, S. Perugia, Objective image quality assessment of 3D synthesized views, Signal Process. Image Commun. 30 (2015) 78–88.
[8] C.T. Hewage, M.G. Martini, Reduced-reference quality assessment for 3D video compression and transmission, IEEE Trans. Consum. Electron. 57 (3) (2011) 1185–1193.
[9] Q. Xu, G. Zhai, M. Liu, K. Gu, Using structural degradation and parallax for reduced-reference quality assessment of 3D images, in: Proc. 2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2014, pp. 1–6.
[10] P. Aflaki, M.M. Hannuksela, M. Gabbouj, Subjective quality assessment of asymmetric stereoscopic 3D video, SIViP 9 (2) (2015) 331–345.
[11] J. Wang, S. Wang, K. Ma, Z. Wang, Perceptual depth quality in distorted stereoscopic images, IEEE Trans. Image Process. 26 (3) (2017) 1202–1215.
[12] X. Geng, L. Shen, K. Li, P. An, A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property, Signal Process. Image Commun. 52 (2017) 54–63.
[13] A.K. Moorthy, C.C. Su, A. Mittal, A.C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Process. Image Commun. 28 (8) (2013) 870–883.
[14] M.J. Chen, L.K. Cormack, A.C. Bovik, No-reference quality assessment of natural stereopairs, IEEE Trans. Image Process. 22 (9) (2013) 3379–3391.
[15] A. Benoit, P. Le Callet, P. Campisi, R. Cousseau, Quality assessment of stereoscopic images, EURASIP J. Image Video Process. 2008 (2009) Art. ID 659024.
[16] Z.M.P. Sazzad, S. Yamanaka, Y. Kawayokeita, Y. Horita, Stereoscopic image quality prediction, in: Proc. Workshop Quality Multimedia Exper., 2009, pp. 180–185.
[17] L. Goldmann, F. De Simone, T. Ebrahimi, Impact of acquisition distortions on the quality of stereoscopic images, in: Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron., 2010, pp. 45–50.
[18] J. Zhou, G. Jiang, X. Mao, et al., Subjective quality analyses of stereoscopic images in 3DTV system, in: Proc. IEEE Vis. Commun. Image Process., 2011, pp. 1–4.
[19] J. Yang, C. Hou, Y. Zhou, Z. Zhang, J. Guo, Objective quality assessment method of stereo images, in: Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video, 2009, pp. 1–4.
[20] J. Wang, A. Rehman, K. Zeng, S. Wang, Z. Wang, Quality prediction of asymmetrically distorted stereoscopic 3D images, IEEE Trans. Image Process. 24 (11) (2015) 3400–3414.
[21] P. Gorley, N. Holliman, Stereoscopic image quality metrics and compression, in: Proc. SPIE, vol. 6803, 2008, pp. 5.
[22] J. Yang, C. Hou, R. Xu, J. Lei, New metric for stereo image quality assessment based on HVS, Int. J. Imag. Syst. Technol. 20 (4) (2010) 301–307.
[23] J. You, L. Xing, A. Perkis, X. Wang, Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis, in: Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron., 2010.
[24] S. Ryu, D.H. Kim, K. Sohn, Stereoscopic image quality metric based on binocular perception model, in: Proc. 2012 19th IEEE International Conference on Image Processing, 2012, pp. 609–612.
[25] M.J. Chen, C.C. Su, D.K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process. Image Commun. 28 (9) (2013) 1143–1155.
[26] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, Perceptual full reference quality assessment of stereoscopic images by considering binocular visual characteristics, IEEE Trans. Image Process. 22 (5) (2013) 1940–1953.
[27] W. Zhou, G. Jiang, M. Yu, F. Shao, Z. Peng, PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality assessment, IEEE Signal Process. Lett. 21 (8) (2014) 1003–1006.
[28] Y.H. Lin, J.L. Wu, Quality assessment of stereoscopic 3D image compression by binocular integration behaviors, IEEE Trans. Image Process. 23 (4) (2014) 1527–1542.
[29] D.V. Meegan, L.B. Stelmach, W.J. Tam, Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery, J. Exp. Psychol.: Appl. 7 (2) (2001) 143–153.
[30] K. Lee, S. Lee, 3D perception based quality pooling: stereopsis, binocular rivalry, and binocular suppression, IEEE J. Sel. Topics Signal Process. 9 (3) (2015) 533–545.

[31] R. Bensalma, M.C. Larabi, A perceptual metric for stereoscopic image quality assessment based on the binocular energy, Multidimens. Syst. Signal Process. 24 (2) (2013) 281–316.
[32] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, Q. Dai, Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties, IEEE Trans. Image Process. 24 (10) (2015) 2971–2983.
[33] S.E. Palmer, Hierarchical structure in perceptual representation, Cogn. Psychol. 9 (4) (1977) 441–474.
[34] E. Wachsmuth, M.W. Oram, 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) (1994) 509–522.
[35] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[36] F. Shao, W. Tian, W. Lin, G. Jiang, Q. Dai, Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions, IEEE Trans. Image Process. 25 (5) (2016) 2059–2074.
[37] Y. Yuan, Q. Guo, X. Lu, Image quality assessment: a sparse learning way, Neurocomputing 159 (2015) 227–241.
[38] W. Levelt, On Binocular Rivalry, The Hague, Mouton, The Netherlands, 1968.
[39] S. Grossberg, F. Kelly, Neural dynamics of binocular brightness perception, Vis. Res. 39 (22) (1999) 3796–3816.
[40] J. Ding, G. Sperling, A gain-control theory of binocular combination, in: Proc. of the National Academy of the United States of America, 2006, pp. 1141–1146.
[41] D. Martin, C. Fowlkes, D. Tal, et al., A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. Eighth IEEE International Conference on Computer Vision, 2001, pp. 416–423.
[42] L. Zhang, Z. Gu, H. Li, SDSP: a novel saliency detection method by combining simple priors, in: Proc. IEEE International Conference on Image Processing (ICIP), 2013, pp. 171–175.
[43] W. Xue, L. Zhang, X. Mou, Learning without human scores for blind image quality assessment, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., 2013, pp. 995–1002.
[44] A. Klaus, M. Sormann, K. Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure, in: Proc. IEEE Int. Conf., Pattern Recog., 2006, pp. 15–18.
[45] S. Ryu, K. Sohn, No-reference quality assessment for stereoscopic images based on binocular quality perception, IEEE Trans. Circ. Syst. Video Technol. 24 (4) (2014) 591–602.
[46] P. Le Callet, F. Autrusseau, Subjective Quality Assessment IR-CCyN/IVC Database, 2005. Available: <http://www.irccyn.ecnantes.fr/ivcdb/>.
[47] Horita, Yuukou, et al., MICT Image Quality Evaluation Database, 2011. <http://mict.eng.u-toyama.ac.jp/mictdb.html> (2012-06-03).