Quality Assessment Metric of Stereo Images Considering Cyclopean Integration and Visual Saliency $\stackrel{\Leftrightarrow}{\sim}$

Jiachen Yang^{a,*}, Yafang Wang^a, Baihua Li^a, Wen Lu^b, Qinggang Meng^c, Zhihan Lv^d, Dezong Zhao^e, Zhiqun Gao^a

^aSchool of Electronic Information Engineering, Tianjin university, Tianjin, 30072, P. R. China

^bSchool of Electronic Engineering, Xidian University, Xi'an 710071, Shaanxi Province, China

^cDepartment of Computer Science, School of Science at Loughborough University, UK

^d University College London, Dept. of Computer Science, 66-72 Gower Street, London, WC1E 6EA, United Kingdom

^e Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom

Abstract

In recent years, there has been great progress in the wider use of threedimensional (3D) technologies. With increasing sources of 3D content, a useful tool is needed to evaluate the perceived quality of the 3D videos/images. This paper puts forward a framework to evaluate the quality of stereoscopic images contaminated by possible symmetric or asymmetric distortions. Human visual system (HVS)studies reveal that binocular combination models and visual saliency are the two key factors for the stereoscopic image quality assessment (SIQA) metric. Therefore inspired by such findings in HVS, this paper proposes a novel saliency map in SIQA metric for the cyclopean image called "cyclopean saliency", which avoids complex calculations and produces good results in detecting saliency regions. Moreover, experimental results show that our metric significantly outperforms conventional 2D quality metrics and yields higher correlations with human subjective judg-

*Corresponding author

Email addresses: yangjiachen@tju.edu.cn (Jiachen Yang),

WangYF0739@tju.edu.cn (Yafang Wang), b.li@lboro.ac.uk (Baihua Li),

luwen@mail.xidian.edu.cn (Wen Lu), q.meng@lboro.ac.uk (Qinggang Meng),

Z.Lu@cs.ucl.ac.uk (Zhihan Lv), D.Zhao2@lboro.ac.uk (Dezong Zhao), zqgao@tju.edu.cn (Zhiqun Gao)

Preprint submitted to Information Sciences

August 28, 2016

ment than the state-of-art SIQA metrics. 3D saliency performance is also compared with "cyclopean saliency" in SIQA. It is noticed that the proposed metric is applicable to both symmetric and asymmetric distortions. It can thus be concluded that the proposed SIQA metric can provide an effective evaluation tool to assess stereoscopic image quality.

Keywords:

Binocular Combination, Human Visual System, Saliency, Stereoscopic Image Quality Assessment, Visual Attention.

1. Introduction

Visual information plays an important role in information acquisition in our daily lives. Over the past few decades, considerable stereoscopic technologies have been invented and commercialized [36, 38, 32]. However, new issues and challenges have also emerged with their development. During stereoscopic content creation, transmission, processing and display, various distortions causing visual fatigue and visual discomfort may be introduced. Consequently, it is necessary to develop an effective tool to measure the quality of stereoscopic images.

SIQA, like the quality assessment of monocular images (IQA), can be categorized into subjective and objective methods. Subjective assessment of stereo images represents the direct reflection of the HVS, and thus it is regarded as the most reasonable assessment method. In recent years, research using subjective experiments has achieved steady development and various factors that may affect stereoscopic perception have been investigated. However, subjective tests are time-consuming and require many duplicate experiments with a large number of participants. Therefore, objective metrics, which can be used to reliably predict the perceived quality of stereo images, has attracted significant attention from scholars and experts.

There are a number of 2D-IQA metrics with competitive performances in IQA. They can be divided into several categories: structural similarity based IQA metrics (e.g. structural similarity (SSIM) [50]; distortion distributionbased structural similarity (ADD-SSIM) and distortion distribution-based structural similarity (ADD-GSIM) [15]); deep learning based blind image quality assessment frameworks [23, 20]; and natural scene statistics based metrics [11, 16]. The multi-scale Geometric Analysis method [12], mimics the HVS and free-energy-based brain theory in IQA, investigating which classical HVS-inspired features could be used to evaluate the image quality [21]. Inspired by that the human visual system (HVS) exhibits obvious orientation selectivity mechanism for visual content extraction, Wu *et al.* proposed a reduced-reference IQA [52]. These metrics can only be used to evaluate the image quality before and after coding from one viewpoint. However, they fail to consider strong correlations with standard disparity from two adjacent viewpoints. Hence, a new objective quality assessment metric for stereoscopic images is needed.

However, it is not easy to design an accurate SIQA metric due to the disparity and depth. The earliest study attempting to evaluate the quality of stereo images simply applied 2D-IQA to their metrics. In our previous work [53], we found that the absolute disparity map approximately reflects the disparity and depth. Therefore we proposed an objective metric for stereoscopic images by combining this discovery. As stereoscopic images are different from plain images due to additional depth information, You *et al.* [56] investigated the possibility of applying some common 2D quality metrics in SIQA, while also taking depth information into consideration. A similar approach was also adopted by Benoit *et al.* [3], in which 2D-IQA metrics were augmented with disparity information.

Other metrics in the literature simulated the perceptual route of our visual system to process input signals. Shao *et al.* [46] divided stereo images into monocular regions, binocular fusion regions and binocular suppression regions. Furthermore, the overall quality was given as the linear summation of the three regions. By simulating monocular and binocular visual perception and analyzing the monocular-binocular feature fidelity induced index, another SIQA algorithm was proposed in his research [44]. Realizing that the neuronal responses are directly implicated in both the control and experience of 3D perception, Park J *et al.* developed a model-based neuronal and statistical framework that automatically predicts the level of visual discomfort [39].

Several metrics that took the theory of binocular combination into consideration were investigated. In the subsequent work of Shao, the binocular receptive field properties in line with human visual perception were introduced into quality assessment [45]. Maalouf *et al.* [34] presented the definition of "cyclopean image". They integrated left and right images into a cyclopean image to simulate brain perception, after which they used contrast sensitivity coefficients of the produced cyclopean image to derive a quality. Moreover, Chen *et al.* [4] proposed a cyclopean full reference metric that is able to account for binocular rivalry. Lin *et al.* [33] adopted three quality components in their metric to evaluate the quality of stereoscopic video, namely the cyclopean view, binocular rivalry, and the scene geometry. Wang *et al.* [49] set up a new subjective SIQA database and proposed a binocular rivalry inspired model to predict the quality of stereoscopic images. In addition, Gu *et al.* [18] introduced a metric based on the analysis of autoregressive model parameters to characterize the sharpness as an index to evaluate image blur, after which the metric was extended to assessing the sharpness of stereoscopic images. These metrics have been shown to outperform the former metrics (e.g. [53, 56, 3]) in predicting the quality of stereoscopic images, especially in the case of asymmetric distortions.

Since HVS is the ultimate assessor of image quality that takes into account a variety of visual characteristics, it has become another important factor affecting how to choose salient features. Li *et al.* developed an unsupervised feature selection algorithm to select the most useful features by integrating cluster and sparse structural analysis into a joint framework, while also experimentally demonstrating the effectiveness of the proposed algorithm [31]. In [30], a Robust Structured Subspace Learning (RSSL) algorithm was presented to uncover an appropriate latent subspace for data representation. Muhammad *et al.* [41] proposed a feature selection algorithm avoiding the positive region to replace the conventional dependency measure. Image understanding and feature learning were first integrated into a joint learning framework. This framework was then applied to several image understanding tasks and shown to achieve good performance.

For a large number of visual characteristics, the saliency map is particularly important. This is because HVS tends to select the most relevant information from a visual scene. Zhang *et al.* has demonstrated that the current soundness of visual saliency modelling is sufficient for IQA to yield a statistically meaningful gain in their performance [59]. Based on this evidence, we consider visual saliency to be a factor in IQA. Meanwhile, we attempted to find an efficient 3D metric to detect saliency regions that would affect the performance of SIQA. However, evaluation metrics that consider the significance of image are based on 3D saliency maps and extraction is both difficult and tedious. In order to overcome this shortcoming, a simple but efficient saliency detection metric is set to be in line with HVS.

In this paper, inspired by previous work on cyclopean images, a full reference quality assessment metric is proposed based on binocular combination and visual attention. In this proposed metric, we take advantage of several existing binocular combination models and propose a novel method that synthesizes two saliency maps to a cyclopean image, thereby predicting the saliency of binocular combination. The main innovations of our paper are as follows: 1) A proposed saliency map for stereoscopic images based on binocular combination models, defined as "cyclopean saliency". The definition of "cyclopean saliency" then leads to the development of a new idea for computing the saliency of stereoscopic images. 2) We propose a framework for stereoscopic image quality assessment based on binocular combination characteristics and "cyclopean saliency". The proposed framework is easy to implement, and the binocular combination model used in our metric is replaceable.

The rest of this paper is organized as follows. Section 2 presents a brief introduction of related work and our motivations. The overall 3D IQA framework is described in Section 3, including the derivation of some practical 3D IQA models. Section 4 presents the experiments conducted on the 3D IQA databases and the analysis of our model's performance. Finally in Section 5, we conclude the paper with a discussion and an outlook on future work.



Figure 1: A stereopair and corresponding saliency map, (a) left view, (b) right view, (c) saliency map based on GBVS of the left view, (d) saliency map based on GBVS of the right view, (e) left view added with corresponding GBVS-based saliency map, (f) left view multiplied by corresponding GBVS-based saliency map.

2. Related Work and Motivation

2.1. Binocular Combination Models

Research on human vision suggests that our two eyes each view the world from a slightly different angle. The resulting small difference perceived between the two views serves as the basis for assessment. Binocular combination refers to the combination of these two retinal images form a single "cyclopean" perceptual image [7]. If the light images on the two retinas are similar, the HVS will combine them into a single percept, while dissimilar images will cause binocular rivalry where only one light image is perceived [26]. The perceived brightness data under binocular viewing conditions serves as an important clue for binocular combination [47]. Some interesting behaviour can arise in binocular brightness combination when the input brightness is asymmetric in both eyes. Examples include binocular brightness summation in response to Ganzfelds theory, and Fechner's paradox where a bright light to one eye may appear less bright when a dim light is shown to the other eye. Many binocular models have been proposed in order to reconcile these apparently contradictory properties, which can be divided into four types [13]:

Eye-weighting models: The cyclopean image is computed as the linear summation of the left and right light images, monocular luminance or quadratic of luminance, described as:

$$C(x,y) = [(\omega_L \cdot I_L(x,y))^{\beta} + (\omega_R \cdot I_R(x+d,y))^{\beta}]^{1/\beta}$$
(1)

where ω_L and ω_R are the weighting factors of the each view respectively, (x, y) is the position of a pixel, d is a disparity index that matches pixels from I_L to those in I_R . And $\beta=1,2$ means monocular luminance or quadratic of luminance respectively. Existing eye-weighting models are simple and the quadratic of luminance model can explain the binocular combination behaviour that the input brightness is symmetric for both eyes. Nonetheless, both models failed in matching with Fechner's paradox and explaining cyclopean perception as a result of ignoring the interaction between two eyes.

Vector summation models: These models suggest that binocular brightness perception is not the result of the simple addition of monocular signals but the sum of two orthogonal vectors with some normalization [35]. These models explain Fechner's paradox and cyclopean perception very well. Neural network models: These models firstly compute the neural response of each eye, before combining the binocular response as follows:

$$C(x, y) = N_L(x, y) + N_R(x + d, y)$$
(2)

where N_L and N_R are the responses of neural cells receiving strong excitation from one eye and weak inhibition from the other eye, respectively. In these models, Fechner's paradox can be explained well while the explanation of cyclopean perception depends on the adopted neural response model.

Gain-control theory models: Gain-control theory is dependent on comparison. For stimuli of ordinary contrast, when either eye is stimulated alone, the predicted cyclopean image is the same as the situation when both eyes are stimulated equally, coinciding with an easily observed property of natural vision [7]. Gain control models are effective in predicting the early stages of binocular combination and they explain Fechner's paradox and cyclopean perception well.

In this paper, only the most commonly used model of each type will be discussed in subsequent sections.

2.2. Saliency Map

Our eyes attempt to search for and recognize particular objects in a scene despite being given a limited amount of information. This information needs to be processed in order to speed up the interpretation of complex scenes in real time. Visual saliency maps try to predict the locations that people are likely to look at and are interested in. In recent years, many saliency computing models have succeeded in highlighting the regions which attract human attention; however, in comparison, these models have become weaker in the parts that the HVS is insensitive to. Such models include: Graph-Based visual saliency model (GBVS) which applies a random walker to detect the salient signals [22], the acuity-based saliency detection model which uses sparse features extracted from image patches for center-surround difference calculation [9], the reconstructed image signature-based model which detects spatially sparse signals embedded in spectrally sparse backgrounds to highlight the sparse salient regions [24], the maximum symmetric surrounding (MSSS) model which exploits features of color and luminance to detect saliency regions [1], and the low-level features-based model which considers local and global contrast based on the likelihood of corresponding features [25]. In [48], a saliency model is presented by integrating the low-level priors, the high-level priors and the Center Bias Prior (CBP). Furthermore,



Figure 2: Existing Quality Assessment Metric Using Saliency

Gu *et al.* introduced a computational Free Energy inspired Saliency detection technique (FES) [18]. Although these saliency detection models perform well in predicting salient areas of 2D images, the development of stereoscopic technologies requires the development of new 3D visual models.

Hence, it is necessary to propose a visual saliency computing model that accounts for both depth information and binocular combination. Based on the above considerations, several 3D saliency models have been put forward. For example, Wang *et al.* proposed a computational model of visual attention for 3D images by combining traditional 2D saliency detection methods with depth information [10]. Fang *et al.* established a stereoscopic saliency detection framework based on color, luminance, texture, and depth features extracted from discrete cosine transform coefficients, which were then used in a contrast calculation [40]. There are also some 3D saliency detection models based on RGB-D, such as [42].

2.3. Existing Quality Assessment Schemes Using Saliency

Saliency maps can predict the areas and objects that the visual system is concerned about, and close attention has accordingly been paid to this issue. Based on the analysis and application of contrast change technology,

Gu et al. not only provided a new subjective database, but also proposed a reduced reference(RR) image assessment and an automatic contrast enhancement model to enhance images [17]. Moreover, Zhang et al. extracted the visual saliency map and gradient magnitude as two basic features, after which they computed their structural similarity separately and combined them to obtain the objective quality [58]. In [28], Chu et al. tried to directly extend 2D visual saliency maps to stereoscopic quality assessment. Their metrics are illustrated in Fig.2. In these frameworks, the images were first merged with the corresponding saliency maps. Next, the structural similarity of the merged images, comprised of a distorted and reference image for both left view and right views, was separately computed for each view. The overall quality is given as a linear combination of each view's quality. In the above methods, visual saliency was used as an additional feature. Thus visual saliency played only a small role in their metrics, and the introduction of visual saliency in their metrics will not lead to great influence on the final results. In this paper, we use saliency maps as not only an additional factor but also a weighting parameter. Using the saliency map as a weighting parameter enables us to assign higher weights to the areas that human eyes are more concerned about. Fig.1c and Fig.1d show the comparison between using addition and multiplication when incorporating saliency maps. As we can see, adding a saliency map to the reference image will not produce dramatic changes in its properties, while multiplying a saliency map clearly highlights the areas or objects of concern.

2.4. Motivation

After better understanding binocular combination models and saliency maps, we intend to propose a quality assessment algorithm for optimally combined stereoscopic images. Although some stereoscopic saliency models [10, 40, 42] for stereopairs exist, it is still difficult to detect stereoscopic saliency. Firstly, there is no ground truth saliency, so the design of precise stereoscopic saliency extracting algorithm remains a difficult issue. Secondly, the introduction of a stereoscopic algorithm will significantly increase the complexity of SIQA algorithms. To solve this problem, a novel concept of a saliency map for cyclopean images called "cyclopean saliency" based on existing 2D saliency computing algorithms is proposed. Furthermore, we extend the "cyclopean saliency" to provide quality prediction for stereoscopic images. There are five models adopted in our algorithm, namely the GBVS model [22], the signature-based model(SGB) [24], the MSSS model [1], the multi-scale wavelet(MSW) transform model [25], and the FES model [18].

The five models are simple and biologically plausible, and perform well when it comes to predicting human visual fixations on plain images. The "cyclopean saliency" is calculated as a binocular combination of the visual saliency of each view. Better performance may be achieved if some more complicated binocular characteristics are considered. However, we try to avoid this precisely because certain complicated characteristics still remain mysterious. The introduction of these characteristics will further dramatically add complexity to the proposed metric.

3. The proposed Metric

As discussed in previous sections, visual saliency and cyclopean perception reflect certain physiological processes of the HVS. Cyclopean perception enables us to simulate what is perceived in our brain, while visual saliency helps us to highlight the areas we pay more attention to. Thus theories and models of visual saliency and cyclopean perception may be valid when applied to quality assessment models. With this inspiration, a cyclopean and saliency-based quality assessment metric for stereoscopic images is proposed. We intend to compute the saliency map and cyclopean image of stereopairs, followed by their combination to derive the metric. Hence, the framework of the proposed metric can be illustrated as Fig.3.

3.1. Synthesizing cyclopean image and cyclopean saliency

Reference stereopairs and distorted stereopairs are first transformed to LAB color space, and only luminance will be adopted for further processing. Thus color is ignored, since the proposed metric focuses on cyclopean perception and the luminance components are what attract our attention.

Saliency maps predict the positions that human eyes tend to be more interested in. A full-reference metric is designed to mark the quality of distorted images by highlighting the difference from the reference images. It will be meaningless to derive saliency maps from distorted stereopairs, as the objects we are interested in remain unchanged, whereas extracted saliency maps may exhibit considerable variation in different saliency detection metrics. In addition only the saliency maps of reference left and right views are computed for further processing. The saliency computing models GB-VS, MSSS, SGB, MSW, and FES are selected based on their simplicity and



Figure 3: Cyclopean and saliency-based framework of the proposed metric

performance. Fig.1c and Fig.1d show the saliency map of the left and right view separately.

We have previously discussed the concept of 2D saliency as a means to obtain separate visual saliency maps for left and right views. A critical issue we now face is how to realize seamless docking of visual saliency and cyclopean images. A straightforward way to compute a saliency map (defined as "cyclopean saliency") is similar to computing a cyclopean image. In this paper, binocular combination models are used to compute cyclopean saliency maps. The cyclopean image and corresponding cyclopean saliency can be expressed as equations (3) and (4):

$$C(x,y) = f_c(I_L(x,y), I_R(x,y+d))$$
(3)

$$CS(x,y) = f_{cs}(S_L(x,y), S_R(x,y+d))$$
(4)

where C represents the synthesized cyclopean image, CS stands for the cyclopean saliency, and I_L and I_R indicate the luminance of left view and right view respectively. Moreover, S_L and S_R are the saliency maps of each view, respectively, d means disparity, and function $f_c(\cdot)$ and $f_{cs}(\cdot)$ refers to the methods of synthesizing the cyclopean image and cyclopean saliency, respectively.

We introduced four binocular combination metrics in section 2. To discuss the superiority and deficiency of each type with regard to predicting

the quality of stereoscopic images, one classical model of each type will be discussed in the following section. Some parameters of these models are modified slightly to match the experiment, with the classical computational models of each type shown as follows:

Eye-weighting model: Engel autocorrelation model (EE) was proposed by Engel [8], where the binocular brightness can be written as:

$$C(x,y) = \sqrt{(\omega_L \cdot E_L(x,y))^2 + (\omega_R \cdot E_R(x+d,y))^2}$$
(5)

Similarly, cyclopean saliency (CS) synthesized using this model can be written as:

$$CS(x,y) = \sqrt{\left(\omega_L \cdot S_L(x,y)\right)^2 + \left(\omega_R \cdot S_R(x+d,y)\right)^2} \tag{6}$$

where E_L and E_R are the luminance and S_L and S_R are the saliency maps of each view, respectively, ω_L and ω_R are the weighting factors of each view which are limited to (0,1), and $\omega_L + \omega_R = 1$. ω_L and ω_R can be determined by finding the integral of the squared autocorrelation function for the pattern in each eye. In our experiment, ω_L and ω_R are all set to 0.5.

The eye-weighting model only considers the luminance characteristic, and the weighting of each view is constant for different stereoscopic images. Therefore, it ignores the interaction of the two eyes, and can't explain the Fechner's paradox and cyclopean perception.

Vector summation model: Curtis and Rule vector summation model (VC) [6]. This model postulated that the monocular input signals being summed were orthogonal and the following model was proposed:

$$C(x,y) = \sqrt{E_L^2(x,y) + E_R^2(x+d,y) + 2 \cdot E_L(x,y) \cdot E_R(x+d,y)}$$
(7)

The cyclopean saliency can be computed using this model and is written as:

$$CS(x,y) = \sqrt{S_L^2(x,y) + S_R^2(x+d,y) + 2 \cdot S_L(x,y) \cdot S_R(x+d,y)}$$
(8)

where E_L and E_R are the luminance, and S_L and S_R are the saliency map of each view, respectively.

This model is not just the simple linear superposition of the two eyes. Unlike the eye-weighting model, it exploits the interaction between visual channels in binocular processing and can explain Fechner's paradox well. Neural network model: the Cogan two-channel model (NC) was proposed in [5] in which separate monocular and binocular channels are employed. This model is formally equivalent to a vector summation model:

$$C(x,y) = \frac{E_L(x,y)}{1+c \cdot E_R(x+d,y)} + \frac{E_R(x+d,y)}{1+c \cdot E_L(x,y)} + k \cdot E_L(x,y) \cdot E_R(x+d,y)$$
(9)

and the cyclopean saliency can be computed as:

$$CS(x,y) = \frac{S_L(x,y)}{1+c \cdot S_R(x,y)} + \frac{S_R(x,y)}{1+c \cdot S_L(x,y)} + k \cdot S_L(x,y) \cdot S_R(x,y)$$
(10)

where E_L and E_R are the luminance, S_L and S_R are the saliency map of each view respectively, and c and k are constants. Here, c = 1.0 and k = 0.1.

However, if the saliency value is very low and close to zero, the corresponding areas will be ignored in later processes. These areas usually appear in the background, but remain important and cannot be ignored. To handle this situation and fit the experimental data, we made a small modification to this equation. An additive constant is introduced to prevent these areas being ignored, and saliency is no longer the dominating factor but a subordinate factor in our metric. Thus, equation (11) and equation (12) can be revised as:

$$C(x,y) = \frac{c_1 + E_L(x,y)}{1 + c_2 \cdot E_R(x+d,y)} + \frac{c_1 + E_R(x+d,y)}{1 + c_2 \cdot E_L(x,y)} + k \cdot E_L(x,y) \cdot E_R(x+d,y)$$
(11)

$$CS(x,y) = \frac{c_1 + S_L(x,y)}{1 + c_2 \cdot S_R(x+d,y)} + \frac{c_1 + S_R(x+d,y)}{1 + c_2 \cdot S_L(x,y)} + k \cdot S_L(x,y) \cdot S_R(x+d,y)$$
(12)

where the additive constant c_1 is set to 1.0 and $c_2 = 1.0$ as before.

Here, the model is more general to account for binocular performance at the differential luminance threshold and in suprathreshold contrast matching. Moreover, by briefly taking into consideration the visual neurophysiology, it offers an explanation for interocular transfer of adaptation.

Gain control model: In Ding and Sperling model (GS) [7], the binocular combination is defined as:

$$C(x,y) = \frac{1 + R_L(x,y)}{1 + R_L(x,y) + R_R(x+d,y)} \cdot E_L(x,y) + \frac{1 + R_R(x+d,y)}{1 + R_L(x,y) + R_R(x+d,y)} \cdot E_R(x+d,y)$$
(13)

where E_L and E_R are the luminance of left view and right view, respectively, R_L and R_R are the energy response sums of all the frequency channels for the left eye and the right eye, respectively.

The cyclopean saliency is defined as:

$$CS(x,y) = \frac{1+R_L(x,y)}{1+R_L(x,y)+R_R(x+d,y)} \cdot S_L(x,y) + \frac{1+R_R(x+d,y)}{1+R_L(x,y)+R_R(x+d,y)} \cdot S_R(x+d,y)$$
(14)

This method is employed for energy calculation with the log-Gabor filter. Log-Gabor filter $G_{s,o}$ (o denotes the orientation index and s denotes the scale index) can decouple the image into a set of responses on different scales and different orientations, denoted as $[r_{s,o}, i_{s,o}]$ (r and i are the real part and the imaginary part of the response, respectively). In the Fourier frequency Domain:

$$G_{s,o}(r,\theta) = \exp\left[-\frac{\left(\log(\omega/\omega_s)\right)^2}{2\sigma_s^2}\right] \exp\left[-\frac{\left(\log\left(\theta - \theta_0\right)\right)^2}{2\sigma_o^2}\right]$$
(15)

where the parameters ω and θ are the normalized radial frequency and the orientation angle of the filter, and ω_s and θ_0 are the corresponding center frequency and orientation of the filter, respectively. The parameters σ_o and σ_s determine the strength of the filter. In our experiment, the design of the filter followed the work in [43]. The local energy response on all scales s and along orientation o is given by:

$$R_o = \sqrt{\sum_s r_{s,o} + \sum_s i_{s,o}} \tag{16}$$

Local energy response R are defined as the maximum value of local energy with every orientation:

$$R = max[R_o] \tag{17}$$

The gain control model is comparatively dependent: very low-contrast stimuli to the left- and right-eye are added linearly to form the predicted cyclopean image, while both binocular combination and binocular rivalry are considered, coinciding with an easily observed property of natural vision. Based on these principles, Fechneros paradox and cyclopean perception are well accounted for.

3.2. Saliency map weighting and overall quality

The cyclopean image for reference stereopairs is denoted as C_r , the cyclopean image for distorted stereopairs is denoted as C_d , and the cyclopean saliency map of reference stereopairs is denoted as CS. Only the cyclopean

saliency map of reference stereopairs is adopted because the saliency calculated from the reference stereopairs is more accurate. With cyclopean image and cyclopean saliency, the next step is to combine them into an overall quality assessment metric. Unlike some papers where saliency is just used as an additional insignificant factor, we also took the cyclopean saliency map as a weighting parameter to highlight the areas in which our eyes are more likely to be interested in. It was realized by multiplying the cyclopean image and the cyclopean saliency map, and thus the overall quality can be described as follows:

$$C_{wr}(x,y) = C_r(x,y) \cdot CS(x,y) \tag{18}$$

$$C_{wd}(x,y) = C_d(x,y) \cdot CS(x,y) \tag{19}$$

The overall quality is obtained by applying a full reference 2D-IQA algorithm to the reference saliency weighted cyclopean image C_{wr} and the test cyclopean image C_{wd} , which can be denoted as:

$$Q = f(C_{wr}, C_{wd}) \tag{20}$$

where Q is the overall quality and $f(\cdot)$ denotes the application of a full reference quality assessment metric.



Figure 4: One of the reference left view images and the corresponding five types of distortion. (a) Reference left view. (b) JPEG compressed. (c) JPEG 2000 compressed. (d) White noise. (e) Blur. (f) Fast fading.

4. Experimental Result and Analysis

4.1. Stereo database

The SIQA database used in our experiment is the LIVE 3D Image Quality Database from The University of Texas at Austin [37]. Distorted stereopairs cover five distortion types, including JPEG, JPEG2000 compression, additive white noise, blur and fast fading (shown in Fig.4). The main reason for using this database is its authority, publicity and availability of subjective evaluation scores.

4.2. Performance measure

To verify the performance of our metric, we choose five 2D metrics and apply them to our framework. The candidates for 2D metrics are SSIM [50], MS-SSIM [51], ADM [29], ADD-SSIM and ADD-GSIM [15]. The predictive performance of each metric is evaluated by four commonly used performance measures: the Pearson linear correlation coefficient (*PLCC*), the root mean squared error (*RMSE*), the Spearman rank-order correlation coefficient (*SROCC*) and Kendall rank-order correlation coefficient (*KROCC*). *PLCC* and *RMSE* are obtained after non-linear regression. For these two measures, the logistic function (21) specified in [2] is used to fit the model predictions to the subjective data. In addition, the *SROCC* and *KROCC* are used to evaluate the prediction monotonicity. A smaller *RMSE* value means a better performance, with bigger values of *PLCC*, *SROCC* and *KROCC* being preferred. For a perfect match of objective and subjective scores, RMSE = 0 and PLCC = SROCC = KROCC = 1.

$$Q' = \frac{\beta_1 - \beta_2}{1 + exp(\frac{Q - \beta_3}{|\beta_4|})} + \beta_2$$
(21)

where β_1 , β_2 , β_3 , and β_4 are determined by using the subjective scores and the objective scores.

4.2.1. Performance in predicting symmetric distortions

In this section, GBVS based CS is used to represent the results under the cyclopean saliency model. In later subsection, we will continue to discuss the performance of different saliency models. The values of PLCC, RMSE, SROCC and KROCC of the proposed framework and 2D-IQA for Live Phase I are given from Table 1 to Table 5. In all metrics, NC MS-SSIM



Figure 5: Performance of MS-SSIM and proposed framework based on GBVS with Live Phase I The performance of (a) 2D baseline MS-SSIM, (b) EE MS-SSIM, (c) VC MS-SSIM, (d) NC MS-SSIM and (e) GS MS-SSIM.

combination performs better than any other combinations and all 2D-IQA metrics. Generally, the proposed framework outperforms the corresponding 2D-IQA; however, some combinations do not boost performance and some even degrade performance. As shown in the results, some specific 2D-IQA may be applicable for predicting the quality of symmetric distorted stere-opairs, though the performance is not always outstanding.

Since our framework using corresponding 2D metric MS-SSIM performs best in the proposed framework, only results using MS-SSIM are adopted for discussion and comparison. Fig. 5 shows the scatter plots of DMOS versus the proposed framework with respect to four different binocular combination models and 2D-IQA MS-SSIM metric on symmetric distorted stereopairs.

From the scatter plots, it can be concluded that the NC MS-SSIM performs the best, while others generally achieve a better convergence than 2D metric MS-SSIM. The proposed framework using binocular models indeed proves to be the better in performance in the case of averaging the 2D scores of the left image and the right image directly.

Table 1: Performance of the proposed framework based on ADM with Live Phase I

Algorithm	PLCC	SROCC	KROCC	RMSE
AVG-ADM	0.9119	0.9083	0.7197	8.9752
EE ADM	0.9239	0.9171	0.7391	8.3640
VC ADM	0.9119	0.9105	0.7256	8.9713
NC ADM	0.8840	0.8835	0.6844	10.2222
GS ADM	0.9135	0.9049	0.7202	8.8971
	Algorithm AVG-ADM EE ADM VC ADM NC ADM GS ADM	Algorithm PLCC AVG-ADM 0.9119 EE ADM 0.9239 VC ADM 0.9119 NC ADM 0.8840 GS ADM 0.9135	AlgorithmPLCCSROCCAVG-ADM0.91190.9083EE ADM 0.92390.9171 VC ADM0.91190.9105NC ADM0.88400.8835GS ADM0.91350.9049	AlgorithmPLCCSROCCKROCCAVG-ADM0.91190.90830.7197EE ADM 0.92390.91710.7391 VC ADM0.91190.91050.7256NC ADM0.88400.88350.6844GS ADM0.91350.90490.7202

Table 2: Performance of the proposed framework based on SSIM with Live Phase I

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-SSIM	0.8721	0.8763	0.6784	10.6999
	EE SSIM	0.8064	0.8123	0.6063	12.9297
GBVS	VC SSIM	0.8362	0.8426	0.6384	11.9889
	NC SSIM	0.9327	0.9312	0.7669	7.8869
	GS SSIM	0.7818	0.7776	0.5698	13.6326

Table 3: Performance of the proposed framework based on ADD-SSIM with Live Phase I

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-ADD-SSIM	0.8995	0.8921	0.6983	9.5519
	EE ADD-SSIM	0.6959	0.6108	0.4318	15.7014
GBVS	VC ADD-SSIM	0.9260	0.9177	0.7416	8.2514
	NC ADD-SSIM	0.9130	0.9172	0.7356	8.9177
	GS ADD-SSIM	0.9337	0.8940	0.7090	7.6281

Table 4: Performance of the proposed framework based on ADD-GSIM with Live Phase I

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-ADD-GSIM	0.8921	0.8888	0.6908	9.8773
	EE ADD-GSIM	0.5873	0.6155	0.4338	20.1222
GBVS	VC ADD-GSIM	0.9304	0.9256	0.7527	8.0124
	NC ADD-GSIM	0.9053	0.9107	0.7320	9.2859
	GS ADD-GSIM	0.9468	0.9109	0.7403	6.8561

Table 5: Performance of the proposed framework based on MS-SSIM with Live Phase I

Saliency Model	Algorithm	PLCC	SROUC	KROCC	RMSE
	AVG-MS-SSIM	0.9252	0.9231	0.7472	8.2976
	EE MS-SSIM	0.9074	0.8977	0.7082	9.1876
GBVS	VC MS-SSIM	0.9085	0.9030	0.7149	9.1356
	NC MS-SSIM	0.9361	0.9342	0.7660	7.6931
	GS MS-SSIM	0.8890	0.8741	0.6774	10.0015

Table 6: Performance of the proposed framework and 2D baseline MS-SSIM with Live Phase I $_$

-	Algorithm	Distortions	PLCC	SROCC	KROCC	RMSE
	NC MS	JPEG	0.7525	0.7259	0.5256	5.7421
	NO MD-	JPEG2000	0.9504	0.9149	0.7418	5.3721
	CDVS	White Noise	0.9284	0.9327	0.7778	8.2412
	GDVS	Blur	0.9500	0.9254	0.7697	6.0284
		Fast Fading	0.8421	0.7821	0.5897	8.9353
		JPEG	0.6859	0.6123	0.4179	6.3445
	MS-	JPEG2000	0.9188	0.8917	0.7006	6.8141
	SSIM-	White Noise	0.9320	0.9320	0.7753	8.0389
	GBVS	Blur	0.9434	0.9261	0.7677	6.3986
		Fast Fading	0.8018	0.7231	0.5371	9.8997

Study has also been conducted to evaluate the performance of each metric in predicting different types of distortions. Table 6 presents the performance of NC MS-SSIM against the 2D baseline of MS-SSIM. From the table we can see that the proposed framework boosts the performance of the corresponding 2D-IQA, except with the distortion of white noise. For stereo images distorted by white noise, though the proposed framework cannot improve the performance, the performance values are still close to that of averaged 2D-IQA. Therefore it demonstrates that the proposed framework is a useful predictor to predict the quality of stereopairs contaminated by various distortions.

4.2.2. Performance in predicting both symmetric and asymmetric distortions

Evaluating the quality of symmetric distorted stereopairs is only one aspect to be considered. Another is to predict the quality of asymmetric distorted stereopairs, as the asymmetric distortions will have an even more

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-ADM	0.7511	0.7242	0.5481	7.4528
	EE ADM	0.7121	0.7020	0.5340	7.9201
GBVS	VC SSIM	0.6689	0.6869	0.5244	8.3908
	NC ADM	0.5604	0.6336	0.4785	9.3484
	GS ADM	0.7320	0.7031	0.5219	7.6904

Table 7: Performance of the proposed framework based on ADM with Live Phase II

complicated impact on our perception. In order to present an intuitive explanation for the importance of binocular combination and cyclopean saliency, the proposed metric is also applied to asymmetric distorted pairs. From Table 7 to Table 11, we can see the performance of each criterion on the Live Phase II database, which contains both symmetric and asymmetric distortions. Fig. 6 shows the scatter plots of DMOS versus the proposed framework with respect to four different binocular combination models and MS-SSIM metric on both symmetric and asymmetric distorted stereopairs. As shown in the results, predicting the quality of both symmetric and asymmetric distorted stereopairs is more difficult than predicting the quality of symmetric distorted data alone. Even though some specific 2D-IQA, i.e., MS-SSIM and ADM are applicable to symmetric distorted stereopairs, they fail to predict the quality of asymmetric distorted data. Furthermore, the corresponding ADM metric performs badly in predicting the quality of asymmetric data. ADM derives the quality by separately evaluating the additive noises and detail losses, but the binocular combination is more complicated for asymmetric distortions. Therefore, the detail losses and additive noises should not be decoupled in this way. The proposed framework significantly boosted the performance of the corresponding SSIM and MS-SSIM, especially the NC model. Our framework also achieved performance improvements for 2D-IQA ADD-SSIM and ADD-GSIM, with the EE model being an exception. As we mentioned previously, the EE model ignores the interaction of the two eyes due to its simple operation. Hence, it can't explain binocular perception and doesn't perform well in either case of symmetrical or asymmetrical distortions. The facts offer an indirect proof of the importance of visual characteristics such as binocular combination and saliency in predicting the quality of stereoscopic images.

Table 8: Performance of the proposed framework based on SSIM with Live Phase II

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-SSIM	0.8020	0.7919	0.6014	6.7429
	EE SSIM	0.8026	0.7926	0.6012	6.7332
GBVS	VC SSIM	0.8298	0.8209	0.6274	6.2989
	NC SSIM	0.8990	0.8796	0.7051	4.9424
	GS SSIM	0.8326	0.8256	0.6388	6.2518

Table 9: Performance of the proposed framework based on ADD-SSIM with Live Phase II

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-ADD-SSIM	0.6681	0.6479	0.4874	8.3983
	EE ADD-SSIM	0.3182	0.3599	0.2553	11.2873
GBVS	VC ADD-SSIM	0.8125	0.8068	0.6207	6.5798
	NC ADD-SSIM	0.8790	0.8680	0.6872	5.3814
_	GS ADD-SSIM	0.8115	0.7842	0.5977	6.7790

Table 10: Performance of the proposed framework based on ADD-GSIM with Live Phase $\scriptstyle\rm II$

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-ADD-GSIM	0.6807	0.6641	0.5072	8.2689
	EE ADD-GSIM	0.3358	0.3431	0.2437	11.2873
GBVS	VC ADD-GSIM	0.7750	0.7746	0.5916	7.1335
	NC ADD-GSIM	0.8418	0.8274	0.6336	6.0913
	GS ADD-GSIM	0.7820	0.7642	0.5841	7.2317

Table 11: Performance of the proposed framework based on MS-SSIM with LIVE Phase $\rm II$

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	AVG-MS-SSIM	0.7824	0.7774	0.6067	7.0293
	EE MS-SSIM	0.7631	0.7635	0.5893	7.2950
GBVS	VC MS-SSIM	0.8061	0.8028	0.6267	6.6799
	NC MS-SSIM	0.9016	0.8885	0.7142	4.8826
	GS MS-SSIM	0.7688	0.7603	0.5768	7.2187



Figure 6: Performance of MS-SSIM and the proposed framework based on GBVS with Live Phase II The performance of (a) 2D baseline MS-SSIM, (b) EE MS-SSIM, (c) VC MS-SSIM, (d) NC MS-SSIM and (e) GS MS-SSIM.

4.2.3. Performance comparison Between 2D Saliency and "Cyclopean Saliency"

In order to elaborate the importance of the proposed cyclopean saliency in our framework, we compare the performance achieved using a saliency map derived from a single image with the proposed cyclopean saliency map. This means that the weighting cyclopean saliency maps will be replaced by the saliency map derived from either view (left or right) only in equation (18) and (19). Fig. 7 shows the performance comparison between the proposed cyclopean saliency and 2D saliency. The performance indicates that cyclopean saliency performs significantly better than 2D saliency. It can be concluded from the results that we have proposed a straightforward method to compute the saliency map for cyclopean images based on existing saliency computing algorithms with minimum cost. The proposed cyclopean saliency matches our framework well and performs much better compared with the corresponding 2D saliency.



Figure 7: Performance comparison Between 2D Saliency and "Cyclopean Saliency" in NC MS-SSIM (a) Performance with Live Phase I (b)Performance with Live Phase II.

4.2.4. Performance comparison with existing 3D-IQA models

In this section, we compare the performance of the proposed metric with five existing 3D quality assessment metrics, i.e., Chen's scheme [4], Lin's scheme [33], Benoit's scheme [3], and Shao's scheme [45] and [44]. Because of the best performance of NC MS-SSIM among all combinations of the proposed metric, we will adopt it for comparison in the following sections.

From Table 12 we can see that though the indices of our model are only lower than Shao's scheme [44], there is a very weak gap between them. Generally, SIQA metrics which consider binocular characteristics perform better than the ones directly extended from 2D-IQA. As listed in Table 13, the performance of the proposed metric is also better than the other metrics in predicting the quality of asymmetric distortions. Though Lin's scheme considered binocular integration behaviors, its performance is mediocre. While the algorithm was specifically designed to predict the quality of 3D image compression, it ignored the importance of visual attention, thus leads to reduce its performance. The NC SSIM and Chen's model yield no significant difference relative to the performance of left-right averaged 2D-IQA using MS-SSIM, while the Benoit's metric performs well in predicting the quality of symmetric distorted stereopairs. However, the metrics of Benoit and Shao [44] failed in predicting asymmetric distorted data. Both the proposed NC MS-SSIM and Chen's model perform well in predicting the quality of symmetric and asymmetric data. Moreover, the PLCC, KROCC and RMSE of the proposed metric are the highest while the SROCC of Chen's model is the higher. The performance of Shao's Scheme [45] is also persuasive, considering the binocular combination or binocular characteristics. Overall, the proposed scheme shows an impressive consistency with human perception.

The proposed NC MS-SSIM yields a higher consistency with subjective scores due to the proposed cyclopean saliency. The result confirms our previous conclusion that cyclopean saliency can effectively predict the visual saliency of stereoscopic images. When compared to the performance where only binocular combination is considered, the performance of the proposed framework also considering cyclopean saliency is significantly improved.

4.2.5. Impacts of the different saliency models

In this section, we discuss the performance based on different cyclopean saliency metrics and the stereoscopic saliency metric. As mentioned in above sections, we select five saliency detection models: GBVS, MSSS, SGB, MSW and FES to form different cyclopean saliency maps. We also observe the per-

Table 12: Performance of the proposed framework NC MS-SSIM against existing 3D quality assessment metrics with Live Phase I

PLCC	SROCC	KROCC	RMSE			
0.9361	0.9342	0.7660	7.6931			
0.9267	0.9257	0.7468	7.6931			
0.8645	0.8559	0.6559	10.9898			
0.8829	0.88862	0.6907	10.2681			
0.9350	0.9251	-	5.8155			
0.9367	0.9365	-	5.7426			
	PLCC 0.9361 0.9267 0.8645 0.8829 0.9350 0.9367	PLCC SROCC 0.9361 0.9342 0.9267 0.9257 0.8645 0.8559 0.8829 0.88862 0.9350 0.9251 0.9367 0.9365	PLCC SROCC KROCC 0.9361 0.9342 0.7660 0.9267 0.9257 0.7468 0.8645 0.8559 0.6559 0.8829 0.88862 0.6907 0.9350 0.9251 - 0.9367 0.9365 -			

Table 13: Performance of the proposed framework NC MS-SSIM against existing 3D quality assessment metrics with Live Phase II

Algorithm	PLCC	SROCC	KROCC	RMSE
NC-MS-SSIM	0.9016	0.8885	0.7142	4.8826
Chen's scheme	0.901	0.893	0.6814	10.58
Lin's scheme	0.6584	0.6375	0.4701	8.4956
Bnoit's scheme	0.745	0.728	0.5357	16.2
Shao's scheme [45]	0.8628	0.8494	-	5.7058
Shao's scheme [44]	0.8601	0.8387	-	5.7581

formance using stereoscopic saliency to replace the cyclopean saliency as the additional feature and weighting parameter. For different cyclopean saliency metrics, the NC MS-SSIM combination performs better than any other combinations, and all 2D-IQA metrics. Therefore, this paper only presents the PLCC, SROCC, KROCC and RMSE in the different cyclopean saliency with the NC MS-SSIM framework. Table 14 and 15 show the results with LIVE database Phase I and II, separately. The best are emphasized in bold. For Phase I, it can be observed that the framework based on the SGB saliency model has the best performance. While the other algorithms under different saliency metrics do not achieve the best results, their performances come quite close to the best algorithm. Therefore, it can be concluded that the proposed cyclopean saliency can achieve a good performance with the majority of saliency detection models. That is to say, the cyclopean saliency model demonstrates generality based on different types of 2D-saliency. For 3D saliency, the proposed framework is efficient for the symmetric distorted images. However, when compared with the proposed cyclopean saliency model, the performance under it is a little bit poorer. The reason may be

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	GBVS	0.9361	0.9342	0.7660	7.6931
Cyclopean-	MSSS	0.9363	0.9340	0.7658	7.6786
Saliency	SGB	0.9365	0.9342	0.7664	7.6692
	MSW	0.9364	0.9341	0.7660	7.6754
	FES	0.9365	0.9342	0.7664	7.6697
3DSaliency	Fang's scheme	0.9107	0.9075	0.7202	9.0311

Table 14: Performance of the proposed framework NC MS-SSIM using different saliency detection model with Live Phase I

Table 15: Performance of the proposed framework NC MS-SSIM using different saliency detection model with Live Phase II

Saliency Model	Algorithm	PLCC	SROCC	KROCC	RMSE
	GBVS	0.9016	0.8885	0.7142	4.8826
Cyclopean-	MSSS	0.9018	0.8886	0.7144	4.8790
Saliency	SGB	0.9016	0.8884	0.7138	4.8830
	MSW	0.9016	0.8885	0.7141	4.8827
	FES	0.9017	0.8885	0.7142	4.8812
3DSaliency	Fang's scheme	0.8503	0.8395	0.6597	5.9405

that although 3D saliency can directly detect the saliency maps from the stereopairs by taking the depth into account, the final assessment algorithms are based on 2D-IQA. For Phase II, similar results are achieved with those in Phase I, while the algorithm based on MSSS model gets the best result. Fig.8 and Fig.9 show the scatter plots of DMOS versus the proposed framework with respect to five different cyclopean saliencies, as well as algorithm based on 3D saliency.

4.2.6. Impacts of the Disparity Algorithms

In the last section, we study the effect of different saliency models on the performance. In this section, investigation was conducted on how disparity algorithms affect the result. We explored three disparity algorithms and compared them with circumstances in which no depth information is used. Table 16 and 17 show the result with LIVE database Phase I and II, separately. MS-SSIM represents the MS-SSIM-based stereo algorithm that uses MS-SSIM to choose the best matches [4], SAD represents the simple sum-of-absolute difference (SAD) luminance matching without a smoothness,



Figure 8: Performance of NC MS-SSIM under the different cyclopean saliency and 3D saliency with Live Phase I The performance of (a) NC MS-SSIM based on GBVS, (b) NC MS-SSIM based on MSSS, (c) NC MS-SSIM based on SGB, (d) NC MS-SSIM based on MSW, (e) NC MS-SSIM based on FES, (f) NC MS-SSIM based on 3D saliency.

as introduced in [27]. For Phase I, we notice that performance with no depth information also achieves a good result. The reason may be that the stereopairs in Phase I are symmetrically distorted, and depth information will not play a decisive role. Also, it is worth noting that the matching error in Klaus's disparity algorithm leads to an undesired result which reminds us to search for a more precise matching algorithm to improve the result. From the performance with Phase II, the results under disparity algorithms are much better than the one with no depth information, while different disparity algorithms can definitely exert quite significant effects on the metric. As shown in the comparison between the two databases, the estimated disparities provide useful information in terms of predicting the quality of the stereo 3D images, especially for asymmetric distortions. On the whole, the MS-SSIM based stereo algorithm performs much better, thus it was applied to the design of the quality assessment metric.



Figure 9: Performance of NC MS-SSIM under the different "cyclopean saliency" and 3D saliency with Live Phase II The performance of (a) NC MS-SSIM based on GBVS, (b) NC MS-SSIM based on MSSS, (c) NC MS-SSIM based on SGB, (d) NC MS-SSIM based on MSW, (e) NC MS-SSIM based on FES, (f) NC MS-SSIM based on 3D saliency.

Table 16: Performance of the proposed framework NC MS-SSIM using different stereo algorithm with Live Phase I

Algorithm	PLCC	SROCC	KROCC	RMSE
MS-SSIM	0.9361	0.9342	0.7660	7.6931
Klaus	0.9198	0.9152	0.7369	8.5808
SAD	0.9365	0.9354	0.7673	7.6641
No depth	0.9345	0.9305	0.7606	7.7834

4.2.7. Testing with other databases

We have enlarged our own database (Stereo Image Database of School of Electronic and Information Engineering, Tianjin University [57]) to include many asymmetric stereo images. In our recent work [54], we proposed an evaluation criteria for stereo camera shooting quality. All reference stereopairs we shot for our database followed our criteria to make sure all of our reference stereopairs are of a good quality. Furthermore, the enlarged database con-

Table 17: Performance of the proposed framework NC MS-SSIM using different stereo algorithm with Live Phase II

	Algorithm	PLCC	SROCC	KROCC	RMSE			
	MS-SSIM	0.9016	0.8885	0.7142	4.8826			
	Klaus	0.8567	0.8443	0.6597	6.2089			
	SAD	0.8423	0.8264	0.6388	6.4886			
	No depth	0.8283	0.8152	0.6248	6.7454			



Figure 10: Performance values test against Tianjin University 3D Image Database

tains 470 distorted stereopairs containing both symmetric and asymmetric distortions and 30 reference stereopairs.

The performance values (PLCC, SROCC, KROCC and RMSE) of 2D metric MS-SSIM and SIQA models on our own stereo database are illustrated in Fig.10. From the figure we can conclude that the proposed metric performs best among all the metrics, which is comparable with Chen's scheme against our database.

5. Conclusions

In this paper, a full reference quality assessment framework is proposed for stereoscopic images. An advantage of the proposed framework is that it is applicable to both symmetric distortions and asymmetric distortions when predicting the image quality. The proposed metric can precisely predict the quality of an image contaminated by different types of distortion. A significant contribution of this work is that it provides a quality assessment structure in which the binocular combination characteristics with the cyclopean saliency are all considered. The binocular combination model and cyclopean saliency synthesis algorithm can both be replaced, while a better performance can be achieved through free combination. Moreover, the "cyclopean saliency" has generality for cooperation with different 2D-saliency maps. In our future work, we will also focus on predicting the quality of asymmetrically distorted stereo images and evaluating the stereo image combined with 3D saliency. Furthermore, enhancement technologies may possibly play a more significant role in future IQA research, since they are able to generate betterlooking images, even outperforming the natural images that are deemed to have the optimal quality. At the same time, further research is needed in screen content images, which are becoming particularly important due to their applications in remote computing and cloud gaming [19, 14, 55].

6. Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No.61471260 and No.61271324), and Natural Science Foundation of Tianjin: 16JCYBJC16000

7. Reference

- Achanta R, Ssstrunk S. Saliency detection using maximum symmetric surround[C]//2010 IEEE International Conference on Image Processing. IEEE, 2010: 2653-2656.
- [2] Antkowiak J, Baina T J. FINAL REPORT FROM THE VIDEO QUAL-ITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF VIDEO QUALITY ASSESSMENT March[J]. ITU-T Standards Contribution COM, 2000.
- [3] Benoit A, Le Callet P, Campisi P, et al. Using disparity for quality assessment of stereoscopic images[C]//2008 15th IEEE International Conference on Image Processing. IEEE, 2008: 389-392.
- [4] Chen M J, Su C C, Kwon D K, et al. Full-reference quality assessment of stereopairs accounting for rivalry[J]. Signal Processing: Image Communication, 2013, 28(9): 1143-1155.

- [5] Cogan A I. Human binocular interaction: Towards a neural model[J]. Vision research, 1987, 27(12): 2125-2139.
- [6] Curtis D W, Rule S J. Binocular processing of brightness information: A vector-sum model[J]. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4(1): 132.
- [7] Ding J, Sperling G. A gain-control theory of binocular combination[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(4): 1141-1146.
- [8] Engel G R. The visual processes underlying binocular brightness summation[J]. Vision research, 1967, 7(9): 753-767.
- [9] Fang Y, Lin W, Fang Z, et al. Visual acuity inspired saliency detection by using sparse features[J]. Information Sciences, 2015, 309: 1-10.
- [10] Fang Y, Wang J, Narwaria M, et al. Saliency detection for stereoscopic images[J]. IEEE Transactions on Image Processing, 2014, 23(6): 2625-2636.
- [11] Gao X, Gao F, Tao D, et al. Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning[J]. Neural Networks and Learning Systems, IEEE Transactions on, 2013, 24(12): 2013-2026.
- [12] Gao X, Lu W, Tao D, et al. Image quality assessment based on multiscale geometric analysis[J]. Image Processing, IEEE Transactions on, 2009, 18(7): 1409-1423.
- [13] Grossberg S, Kelly F. Neural dynamics of binocular brightness perception. Vision Research, 1999, 39(22): 3796-3816.
- [14] Gu K, Wang S, Yang H, et al. Saliency-Guided Quality Assessment of Screen Content Images[J].
- [15] Gu K, Wang S, Zhai G, et al. Analysis of Distortion Distribution for Pooling in Image Quality Prediction[J]. 2016.
- [16] Gu K, Wang S, Zhai G, et al. Blind Quality Assessment of Tone-Mapped Images via Analysis of Information, Naturalness and Structure[J]. 2016.

- [17] Gu K, Zhai G, Lin W, et al. The analysis of image contrast: From quality assessment to automatic enhancement[J]. Cybernetics, IEEE Transactions on, 2016, 46(1): 284-297.
- [18] Gu K, Zhai G, Lin W, et al. No-reference image sharpness assessment in autoregressive parameter space[J]. Image Processing, IEEE Transactions on, 2015, 24(10): 3218-3231.
- [19] Gu K, Zhai G, Lin W, et al. Learning a blind quality evaluation engine of screen content images[J]. Neurocomputing, 2016.
- [20] Gu K, Zhai G, Yang X, et al. Deep learning network for blind image quality assessment[C]//Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014: 511-515.
- [21] Gu K, Zhai G, Yang X, et al. Using free energy principle for blind image quality assessment[J]. Multimedia, IEEE Transactions on, 2015, 17(1): 50-63.
- [22] Harel J, Koch C, Perona P. Graph-based visual saliency. Advances in neural information processing systems, 2007, 19: 545.
- [23] Hou W, Gao X, Tao D, et al. Blind image quality assessment via deep learning[J]. Neural Networks and Learning Systems, IEEE Transactions on, 2015, 6(26): 1275-1286.
- [24] Hou X, Harel J, Koch C. Image signature: Highlighting sparse salient regions[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(1): 194-201.
- [25] Imamoglu N, Lin W, Fang Y. A saliency detection model using low-level features based on wavelet transform[J]. Multimedia, IEEE Transactions on, 2013, 15(1): 96-105.
- [26] Kaplan I T, Metlay W. Light intensity and binocular rivalry. Journal of Experimental Psychology, 1964, 67(1): 22.
- [27] Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure[C]//Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. IEEE, 2006, 3: 15-18.

- [28] Li L, Lin W, Wang X, et al. No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments[J]. Cybernetics, IEEE Transactions on, 2016, 46(1): 39-50.
- [29] Li S, Zhang F, Ma L, et al. Image quality assessment by separately evaluating detail losses and additive impairments. Multimedia, IEEE Transactions on, 2011, 13(5): 935-949.
- [30] Li Z, Liu J, Tang J, et al. Robust structured subspace learning for data representation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2015, 37(10): 2085-2098.
- [31] Li Z, Liu J, Yang Y, et al. Clustering-guided sparse structural learning for unsupervised feature selection[J]. Knowledge and Data Engineering, IEEE Transactions on, 2014, 26(9): 2138-2150.
- [32] Liang R, Shen W, Li X X, et al. Bayesian multi-distribution-based discriminative feature extraction for 3D face recognition[J]. Information Sciences, 2015, 320: 406-417.
- [33] Lin Y H, Wu J L. Quality assessment of stereoscopic 3D image compression by binocular integration behaviors[J]. Image Processing, IEEE Transactions on, 2014, 23(4): 1527-1542.
- [34] Maalouf A, Larabi M C. CYCLOP: A stereo color image quality assessment metric//Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, 2011: 1161-1164.
- [35] MacLeod D I A. The Schrodinger equation in binocular brightness combination. Perception, 1972, 1: 321-324.
- [36] Manap R A, Shao L. Non-distortion-specific no-reference image quality assessment: A survey[J]. Information Sciences, 2015, 301: 141-160.
- [37] Moorthy A K, Su C C, Mittal A, et al. Subjective evaluation of stereoscopic image quality[J]. Signal Processing Image Communication, 2013, 28(8):870-883.
- [38] Mrozek D, Dani?owicz P, Ma?ysiak-Mrozek B. HDInsight4PSi: Boosting performance of 3D protein structure similarity searching with HDInsight clusters in Microsoft Azure cloud[J]. Information Sciences, 2016, 349: 77-101.

- [39] Park J, Oh H, Lee S, et al. 3D visual discomfort predictor: Analysis of disparity and neural activity statistics[J]. Image Processing, IEEE Transactions on, 2015, 24(3): 1101-1114.
- [40] Qi F, Zhao D, Liu S, et al. 3D visual saliency detection model with generated disparity map[J]. Multimedia Tools and Applications, 2016: 1-17.
- [41] Raza M S, Qamar U. An incremental dependency calculation technique for feature selection using rough sets[J]. Information Sciences, 2016, 343: 41-65.
- [42] Ren J, Gong X, Yu L, et al. Exploiting global priors for RGB-D saliency detection[C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on. IEEE, 2015: 25-32.
- [43] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense twoframe stereo correspondence algorithms. International Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002.
- [44] Shao F, Li K, Jiang G, et al. MonocularCbinocular feature fidelity induced index for stereoscopic image quality assessment[J]. Applied Optics, 2015, 54(33): 9671-9680.
- [45] Shao F, Li K, Lin W, et al. Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties[J]. Image Processing, IEEE Transactions on , 2015, 24(10): 2971-2983.
- [46] Shao F, Lin W, Gu S, et al. Perceptual Full-Reference Quality Assessment of Stereoscopic Images by Considering Binocular Visual Characteristics. Image Processing, IEEE Transactions on , 2013, 22(5): 1940-1953.
- [47] Sohn H, Jung Y J, Lee S, et al. Predicting visual discomfort using object size and disparity information in stereoscopic images. Broadcasting, IEEE Trans on. 2013, 59(1): 28-37.
- [48] Song M, Chen C, Wang S, et al. Low-level and high-level prior learning for visual saliency estimation[J]. Information Sciences, 2014, 281: 573-585.

- [49] Wang J, Zeng K, Wang Z. Quality prediction of asymmetrically distorted stereoscopic images from single views[C]//Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE, 2014: 1-6.
- [50] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on, 2004, 13(4): 600-612.
- [51] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment//Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on. IEEE, 2003, 2: 1398-1402.
- [52] Wu J, Lin W, Shi G, et al. Orientation selectivity based visual pattern for reduced-reference image quality assessment[J]. Information Sciences, 2016, 351: 18-29.
- [53] Yang J, Hou C, Zhou Y, et al. Objective quality assessment method of stereo images//3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009. IEEE, 2009: 1-4.
- [54] Yang J, Liu Y, Meng Q and Chu R (2015). Objective Evaluation Criteria for Stereo Camera Shooting Quality under Different Shooting Parameters and Shooting Distances. Sensors Journal, IEEE, 2015, 15(8): 4508-4521.
- [55] Yang J, Shi Y. Towards finger-vein image restoration and enhancement for finger-vein recognition[J]. Information Sciences, 2014, 268: 33-52.
- [56] You J, Xing L, Perkis A, et al. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis//Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA. 2010.
- [57] Yun Nan, Feng Z, Yang J, et al. et al. "The objective quality assessment of stereo image." Neurocomputing 120 (2013): 121-129.
- [58] Zhang L, Shen Y, Li H. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. Image Processing, IEEE Transactions on, 2014, 23(10).

[59] Zhang W, Borji A, Wang Z, et al. The application of visual saliency models in objective image quality assessment: A statistical evaluation[J]. 2015.