

Binocular vision based objective quality assessment method for stereoscopic images

Gangyi Jiang · Junming Zhou · Mei Yu · Yun Zhang ·
Feng Shao · Zongju Peng

Received: 23 July 2013 / Revised: 1 April 2014 / Accepted: 22 April 2014 / Published online: 18 May 2014
© Springer Science+Business Media New York 2014

Abstract Human visual system (HVS) can perceive the difference between two retinal images to create a mental image with depth perception, which is the result of two binocular interactions, i.e., binocular fusion and suppression. According to perceptual attributes of binocular interactions, in this paper, a full-reference stereoscopic image quality assessment (SIQA) method is proposed based on the mechanisms of binocular fusion and suppression. There are two kinds of information in stereoscopic images: monocular information which is visible in only one view, and binocular information which is visible in two views. HVS adopts two ways to deal with the binocular information, one is binocular fusion which deals with the information with similar content and small disparity, the other is binocular suppression which deals with the information with dissimilar content or large disparity. Therefore, the proposed method firstly divides a distorted stereoscopic image into occluded, pseudo-binocular fusion and pseudo-binocular suppression regions. Then three methods are respectively adopted to assess the quality of the three regions and the three quality indices combine into one to represent the overall quality of the distorted stereoscopic image. Finally, the predictive performance of the proposed method is evaluated and compared with existing methods in terms of consistency, cross-image and cross-distortion, and robustness. Experimental results show that the proposed SIQA method outperforms other methods and can predict human visual perception of stereoscopic image more accurately.

Keywords Stereoscopic image · Objective quality assessment · Human visual system · Binocular fusion · Binocular suppression

G. Jiang (✉) · J. Zhou · M. Yu · F. Shao · Z. Peng
Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
e-mail: jianggangyi@126.com

G. Jiang · J. Zhou
Institute of Computing Technology, Chinese Academic of Sciences, Beijing 100080, China

Y. Zhang
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

1 Introduction

With the development of stereoscopic display and network technologies, three-dimensional (3D) image processing technologies have attracted public attentions and have a widespread prospect of applications [11, 35]. 3D imaging technology presents two slightly different images of one scene to the left and the right eyes, allowing the brain to reconstruct the original scene via binocular disparity [18]. As the amount of data to store or transmit 3D images may be double or even more compared with the traditional two-dimensional (2D) image [24], a series of 3D (including stereoscopic and multi-view) video coding schemes have been proposed [5, 12, 19]. How to assess coding distortions directly affects the performance evaluation of the coding schemes. Therefore, stereoscopic (or 3D) image quality assessment (SIQA) has been becoming an important issue.

Generally, SIQA methods can be categorized as subjective and objective methods. Subjective methods have been standardized by the International Telecommunications Union (ITU) [8–10]. As observers are the ultimate receivers of visual information, the results of subjective methods are reasonable and reliable, which can be exploited to analyze the effects on the perceived quality of stereoscopic images and evaluate the predictive performance of objective methods. Jjsselsteijn [7], Tam [23], and Wang [27] analyzed the effects of camera parameters, display duration and quality-asymmetric coding on the perceived quality of stereoscopic images, respectively. However, subjective methods are not only inconvenient and time-consuming, but also cannot be performed for many scenarios, e.g., real-time video systems. Therefore, the goal of SIQA research is to design an efficient algorithm for objective assessment of image quality in a certain way that is consistent with human visual perception.

Objective SIQA methods mainly depend on a number of quantified image features to measure the quality of stereoscopic images. Most perceptual evaluations of 3D television systems are currently performed with assessment concepts based on 2D image quality assessment (2D-IQA) methods, which directly apply 2D-IQA methods to evaluate the overall quality of stereoscopic images as the mean quality of two views [30, 33]. Peak Signal-to-Noise Ratio (PSNR) was applied to predict the quality of stereoscopic images in [30], and a joint method of PSNR and Structural SIMilarity index (SSIM) [26] was applied in [33]. It is a well-known fact that these methods have not taken binocular perceptions of human visual system (HVS) into consideration. Yasakethu et al. described stereoscopic image quality as a combination of several perceptual attributes, including image quality, perceived depth, presence, naturalness and eye strain [29]. Some researchers simplified stereoscopic image quality as a combination of two dominant perceptual attributes, i.e., image quality and depth perception [1, 17, 25, 28, 31, 32]. Usually, the image quality component assesses the ordinary monoscopic perceived distortions caused by blur, noise, contrast change etc., while the depth perception component assesses the degradation of depth perception (or stereo sense) via depth or disparity map. Yang et al. assessed stereoscopic image quality from the perspective of image quality and depth perception [28]. Zhang et al. proposed a multi-view video quality assessment method based on disparity and SSIM from both image quality and depth perception [32]. Sazzad et al. proposed a no-reference quality assessment for JPEG coded stereoscopic images based on segmented local features of artifacts and disparity [17]. You et al. compared several combinations of disparity quality and image quality obtained by some well-known 2D-IQA methods and gave a good combination scheme using SSIM and Universal Quality Index (UQI) [25] in SIQA [31]. Benoit et al. also presented a linear combination scheme for disparity distortion and the measurement of 2D image quality on both views [1]. It is obvious that these approaches are the extensions or improvements of 2D-IQA methods. Even though the predictive performance of these methods is remarkably improved compared with the methods

without depth perception, they have not taken into account binocular interactions between two eyes in HVS and the relationship between image content and depth perception. As a result, the predictive results of existing objective SIQA methods are not quite in line with HVS.

Psychovisual researchers have indicated that HVS relies on binocular fusion and suppression to solve the different information from two retinal images and achieve binocular single vision [4, 14]. The difference between the two retinal images, which is called binocular disparity, contributes critically to depth perception [3]. Therefore, the generation of binocular single vision involves not only the interactions between two retinal images, but also the creation of depth perception. According to the way of HVS dealing with two retinal images, we propose a full-reference SIQA method based on binocular vision in the paper. The paper is structured as follows: Section II analyzes the problems of existing SIQA methods and presents the mechanism of HVS dealing with a stereoscopic image; Section III proposes an objective SIQA method based on binocular vision; Section IV optimizes the method and analyzes the predictive performance of the proposed method compared with existing SIQA methods; Finally, both conclusion and future work are given in Section V.

2 Problem description and motivations

It is well known that there is a binocular interval of 60–65 mm between human's two eyes, which results in two slightly different retinal images of the same scene. However, the phenomena of diplopia (one object, two images) will not occur in normal vision. This is benefited from binocular fusion and suppression, by which HVS is able to process the different information from the two retinal images and achieve depth perception in binocular single vision [4, 14]. Therefore, the processing of a stereoscopic image in HVS is the generation of binocular single vision, which involves the interactions between two views and the creation of depth perception. However, existing objective SIQA methods may not take binocular perceptual attributes into consideration and have the following two main problems:

- i) Existing SIQA methods may be unable to assess some unique distortions occurring in stereoscopic images, such as crosstalk, cardboard, keystone, etc. [3]. As these distortions involve the interactions between two views of stereoscopic images rather than influence each view separately, any combination of image quality assessment over separated views may fail to predict the perceptual gradation from distortions. In addition, existing SIQA methods mainly follow the way of traditional 2D-IQA methods, i.e., they are from the perspective of distortions and established based on statistical analysis of distortions on the perceived quality of image content and depth perception. As a result, they may predict some types of degradations well, but fail to predict other degradations. However, objective SIQA method should be suitable to any distortion rather than just one or two distortions.
- ii) They may not truly assess depth perception of stereoscopic images. It is well-known that stereoscopic images provide us not only image contents, but also depth perception. Compared with 2D-IQA methods, the quality of depth perception needs to be assessed in SIQA. However, depth perception is a mental sense yielded in brain, and it is hard to be described using mathematical models. Thus, a depth map (or disparity map) of stereoscopic images is used as a substitute for describing depth perception, which is assessed by traditional 2D-IQA methods. In fact, there are significant differences between depth map and 2D image. Depth map represents information relating to the distance from the surfaces of objects to the camera, while 2D image represents color and brightness information of objects. Therefore, 2D-IQA methods may not truly reflect the changes

of depth perception. In addition, depth perception can not exist independently, it accompanies and changes with the content of stereoscopic image. Existing SIQA methods which are thoughtless of the relationship between depth perception and image content may be unreasonable.

To solve the problems mentioned above, the mechanism of HVS dealing with two retinal images should be taken into account in SIQA design. Fig. 1 shows the way of HVS dealing with two retinal images. According to researches on human perception and psychophysics, HVS firstly searches for matching local features between two retinal images. If the two retinal images (or local features) are similar with small disparity, binocular fusion will fuse the two retinal images into a single binocular percept [22]. However, when attempting to fuse two dissimilar images or two similar images with large disparity, HVS is facing with conflicting data from two eyes—a situation known as rivalry. During rivalry, the initial perception will be diplopia or confusion. However, HVS normally cannot tolerate rivalry for long. It usually reconciles the conflicting data by suppressing one. The entire image from one retina may be suppressed, but in most cases, parts of the right eye's visual field will be suppressed while other parts of the left eye's visual field are suppressed. Sometimes HVS may solve the rivalry problem by alternately suppressing either eye.

On the other hand, in full-reference SIQA, an original stereoscopic image is used as benchmark for reference, and it is regarded as the image that is capable of presenting a perfect perception of a real-world scene for viewers. As binocular suppression usually accompanies visual discomfort and fatigue [13] while the reference stereoscopic image does not bring any visual discomfort and fatigue, we assume that binocular suppression will not occur when HVS deals with the reference stereoscopic image. Thus, the reference stereoscopic image can be categorized into two regions: monocular region dealt with by monocular vision, i.e., occluded (or disoccluded) region, and binocular region dealt with by binocular fusion. However, after the injection of distortions, the reference stereoscopic image turns into a distorted stereoscopic image. The distortions may affect not only the identification of some objects, but also depth perception of some objects. For example, some

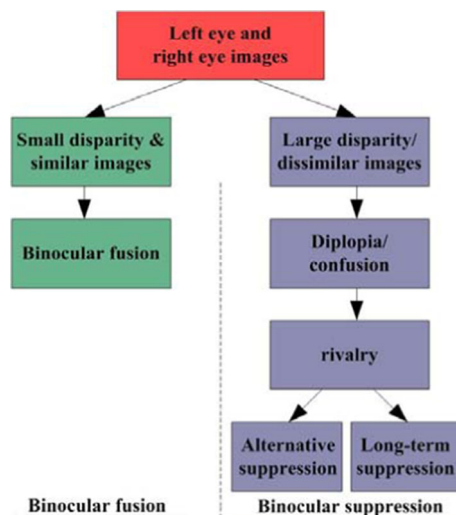


Fig. 1 The way that two retinal images may be dealt with in HVS [22]

parts may turn into a new occluded region because of the disappearance of the similarity between the parts in two views; some parts' disparities may become larger or even result in the occurrence of binocular suppression; some parts' disparities may become smaller; in other case, some occluded parts may not exist any more. Therefore, the distorted stereoscopic image may comprise of three kinds of regions: occluded region dealt with by monocular vision, binocular region dealt with by binocular fusion, and binocular region dealt with by binocular suppression.

3 The proposed binocular fusion and suppression based objective stereoscopic image quality assessment (siqa) method

As mentioned above, three regions, i.e., occluded, binocular fusion, and binocular suppression regions, may coexist in a distorted stereoscopic image. The three regions are dealt with by HVS in different ways, thus it is better to assess them in different manners. However, except the occluded region which has no corresponding region in another view, it is hard to accurately identify the binocular fusion and suppression regions. Shao et al. proposed a method to segment these three kinds of regions based on which a perceptual full-reference quality metric is further proposed [6]. In literature [6], left–right consistency check and matching error between the corresponding pixels are utilized to indentify the binocular suppression region, and the binocular fusion region is defined as the areas excluding the occluded region and binocular suppression region. After that, the local phase and local amplitude maps are extracted from the original and distorted stereoscopic images as features in quality assessment. Then, each region is evaluated independently, and all evaluation results are integrated into an overall score. In this paper, to simplify the segmentation, disparity information of the reference stereoscopic image is used as benchmark to distinguish the two regions, based on the assumption that reference stereoscopic image does not have binocular suppression regions because it does not result in any visual discomfort and fatigue which are usually brought by binocular suppression. The framework of the proposed binocular vision based SIQA method is shown in Fig. 2. A distorted stereoscopic image is firstly divided into occluded region (unmatched region) and binocular region (matching region with disparity) according to stereo matching between two views. Compared with the corresponding disparity of the reference stereoscopic image, the binocular region with smaller or invariable disparity in the distorted stereoscopic image is regarded to be dealt with by binocular fusion, while the binocular region with larger disparity is regarded to be dealt with by binocular suppression. Therefore, the binocular region in the distorted stereoscopic image is subdivided into pseudo-binocular fusion region and pseudo-binocular suppression region according to the changes of the disparity in the corresponding binocular region. Then the visual quality indices of the three regions are obtained through simulating the ways of the three regions dealt with by HVS. According to the contributions of the three regions to the overall visual quality of the distorted stereoscopic image, the weighted sum of the three quality indices are used to represent the overall visual quality of the distorted stereoscopic image.

3.1 Region classification of distorted stereoscopic image

The classification of the occluded, pseudo-binocular fusion, and pseudo-binocular suppression regions are implemented as follows.

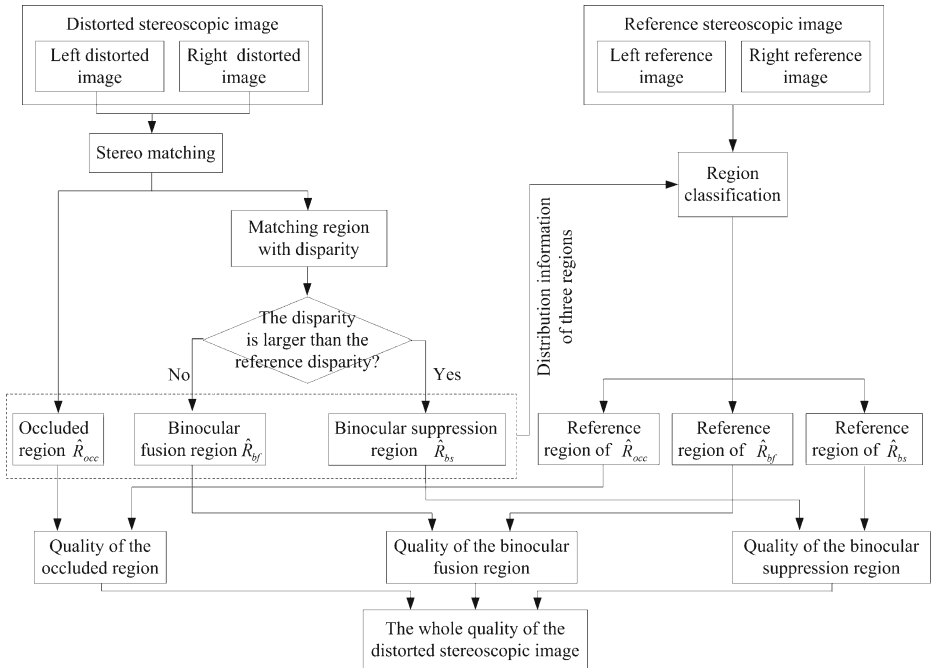


Fig. 2 The framework of the proposed SIQA method

1) Ocluded region detection: For a given distorted stereoscopic image \hat{I} with the resolution of $m \times n$, stereo matching is performed between the two views. Let $\hat{p}_{x,y}^l$ denote a pixel at the position (x, y) of the left view \hat{I}_l , $\hat{p}_{s,t}^r$ be a pixel at the position (s, t) of the right view \hat{I}_r matching with $\hat{p}_{x,y}^l$. Then there are three cases denoted as $(\hat{p}_{x,y}^l, \hat{p}_{s,t}^r)$, $(\hat{p}_{x,y}^l, \phi)$, $(\phi, \hat{p}_{s,t}^r)$, and ϕ means that there is no matching pixel in the corresponding view for the pixel in the other view. Then the occluded region \hat{R}_{occ} of the distorted stereoscopic image is a set of pixels that have no matching pixels, which can be defined as

$$\hat{R}_{occ} = \hat{R}_{occ}^l \cup \hat{R}_{occ}^r \tag{1}$$

$$\hat{R}_{occ}^l = \left\{ \hat{p}_{x,y}^l \mid \left(\hat{p}_{x,y}^l, \phi \right) \wedge \hat{p}_{x,y}^l \in \hat{I}_l, 0 \leq x < m, 0 \leq y < n \right\} \tag{2}$$

$$\hat{R}_{occ}^r = \left\{ \hat{p}_{s,t}^r \mid \left(\phi, \hat{p}_{s,t}^r \right) \wedge \hat{p}_{s,t}^r \in \hat{I}_r, 0 \leq s < m, 0 \leq t < n \right\} \tag{3}$$

where \hat{R}_{occ}^l denotes the occluded region in \hat{I}_l , and \hat{R}_{occ}^r denotes the occluded region in \hat{I}_r .

2) Pseudo-binocular suppression region detection: In the distorted stereoscopic image, the pseudo-binocular suppression region \hat{R}_{bs} is defined as a set of pixels whose disparities between the matching pixels are larger than the original disparities in the reference stereoscopic image. Let \hat{R}_{bs}^l denote the corresponding region to \hat{R}_{bs} in \hat{I}_l , \hat{R}_{bs}^r be the

corresponding region to \widehat{R}_{bs} in \widehat{I}_r . \widehat{R}_{bs} is regarded to be dealt with by binocular suppression for which the perceived quality is dominated by either \widehat{R}_{bs}^l or \widehat{R}_{bs}^r depending on the better quality of \widehat{R}_{bs}^l and \widehat{R}_{bs}^r [3], which can be defined as

$$\widehat{R}_{bs} = Sup\left\{\widehat{R}_{bs}^l, \widehat{R}_{bs}^r\right\} \tag{4}$$

$$\widehat{R}_{bs}^l = \left\{\widehat{p}_{x,y}^l \mid \widehat{p}_{x,y}^l \in \widehat{I}_l \wedge \widehat{p}_{s,t}^r \in \widehat{I}_r \wedge \left(\widehat{p}_{x,y}^l, \widehat{p}_{s,t}^r\right) \wedge \left|\widehat{d}_{x,y}^h\right| + \left|\widehat{d}_{x,y}^v\right| > \left|d_{x,y}^h\right| + \left|d_{x,y}^v\right|, 0 \leq x < m, 0 \leq y < n\right\} \tag{5}$$

$$\widehat{R}_{bs}^r = \left\{\widehat{p}_{s,t}^r \mid \widehat{p}_{x,y}^l \in \widehat{R}_{bs}^l \wedge \widehat{p}_{s,t}^r \in \widehat{I}_r \wedge \left(\widehat{p}_{x,y}^l, \widehat{p}_{s,t}^r\right), 0 \leq s < m, 0 \leq t < n\right\} \tag{6}$$

where $Sup\{\}$ denotes the way that HVS deals with the binocular region, which will be described in detail in subsection 3.2, $\widehat{d}_{x,y}^h$ and $\widehat{d}_{x,y}^v$ are the horizontal and vertical left-to-right disparities of the pixel $\widehat{p}_{x,y}^l$, respectively. $d_{x,y}^h$ and $d_{x,y}^v$ are the horizontal and vertical left-to-right disparities of the corresponding position in the reference stereoscopic image, respectively. All $\widehat{d}_{x,y}^h$, $\widehat{d}_{x,y}^v$, $d_{x,y}^h$, and $d_{x,y}^v$ are obtained through a stereo matching algorithm [21] provided by Cornell University.

3) Pseudo-binocular fusion region detection: In the distorted stereoscopic image, the pseudo-binocular fusion region \widehat{R}_{bf} is a set of pixels whose disparities between the matching pixels are smaller than or equal to the original disparities in the reference stereoscopic image. Let \widehat{R}_{bf}^l denote the corresponding region to \widehat{R}_{bf} in \widehat{I}_l , and \widehat{R}_{bf}^r be the corresponding region to \widehat{R}_{bf} in \widehat{I}_r . \widehat{R}_{bf} is regarded to be dealt with by binocular fusion for which the perceived quality is determined by both \widehat{R}_{bf}^l and \widehat{R}_{bf}^r [22], which can be defined as

$$\widehat{R}_{bf} = Fus\left\{\widehat{R}_{bf}^l, \widehat{R}_{bf}^r\right\} \tag{7}$$

$$\widehat{R}_{bf}^l = \left\{\widehat{p}_{x,y}^l \mid \widehat{p}_{x,y}^l \in \widehat{I}_l \wedge \widehat{p}_{s,t}^r \in \widehat{I}_r \wedge \left(\widehat{p}_{x,y}^l, \widehat{p}_{s,t}^r\right) \wedge \left|\widehat{d}_{x,y}^h\right| + \left|\widehat{d}_{x,y}^v\right| \leq \left|d_{x,y}^h\right| + \left|d_{x,y}^v\right|, 0 \leq x < m, 0 \leq y < n\right\} \tag{8}$$

$$\widehat{R}_{bf}^r = \left\{\widehat{p}_{s,t}^r \mid \widehat{p}_{x,y}^l \in \widehat{R}_{bf}^l \wedge \widehat{p}_{s,t}^r \in \widehat{I}_r \wedge \left(\widehat{p}_{x,y}^l, \widehat{p}_{s,t}^r\right), 0 \leq s < m, 0 \leq t < n\right\} \tag{9}$$

where $Fus\{\}$ is the way that HVS deals with the pseudo-binocular fusion region involving the processing of binocular summation [2], which will be described in detail in subsection 3.2.

3.2 Binocular vision based quality assessment

For an image, its luminance component I can be considered as a matrix with an integer value corresponding to each pixel. I can be decomposed into a product of three matrices

$$I = USV^T \tag{10}$$

where U and V are orthogonal matrices, and $S = diag(s_1, s_2, \dots)$. The diagonal entries of S are

called singular values of I . It is well known that singular values from singular value decomposition (SVD) are sensitive to perturbations [20]. Adding distortions to an image will modify the structural information of the image, resulting in the perturbations of singular values. Since HVS is sensitive to structural changes, using singular values to quantify structural distortions provides a significant basis for assessing image quality. In the paper, singular values are used as features for gauging structural changes in stereoscopic images.

In order to reduce computational complexity, both the left and the right views of the distorted stereoscopic image are segmented into non-overlapped blocks with the size of $k \times k$, where k is an integer. Before the SVD is applied to each block, all blocks of stereoscopic image are classified into occluded block, pseudo-binocular suppression block, and pseudo-binocular fusion block. The classification method is as follows: for any block, if the block contains a pixel which belongs to the occluded region, the block will be considered as a occluded block; otherwise, if the block contains a pixel which belongs to the pseudo-binocular suppression region, it will be considered as a pseudo-binocular suppression block; or else, the block will be considered as a pseudo-binocular fusion block. According to three kinds of block-wise regions of the distorted stereoscopic image, three corresponding regions of the reference stereoscopic image are updated to three kinds of block-wise regions.

Since HVS deals with the occluded, binocular suppression and binocular fusion regions in different ways, different quality assessment methods will be adopted for the three regions. SVD is used to each block in each region and the local error in that block is computed to obtain all the local errors in the blocks of the region. The quality assessments of these different regions are described in the following subsections.

- 1) Quality assessment for occluded region: As pixels in the occluded region exist either in the left view or in the right view, the occluded region of the distorted stereoscopic image only refers to monocular vision and can be assessed by 2D-IQA methods. Let \widehat{B}_i denote the i -th block in \widehat{R}_{occ} , and B_i be the reference block of \widehat{B}_i in the reference stereoscopic image, then the distance $D_{occ}(i)$ between singular values of \widehat{B}_i and B_i is defined as

$$D_{occ}(i) = \sqrt{\sum_{j=1}^k (s_{i,j} - \widehat{s}_{i,j})^2} \tag{11}$$

where $s_{i,j}$ is the j -th singular value of B_i , $\widehat{s}_{i,j}$ is the j -th singular value of \widehat{B}_i , and k is the block size.

Let Q_{occ} be the global error of the occluded region, which can be defined as

$$Q_{occ} = \frac{1}{N_{occ}} \sum_{i=1}^{N_{occ}} |D_{occ}(i) - D_{occ}^m| \tag{12}$$

where N_{occ} is the number of blocks in the occluded region, and D_{occ}^m is the median value of $\{D_{occ}(1), D_{occ}(2), \dots, D_{occ}(N_{occ})\}$.

- 2) Quality assessment for pseudo-binocular suppression region: Let \widehat{B}_i be the i -th block in \widehat{R}_{bs} , \widehat{B}_i^l be the corresponding block in \widehat{R}_{bs}^l , \widehat{B}_i^r be the corresponding block in \widehat{R}_{bs}^r , B_i^l be the reference block of \widehat{B}_i^l in the left view of the reference stereoscopic image, and B_i^r be the reference block of \widehat{B}_i^r in the right view of the reference stereoscopic image, then the distance $D_{bs}^l(i)$ between singular values of \widehat{B}_i^l and B_i^l , and the distance $D_{bs}^r(i)$ between

singular values of \widehat{B}_i^r and B_i^r can be similarly calculated by Eq. (11). The overall visual quality of the binocular suppression region is dominated by the better quality view [13], therefore, the quality $D_{bs}(i)$ of \widehat{B}_i is defined as

$$D_{bs}(i) = \min\{D_{bs}^l(i), D_{bs}^r(i)\} \tag{13}$$

Let Q_{bs} be the global error of the pseudo-binocular suppression region, which can be defined as

$$Q_{bs} = \frac{1}{N_{bs}} \sum_{i=1}^{N_{bs}} |D_{bs}(i) - D_{bs}^m| \tag{14}$$

where N_{bs} is the number of blocks in the pseudo-binocular suppression region, and D_{bs}^m is the median value of $\{D_{bs}(1), D_{bs}(2), \dots, D_{bs}(N_{bs})\}$.

- 3) Quality assessment for pseudo-binocular fusion region: Let $D_{bf}^l(i)$ denote the distance between singular values of the i -th block in \widehat{R}_{bf}^l and the corresponding block in R_{bf}^l , and $D_{bf}^r(i)$ be the distance between singular values of the i -th block in \widehat{R}_{bf}^r and the corresponding block in R_{bf}^r , then $D_{bf}^l(i)$ and $D_{bf}^r(i)$ can be similarly calculated by using Eq. (11), respectively. The global error Q_{bf}^l of \widehat{R}_{bf}^l and the global error Q_{bf}^r of \widehat{R}_{bf}^r can be similarly calculated by using Eq. (12), respectively. The overall visual quality of the binocular fusion region is determined by both the left and right views' global errors in binocular fusion regions [22]. In addition, according to the property of binocular summation in the binocular fusion region, binocular acuity is approximately 1.4 times better than individual monocular acuities [22]. Therefore, the global error Q_{bf} of \widehat{R}_{bf} is defined as

$$Q_{bf} = 1.4 \times \frac{Q_{bf}^l + Q_{bf}^r}{2} \tag{15}$$

- 4) Quality fusion for distorted stereoscopic image: The overall visual quality of the distorted stereoscopic image is decided by \widehat{R}_{occ} , \widehat{R}_{bs} and \widehat{R}_{bf} . As distortions in these three regions are independent of one another, the global error Q of the distorted stereoscopic image is obtained by a linear regression equation of the quality indices of the three regions, which can be defined as

$$Q = a \cdot Q_{occ} + b \cdot Q_{bs} + c \cdot Q_{bf} \tag{16}$$

where a , b and c are the weights of the three regions in the overall quality and restricted by

$$\begin{cases} a + b + c = 1 \\ 0 \leq a \leq 1 \\ 0 \leq b \leq 1 \\ 0 \leq c \leq 1 \end{cases} \tag{17}$$

4 Experimental results and analyses

In this section, we will optimize the proposed method, and then compare its predictive performance with existing SIQA methods based on the following experiments: consistency

test, cross-image and cross-distortion tests, robustness test. At the beginning of this section, we firstly brief the database of SIQA.

4.1 Database of SIQA

A database of nine reference stereoscopic images and their 234 corresponding distorted stereoscopic images, which is established in our previous research [34], is used to evaluate the predictive performance of the proposed SIQA method. The 234 distorted stereoscopic images are generated with five distortions: Gaussian blurring (Gblur), white Gaussian noise (Wn), JPEG compression, and JPEG2000 compression with five quality levels, and H.264 compression with six quality levels. All the reference stereoscopic images in the database are captured by parallel cameras with the spacing of 50~75 mm, which is consistent with interpupillary distance of human eyes. Fig. 3 shows the left views of stereoscopic images in the SIQA database. According to Double Stimulus Continuous Quality Scale (DSCQS) testing method described in ITU-R recommendation BT.500-11 [9], the subjective ratings of the distorted images are conducted by a linear polarization stereoscopic display system [16]. Thus, a total of 243 stereoscopic images with varied types and amounts of distortions have tested as a way to demonstrate the general applicability of the proposed method.

4.2 Optimization of the proposed method

In this paper, the SIQA database is divided into two parts for training and testing. The training data set consists of four reference stereoscopic images (i.e., ‘Alt Moabit’, ‘Door flowers’, ‘Kendo’, and ‘Newspaper’) and their all distorted versions, which involves indoor and outdoor, close and long shot, complex and simple, strong and weak depth perception scenes. Therefore, the training data set has a wide variety of image contents, and it can be regarded as a comprehensive training data set. The testing data set consists of the other five reference stereoscopic images (i.e., ‘Akko & Kayo’, ‘Leaving Laptop’, ‘Balloons’, ‘Lovebird1’, and ‘Xmas’) and their all distorted versions. And there is no overlapping between training and testing. In the proposed method, the block size k and the weights of the three regions (i.e., a , b , and c) are obtained with all distorted images in the training data set. A nonlinear mapping is firstly employed between the output of the proposed method and the subjective quality score following the validation method in [9]. The nonlinearity chosen for regression for the proposed method is a five-parameter logistic function (a logistic function with an added linear term, constrained to be monotonic), which is given by

$$Quality(x) = \beta_1 \cdot \text{logistic}(\beta_2, (x - \beta_3)) + \beta_4 \cdot x + \beta_5 \quad (18)$$

$$\text{logistic}(\tau, x) = 1/2 - 1/(1 + \exp(\tau \cdot x)) \quad (19)$$

This nonlinearity is applied to the output of the proposed method or its logarithm, whichever gives a better fit for all data. As all k , a , b , and c in the proposed method are unknown, the output of the proposed method cannot be directly obtained. Thus, we put the parameter optimization into the fitting between the output of the proposed method and the subjective quality score.

In the proposed method, k denotes the size of segmented block in the left and right views of stereoscopic images. The larger the block size is, the higher the computational complexity of the SVD is. Therefore, k should be as small as possible, and four conditions of block size $k = \{4, 8, 12, 16\}$ are analyzed. Parameters (including a , b , and c) are computed with the Levenberg Marquardt (LM) and Universal Global Optimization (UGO), using the



Fig. 3 The left views of stereoscopic images in the SIQA database

mathematical software 1stOpt (First Optimization) Pro. v1.5 [15]. Table 1 shows the optimum values of a , b , c over four different k values, and the corresponding Pearson correlation coefficients (CC) of the proposed method with training and testing data sets are also given. The closer the value of CC is to 1, the better the performance of the proposed method is. As shown in Table 1, as for the training data set, when the value of k is 4, the corresponding value of CC is 0.943; when the value of k is 8, the corresponding value of CC is 0.947; there is no obvious difference between two CC values. However, as for the testing data set, when the value of k is 4, the corresponding value of CC is 0.936; when the value of k is 8, the corresponding value of CC is 0.922; the former is much larger than the latter. Meanwhile, as the value of k further increases (when k is larger than 8), the corresponding value of CC decreases rapidly. That is because the larger the value of k is, the more inaccurate the region classification of the distorted stereoscopic image is. Fig. 4 shows the proportions of the three regions (i.e., the occluded region, the pseudo-binocular suppression region, and the pseudo-binocular fusion region) in a distorted ‘Akko & Kayo’ varying with k . It is clear that, as k increases, the proportion of the occluded region increases, the proportion of the pseudo-binocular fusion region decreases, and the proportion of the pseudo-binocular suppression region has no evident changes. It means that parts of pseudo-binocular suppression region and pseudo-binocular fusion region are treated as the occluded region, and parts of pseudo-binocular fusion region are treated as pseudo-binocular suppression region. As a result, CC decreases as k increases, the predictive performance of the proposed method become worse. Based on the analyses above, k is set as 4. When k value is 4, the optimum values of a , b , c are 0, 0.440, 0.560, respectively. The value of a is 0, which indicates that human visual attention

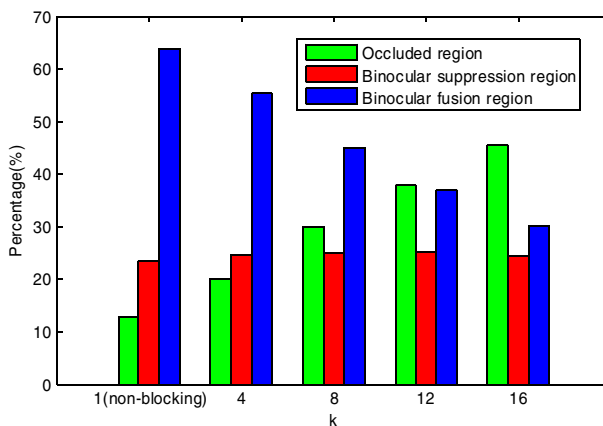
Table 1 The results of a , b , c , CC over four k values

Block size k	a	b	c	CC (Training)	CC (Testing)
4	0	0.440	0.560	0.943	0.936
8	0	0	1	0.947	0.922
12	0	0.594	0.406	0.917	0.882
16	0	0.714	0.286	0.897	0.846

focuses on binocular information, and the occluded region belongs to monocular information whose quality can not be represented by binocular information of stereoscopic images.

4.3 Comparison with the existing methods

We will compare the predictive performance of the proposed method with five existing SIQA methods, namely, PSNR-based method, SSIM-based method, MSVD-based (the block size is set as 4) method [20], SSIM d_1 [1], OQM [31]. Although the list of methods reported in the paper is not exhaustive, it is representative of existing SIQA methods. The PSNR-based, SSIM-based, and MSVD-based methods respectively apply PSNR, SSIM, MSVD to estimate the quality of each view separately and then represent the overall stereoscopic image quality as the mean of the results for each view. SSIM d_1 assesses the quality of stereoscopic image through combining the disparity and the averaged left and right view distortions, which is given by $d_1 = M \cdot \sqrt{D_d}$, where M is the average results of the left and right views using SSIM, D_d is the quality result of the disparity which is computed using global correlation coefficient between reference and distorted disparity maps. OQM combines the quality of image and disparity map by $OQM = \sqrt{IQM} + \sqrt{DQM} + \sqrt{IQM \cdot DQM}$, where IQM is the average result of the left and right views using SSIM, DQM is the quality result of the disparity map using UQI. We will analyze the predictive performances of the proposed method from consistency test, cross-image and cross-distortion tests, robustness test, compared with the five existing SIQA methods.

**Fig. 4** The proportions of three regions over different k values

- 1) Consistency Test: Experimental results are reported in terms of four criteria used for performance comparison, namely, Pearson linear correlation coefficient CC (for prediction accuracy), Spearman rank order correlation coefficient SROCC (for monotonicity), root mean squared error RMSE (for prediction accuracy), and outlier ratio OR (for prediction consistency), between subjective and objective scores. For a perfect match between objective and subjective scores, $CC = SROCC = 1$, and $RMSE = OR = 0$. Test results for all SIQA methods being compared are given as benchmark in Tables 2–5. Fig. 5 shows the scatter plots of the Difference Mean Opinion Score (DMOS) by subjective evaluation versus the predicted score by the six objective SIQA methods after the nonlinear mapping. The density of data points closely to the line $y = x$ represents the consistency of the predictive method and the subjective evaluation.

The results of Tables 2, 3, 4, 5 show that all the accuracy, the monotonicity, and the consistency of the proposed method are better than the other five methods in overall terms. The results also demonstrate that the perceived quality of stereoscopic image is quite different from that of 2D images. For example, human eyes are sensitive to the structure information of images, resulting in a good predictive performance of SSIM to 2D images. However, as to stereoscopic image, the binocular single vision of a stereoscopic image from the left and right views in HVS is more important. Distortions from any view of a stereoscopic image will directly affect the generation of the binocular single vision. PSNR reflecting signal errors in images can predict signal errors of a stereoscopic image to some extent. Therefore, the predictive performance of PSNR-based method is better than SSIM-based method. Singular values from the SVD are sensitive to perturbations resulting in the predictive performance of MSVD-based method is better than PSNR-based method. As all the PSNR-based, SSIM-based, and MSVD-based methods do not take into account the generation of the binocular single vision in HVS, the predictive performances of these methods are far worse than the proposed method, especially for all data. Even though both SSIM d_1 and OQM take into account depth perception, the combination methods between image quality and depth perception may be unreasonable resulting in worsening their performances for some distortions compared

Table 2 Linear correlation coefficient (CC) after nonlinear regression

Training (Total 104 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.970	0.763	0.829	0.985	0.913	0.885
SSIM-based	0.954	0.756	0.806	0.960	0.834	0.813
MSVD-based	0.955	0.923	0.953	0.985	0.971	0.936
SSIM d_1 [1]	0.935	0.887	0.805	0.950	0.725	0.748
OQM [31]	0.981	0.961	0.881	0.919	0.951	0.886
The proposed	0.964	0.938	0.938	0.984	0.968	0.943
Testing (Total 130 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.961	0.927	0.813	0.967	0.961	0.886
SSIM-based	0.938	0.915	0.753	0.959	0.945	0.817
MSVD-based	0.979	0.978	0.923	0.967	0.937	0.890
SSIM d_1 [1]	0.886	0.937	0.839	0.952	0.948	0.811
OQM [31]	0.927	0.852	0.771	0.930	0.917	0.800
The proposed	0.981	0.975	0.934	0.966	0.977	0.936

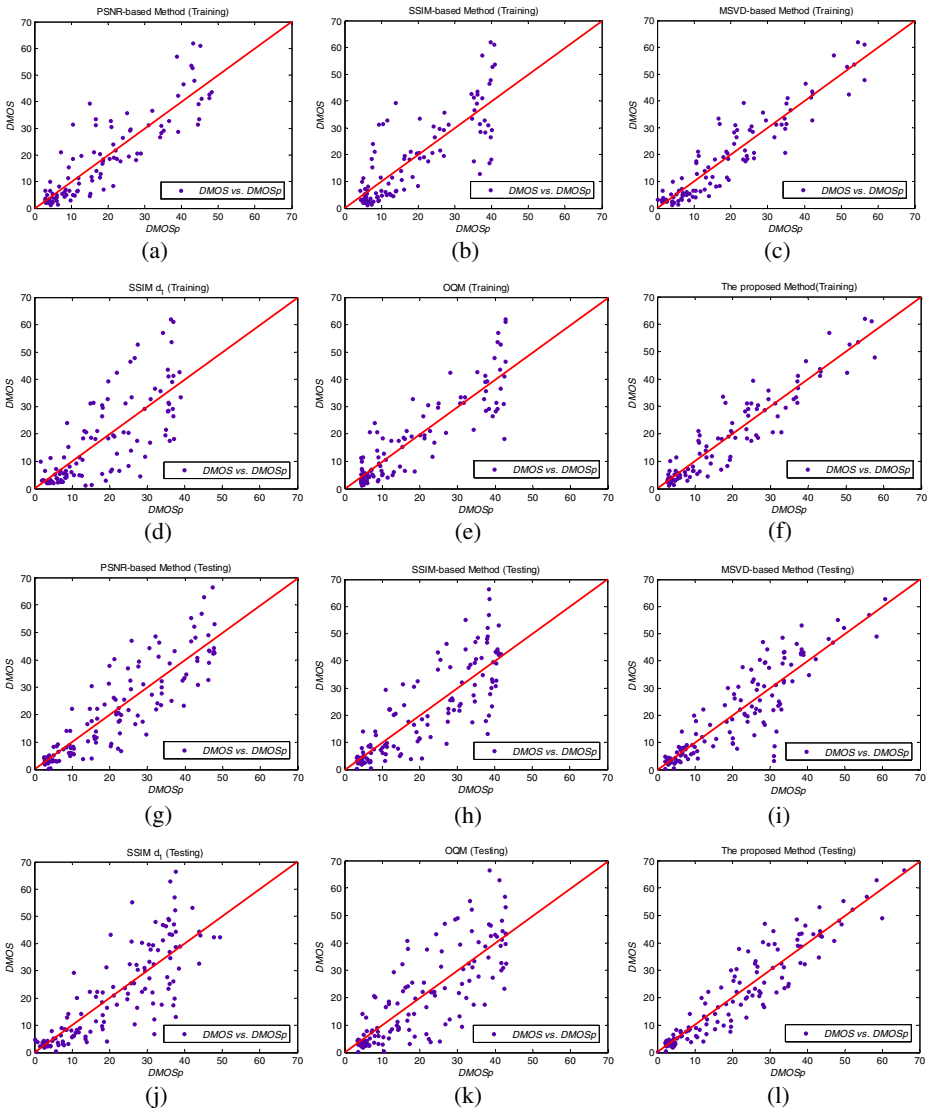


Fig. 5 Scatter plots for the quality predictions by the six methods with the training and the testing data sets. Training: (a) PSNR-based, (b) SSIM-based, (c) MSVD-based, (d) SSIM d_1 , (e) OQM, (f) The proposed; Testing: (g) PSNR-based, (h) SSIM-based, (i) MSVD-based, (j) SSIM d_1 , (k) OQM, (l) The proposed

with SSIM-based method. These two methods also do not consider the generation of the binocular single vision in HVS. As a result, the predictive results of SSIM d_1 and OQM are inconsistent to subjective quality evaluation. Besides, Experimental results show that the predictive performances of PSNR-based and MSVD-based methods are better than the proposed method for Wn distortion. That is because Wn is an additive noise and mainly affects non-edge areas of images rather than the generation of the binocular single vision, the subjective quality of stereoscopic images is determined by the quantity of noises injected into the stereoscopic images.

Table 6 additionally gives the comparison results between the proposed method and the method of literature [6]. Since literature [6] used the total 9 stereoscopic images shown in Fig. 3

- as the test images, different from Tables 2, 3, 4, 5, the results in Table 6 achieved with the proposed method also correspond to all of the 9 images. From Table 6, it is seen that the proposed method is good at Gblur, Wn and H.264 distortions, while the method of literature [6] is appropriate for JPEG and JPEG 2000 distortions. But for all data which crosses all five kinds of distortions, the proposed method is a little bit superior to the compared one.
- 2) Cross-image and Cross-distortion Tests: Many SIQA methods have been shown to be consistent when applied to distorted images generated from the same reference image by using the same distortion type. However, the effectiveness of these methods degrades significantly when applied to a set of images originating from different reference images involving a variety of different distortions. To further evaluate the effectiveness of the proposed method, the cross-image and cross-distortion tests are conducted, which are critical in evaluating the effectiveness of a quality assessment method. As shown in Fig. 5 and Tables 2, 3, 4, 5, the proposed method performs better than the other methods. Some methods are sensitive to image content, such as SSIM d_1 , there is a significant difference in the predictive performance on H.264 between the training and testing data sets, and the corresponding CC values are 0.725, 0.948, respectively. While other methods are independent of image content, but the predictive performance is good in the case of individual distortion but not good for all distortion existing (all data). Such as MSVD-based method, although all the CC values of the five individual distortion types are larger than 0.920, the CC value of all data is only 0.890 for the testing data set. All the CC values of the proposed method for the five individual distortions are larger than 0.933, and for all data the CC value of the proposed method is larger than 0.935 and the SROCC value is larger than 0.930. This fully demonstrates that the proposed method is a good predictive method to the perceived quality of stereoscopic image.
 - 3) Robustness Test: We choose some distorted stereoscopic images whose PSNR value of left images are closely to 28 dB from the SIQA database to test the robustness of the six methods. Although stereoscopic images are injected almost the same quantity of errors, the perceived quality differs greatly. Table 7 shows the information of these distorted stereoscopic images and predictive scores ($DMOS_p$) obtained with six SIQA methods. These distorted images are relating to five image contents and five distortion types. The

Table 3 Spearman rank order correlation coefficient (SROCC) after nonlinear regression

Training (Total 104 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.944	0.743	0.851	0.985	0.876	0.884
SSIM-based	0.922	0.749	0.823	0.929	0.724	0.808
MSVD-based	0.946	0.908	0.958	0.968	0.962	0.924
SSIM d_1 [1]	0.923	0.818	0.824	0.886	0.657	0.774
OQM [31]	0.958	0.934	0.850	0.863	0.937	0.876
The proposed	0.940	0.932	0.943	0.965	0.964	0.931
Testing (Total 130 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.945	0.924	0.838	0.943	0.915	0.901
SSIM-based	0.952	0.914	0.840	0.909	0.856	0.845
MSVD-based	0.975	0.962	0.912	0.950	0.818	0.888
SSIM d_1 [1]	0.908	0.962	0.856	0.879	0.892	0.846
OQM [31]	0.924	0.908	0.822	0.850	0.809	0.819
The proposed	0.983	0.964	0.928	0.940	0.964	0.943

Table 4 Root-mean-squared error (RMSE) after nonlinear regression

Training (Total 104 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	4.782	7.773	4.739	2.356	4.659	7.292
SSIM-based	5.871	7.862	5.014	3.866	6.299	9.115
MSVD-based	5.771	4.636	2.564	2.403	2.737	5.520
SSIM d_1 [1]	6.904	5.555	5.029	4.338	7.868	10.381
OQM [31]	3.764	3.344	4.013	5.458	3.523	7.256
The proposed	5.175	4.186	2.932	2.432	2.879	5.219
Testing (Total 130 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	5.572	5.638	7.238	4.037	4.186	7.684
SSIM-based	6.955	6.061	8.189	4.438	5.397	9.544
MSVD-based	4.070	3.154	4.797	3.990	5.952	7.554
SSIM d_1 [1]	9.276	5.276	6.767	4.827	5.103	9.685
OQM [31]	7.505	7.875	7.933	5.796	6.815	9.947
The proposed	3.882	3.336	4.432	4.073	2.645	5.827

distorted version of ‘Newspaper’ with Wn distortion brings the best perceived quality, and the corresponding DMOS is 12.957. While the distorted ‘Akko & kayo’ with Gblur brings the worst perceived quality, and the corresponding DMOS is 32.217. Therefore, the perceived quality of stereoscopic image is sensitive to distortion type. As shown in Table 7, only the proposed method can tell the differences of the effects of these five distortion types on the perceived quality of stereoscopic image, and the quality rank order obtained from the proposed method coincides with that of DMOS. The robustness of objective method to distortion types is very important to predict the perceived quality of stereoscopic image, and the proposed method meets the requirement.

Table 5 Outlier ratio (OR) after nonlinear regression

Training (Total 104 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.000	0.000	0.000	0.000	0.000	0.010
SSIM-based	0.000	0.022	0.022	0.000	0.019	0.019
MSVD-based	0.000	0.000	0.000	0.000	0.000	0.000
SSIM d_1 [1]	0.000	0.000	0.022	0.000	0.037	0.058
OQM [31]	0.000	0.000	0.000	0.000	0.000	0.010
The proposed	0.000	0.000	0.000	0.000	0.000	0.000
Testing (Total 130 distorted stereoscopic images)						
Methods/Distortion types	Gblur	JPEG	JPEG2000	Wn	H.264	All data
PSNR-based	0.000	0.000	0.022	0.000	0.000	0.023
SSIM-based	0.000	0.000	0.000	0.000	0.000	0.046
MSVD-based	0.000	0.000	0.000	0.000	0.019	0.031
SSIM d_1 [1]	0.044	0.000	0.022	0.000	0.000	0.038
OQM [31]	0.022	0.022	0.022	0.000	0.019	0.069
The proposed	0.000	0.000	0.000	0.000	0.000	0.000

Table 6 Outlier ratio (OR) after nonlinear regression

Methods/Distortion types		Gblur	JPEG	JPEG2000	Wn	H.264	All data
CC	The proposed	0.966	0.949	0.927	0.969	0.963	0.938
	Literature [6]	0.928	0.974	0.942	0.921	0.959	0.917
SROCC	The proposed	0.954	0.943	0.937	0.947	0.956	0.938
	Literature [6]	0.958	0.982	0.956	0.957	0.967	0.932
RMSE	The proposed	5.139	4.380	4.180	3.700	3.284	5.609
	Literature [6]	7.406	3.162	3.750	5.835	3.604	6.503

4.4 Summary

Compared with the five existing SIQA methods in terms of the consistency test, the cross-image and cross-distortion tests, and the robustness test, the proposed method has the following advantages.

Overall performance The proposed method simulates the processing of a stereoscopic image in HVS, applying the properties of binocular suppression, binocular fusion, and binocular summation into the objective quality assessment of stereoscopic image. The proposed method outperforms the other methods in terms of the accuracy, the monotonicity, the consistency.

Cross-image and cross-distortion The proposed method is based on the mechanisms of binocular vision, and the predictive performance of the proposed method is only related to the generation of the binocular single vision. Therefore, the proposed method is an objective quality assessment method crossing image content and distortion type, and its predictive results conform to subjective evaluations well.

Robustness In order to evaluate the robustness of the proposed method, the five distorted stereoscopic images whose left images are with almost the same PSNR value are chosen. According to the difference comparisons between DMOS scores and predictive scores by the six SIQA methods, the proposed method is robust to distortion type and can predict human visual perception well.

Table 7 The information of distorted stereoscopic images used in robustness test and their predictive scores by six SIQA methods

Stereoscopic image (distortion type/distortion parameter/PSNR value of left image)	Perceived quality rank order	PSNR-based	SSIM-based	MSVD-based	SSIM d_1	OQM	The proposed	DMOS
Newspaper (Wn/10/28.394)	1	20.178	38.275	11.735	27.918	15.698	11.090	12.957
Altmobit(JPEG/10/28.612)	2	24.042	24.320	21.702	27.274	23.203	18.740	19.174
Leaving Laptop (H.264/47/28.277)	3	28.414	29.633	28.696	34.906	38.345	33.264	27.348
Lovebird1 (JPEG2000/0.05/27.709)	4	26.253	30.064	29.053	26.500	30.837	33.740	31.261
Akko & Kayo (Gblur/2.5/28.158)	5	31.043	26.118	29.045	24.539	10.797	39.267	32.217

5 Conclusions

According to the processing of a stereoscopic image in human visual system (HVS), in the paper, we have proposed a novel objective stereoscopic image quality assessment method based on binocular vision. We firstly analyzed the generation of the binocular single vision in HVS, and classified the distorted stereoscopic image into three regions, i.e., occluded region dealt with by monocular vision, pseudo-binocular fusion region simulating the region dealt with by binocular fusion, and pseudo-binocular suppression region which is assumed to be dealt with by binocular suppression. As the occluded region referring to monocular vision, we adopted a two-dimensional image quality assessment method to predict its quality. As both pseudo-binocular fusion region and pseudo-binocular suppression region relating to binocular vision, we assessed these two regions by different methods according to the mechanisms of binocular fusion and suppression. Then we combined the three quality indices into one to represent the overall visual quality of the stereoscopic image. Finally, compared with the existing objective quality assessment methods, the predictive performance of the proposed method has been analyzed in terms of the consistency test, the cross-image and cross-distortion tests, and the robustness test. Experimental results show the proposed method outperforms the other methods and is in line with the human visual perception. In the paper, we only consider the generation of the binocular single view, while there are other perceptual attributes of binocular vision need to be considered in the future work. Additionally, the region segmentation in this paper which simply uses disparity of original image as a benchmark is just a rough processing, how to segment the three kinds of regions more reasonably and accurately is also worth to be considered in the future.

Acknowledgments This work was supported by the Natural Science Foundation of China (grant U1301257, 61271270, 61171163, 61271021) and the K. C. Wong Magna Fund in Ningbo University.

References

1. Benoit A, Le Callet P, Campisi P, and Cousseau R (2008) “Using disparity for quality assessment of stereoscopic images,” In: Proc. of IEEE International conference on image processing, San Diego, CA, USA, pp 389–392
2. Blake R, Sloane M FR (1981) Further developments in binocular summation. *Percept Psychophys* 30(3): 266–276
3. Boev A, Hollosi D, Gotchev A (2008) “Classification of stereoscopic artefacts”. Technical report D5.1, available at <http://sp.cs.tut.fi/mobile3dtv/>
4. Braddick OJ (1979) “Binocular single vision and perceptual processing”. In: Proc. of royal society of London—biological sciences, 204(1157) pp 503–512
5. Domanski M SO, Wegner K, Kurc M KJ, Siast J, Stankowski J, Ratajczak R, Grajek T (2013) High efficiency 3D video coding using new tools based on view synthesis. *IEEE Trans Image Process* 22(9):3517–3527
6. Feng S, Weisi L, Shanbo G, Gangyi J, Thambipillai S (2013) Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Trans Image Process* 22(5):1940–1953
7. IJsselstein WA, De Ridder H, Vliegen J (2000) Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. *IEEE Trans Circ Syst Technol* 10(2):225–233

8. ITU-R Recommendation BT.1438 (2000) Subjective assessment of stereoscopic television pictures
9. ITU-R Recommendation BT.500-11 (2002) Methodology for the subjective evaluation of the quality of television pictures
10. ITU-T Recommendation (2008) Subjective video quality assessment methods for multimedia applications. 910
11. Jangwon L, Kugjin Y, Kyuheon K (2013) A 3DTV broadcasting scheme for high-quality stereoscopic content over a hybrid network. *IEEE Trans Broadcast* 59(2):281–289
12. Kang MK, Ho YS (2012) Depth video coding using adaptive geometry based intra prediction for 3D video system. *IEEE Trans Multimed* 14(1):121–128
13. Kooi FL, Toet A (2004) Visual comfort of binocular and 3D displays. *Displays* 25(2–3):99–108
14. Ono H, Angus R, Gregor P (1977) Binocular single vision achieved by fusion and suppression. *Percept Psychophys* 21(6):513–521
15. Optimization software 1st Opt (7D-Soft High Technology Inc.), available at <http://www.7d-soft.com>
16. Peinsipp-Byma E, Rehfeld N, Eck R (2009) “Evaluation of stereoscopic 3D displays for image analysis tasks”. In: Proc. of SPIE. 7237/72370 L)
17. Sazzad ZMP, Yamanaka S, Kawayoke Y, Horita Y (2009) “Stereoscopic image quality prediction”. In: Proc. of international conference on quality of multimedia experience (QoMEX). San Diego, CA, USA, pp 180–185
18. Serrano-Pedraza I, Read JCA (2009) Stereo vision requires an explicit encoding of vertical disparity. *J Vis* 9(13):11, 1–37
19. Shao F, Jiang G, LinW MY, Dai Q (2013) Joint bit allocation and rate control for coding multi-view video plus depth based 3D video. *IEEE Trans Multimed* 15(8):1843–1854
20. Shnayderman A, Gusev A, Eskicioglu AM (2006) An SVD-based grayscale image quality measure for local and global assessment. *IEEE Trans Image Process* 15(2):422–429
21. Software for computing dense correspondence (disparity map) between two images using graph cuts, available at <http://www.cs.cornell.edu/People/vnk/recon.html>
22. Steinman SB, Steinman BA, Garzia RP (2000) Foundations of binocular vision: A clinical perspective. The McGraw Companies, New York
23. Tam WJ (2007) Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV. JVT-W094
24. Vetro A, Tourapis AM, Muller K, Chen T (2011) 3D-TV content storage and transmission. *IEEE Trans Broadcast* 57(2):384–394
25. Wang Z, Bovik A (2002) A universal image quality index. *IEEE Sig Process Lett* 9(3):81–84
26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
27. Wang X, YuM, Yang Y, Jiang G (2009) Research on subjective stereoscopic image quality assessment. In: Proc. of SPIE. 725501(725509)
28. Yang J, Hou C, Zhou Y, Zhang Z, Guo J (2009) Objective quality assessment method of stereo images. In: Proc. of IEEE international conference for 3DTV (3DTV-CON). Potsdam, Germany, pp 1–4
29. Yasakethu SLP, FernandoWAC KB, Kondoz A (2009) Analyzing perceptual attributes of 3d video. *IEEE Trans Consum Electron* 55(2):864–872
30. Yasakethu SLP, Hewage CTER, Fernando WAC et al (2008) Quality analysis for 3D video using 2D video quality models. *IEEE Trans Consum Electron* 54(4):1969–1976
31. You J, Jiang G, Xing L, Perkis A (2010) “Quality of visual experience for 3D presentation: stereoscopic image”. High-quality visual experience: Creation, processing and interactivity of high-resolution and high-dimensional video signals. Springer, pp 51–77
32. Zhang Y, An P, Wu Y, Zhang Z (2010) A multiview video quality assessment method based on disparity and SSIM. In: Proc. of IEEE international conference on signal processing (ICSP), no. 5655900. Beijing, China, pp 1044–1047
33. Zhang Y, Jiang G, Yu M, Yang Y, Peng Z, Chen K (2010) Depth perceptual region-of-interest based multiview video coding. *J Vis Commun Image Represent* 21(5–6):498–512
34. Zhou J, Jiang G, Mao X, Yu M, Shao F et al (2011) “Subjective quality analyses of stereoscopic images in 3DTV system”. In: Proc. of IEEE visual communications and image processing (VCIP), Tainan, Taiwan, pp 1–4
35. Zinger Svitlana, Do Luat, deWith PHN (2012) “Recent developments in free-viewpoint interpolation for 3DTV”. *3D Research* 3(4). doi: [10.1007/3DRes.01\(2012\)4](https://doi.org/10.1007/3DRes.01(2012)4)



Gangyi Jiang received his M.S. degree in electronics engineering from Hangzhou University in 1992, and received his Ph. D. degree in electronics engineering from Ajou University, Korea, in 2000. He is now a professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include video compression, stereoscopic/multi-view video coding, image quality assessment, etc.. He has published over 100 technical articles in refereed journals and proceedings in these fields.



Junming Zhou received the Ph. D. degree in the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2012. He is currently a researcher with the video technology research group at Tencent Inc., focusing on the development of HEVC. His current research interests include video codec and image processing.



Mei Yu received the M.S. degree from Hangzhou Institute of Electronics Engineering, Hangzhou, China, in 1993, and the Ph. D. degree from Ajou University, Suwon, Korea, in 2000. She is now a Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. Her research interests include image/video coding and video perception.



Yun Zhang received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph. D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2013, he was a Visiting Scholar with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. In 2010, he joined in the Shenzhen Institutes of Advanced Technology (SIAT), CAS, as an Assistant Professor. Since 2012, he serves as Associate Professor. His research interests are multiview video coding, video object segmentation and content based video processing.



Feng Shao received the B.S. and P.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in electronic science and technology. He is currently an Associate Professor in the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His research interests include video coding, image processing and perception, etc.



Zongju Peng received his B.S. degree from Sichuan Normal College, China, in 1995, and M.S. degree from Sichuan University, China, in 1998, and received his Ph. D. degree from Institute of Computing Technology, Chinese Academy of Science in 2010. He is now an associate professor in Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include image/video compression, multi-view video coding and video perception.