

Article

A Novel No-Reference Quality Assessment Metric for Stereoscopic Images with Consideration of Comprehensive 3D Quality Information

Liquan Shen ^{1,*}, Yang Yao ^{1,†}, Xianqiu Geng ¹, Ruigang Fang ^{2,‡} and Dapeng Wu ^{2,‡}

¹ Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

² Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32603, USA

* Correspondence: jsslq@163.com

† These authors contributed equally to this work.

‡ These authors contributed equally to this work.

Abstract: Recently, stereoscopic image quality assessment has attracted a lot attention. However, compared with 2D image quality assessment, it is much more difficult to assess the quality of stereoscopic images due to the lack of understanding of 3D visual perception. This paper proposes a novel no-reference quality assessment metric for stereoscopic images using natural scene statistics with consideration of both the quality of the cyclopean image and 3D visual perceptual information (binocular fusion and binocular rivalry). In the proposed method, not only is the quality of the cyclopean image considered, but binocular rivalry and other 3D visual intrinsic properties are also exploited. Specifically, in order to improve the objective quality of the cyclopean image, features of the cyclopean images in both the spatial domain and transformed domain are extracted based on the natural scene statistics (NSS) model. Furthermore, to better comprehend intrinsic properties of the stereoscopic image, in our method, the binocular rivalry effect and other 3D visual properties are also considered in the process of feature extraction. Following adaptive feature pruning using principle component analysis, improved metric accuracy can be found in our proposed method. The experimental results show that the proposed metric can achieve a good and consistent alignment with subjective assessment of stereoscopic images in comparison with existing methods, with the highest SROCC (0.952) and PLCC (0.962) scores being acquired on the LIVE 3D database Phase I.

Keywords: no reference; stereoscopic image quality assessment; spatial domain; transform domain; stereo visual information; natural scene statistics; machine learning



Citation: Shen, L.; Yao, Y.; Geng, X.; Fang, R.; Wu, D. A Novel No-Reference Quality Assessment Metric for Stereoscopic Images with Consideration of Comprehensive 3D Quality Information. *Sensors* **2023**, *23*, 6230. <https://doi.org/10.3390/s23136230>

Academic Editors: Guangtao Zhai, Massimo Mischi, Xiongkuo Min, Menghan Hu and Wei Zhou

Received: 7 December 2022

Revised: 24 March 2023

Accepted: 29 March 2023

Published: 7 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of entertainment, military and industry, stereoscopic image/video technology has been widely used [1–4]. Meanwhile, the demand for high-definition and high-quality stereoscopic images/videos has become urgent, which leads to perceptual stereoscopic image quality assessment (SIQA) being an important issue for evaluating the performance of 3D technologies [5]. Stereoscopic images consist of two images, the left image and the right image, which are captured by two cameras [6–8]. Two channels of information need to be dealt with simultaneously to assess the quality of stereoscopic images. SIQA is widely used to assess the performance of stereoscopic image processing algorithms, such as stereoscopic image compression and denoising. Furthermore, a good SIQA algorithm can also be utilized to guide stereoscopic image processing, as well as the implementations and optimization procedures of the image [9].

Intrinsically, stereoscopic image quality is determined not only by the spatial domain factors, but also by the transformed domain factors and stereo perceptual factors, such as binocular disparity/depth perception and binocular rivalry/combination [4,10–13]. Such

dual-image and perceptual factors should be simultaneously considered in the process of feature description in SIQA cases, in addition to cyclopean features. However, almost all the existing feature extraction-based SIQA algorithms have not taken such features into consideration, with only the cyclopean features being exploited. Many assessment methods utilize cyclopean images in a different way. In Chen et al. [14], the natural scene statistics (NSS) model is used to extract spatial features from a cyclopean image, followed by support vector regression (SVR) score fitting. Recently, also based on the generated cyclopean image, Su et al. [15] utilized the spatial domain univariate NSS features, the wavelet domain univariate NSS features, and the bivariate density and correlation NSS features to create another assessment model. In addition, Zhou et al. [16] proposed an assessment model which utilizes the complementary local patterns of binocular energy response (BER) and binocular rivalry response (BRR), both extracted from the cyclopean image, to simulate the binocular visual perception.

In this paper, a perceptual quality assessment metric for stereoscopic images is proposed. The important effort of this work is to take comprehensive dual-image/3D relevant information into consideration of the stereoscopic quality to assess its image quality. Specifically, the traditional goal of an SIQA algorithm is to predict the quality of the cyclopean image formed within an observer's mind when left and right images are presented stereoscopically. Therefore, like many assessment methods, in our method, features of the cyclopean images in both the spatial domain and transformed domain are extracted based on NSS. Furthermore, to better comprehend the intrinsic properties of the stereoscopic image, in our method, the binocular rivalry effect and other 3D visual properties are also considered in the process of feature extraction. The main contributions of this work are as follows: (1) new 3D perceptual features including binocular rivalry-based NSS features, binocular disparity matching error-based NSS features and binocular disparity consistency-based NSS features, are extracted for binocular perception representation; (2) neighbor difference and neighbor product-based features are optimized to reflect the high correlation between surrounding pixels in the spatial domain to assess the cyclopean image's quality; (3) gradient magnitude, phase congruency and Log-Gabor response features are designed in the transformed domain to assess the cyclopean image's quality; and finally, (4) 2D features from the spatial domain and the transformed domain and 3D visual perceptual features are jointly utilized to assess 3D image quality, and principal component analysis (PCA) [17] is utilized to reduce the dimensions of the feature vectors and enhance the performance of the quality prediction.

The proposed paper has several differences from our previously published work, called "No-Reference Stereoscopic Image Quality Assessment Based on Image Distortion and Stereo Perceptual Information" [18]. The differences can be listed as follows: (1) this paper focuses on both 2D and 3D statistics about the given stereoscopic image pairs, such as coherent and rivaling 3D perceptions of the human visual system, while the work in [18] only emphasizes the intrinsic image distortions of 2D image pairs, such as blurriness, noisiness and blockiness; (2) in this work, to exploit more fundamental 3D distortion characteristics, before the machine learning process, a principle component analysis (PCA) is also applied to adaptively recognize the notable features, giving useful attentions for the following machine learning.

The remainder of this paper is organized as follows. Section 2 reviews previous work on SIQAs. We present our model, describing in detail the generation of cyclopean images, the NSS-based feature extraction and quality prediction mechanisms in Section 3. Experimental results are analyzed in Section 4, and finally, conclusions are drawn in Section 5.

2. Previous Work on SIQA

Typically, there are two types of stereoscopic image quality assessment (SIQA) methods, which are the subjective SIQA and the objective SIQA. For the subjective one, such assessment is conducted by human observers [19], thus being unstable and time-consuming,

hindering its industrial application. Contrarily, the objective SIQA method is conducted by computer algorithms, making it an appropriate solution for industries. Basically, such methods are divided into three types: the full reference (FR) [20–24], the reduced reference (RR) [25,26] and the no reference (NR) [27–30]. In FR-SIQA methods, evaluations are carried out with both the distorted and the original images provided. In RR-SIQA ones, only partial information from the original image, along with the distorted image, is provided, while in NR-SIQA methods, only the distorted image is given.

For FR-SIQA metrics, four categories altogether can be classified as follows. (1) First is directly applying the traditional image quality assessment algorithms on both the left and right stereo images, and then fusing these scores into the final stereoscopic score [22,23]. Such traditional methods include the peak signal-to-noise ratio (PSNR) [31], the structural similarity index (SSIM) [32] and the MS-SSIM [33]; (2) Second is calculating the traditional IQA scores on both the left and right images and simultaneously taking the stereo perceptual information, such as stereo disparity/depth [34], into consideration. For example, You et al. [35] applied different 2D quality metrics on stereo image pairs as well as the generated disparity map, finding a better combination for good performance. Meanwhile, Benoit et al. [36] computed quality scores of both stereo images and their disparity map, combining them to produce a final score. (3) Thirdly, in addition to evaluating stereo pairs, biological models can be used to simulate binocular combination behaviors [37]. Such metrics usually generate cyclopean images based on the binocular combination model and then evaluate the quality of cyclopean images using 2D-IQA metrics, such as the stereoscopic image quality. For example, Chen et al. [38] formulated binocular rivalry problems by modeling binocular combination behaviors based on a cyclopean map and the estimated disparity map. (4) Fourth is simulating human cognition by building human perception models through the mechanism of the human vision system (HVS) or the binocular visual characteristics, such as binocular fusion, binocular rivalry and binocular suppression [39]. However, due to the original non-distorted reference being missing in most cases, FR-SIQA methods are rarely used in real-world scenarios.

Recently, only a few works have focused on the development of RR and NR-SIQA models. Hewage et al. [40] proposed an RR-SIQA algorithm, in which edges are first computed from depth maps, and then the PSNR between the reference and edge maps is used to analyze the 3D image quality. Akhter et al. [41] proposed an NR-SIQA algorithm, which extracts segmented local features of artifacts from stereo-pairs and the estimated disparity map. In ref. [42], Ryu et al. computed the perceptual blurriness and blockiness scores of left and right images. The two scores are then combined into an overall quality index using the binocular quality perception model. Shao's methods [27,43] conduct NR tasks without learning based on human opinion scores. Shao's method [44] specifies the task of SIQA as monocular feature encoding and binocular feature combination based on sparse coding theory. Wang's scheme [45] extracts features from binocular energy response, and machine learning is used to learn a visual quality prediction model; however, it does not consider any 3D visual characteristics. Fang's method [46] combines blurriness- and noisiness-based features and information weight features to assess the quality of stereo-pairs. In ref. [47], NSS-based features are extracted from two synthesized images from the cyclopean image and fed into a stacked neural network model for quality prediction.

Recently, with the growth of the deep learning technology, several convolution neural network (CNN)-based methods have been proposed in pursuit of higher performance gains. For example, Du's method [48] adopts an unsupervised deep network to map images to hierarchical representations. In ref. [49], a general active learning framework is proposed by combining the representativeness and informativeness of samples for active sampling. In ref. [50], an iterative heterogeneous transfer learning framework is proposed that learns a shared feature space for the source and target images and utilizes a novel iterative reweighting strategy to fuse the source samples. Recently, with the advancement of CNNs, several methods also try to implicitly model the feature relations with improved CNN structures. For example, in Zhou et al. [51], a deformable feature extraction block is

explored to simulate the function of the primary visual cortex cells, and in ref. [52], a new network block, called the fusion weights assignment module (FWAM), is proposed to adaptively weigh features to guide the fusion of the stereo image features. The solution in [53] analyzes the features using CNN in the wavelet domain. The method in [54] models the relationship between the extracted features and subjective scores using the general regression neural network (GRNN).

Table 1 depicts the considered binocular characteristics and some implementation details of the state-of-the-art SIQA metrics according to the characteristics of HVS. The FR metrics are in italics. “CYC”, “BD”, “BF”, “BR”, “G/P”, “RM”, “SD” and “TD” denote “Cyclopean”, “Binocular Disparity”, “Binocular Fusion”, “Binocular Rivalry”, “Global/Patch”, “Regression Method”, “Spatial Domain” and “Transformed Domain”, respectively. “Global/Patch” means that the metric is conducted globally across whole images or in patches. “O” and “X” represent “including” and “not including” in the corresponding scheme. The FR metrics are in italics. “NN” denotes a neural network fitting method. It can be seen that the proposed method takes all the listed binocular properties into consideration, while the other metrics lack some of them.

Table 1. The categorization of SIQA studies according to the considered visual characteristics and some implementation details.

Algorithm	CYC	NSS	BD	BF	BR	G/P	RM	SD	TD
Gorley [22]	X	X	X	X	X	G	-	O	X
You [35]	X	X	O	X	X	G	-	O	X
Benoit [36]	X	X	O	X	X	G	-	O	X
Chen [38]	O	X	O	O	O	G	-	O	X
Chen [14]	O	O	O	O	O	G	SVR	O	X
Akhter [41]	X	X	O	X	X	P	-	O	X
Ryu [42]	X	X	X	O	O	G	-	O	O
Wang [45]	X	X	O	X	X	P	SVR	X	O
Shao [27]	X	X	O	O	O	P	-	X	O
Shao [43]	X	X	O	O	O	P	-	X	O
Shao [44]	X	X	X	X	X	P	SVR	X	O
Fang [46]	O	O	O	X	O	G	NN	O	X
Karimi [47]	O	O	O	O	X	G	NN	O	X
Proposed	O	O	O	O	O	G	SVR	O	O

3. The Proposed Algorithm

The flowchart of the proposed NR-SIQA method is shown in Figure 1. Binocular disparity is firstly estimated according to an improved-SSIM-based stereo-matching algorithm. Then, a cyclopean image will be generated using stereoscopic images based on a binocular combination model. Next, both 2D NSS features based on neighbor difference and neighbor product in the spatial domain and gradient magnitude, phase congruency and log-Gabor filter responses in the transformed domain are extracted from the cyclopean image. Meanwhile, 3D NSS features are extracted by considering the binocular rivalry, binocular disparity matching error and binocular consistency. Finally, PCA and machine learning are applied for image score computation through training and testing procedures on these extracted feature vectors.

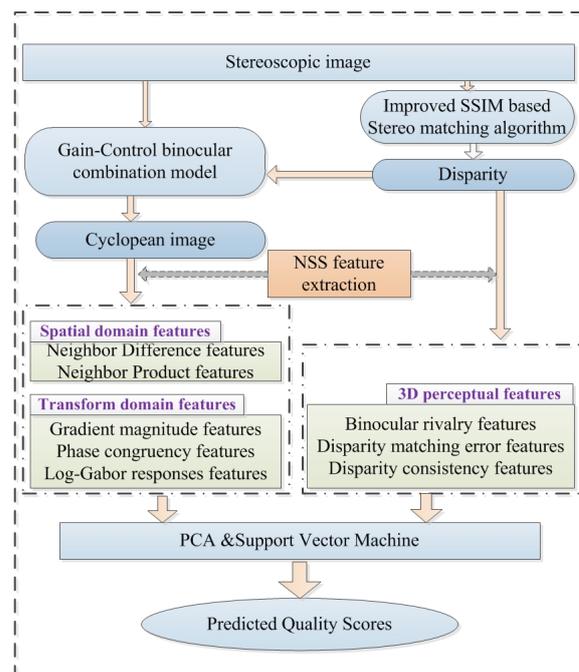


Figure 1. Framework of the proposed NR-SIQA algorithm.

3.1. Disparity Estimation

Binocular disparity is an important factor in stereo visual perception and is utilized to reflect the consistency of left and right images. In the past few years, many disparity prediction methods have been proposed, based on the Sum of Absolute Difference (SAD) [38] or SSIM [32]. For SAD-based method, the disparity is predicted by minimizing the SAD between current pixel and its shifted correspondence in the other view. For SSIM-based methods, the disparity is estimated by maximizing the SSIM scores between two such pixels. However, all these methods only achieve satisfactory results when the input quality is high. In this paper, the improved Gaussian average scheme is added to achieve a better SSIM-based disparity estimation performance in low-quality scenarios.

Specifically, both the pixel-matching condition and that for its neighbors are considered when we search one point in other image. For any point in the left image $[x_l, y_l]$, its corresponding point in the right image is searched within $[x_l - range, y_l]$ to $[x_l + range, y_l]$, and the *range* could be changed with specific settings. The searched-for point having the largest calculated similarity with current point in the left image is chosen to be the matched point. Specifically, the SSIM values between those two points are calculated, and all the SSIM values are merged together with a Gaussian weighted sum. The Gaussian weighted sum of SSIM values, considered as the similarity, is used to search for the best matched point. Figure 2 shows the disparity maps of reference stereo-pair ‘im20_l.bmp’ and ‘im20_r.bmp’ from the LIVE 3D IQA database Phase I [55], estimated using the three above-mentioned methods. It can be seen from Figure 2 that the disparity map estimated by the proposed method has less no-texture and disclosure areas, and the rate of error matching of discontinuous areas is also low.

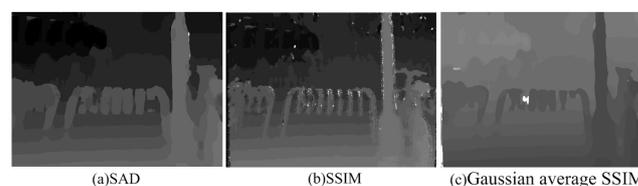


Figure 2. Disparities estimated by different disparity estimation algorithms: (a) SAD-based algorithm, (b) SSIM-based algorithm, (c) Gaussian average SSIM-based algorithm.

3.2. Cyclopean Image

When a human processes the visual stereoscopic images, internally, these signals will be composited into a fusion image, i.e., the cyclopean image, as is analyzed by the binocular vision combination model. To analytically interpret such a process, the Gain-Control Theory model [56] is utilized to conduct binocular fusion and cyclopean perception. Meanwhile, as is suggested in [57], Gabor filter can model the visual processing mechanism of the cells in the primary visual cortex of the human eye in high performance, the normalized Gabor filters are used to weight the paired stereoscopic images. A complex 2D Gabor filter is defined as

$$\begin{aligned}
 GE(x, y, \sigma_x, \sigma_y, \xi_x, \xi_y, \theta) &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left[\left(\frac{R_1}{\sigma_x}\right)^2 + \left(\frac{R_2}{\sigma_y}\right)^2\right]\right) \exp(i(x\xi_x + y\xi_y)) \\
 R_1 &= x\cos\theta + y\sin\theta \\
 R_2 &= -x\sin\theta + y\cos\theta
 \end{aligned} \tag{1}$$

where σ_x and σ_y are the standard deviations of an elliptical Gaussian envelope along x and y axes, ξ_x and ξ_y are the spatial frequencies of filter and θ denotes the orientation of the filter.

In this paper, the first step of cyclopean image generation is to decide the base image between the left image and the right image. In our implementation, the competitive 2D NR-IQA metric, BRISQUE [58] is utilized to evaluate the quality of the left and right images. The image with better quality is selected as the base image, while the other one is chosen as the aid image.

Case 1: The left image is selected as the base image, and the synthesized cyclopean image $I(x, y)$ is calculated as

$$\begin{aligned}
 I(x, y) &= \omega_L(x, y) \cdot I_L(x, y) \\
 &+ \omega_R(x - D_L(x, y), y) \cdot I_R(x - D_L(x, y), y)
 \end{aligned} \tag{2}$$

where

$$\omega_L = \frac{GE_L(x, y)}{GE_L(x, y) + GE_R(x - D_L(x, y), y)} \tag{3}$$

$$\omega_R = \frac{GE_R(x - D_L(x, y), y)}{GE_L(x, y) + GE_R(x - D_L(x, y), y)} \tag{4}$$

Case 2: The right image is selected as base image, and the synthesized cyclopean image $I(x, y)$ can be calculated as

$$\begin{aligned}
 I(x, y) &= \omega_L(x + D_R(x, y), y) \cdot I_L(x + D_R(x, y), y) \\
 &+ \omega_R(x, y) \cdot I_R(x, y)
 \end{aligned} \tag{5}$$

where

$$\omega_L = \frac{GE_L(x + D_R(x, y), y)}{GE_L(x + D_R(x, y), y) + GE_R(x, y)} \tag{6}$$

$$\omega_R = \frac{GE_R(x, y)}{GE_L(x + D_R(x, y), y) + GE_R(x, y)} \tag{7}$$

From Equations (2)–(7), D_L or D_R is the disparity map estimated by fixing the left or right image and shifting right or left image, respectively. ω_L and ω_R are weighting maps computed using the overall response magnitudes GE_L of left image and GE_R of right image on all scales and all orientations, respectively.

Figure 3 shows examples of synthesized cyclopean images. The distorted stereoscopic image is asymmetrically distorted by white noise, as shown in Figure 3a,b. Since

white noise distortion can increase the right stimulus strength, the cyclopean image in Figure 3c is dominated by the right view with white noise. In the other case, when the stereoscopic image is asymmetrically distorted by Gaussian blur, as shown in Figure 3d,e, the simulated cyclopean image in Figure 3f is determined by the good image, i.e., the left image. This is because blur can decrease the Gabor filter response to a great extent, which causes the weight of the right image to be much smaller than that of the left image. This phenomenon is in accord with the subjective/human perception when stereoscopic images are asymmetrically distorted by different types of distortions.

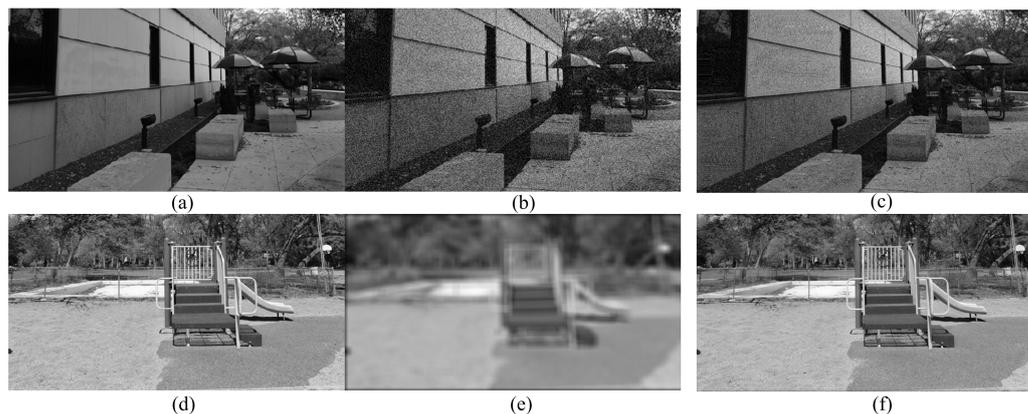


Figure 3. Cyclopean images synthesized by distorted stereo-pairs. (a,b) and (d,e) are left and right views of stereo-pairs, respectively. (c,f) are their respective cyclopean images.

3.3. Feature Extraction

Research on NSS has demonstrated that natural images have certain statistical properties [59], i.e., the probability density distribution of luminance in an image exhibits a Gaussian-like appearance after a local non-linear pre-processing operation called mean subtracted contrast normalization (MSCN) is performed. Applying the local non-linear operation to log-contrast luminances can remove local mean displacements from zero log-contrast and normalize the local variance of the log contrast, thus producing a decorrelating effect [59]. Using these kind of NSS features, several highly competitive 2D NR-IQA models, such as BRISQUE [58], DIIVINE [60] and BLIINDS-II [61], have been developed.

Given a pixel $I_{CC}(x,y)$ of a cyclopean image, the normalized coefficient $\hat{I}_{CC}(x,y)$ can be calculated through such an MSCN operation:

$$\hat{I}_{CC}(x,y) = \frac{I_{CC}(x,y) - \mu(x,y)}{\sigma(x,y) + C} \quad (8)$$

where C is a small constant to increase the stability of $\hat{I}_{CC}(x,y)$ when the denominator tends to zero. In our implementation, a 7×7 Gaussian weighting matrix with a sigma of 1.17 pixels is utilized to compute μ and σ , and C is fixed to 1.

A zero-mean generalized Gaussian distribution (GGD) model is used to fit the probability density distributions of $\hat{I}_{CC}(x,y)$ as

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (9)$$

where $\beta = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$ and $\Gamma(\cdot)$ is the gamma function.

The shape parameter α controls the real ‘shape’ of the distribution, while σ^2 controls the variance. The MSCN operation can show the statistics properties of images with different shape, kurtosis and spread in the presence of distortion. This means that it is possible to predict the type of distortion as well as its perceptual quality through quantifying these changes. The parameters α and σ^2 are effective features, which can be reliably estimated using the moment-matching-based approach in [62].

It also can be found that statistical relationships exist among neighboring pixels along different orientations [58]. The empirical distribution of the statistical relationships between pixels along different orientations is not logically symmetric about zero. To better fit the empirical histograms of the relationships, the very general asymmetric generalized Gaussian distribution (AGGD) model [63] is adopted. The AGGD with zero mode is given by

$$f(x; v, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{v}{(\beta_l + \beta_r)\Gamma(1/v)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^v\right), & x < 0 \\ \frac{v}{(\beta_l + \beta_r)\Gamma(1/v)} \exp\left(-\left(\frac{-x}{\beta_r}\right)^v\right), & x \geq 0 \end{cases} \quad (10)$$

where $\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}}$ and $\beta_r = \sigma_r \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}}$

The shape parameter v controls the real ‘shape’ of the distribution, while σ_l^2 and σ_r^2 are scale parameters that control the variance on each side of the mode. The parameters of the AGGD distribution are also estimated using the moment-matching-based approach in [62]. All four parameters ($\eta, v, \sigma_l^2, \sigma_r^2$) of AGGD are extracted as features where η is computed by

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(2/v)}{\Gamma(1/v)} \quad (11)$$

Inspired by the success of NSS in image quality assessment, a large amount of images with different contents and different distortions are selected from the LIVE database [64], and a series of experiments are conducted based on these images to prove that there is identical statistical regularity between different contents of images and that the distortions will remarkably disturb the statistical distributions of the images. The reference (REF) image “*building2.bmp*” and its five distorted versions, including such distortions as jpeg2000 (JP2K), JPEG, white noise (WN), Gaussian blur (BLUR) and fast fading (FF), are chosen as examples in Figures 4 and 5a–c to show the experimental results. The NSS feature extraction and explanation of Figures 4 and 5a–c are specified as follows.

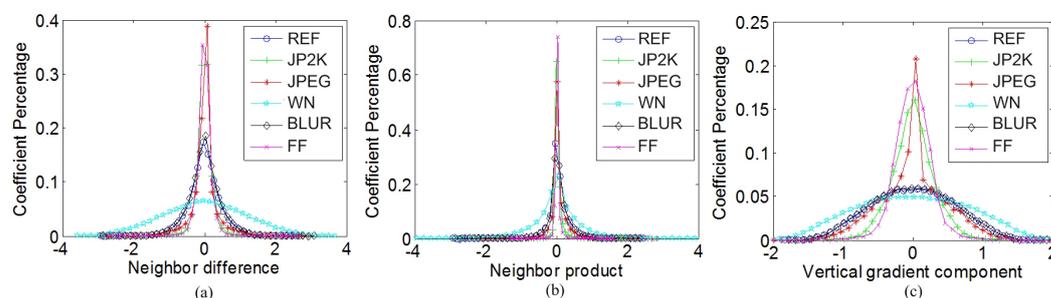


Figure 4. Probability density distributions of (a) neighbor difference, (b) neighbor product and (c) vertical gradient magnitude for six natural cyclopean images synthesized by a reference stereo-pair and its five distorted versions.

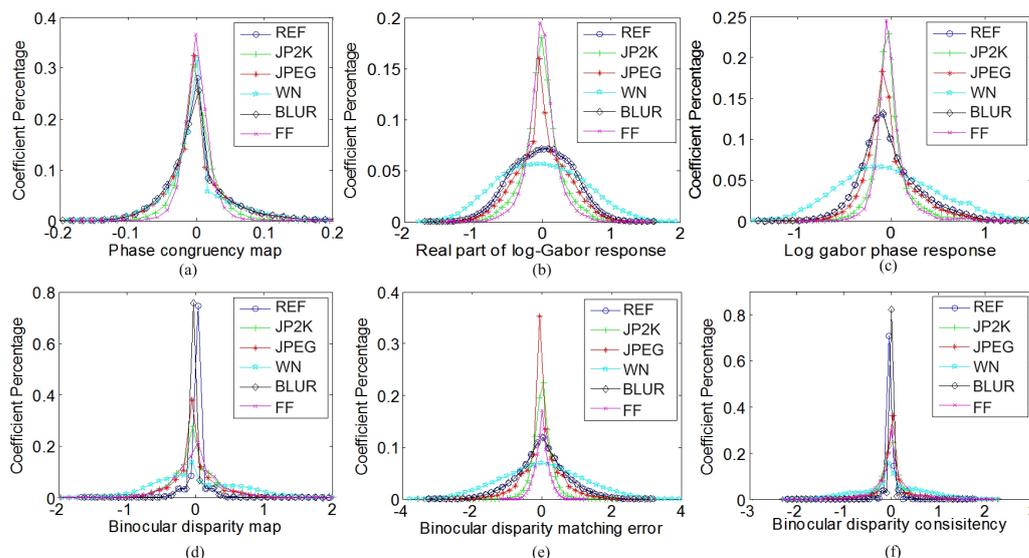


Figure 5. Probability density distributions of (a) phase congruency, (b) real part of log-Gabor response, (c) log-Gabor phase response, (d) binocular disparity, (e) binocular disparity matching error and (f) binocular disparity consistency computed from a reference stereo-pair and its five distorted versions.

3.3.1. Neighbor Difference-Based 2D Features

There is high correlation between surrounding pixels of an image in the spatial domain [58]. The difference between neighboring pixels and the product of neighboring pixels can effectively capture information on the image quality.

The difference between adjacent MSCN coefficients $I_{\hat{C}C}(x, y)$ of a cyclopean image are computed at a distance of one pixel along four orientations: horizontal (H), vertical (V), main-diagonal (D1) and secondary-diagonal (D2), which is illustrated in Figure 6. Specifically,

$$\begin{aligned}
 dif_H(i, j) &= I_{\hat{C}C}(i, j) - I_{\hat{C}C}(i, j + 1) \\
 dif_V(i, j) &= I_{\hat{C}C}(i, j) - I_{\hat{C}C}(i + 1, j) \\
 dif_{D1}(i, j) &= I_{\hat{C}C}(i, j) - I_{\hat{C}C}(i + 1, j + 1) \\
 dif_{D2}(i, j) &= I_{\hat{C}C}(i, j) - I_{\hat{C}C}(i + 1, j - 1)
 \end{aligned} \tag{12}$$

where $i \in \{2, \dots, M - 1\}$ and $j \in \{2, \dots, N - 1\}$ and M, N are the image width and height, respectively.

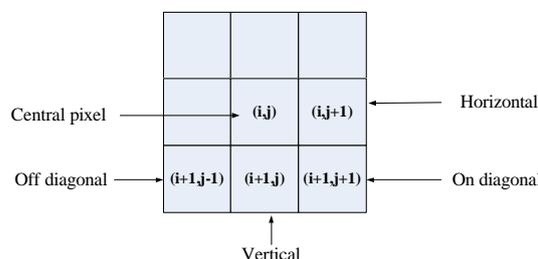


Figure 6. Various pairwise differences computed to quantify neighboring pixel statistical correlation. Neighboring pixel differences are computed along four orientations—horizontal (H), vertical (V), main(on)-diagonal (D1) and secondary(off)-diagonal (D2).

Figure 4a shows the probability density distributions of neighbor differences along the horizontal orientation. Neighbor differences along these four orientations are parameterized using the GGD model to effectively capture the image statistics in the presence

of distortion. By estimating two parameters, α and σ^2 , in each orientation as the spatial domain features, eight features are obtained overall.

3.3.2. Neighbor Product-Based 2D Features

In Figure 7, we also model the statistical correlation of pairwise products of neighbor pixels at a distance of two pixels along eight orientations, i.e., 0° , 22.5° , 45° , 67.5° , 90° , 112.5° , 135° and 157.5° . Specifically,

$$\begin{aligned}
 pro_1(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i, j + 2) \\
 pro_2(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 1, j + 2) \\
 pro_3(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 2, j + 2) \\
 pro_4(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 2, j + 1) \\
 pro_5(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 2, j) \\
 pro_6(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 2, j - 1) \\
 pro_7(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 2, j - 2) \\
 pro_8(i, j) &= \hat{I}_{CC}(i, j) \cdot \hat{I}_{CC}(i + 1, j - 2)
 \end{aligned} \tag{13}$$

where $i \in \{3, \dots, M - 2\}$ and $j \in \{3, \dots, N - 2\}$ and M, N are the image width and height, respectively.

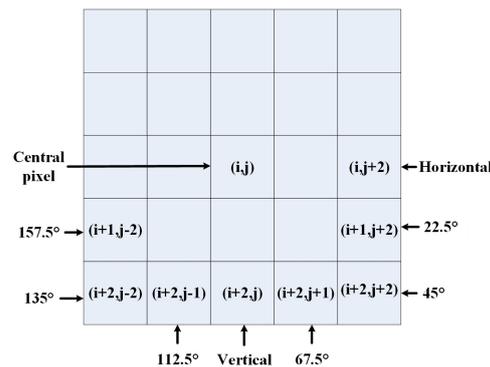


Figure 7. Various pairwise products computed to quantify neighboring pixels' statistical correlations. Neighboring pixel products are computed along eight orientations— 0° , 22.5° , 45° , 67.5° , 90° , 112.5° , 135° and 157.5° —at a distance of 2 pixels from central pixel.

Figure 4b shows examples of the probability density distributions of neighbor products along the horizontal orientation, i.e., $pro_1(i, j)$. We utilize the AGGD model to fit the empirical histograms of the paired neighbor products, and the parameters $(\eta, v, \sigma_l^2, \sigma_r^2)$ are estimated as the spatial domain features. Finally, 32 features are extracted for these 8 paired products.

3.3.3. Gradient Magnitude-Based 2D Features

As suggested in [65], image gradient magnitudes (GM) can reflect some properties of images. Specifically, GM is computed to be the root mean square of the directional gradients in the horizontal and vertical directions. In this paper, the Scharr operators [66] are applied along the horizontal (x) and vertical (y) directions, which are defined as

$$h_x = \frac{1}{16} \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix}, h_y = \frac{1}{16} \begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \tag{14}$$

Convolving h_x and h_y with cyclopean image I_{CC} produces the horizontal gradient $GM_x(x)$ and vertical gradient $GM_y(y)$. The gradient magnitude of the cyclopean image is defined as

$$\begin{aligned} GM(x) &= \sqrt{(I_{CC} \otimes h_x)^2 + (I_{CC} \otimes h_y)^2} \\ &= \sqrt{GM_x^2(x) + GM_y^2(y)} \end{aligned} \quad (15)$$

where the symbol “ \otimes ” denotes the convolution operation.

Figure 4c shows the probability density distributions of $GM_y(y)$ after the MSCN operation. The GGD model is utilized to fit the distributions of $GM(x)$, $GM_x(x)$ and $GM_y(y)$ after MSCN, and the parameters α and σ^2 are estimated as the NSS-based transformed domain features.

3.3.4. Phase Congruency-Based 2D Features

Phase congruency (PC) reflects the image changes in brightness or contrast, which represents the spatial and phase locations where the maximum Fourier response is received [67]. This can be considered as the coherency degree of the image’s local frequencies. In this paper, PC is computed by Kovess’s salient work [67] on the cyclopean image.

For cyclopean image I_{CC} , the multi-direction and multi-scale complex Gabor filters are applied. M_n^e and M_n^o denote the even-symmetric (cosine) and odd-symmetric (sine) filters at scale n , respectively, forming a quadrature pair. By convolving I_{CC} with these filters, the responses will form a vector at position x , on scale n and orientation θ :

$$[e_{n,\theta}(x), o_{n,\theta}(x)] = [I_{CC} \otimes M_n^e, I_{CC} \otimes M_n^o] \quad (16)$$

where the symbol “ \otimes ” denotes the convolution operation and θ is the orientation angle of the filters. The local amplitude of point x on scale n and orientation θ is computed by

$$A_{n,\theta}(x) = \sqrt{e_{n,\theta}^2(x) + o_{n,\theta}^2(x)}. \quad (17)$$

The local energy along orientation θ is computed by

$$E_\theta(x) = \sqrt{F_\theta^2(x) + H_\theta^2(x)} \quad (18)$$

where $F_\theta(x) = \sum_n e_{n,\theta}(x)$, $H_\theta(x) = \sum_n o_{n,\theta}(x)$. Then, the phase congruency of the cyclopean image at point x is defined by

$$PC(x) = \frac{\sum_j E_\theta(x)}{\epsilon + \sum_n \sum_j A_{n,\theta}(x)} \quad (19)$$

where ϵ is a small positive constant, j is the orientation angle number and n is the scale number.

Figure 5a shows the probability density distributions of the phase congruency map after MSCN is performed. To extract features from the phase congruency map, the AGGD model is used to fit the MSCN coefficients of phase congruency and AGGD model parameters $(\eta, v, \sigma_r^2, \sigma_f^2)$ are extracted as the transformed domain features.

3.3.5. Log-Gabor Response-Based 2D Features

Since neurons in the visual cortex selectively respond to stimuli in different orientations and frequencies, filters with multiple scales and multiple orientations are useful to simulate this physiological phenomenon [57]. These filters can closely model frequency–orientation decompositions in the primary visual cortex and can capture energy in a highly localized manner in both space and frequency domains [68]. Such filter responses of an image are

also useful for generating quality-aware NR-IQA features. Here, perceptually relevant log-Gabor filters are exploited to accomplish multi-scale and multi-orientation filtering.

A 2D log-Gabor filter with multiple frequencies and multiple orientations in the Fourier domain can be expressed as

$$G_{log}(\omega, \theta) = \exp\left(-\frac{(\log(\omega/\omega_0))^2}{2\sigma_r^2}\right) \cdot \exp\left(-\frac{(\theta - \theta_j)^2}{2\sigma_\theta^2}\right) \quad (20)$$

where $\theta_j = j\pi/J, j = \{0, 1, \dots, J-1\}$ is the orientation of the log-Gabor filter, J is the overall number of orientations, ω_0 is the center frequency of the filter, σ_r determines the filter's radial bandwidth and σ_θ controls the filter's angular bandwidth. After convolving with the cyclopean image I_{CC} , the local responses of the real part, $e_{n,\theta_j}(x)$, and imaginary part, $o_{n,\theta_j}(x)$, of the log-Gabor filter are obtained at position x on scale n and orientation θ_j .

The log-Gabor filter response of cyclopean image I_{CC} is expressed as

$$G_{log}(x) = \sum_n \sum_j \sqrt{e_{n,\theta_j}^2(x) + o_{n,\theta_j}^2(x)}. \quad (21)$$

The real and imaginary parts of the log-Gabor response of cyclopean image I_{CC} are respectively expressed as follows:

$$\begin{aligned} Gx_{log}(x) &= \sum_n \sum_j e_{n,\theta_j}^2(x) \\ Gy_{log}(x) &= \sum_n \sum_j o_{n,\theta_j}^2(x) \end{aligned} \quad (22)$$

The log-Gabor filter phase of cyclopean image I_{CC} is expressed as

$$GP_{log}(x) = \sum_n \sum_j \arctan(e_{n,\theta_j}(x), o_{n,\theta_j}(x)) \quad (23)$$

Figure 5b,c show the probability density distributions of $Gx_{log}(x)$ and $GP_{log}(x)$ after MSCN operation, respectively. The probability density distributions of $G_{log}(x)$, $Gx_{log}(x)$, $Gy_{log}(x)$ and $GP_{log}(x)$ after MSCN is performed can be fitted using the model of GGD, and the parameters α and σ^2 are included in the feature set.

3.3.6. Binocular Rivalry-Based 3D Features

Binocular disparity or depth is an important factor in stereo visual perception. Binocular disparity is computed by the distance between the corresponding pixels in the left and right images. The binocular disparity can be regarded as the measurement that the human brain senses the depth information with. In general, it can lead to binocular rivalry and binocular suppression. As indicated by [35,38], the disparity or depth information has an impact on stereoscopic image quality. Therefore, the binocular disparity can be used to generate the binocular rivalry properties and enhance the detection accuracy of the saliency map.

According to visual psychophysical research, binocular combination will integrate two retinal points into a single binocular perception when similar contents of the left and right images fall on the corresponding retinal points [69]. Binocular rivalry happens when the content of the left and right images is relatively different. Therefore, the left and right eyes view mismatched images at the same retinal locations. Numerous studies suggest that binocular rivalry is the result of competition between human eyes [70]. This competition is caused by the mismatched content of the left and right images. Binocular suppression is a special case of binocular rivalry. When binocular suppression happens, the entire image from one retina is partly suppressed [1].

We use the binocular disparity estimated in Section 3.1 to describe the binocular rivalry. A variety of stereo-pairs with different distortions and contents are selected from the LIVE 3D database Phase I [55] to validate the effectiveness of the binocular rivalry-based 3D features and the following 3D features. The statistical distributions of the stereo-pair “*im17_l.bmp*” and “*im17_r.bmp*” are chosen as an example in Figure 5d–f. Figure 5d shows the probability density distributions of the binocular disparity map after MSCN operation. We use parameters α and σ^2 of the GGD model to represent the statistical features of binocular disparity.

3.3.7. Binocular Disparity Matching Error-Based 3D Features

Except for the estimated binocular disparity, the disparity matching error caused by the improved SSIM-based stereo matching algorithm is also a useful feature related to the stereo image quality. One pixel fails to find the corresponding pixel in the other image in a stereo-pair when large matching error exists. The disparity matching error is defined as follows:

$$\begin{aligned} D_{Lerror}(x, y) &= I_L(x, y) - I_R(x - D_L(x, y), y) \\ D_{Rerror}(x, y) &= I_L(x + D_R(x, y), y) - I_R(x, y) \end{aligned} \quad (24)$$

where D_L and D_R are disparity maps estimated using the left and right images as the base image, respectively. $I_L(x + D_R(x, y), y)$ and $I_R(x - D_L(x, y), y)$ are the left and right disparity-compensated images, respectively.

Figure 5e shows the probability density distribution of binocular disparity matching error map after MSCN operation. Then, the GGD model is utilized to fit the probability density distribution, and the parameters (α, σ^2) are estimated as 3D features.

3.3.8. Binocular Disparity Consistency-Based 3D Features

Like normal natural images, there is high correlation between neighboring pixels, except in some edge regions in a disparity map. Thus, one point in a disparity map is usually similar to its surrounding points when the disparity map is computed from an original, undistorted stereoscopic image. We describe the disparity consistency by using a high-pass filter to convolve with the estimated disparity map $D(x, y)$. The disparity consistency is defined as

$$D_{consist}(x, y) = D(x, y) \otimes \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & -1 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{bmatrix} \quad (25)$$

The value of $D_{consist}(x, y)$ would tend to zero if the stereoscopic image has no distortion or little distortion. Figure 5f shows the probability density distributions of disparity consistency maps after MSCN operation. These six corresponding disparity maps are extracted from the stereo-pair *im17.bmp* and its five distorted versions from the LIVE 3D database Phase I. It can be seen that the coefficient values from a disparity consistency map mostly concentrate to zero. However, when the disparity is obtained from a distorted stereo-pair, the coefficient values of its disparity consistency map are more likely to spread out. In addition, the probability of coefficient values equal to zero is much lower.

The GGD model is used to fit the empirical distribution of $D_{consist}(x, y)$ after the MSCN operation is performed, and the GGD parameters (α, σ^2) are extracted as features.

3.4. Machine Learning

To adaptively choose a salient feature set for better regression model fitting, the principal component analysis (PCA) is applied, and the chosen salient features are sent into the support vector regressor (SVR) for the final fitting of the objective quality score. In the training stage, the generated features and the labeled subjective quality scores are sent to the regression model, and the training epoch is 1000. In the testing stage, the predicted

objective scores are obtained from the trained regressor. Our implementation is based on the MATLAB platform, and the SVR is conducted using the LIBSVM package [2]. For the SVR's settings, the epsilon-SVR with a radial basis function (RBF) kernel is used, and the SVR parameters C and γ are set to 512 and 0.015625, respectively, according to extensive experiments. The details about the parameter selection can be found in Section 4.2.

4. Experimental Results and Analysis

4.1. Databases and Evaluation Criteria

For training and testing datasets, we adopt the LIVE 3D IQA database Phase I [1] and Phase II [14] and the Waterloo IVC (UW/IVC) 3D IQA database [3] for testing the proposed algorithm and other SIQA algorithms. For LIVE 3D Phase I, 20 reference stereoscopic images and 365 distorted stereoscopic pairs are included, each having 5 types of distortions, i.e., JP2K compression distortion, JPEG compression distortion, white noise (WN), fast fading (FF) and blurring (BLUR). All distortions are symmetric in nature. Similarly, for LIVE 3D Phase II, 8 reference stereoscopic images, 120 symmetrically distorted stereoscopic pairs and 240 asymmetrically distorted stereoscopic pairs generated from those 8 references are included. The Waterloo IVC 3D IQA database also contains two data phases, created from 6 and 10 pristine stereo-pairs, respectively. Both of these phases include symmetrical and asymmetrical distortions, including additive white Gaussian noise, Gaussian blur and JPEG compression. Each type of distortion has four levels. Altogether, a total of 78 single-view images and 330 stereoscopic images are contained in Phase I, and a total of 130 single-view images and 460 stereoscopic images are included in Phase II.

To evaluate the overall accuracy of the stereoscopic objective assessment, three criteria are used: the Pearson linear correlation coefficient (PLCC), the Spearman rank order correlation coefficient (SROCC) and the root mean squared error (RMSE). Traditionally, the PLCC and RMSE are used to evaluate the prediction accuracy of SIQA metrics, and the SROCC detects the prediction's monotonicity. Note that the higher the SIQA's accuracy, the higher the values PLCC and SROCC will be, while the value of the RMSE should be lower. Before the evaluation, a non-linear regression analysis, suggested by the video quality experts group [4], is utilized to provide mappings between the objective scores and subjective mean opinion scores (MOSs). Specifically, a five-parameter logistic function [10] is adopted as such a non-linear mapping:

$$f(x) = \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x - \beta_3)}} \right) + \beta_4 \cdot x + \beta_5 \quad (26)$$

where $\beta_i = 1, 2, \dots, 5$ are parameters determined by the subjective scores and the objective scores.

4.2. Implementation Details

The implementation of the whole procedure is conducted on MATLAB, and the support vector regressor is conducted using LibSVM. The PCA is introduced and its effectiveness is validated by various experiments. The following experiments are also conducted to select the SVR parameters (C , γ) and prove that almost no over-fitting exists in our method.

4.2.1. PCA for Feature Dimension Reduction

In the proposed algorithm, more than 60 features are extracted from the 2D cue and 3D cue for each image. Inevitably, information redundancy may be contained in these NSS features (i.e., the shape feature). To highlight the main component and reduce the dimensions of the feature vectors, PCA is utilized, resulting in a more efficient quality predictor. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ denote the feature matrix extracted from n images, and d is the number of features extracted from each image. By applying PCA to \mathbf{X} , a transform matrix, $\Phi \in \mathbb{R}^{d \times m}$, will be obtained by an eigenvalue decomposition conducted on the covariance matrix of the feature matrix \mathbf{X} . The transform matrix is formed by the m ($m < d$) principle

vectors associated with the m most significant eigenvalues of the covariance matrix of \mathbf{X} . Then, the dimension of each feature vector \mathbf{x}_i can be reduced by transformation as follows:

$$\mathbf{x}'_i = \mathbf{\Phi}^T \mathbf{x}_i, \mathbf{x}'_i \in \mathbb{R}^{m \times 1} \quad i = 1, 2, \dots, n \quad (27)$$

The PLCC with various PCA dimensions in LIVE Phase I and II is shown in Figure 8. As the number of feature dimensions is set to 44, the performance is the best. m is set to 44 in our implementation.

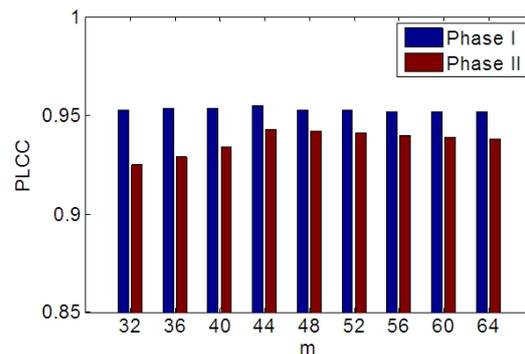


Figure 8. PLCC of the proposed metric with various feature dimensions from LIVE Phase I and Phase II.

4.2.2. SVR Parameters (C, γ) Selection

A cross-validation experiment is conducted to choose the value of (C, γ) . During this experiment, samples from LIVE 3D Phase I and II are randomly partitioned into two sets: the training set and the testing set. For each phase, 80% of the images are for training and the remaining 20% of the images are for testing. The training epoch is 1000, and the median results are reported for performance evaluation. Parameters (C, γ) delivering the best median result are chosen as the best. The illustration of this experimental process can be found in Figure 9. The number of the level contour indicates the SROCC values for the cross-validation. Apparently, there exists a circular region surrounding the center, where the proposed metric obtains the best SROCC results on the two databases. This indicates that this learning model is more robust to different databases of samples. According to the results, the optimal parameters (C, γ) are set to be 512, 0.015625.

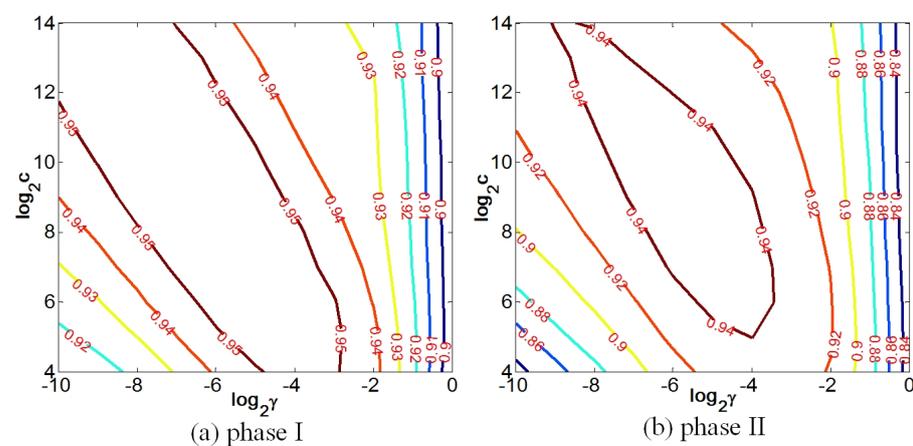


Figure 9. The SVR parameters' (C, γ) selection process on (a) LIVE 3D Database Phase I and (b) Phase II. The number on the level contour denotes the SROCC value of the cross-validation.

4.2.3. Two-Fold Cross-Validation

Each database is randomly split into two sets: the training set and the testing set, with a 50:50 splitting ratio. Performance evaluations are conducted as follows. First, the optimal regression model is learned from the training set, and then tested on the training set and testing set, respectively. Next, the other regression model is learned from the testing set, and then tested on the training set and testing set, respectively. Performance evaluations on LIVE Phase I and II are listed in Table 2 (“MTrain→Testing set” means that the model is learned from the training set is tested on the testing set, and “MTest→Training set” represents that the model learned from the testing set is tested on the training set). As is seen on the table, regardless of whether the regression model is learned from the training set or the testing set, no obvious difference exists between the evaluated performance, indicating that almost no over-fitting exists in the proposed metric.

Table 2. Performance on the training set and testing set as the regression model learned from the TRAINING set and TESTING set, respectively.

Database Training/Testing	LIVE Phase I			LIVE Phase II		
	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
MTrain→Training set	0.978	0.979	3.265	0.971	0.976	2.452
MTrain→Testing set	0.948	0.958	4.807	0.933	0.939	3.929
MTest→Training set	0.948	0.956	4.826	0.934	0.940	3.881
MTest→Testing set	0.977	0.980	3.282	0.971	0.976	2.482

4.2.4. Features with Noise-Addition

Here, we test the impact of Gaussian white noise (WN) on the model’s fitting performance. Similarly, LIVE 3D Phase I and II are randomly partitioned into the training set and the testing set with the ratio 80:20. All the features of the samples in the training set are applied with white noise before the experiment. The variance of the white noise is set to 0.01. Performance evaluations are conducted on the training set affected or not affected by white noise, and then tested on the non-affected training set and testing set, respectively. The results are listed in Table 3, where “WN→Training set” means that the model is learned from the WN-affected training set and is tested on the non-affected training set, and “WN/F→Testing set” denotes that the model is learned from the non-affected training set and is tested on the non-affected testing set. As is seen in the table, there is no big difference between the model’s performance when tested on the training set or testing set. Moreover, the results also indicate that white noise has barely any impact on the model’s fitting process. Therefore, we can make a conclusion that almost no over-fitting exists in the proposed metric.

Table 3. Performance on training set and testing set using a regression model learned from the features of a training set with Gaussian white noise added or with no noise.

Database Training/Testing	LIVE Phase I			LIVE Phase II		
	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
WN→Training set	0.962	0.967	4.234	0.950	0.954	3.386
WN→Testing set	0.944	0.954	4.956	0.924	0.936	4.000
WN/F→Training set	0.976	0.978	3.401	0.969	0.973	2.259
WN/F→Testing set	0.952	0.962	4.493	0.940	0.950	3.546

4.3. Performance Comparison

In the proposed method, 80% images of each database are chosen randomly as training set and the rest is for testing. The training-testing process is repeated 1000 times, and the average of performance is reported for comparison.

4.3.1. Overall Performance Comparison

The overall performance comparisons of the proposed algorithm and other NR-SIQA methods on both LIVE Phase I and II are shown in Table 4. Shao [27,44] split the LIVE Phase II into symmetrically and asymmetrically distorted dataset, and thus the overall performance on whole LIVE Phase II is unavailable, which represented by “-”. It can be seen that the proposed method achieves much higher performance than other state-of-the-art metrics. Besides, it also can be seen from Table 4 that the proposed metric and other NR-SIQA metrics achieve a better performance on Phase I than those on Phase II. The reason is that Phase I only contains symmetric distortions, while Phase II contains both symmetric and asymmetric distortions. It is much more difficult to evaluate the stereoscopic image quality with asymmetric distortions due to the limited understanding of HVS when human watch asymmetrically distorted stereoscopic images.

Table 4. Performance on LIVE 3D database Phase I and II compared with the existing NR SIQA methods.

Database	LIVE Phase I			LIVE Phase II		
	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
Akhter [41]	0.383	0.626	14.827	0.543	0.568	9.249
Chen [14]	0.891	0.895	7.247	0.880	0.895	5.102
Wang [45]	0.828	0.885	7.238	0.794	0.784	7.326
Shao [27]	0.894	0.899	-	-	-	-
Shao [44]	0.950	0.957	-	-	-	-
Fang [46]	0.932	0.936	-	0.931	0.936	-
Proposed	0.952	0.962	4.493	0.940	0.950	3.546

Table 5 presents the comparison results between the proposed method and other FR and NR metrics on Waterloo IVC 3D database. In Table 5, the FR methods are marked in italic. Specifically, Mittal et al. [58] is a 2D NR metric; therefore, its performance is evaluated on the both left and right views, resulting two objective scores. The final score for this method is the average of these two scores. It can be seen that the proposed method outperforms the other metrics. The reason that Chen’s [14] method performs worse than Mittal’s [58] may be because its stereo-matching algorithm fails to generate precise disparity.

Table 5. Performance on Waterloo IVC 3D database Phase I and II compared with the existing FR and NR SIQA methods.

Database	UW/IVC Phase I		UW/IVC Phase II	
	SROCC	PLCC	SROCC	PLCC
You [35]	0.597	0.713	0.587	0.682
Chen [38]	0.682	0.734	0.578	0.613
Mittal [58]	0.845	0.869	0.794	0.849
Chen [14]	0.708	0.715	0.547	0.551
Proposed	0.906	0.919	0.852	0.863

4.3.2. Performance Comparison of Each Distortion Type

Here, we discuss the performance comparisons over different distortions. Several state-of-the-art FR-SIQA metrics (i.e., Benoit [36], You [35], Gorley [22] and Chen [38]), along with our method, are tested on LIVE Phase I and II, and the results are listed in Table 6. Note that the best scores are marked in bold in that table. According to the table, the proposed algorithm achieves a slightly higher performance than the algorithm in [44], while performing better than other FR-SIQA and NR-SIQA metrics. As FR-SIQA metrics, Gorley et al. [22] predicts the overall quality by averaging the scores of left and right views generated from 2D IQA algorithms. Its performance is much worse than other SIQA

methods because it does not take the binocular visual properties into account. In addition to the dual-2D quality scores for both views, You et al. [35] and Benoit et al. [36] simultaneously consider disparity information, and the performances of these two methods are better than Gorley et al., as their SROCC and PLCC scores reach up to 0.90 on LIVE Phase I. Chen et al.'s model [38] applies the SSIM to assess the quality of cyclopean images, making its SROCC and PLCC scores reach up to 0.920 on LIVE Phase I. For NR-SIQA metrics, Shao et al. [27] may be opinion-unaware; however, its performance is inferior to other methods based on prior subjective human opinion scores. Shao et al. [44] achieves better performance than the other methods, except for the proposed one, as it lacks the binocular properties of our method. The performance of method [41] is barely satisfactory on JP2K and JPEG distortion in the LIVE Phase I database. For Chen et al.'s method [14], it extracts NSS-based features of the cyclopean image in the spatial domain, but it lacks considerations of the binocular rivalry effect and other 3D visual perceptions. Its performance is worse than our proposed algorithm, as our method sufficiently extracts 2D and 3D features, and takes 3D visual characteristics into consideration.

Table 6. Performance on each type of distortion on LIVE 3D database Phase I and Phase II.

Database		LIVE Phase I						LIVE Phase II					
Algorithm		JP2K	JPEG	WN	BLUR	FF	ALL	JP2K	JPEG	WN	BLUR	FF	ALL
SROCC	Benoit [36]	0.910	0.603	0.930	0.931	0.699	0.899	0.751	0.867	0.923	0.455	0.773	0.728
	You [35]	0.860	0.439	0.940	0.882	0.588	0.878	0.894	0.795	0.909	0.813	0.891	0.786
	Gorley [22]	0.015	0.569	0.741	0.750	0.366	0.142	0.110	0.027	0.875	0.770	0.601	0.146
	Chen [38]	0.888	0.530	0.948	0.925	0.707	0.916	0.814	0.843	0.940	0.908	0.884	0.889
	Shao [27]	0.900	0.607	0.926	0.924	-	0.894	-	-	-	-	-	-
	Shao [44]	0.936	0.818	0.935	0.927	0.814	0.950	-	-	-	-	-	-
	Akhter [41]	0.914	0.866	0.675	0.555	0.640	0.383	0.714	0.724	0.649	0.682	0.559	0.543
	Chen [14]	0.919	0.863	0.617	0.878	0.652	0.891	0.950	0.867	0.867	0.900	0.933	0.880
	Proposed	0.937	0.779	0.959	0.921	0.851	0.952	0.946	0.903	0.831	0.912	0.939	0.940
PLCC	Benoit [36]	0.939	0.640	0.925	0.948	0.747	0.902	0.784	0.853	0.926	0.535	0.807	0.748
	You [35]	0.877	0.487	0.941	0.919	0.730	0.881	0.905	0.830	0.912	0.784	0.915	0.800
	Gorley [22]	0.485	0.312	0.796	0.852	0.364	0.451	0.372	0.322	0.874	0.934	0.706	0.515
	Chen [38]	0.912	0.603	0.942	0.942	0.776	0.917	0.834	0.862	0.957	0.963	0.901	0.900
	Shao [27]	0.872	0.597	0.916	0.923	-	0.899	-	-	-	-	-	-
	Shao [44]	0.949	0.796	0.938	0.986	0.837	0.957	-	-	-	-	-	-
	Akhter [41]	0.904	0.905	0.729	0.617	0.503	0.626	0.722	0.776	0.786	0.795	0.674	0.568
	Chen [14]	0.917	0.907	0.695	0.917	0.735	0.895	0.947	0.899	0.901	0.941	0.932	0.895
	Proposed	0.958	0.801	0.971	0.965	0.883	0.962	0.974	0.922	0.858	0.977	0.949	0.950
RMSE	Benoit [36]	4.426	5.022	6.307	4.571	8.257	7.061	6.096	3.787	4.028	11.763	6.894	7.490
	You [35]	6.206	5.709	5.621	5.679	8.492	7.746	4.186	4.086	4.396	8.649	4.649	6.772
	Gorley [22]	11.323	6.211	10.197	7.562	11.569	14.635	9.113	6.940	5.202	4.988	8.155	9.675
	Chen [38]	5.320	5.216	5.581	4.822	7.837	6.533	5.562	3.865	3.368	3.747	4.966	4.987
	Shao [27]	-	-	-	-	-	-	-	-	-	-	-	-
	Shao [44]	-	-	-	-	-	-	-	-	-	-	-	-
	Akhter [41]	7.092	5.483	4.273	11.387	9.332	14.827	7.416	6.189	4.535	8.450	8.505	9.249
	Chen [14]	6.433	5.402	4.523	5.898	8.322	7.247	3.513	4.298	3.342	4.725	4.180	5.102
	Proposed	3.938	3.908	3.912	4.055	5.812	4.493	2.575	3.773	2.843	3.207	3.606	3.546

4.4. Performance Evaluation

4.4.1. Performance Evaluation of Each Feature Type

In our metric, features including neighbor difference (ND)-based NSS features and neighbor product (NP)-based NSS features in the spatial domain and gradient magnitude (GM)-based NSS features, phase congruency (PC)-based NSS features, log-Gabor filter response (GR)-based NSS features in the transformed domain and 3D information (3DI), such as binocular rivalry, binocular disparity matching error and binocular disparity consistency NSS features, are extracted. In order to verify whether some of these features

could be ignored, we analyze the gain of each type of feature in Table 7. Note that this experiment is conducted with an incomplete feature set; thus, the PCA process is not used to reduce the feature dimensions. We can see from Table 7 that each type of feature can improve the performance of our model.

Table 7. Performance with feature type increasing on LIVE 3D database Phase I and Phase II.

Database	LIVE Phase I			LIVE Phase II			
	Feature Set	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
PC		0.868	0.879	7.315	0.736	0.775	7.057
PC+GM		0.893	0.912	6.812	0.793	0.857	5.852
PC+GM+GR		0.903	0.922	6.201	0.832	0.869	5.031
PC+GM+GR+ND		0.931	0.943	5.815	0.874	0.891	4.764
PC+GM+GR+ND+NP		0.938	0.948	5.128	0.907	0.918	4.197
PC+GM+GR+ND+NP+3DI		0.948	0.957	4.914	0.932	0.939	3.867

4.4.2. Cross-Database Performance Evaluation

The evaluation strategy used in Section 4.3 is inadequate to evaluate the generalization capability of a NR-SIQA model, since the random partition method on a single database leads to the training set and the testing set containing the same types of distortions. Therefore, to demonstrate the efficiency of an IQA model on other possibly unknown distortions, and to verify such methodologies are actually “completely blind”, we test the generalization capability of the proposed method by training the regression model on one database, i.e., LIVE Phase I, and testing the regression model on another database, i.e., LIVE Phase II.

In our implementation, we train the regression model on LIVE Phase I and test it on LIVE Phase II on one hand, and on the other hand, we train the model on LIVE Phase II and test it on LIVE Phase I. The results are reported in Tables 8 and 9, from which we can draw the conclusion that the regression model that is trained on one database (i.e., LIVE Phase I) is not completely efficient for testing on the other database (i.e., LIVE Phase II), since the contents in the two databases are quite different. Phase II includes both symmetrically and asymmetrically distorted stereo-pairs, while Phase I only contains symmetrically distorted stereo-pairs.

Table 8. Performance on LIVE 3D database Phase II training on Phase I.

Criteria	JP2K	JPEG	WN	BLUR	FF	ALL
SROCC	0.864	0.573	0.888	0.883	0.872	0.802
PLCC	0.859	0.581	0.908	0.959	0.889	0.822
RMSE	5.029	5.967	4.625	3.948	5.261	6.427

Table 9. Performance on LIVE 3D database Phase I training on Phase II.

Criteria	JP2K	JPEG	WN	BLUR	FF	ALL
SROCC	0.886	0.540	0.879	0.915	0.787	0.867
PLCC	0.928	0.573	0.889	0.944	0.835	0.872
RMSE	4.843	5.359	7.618	4.779	6.845	8.014

4.4.3. Cross-Distortion Performance Evaluation

Here, we verify the cross-distortion performance comparisons between the proposed method and the other methods. Specifically, each tested method is trained on one type of distortion but tested on another type, with the same training sample sources. The results are displayed in Tables 10 and 11. According to the row elements of the tables, we can see that the model trained on one type of distortion has some influence on the performance on other types of distortion, but with less influence on the overall performance. This demonstrates

that the proposed scheme is a good metric for the perceived quality prediction of the stereoscopic image.

Table 10. SROCC on LIVE 3D database Phase I for cross-distortion performance evaluation.

Train \ Test	JP2K	JPEG	WN	BLUR	FF	ALL
JP2K	-	0.758	0.911	0.936	0.664	0.898
JPEG	0.586	-	0.926	0.764	0.498	0.746
WN	0.805	0.552	-	0.888	0.724	0.848
BLUR	0.878	0.553	0.921	-	0.737	0.889
FF	0.838	0.470	0.933	0.884	-	0.874

Table 11. SROCC on LIVE 3D database Phase II for cross-distortion performance evaluation.

Train \ Test	JP2K	JPEG	WN	BLUR	FF	ALL
JP2K	-	0.757	0.945	0.754	0.906	0.793
JPEG	0.943	-	0.812	0.693	0.700	0.758
WN	0.863	0.588	-	0.621	0.833	0.778
BLUR	0.945	0.386	0.922	-	0.883	0.818
FF	0.930	0.782	0.922	0.841	-	0.887

4.4.4. Image Content-Based Performance Evaluation

In this experiment, we discuss the database partitioning strategies, which are split by content or split randomly. During this experiment, we took 80% of the samples for training and the rest for testing. Specifically, for the LIVE 3D database Phase I, 16 reference images and their corresponding distorted images were selected as the training sets, while the others were chosen as the testing sets. For LIVE 3D Phase II, six samples were chosen for training, and the other two were used for testing. The comparison results between the content-based partitioning and random partitioning are listed in Table 12. It can be seen that the performance using image content-based partitioning is slightly worse than that using random partitioning. This is because the regression model learned from the content-based partitioned samples is a “content-aware” model, and thus the performance is a little worse than that of the “content-unaware” model.

Table 12. Performance evaluation of random database partitioning and image content-based partitioning.

Database	LIVE Phase I			LIVE Phase II			
	Partition Method	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
Random-based partition		0.952	0.962	4.493	0.940	0.950	3.546
Content-based partition		0.908	0.928	6.752	0.867	0.866	6.685

4.5. Influence of Disparity Estimate Algorithm

In order to verify whether our metric is method-independent on disparity estimation, we compare the performance of the improved SSIM-based disparity search algorithm in our method with those of the SAD-based and SSIM-based disparity estimation algorithms. The performance comparisons are listed in Table 13. It can be seen from Table 13 that different disparity estimation algorithms lead to similar performance, which means our metric is method-independent. It also can be seen that the performance from the improved SSIM-based disparity search algorithm is slightly better.

Table 13. Performance evaluation of different disparity estimation algorithms.

Database	LIVE Phase I			LIVE Phase II		
	Algorithm	SROCC	PLCC	RMSE	SROCC	PLCC
SAD-based	0.952	0.960	4.581	0.927	0.941	3.846
SSIM-based	0.946	0.956	4.841	0.937	0.945	3.727
Improved SSIM-based	0.952	0.962	4.493	0.940	0.950	3.546

4.6. Complexity Analysis

The time–cost ratios for each step of our algorithm are listed in Table 14. As is described in the table, the biggest share of complexity in our method is from the PCA and SVM fitting steps, which cost more than half of our prediction time, followed by the SSIM matching step that costs 36% of our prediction time.

Table 14. Complexity analysis for the proposed method.

Steps	Time Ratios
SSIM-Based Matching and Disparity Generation	36.37%
Cyclopean Generation	2.48%
Spatial Domain Features	1.95%
Transform Domain Features	7.42%
3D Perceptual Features	0.66%
PCA and SVM	51.12%

5. Conclusions

In this paper, we propose a novel natural scene statistics-based, no-reference quality assessment algorithm for stereoscopic images, in which both the quality of the cyclopean image is considered and the binocular rivalry and other 3D visual intrinsic properties are exploited. The proposed algorithm mainly consists of two steps. First, 2D features are extracted from the cyclopean image in the spatial domain and transformed domain, and then 3D quality-relevant perceptual features are extracted from stereo visual perceptual information, including the binocular rivalry, the disparity matching error and the disparity consistency. Next, followed by NSS modeling and adaptive PCA feature pruning, SVR is utilized to fit an optimum regression model. Compared with other FR-SIQA metrics and NR-SIQA metrics on the LIVE Phase I and LIVE Phase II databases, the proposed stereoscopic image quality metric achieves the best performance compared to other state-of-the-art SIQA metrics.

Author Contributions: Conceptualization, Y.Y. and X.G.; methodology, Y.Y. and L.S.; software, Y.Y.; validation, Y.Y.; formal analysis, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y. and L.S.; visualization, Y.Y. and D.W.; supervision, L.S. and R.F.; project administration, L.S.; funding acquisition, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China under grant No.61422111, 61671282 and U1301257 and was sponsored by the Shanghai Shuguang Program (15pjd015) and the Innovation Program of Shanghai Municipal Education Commission (17SG37).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, H.; Wang, S.; Zhao, Y.; Wei, J.; Piao, M. Large-scale elemental image array generation in integral imaging based on scale invariant feature transform and discrete viewpoint acquisition. *Displays* **2021**, *69*, 102025. [[CrossRef](#)]
2. Deng, L.; Pu, Y. Analysis of college martial arts teaching posture based on 3D image reconstruction and wavelet transform. *Displays* **2021**, *69*, 102044. [[CrossRef](#)]
3. Qi, S.; Ning, X.; Yang, G.; Zhang, L.; Long, P.; Cai, W.; Li, W. Review of multi-view 3D object recognition methods based on deep learning. *Displays* **2021**, *69*, 102053. [[CrossRef](#)]
4. Wang, X.; Wang, C.; Liu, B.; Zhou, X.; Zhang, L.; Zheng, J.; Bai, X. Multi-view stereo in the Deep Learning Era: A comprehensive review. *Displays* **2021**, *70*, 102102. [[CrossRef](#)]
5. Winkler, S.; Min, D. Stereoscopic image quality compendium. In Proceedings of the 2011 8th International Conference on Information, Communications & Signal Processing, Singapore, 13–16 December 2011; pp. 1–5.
6. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
7. Yang, H.; Li, J. Application of multivariate statistics and 3D visualization analysis in tacit knowledge diffusion map. *Displays* **2021**, *69*, 102062. [[CrossRef](#)]
8. Duan, Z.; Chen, Y.; Yu, H.; Hu, B.; Chen, C. RGB-Fusion: Monocular 3D reconstruction with learned depth prediction. *Displays* **2021**, *70*, 102100. [[CrossRef](#)]
9. Brandao, T.; Queluz, M.P. No-reference image quality assessment based on DCT-domain statistics. *Signal Process.* **2008**, *88*, 822–833. [[CrossRef](#)]
10. Lu, B.; Sun, L.; Yu, L.; Dong, X. An improved graph cut algorithm in stereo matching. *Displays* **2021**, *69*, 102052. [[CrossRef](#)]
11. Ye, P.; Wu, X.; Gao, D.; Deng, S.; Xu, N.; Chen, J. DP3 signal as a neuro-indicator for attentional processing of stereoscopic contents in varied depths within the ‘comfort zone’. *Displays* **2020**, *63*, 101953. [[CrossRef](#)]
12. Lu, B.; He, Y.; Wang, H. Stereo disparity optimization with depth change constraint based on a continuous video. *Displays* **2021**, *69*, 102073. [[CrossRef](#)]
13. Gao, Z.; Zhai, G.; Deng, H.; Yang, X. Extended geometric models for stereoscopic 3D with vertical screen disparity. *Displays* **2020**, *65*, 101972. [[CrossRef](#)]
14. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised feature learning framework for no-reference image quality assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1098–1105.
15. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [[CrossRef](#)]
16. You, J.; Xing, L.; Perki, A.; Wang, X. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In Proceedings of the Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA, 13–15 January 2010; pp. 4033–4036.
17. Howard, I.P.; Rogers, B.J. *Seeing in Depth*; Oxford University Press: New York, NY, USA, 2008.
18. Shen, L.; Fang, R.; Yao, Y.; Geng, X.; Wu, D. No-Reference Stereoscopic Image Quality Assessment Based on Image Distortion and Stereo Perceptual Information. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *3*, 59–72. [[CrossRef](#)]
19. Lebreton, P.; Raake, A.; Barkowsky, M.; Callet, P.L. Evaluating depth perception of 3D stereoscopic videos. *IEEE J. Sel. Top. Signal Process.* **2012**, *6*, 710–720. [[CrossRef](#)]
20. Mikkola, M.; Jumisko-Pyykko, S.; Strohmeier, D.; Boev, A.; Gotchev, A. Stereoscopic depth cues outperform monocular ones on autostereoscopic display. *IEEE J. Sel. Top. Signal Process.* **2012**, *6*, 698–709. [[CrossRef](#)]
21. Tam, W.J.; Speranza, F.; Yano, S.; Shimono, K.; One, H. Stereoscopic 3D-TV: Visual comfort. *IEEE Trans. Broadcast.* **2011**, *57*, 335–346. [[CrossRef](#)]
22. Jung, C.; Liu, H.; Cui, Y. Visual comfort assessment for stereoscopic 3D images based on salient discomfort regions. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 4047–4051.
23. Coria, L.; Xu, D.; Nasiopoulos, P. Quality of experience of stereoscopic content on displays of different sizes: A comprehensive subjective evaluation. In Proceedings of the 2011 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 9–12 January 2011; pp. 755–756.
24. Liu, H.; Heynderickx, I. Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 971–982.
25. Kim, H.; Lee, S. Transition of Visual Attention Assessment in Stereoscopic Images with Evaluation of Subjective Visual Quality and Discomfort. *IEEE Trans. Multimedia* **2015**, *17*, 2198–2209. [[CrossRef](#)]
26. Gorley, P.; Holliman, N. Stereoscopic image quality metrics and compression. *Proc. SPIE* **2008**, *6803*, 45–56.
27. Benoit, A.; Callet, P.L.; Campisi, P.; Cousseau, R. Quality Assessment of Stereoscopic Images. *EURASIP J. Image Video Process.* **2008**, *2008*, 659024. [[CrossRef](#)]
28. Wang, S.; Shao, F.; Li, F.; Yu, M.; Jiang, G. A simple quality assessment index for stereoscopic images based on 3d gradient magnitude. *Sci. World J.* **2014**, *2014*, 890562. [[CrossRef](#)] [[PubMed](#)]

29. Lin, Y.-H.; Wu, J.-L. Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. *IEEE Trans. Image Process.* **2014**, *23*, 1527–1542. [[CrossRef](#)] [[PubMed](#)]
30. Chen, M.J.; Su, C.C.; Kwon, D.K.; Cormack, L.K.; Bovik, A.C. Full-reference quality assessment of stereo pairs accounting for rivalry. *Signal Process. Image Commun.* **2013**, *28*, 1143–1155. [[CrossRef](#)]
31. Bensalma, R.; Larabi, M.-C. A perceptual metric for stereoscopic image quality assessment based on the binocular energy. *Multidimens. Syst. Signal Process.* **2013**, *24*, 281–316. [[CrossRef](#)]
32. Zhang, Y.; Chandler, D.M. 3D-MAD: A Full Reference Stereoscopic Image Quality Estimator Based on Binocular Lightness and Contrast Perception. *IEEE Trans. Image Process.* **2015**, *24*, 3810–3825. [[CrossRef](#)]
33. Li, F.; Shen, L.; Wu, D.; Fang, R. Full-reference quality assessment of stereoscopic images using disparity-gradient-phase similarity. In Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP) Chengdu, China, 12–15 July 2015; pp. 658–662.
34. Jiang, G.; Xu, H.; Yu, M.; Luo, T.; Zhang, Y. Stereoscopic Image Quality Assessment by Learning Non-negative Matrix Factorization-based Color Visual Characteristics and Considering Binocular Interactions. *J. Vis. Commun. Image Represent.* **2017**, *46*, 269–279. [[CrossRef](#)]
35. Geng, X.; Shen, L.; Li, K.; An, P. A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property. *Signal Process. Image Commun.* **2017**, *52*, 54–63. [[CrossRef](#)]
36. Shao, F.; Li, K.; Lin, W.; Jiang, G.; Dai, Q. Learning blind quality evaluator for stereoscopic images using joint sparse representation. *IEEE Trans. Multimedia* **2016**, *18*, 2104–2114. [[CrossRef](#)]
37. Akhter, R.; Baltes, J.; Sazzad, Z.M.P.; Horita, Y. No reference stereoscopic image quality assessment. *Proc. SPIE* **2010**, *7524*, 75240T.
38. Chen, M.J.; Cormack, L.K.; Bovik, A.C. No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.* **2013**, *22*, 3379–3391. [[CrossRef](#)] [[PubMed](#)]
39. Wang, S.; Shao, F.; Jiang, G. Supporting binocular visual quality prediction using machine learning. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Chengdu, China, 14–18 July 2014; pp. 1–6.
40. Ryu, S.; Sohn, K. No-Reference Quality Assessment for Stereoscopic Images Based on Binocular Quality Perception. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 591–602.
41. Shao, F.; Lin, W.; Wang, S.; Jiang, G.; Yu, M. Blind Image Quality Assessment for Stereoscopic Images Using Binocular Guided Quality Lookup and Visual Codebook. *IEEE Trans. Broadcast.* **2015**, *61*, 154–165. [[CrossRef](#)]
42. Shao, F.; Lin, W.; Wang, S.; Jiang, G.; Yu, M.; Dai, Q. Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images. *IEEE Trans. Cybern.* **2016**, *46*, 730–743. [[CrossRef](#)] [[PubMed](#)]
43. Su, C.C.; Cormack, L.K.; Bovik, A.C. Oriented Correlation Models of Distorted Natural Images With Application to Natural Stereopair Quality Evaluation. *IEEE Trans. Image Process.* **2015**, *24*, 1685–1699. [[CrossRef](#)] [[PubMed](#)]
44. Wang, J.; Zeng, K.; Wang, Z. Quality prediction of asymmetrically distorted stereoscopic images from single views. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Chengdu, China, 14–18 July 2014; pp. 1–6.
45. Wang, J.; Rehman, A.; Zeng, K.; Wang, S.; Wang, Z. Quality Prediction of Asymmetrically Distorted Stereoscopic 3D Images. *IEEE Trans. Image Process.* **2015**, *24*, 3400–3414. [[CrossRef](#)]
46. Zhao, Y.; Zhang, Y.; Yu, L. Subjective study of binocular rivalry in stereoscopic images with transmission and compression artifacts. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; pp. 132–135.
47. Zhang, L.; Zhang, L.; Tao, D.; Huang, X.; Du, B. Compression of hyperspectral remote sensing images by tensor approach. *Neurocomputing* **2015**, *147*, 358–363. [[CrossRef](#)]
48. Du, B.; Zhang, M.; Zhang, L.; Hu, R.; Tao, D. PLTD: Patch-Based Low-Rank Tensor Decomposition for Hyperspectral Images. *IEEE Trans. Multimed.* **2017**, *19*, 67–79. [[CrossRef](#)]
49. Zhou, W.; Qiu, W.; Wu, M. Utilizing dictionary learning and machine learning for Blind Quality Assessment of 3-D Images. *IEEE Trans. Broadcast.* **2017**, *63*, 404–415. [[CrossRef](#)]
50. Samek, W.; Binder, A.; Montavon, G.; Lapuschkin, S.; Müller, K. Evaluating the visualization of what a deep neural network has learned. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2660–2673. [[CrossRef](#)]
51. Zhou, M.; Li, S. Deformable Convolution Based No-Reference Stereoscopic Image Quality Assessment Considering Visual Feedback Mechanism. In Proceedings of the 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 5–8 December 2021; pp. 1–5.
52. Jinhui, F.; Li, S.; Chang, Y. No-Reference Stereoscopic Image Quality Assessment Considering Binocular Disparity and Fusion Compensation. In Proceedings of the 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 5–8 December 2021; pp. 1–5.
53. Bourbia, S.; Karine, A.; Chetouani, A.; Hassoun, M.E. A Multi-Task Convolutional Neural Network For Blind Stereoscopic Image Quality Assessment Using Naturalness Analysis. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 1434–1438.
54. Sandić-Stanković, D.D.; Kukolj, D.D.; Callet, P.L. Quality Assessment of DIBR-Synthesized Views Based on Sparsity of Difference of Closings and Difference of Gaussians. *IEEE Trans. Image Process.* **2022**, *31*, 1161–1175. [[CrossRef](#)] [[PubMed](#)]

55. Gu, K.; Tao, D.; Qiao, J.; Lin, W. Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1301–1313. [[CrossRef](#)] [[PubMed](#)]
56. Bosse, S.; Maniry, D.; Müller, K.; Wiegand, T.; Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* **2018**, *27*, 206–219. [[CrossRef](#)]
57. Oh, H.; Ahn, S.; Kim, J.; Lee, S. Blind deep S3D image quality evaluation via local to global feature aggregation. *IEEE Trans. Image Process.* **2017**, *26*, 4923–4936. [[CrossRef](#)]
58. Hoa, D.K.; Dung, L.; Dzung, N.T. Efficient determination of disparity map from stereo images with modified sum of absolute differences (SAD) algorithm. In Proceedings of the 2013 International Conference on Advanced Technologies for Communications (ATC 2013), Ho Chi Minh City, Vietnam, 16–18 October 2013; pp. 657–660.
59. Howard, P.I.; Rogers, B.J. *Binocular Vision and Stereopsis*; Oxford University Press: Oxford, UK, 1995.
60. Levelt, W.J.M. On Binocular Rivalry. Ph.D. Thesis, Leiden University, Leiden, The Netherlands, 1965.
61. Hubel, D.H. The visual cortex of the brain. *Sci. Am.* **1963**, *209*, 54–63. [[CrossRef](#)]
62. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)]
63. Fang, R.; Al-Bayaty, R.; Wu, D. BNB Method for No-Reference Image Quality Assessment. *IEEE Trans. Circuits Syst. Video Tech.* **2017**, *27*, 1381–1391. [[CrossRef](#)]
64. Sheikh, H.R.; Bovik, A.C.; de Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans Image Process.* **2005**, *14*, 2117–2128. [[CrossRef](#)] [[PubMed](#)]
65. Fang, Y.; Ma, K.; Wang, Z.; Lin, W.; Fang, Z.; Zhai, G. No-Reference Quality Assessment of Contrast-Distorted Images Based on Natural Scene Statistics. *IEEE Signal Process. Lett.* **2015**, *22*, 838–842. [[CrossRef](#)]
66. Md, S.K.; Appina, B.; Channappayya, S.S. Full-Reference Stereo Image Quality Assessment Using Natural Stereo Scene Statistics. *IEEE Signal Process. Lett.* **2015**, *22*, 1985–1989.
67. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. 2001. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed on 1 January 2020).
68. Moorthy, A.K.; Su, C.-C.; Mittal, A.; Bovik, A.C. Subjective evaluation of stereoscopic image quality. *Signal Process. Image Commun.* **2013**, *28*, 870–883. [[CrossRef](#)]
69. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [[CrossRef](#)] [[PubMed](#)]
70. Gholipour, A.; Araabi, B.N.; Lucas, C. Predicting chaotic time series using neural and neurofuzzy models: A comparative study. *Neural Process. Lett.* **2006**, *24*, 217–239. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.