ADABOOST NEURAL NETWORK AND CYCLOPEAN VIEW FOR NO-REFERENCE STEREOSCOPIC IMAGE QUALITY ASSESSMENT

Oussama Messai[a] *Fella Hachouf*[a] *Zianou Ahmed Seghir*[b]

[a] Laboratoire Automatique et Robotique, Université des Frères Mentouri Constantine 1, Algeria.
[b] Computing Department, University of Abbes laghrour Khenchela, Algeria.
mr.oussama.messai@gmail.com, hachouf.fella@gmail.com, zianou_ahmed_seghir@yahoo.fr

ABSTRACT

Stereoscopic imaging is widely used in many fields. In many scenarios, stereo images quality could be affected by various degradations, such as asymmetric distortion. Accordingly, to guarantee the best quality of experience, robust and accurate reference-less metrics are required for quality assessment of stereoscopic content. Most existing stereo no-reference Image Quality Assessment (IQA) models are not consistent with asymmetrical distortions. This paper presents a new no-reference stereoscopic image quality assessment metric using a human visual system (HVS) modeling and an advanced machine-learning algorithm. The proposed approach consists of two stages. In the first stage, cyclopean image is constructed considering the presence of binocular rivalry in order to cover the asymmetrically distorted part. In the second stage, gradient magnitude, relative gradient magnitude, and gradient orientation are extracted. These are used as a predictive source of information for the quality. In order to obtain the best overall performance against different databases, Adaptive Boosting (AdaBoost) idea of machine learning combined with artificial neural network model has been adopted. The benchmark LIVE 3D phase-I, phase-II, and IRCCyN/IVC 3D databases have been used to evaluate the performance of the proposed approach. Experimental results have demonstrated that the proposed metric performance achieves high consistency with subjective assessment and outperforms the blind stereo IQA over various types of distortion.

***Index Terms*—** Stereoscopic quality assessment, no-reference, binocular rivalry, cyclopean view, neural network, AdaBoost.

1. INTRODUCTION

According to the latest theatrical market statistics collected by Motion Picture Association of America (MPAA), a number of worldwide 3D screens continued to grow in 2016 at a faster pace (17%) than 2015 (15%) [1]. 3D film production methods have been improved in recent years. This 3D development will undoubtedly introduce more successful stereoscopic/3D films, such as Avatar in 2009. (The Avatar 2 is expected by late 2020).

The stereoscopic images and videos have not been limited to the entertainment industry. Stereo visualization concerns many applications, such as remote education [2], medical body exportation [3], robot navigation [4] and so forth. Therefore, it is reasonable to believe that the amount of stereo content will continue growing throughout the next few years. Since more perceptual issues such as visual discomfort/fatigue should be considered, the quality assessment of stereo images and videos is becoming more complex. Furthermore, our understanding of the perceptual components that determine the quality of the stimulus remains limited. In most cases, the fatigue experienced by a viewer is caused by asymmetrical distortions, also known as binocular rivalry/suppression. This paper mainly takes this annoying experience into account as a previous work did [5].

There are numerous aspects that affect the stereo/2D image quality during scene capture, such as light, color, structure and noise. In addition, images can be distorted in some applications by compression, transmission, rendering and display. All these factors influence our decision when we assess the quality of an image.

The Image Quality Assessment (IQA) is critically important in various image processing applications as acquisition, compression and specifically for image enhancement. In general, the quality evaluation can be divided into two classes: subjective and objective methods. The subjective evaluation is based on an opinion score of human observers, mostly expressed in terms of Mean Opinion Score (MOS) or Difference Mean Opinion Score (DMOS). This type of methods is effective and reliable to evaluate perceptual quality but it has a number of limitations such as: time-consuming, high costs, and they cannot be applied in online applications. On the other hand, objective evaluation aims to design a computational model that predicts human perceived image quality, accurately and automatically. Automatic evaluation (objective evaluation) metrics offer many advantages, such as rapid assessment, low cost. They are easy to incorporate into image processing systems/applications. For this reason, a significant amount of research has been devoted to the development of objective evaluation metrics that bring us closer to the ultimate goal of simulating the perceptual process of the human visual system (HVS).

The objective methods can be further categorized into three categories; Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) metrics. In FR based methods the algorithm has an access to a perfect version of the pristine stimuli. Which is then compared with the deformed version. In many practical applications, the original stimuli are not fully available. Thus the RR and NR methods are needed. The reduced-reference metrics utilize only partial information, on pristine stimuli, while the NR methods, called blind reference, do not have access to the reference, or it can not be available.

A lot of efforts have been dedicated to understand the HVS to apply it for image processing applications. The main goal of an objective IQA is to create an approach that mimics the HVS, then, a perceived quality of the image is automatically predicted. To assess this objective, it is important to compare the performance of the metric with the subjective evaluation. Which is the perceptual human judgment.

In this paper, a new no-reference model is designed. It will be

useful for practical applications, such as compression and quality improvement. The remaining sections of the paper are organized as follows: Section 2 presents the related works. Afterward, the overall framework of the proposed model is described in section 3. Experimental results are discussed in section 4. Finally, section 5 concludes the paper and future work is given.

2. RELATED WORKS

Generally, stereo IQA can be classified into two classes method. The first method class does not include disparity information. It applies the 2D IQA model directly to the stereo IQA problem by computing the mean quality of left and right views. For instance, Moorthy *et al* in [6] improved their blind 2D IQA method (DIIVINE) for stereo images. The NR 2D-extended metrics usually extract feature vectors separately for the left and right images. They are weight-averaged to obtain the final feature vector for training. A novel FR QA metric for stereopairs called binocular energy quality metric (BEQM) has been proposed by Bensalma *et al* [7]. It estimates the stereo image quality by computing the binocular energy difference between the original and distorted stereopairs.

The second class is based on the belief that 3D quality cannot be accurately deduced from the average of the two views, but from the binocular perceptual view (cyclopean view) in which disparity information is used. However, research has shown that the 3D viewing quality is correlated with disparity quality. The quality-aware features of stereo images are clearly different from those of their 2D counterparts [8, 9]. For example, Akhter *et al*. [10] have designed a no-reference stereo IQA algorithm. They extracted features from disparity map and stereo image pairs. Benoit *et al.* [11] proposed a FR stereo QA model that employs two measures. The first is the difference between left and right reference pictures and distorted images. The second is the difference between disparity map of the original stereo image pairs and the distorted ones. A similar FR stereo IQA has been suggested by You *et al.* [12]. They used a variety of 2D IQA models for stereo images and tried to combine the predicted quality scores from both disparity map and stereo pair. The authors of [13, 14] proposed a PSNR-based stereo QA metrics. Hewage *et al*. [14] computed the edges from the disparity map, then PSNR has been used between the pristine and test edge maps to predict the quality.

Meanwhile, Gorley *et al*. [13] did not use or measure disparity/depth information. They compute quality scores on matched feature points delivered by SIFT [15] and RANSAC [16] applied to the left and right views.

Recently, a considerable amount of research has been carried out on how the visual system processes the signals perceived by the two eyes, and using these researches for stereo IQA problem in the meantime. Chen *et al.* [17] proposed a FR quality assessment model that utilized the linear expression of cyclopean view [18] influenced by binocular suppression/rivalry between left and right views. An extended version of this framework has been used to create a NR model using natural scene statistics features extracted from stereo image pairs [19]. In [20], another FR metric has been adopted by Hachicha *et al.* for stereoscopic images. They used Binocular Just Noticeable Difference (BJND) [21] approach to model the binocular rivalry theory. Fang *et al.* [22]. proposed an unsupervised blind model for stereoscopic images. From the monocular and cyclopean view patches, they extracted various quality-aware features in spatial and frequency domains. Then, Bhattacharyya-like distance has been used to produce a quality score. Another referenceless dictionary learning (DL)-based metric has been proposed in [23]. The authors simulated the main functional structure of binocular vision. A log-Gabor filter is then used to extract features and k-nearest-neighbors (KNN) has been deployed to map the quality score.

In particular, the natural scene statistics (NSS) have proven to be reliable in IQA algorithms. For example, Su et al. [24] built a NR stereo IQA framework. They synthesized a cyclopean view and then they extracted bivariate and generalized univariate NSS as features. More NSS-based framework has been conducted by Appina *et al* [25]. Lv *et al* [26] also developed a NR stereo IQA metric. Their scheme computes binocular self-similarity and binocular integration using NSS features.

In the literature, compared to the number of full-reference metrics, a small number of no-reference stereo IQA metrics have been proposed. In addition, performance of most metrics is not consistent with asymmetric distortions. Consequently, the blind stereo IQA is still in its initial development phase. To solve these problems, this paper proposes an automatic no-reference metric for stereoscopic images based on human visual perception, taking into account the presence of binocular rivalry/suppression. The built model also uses advanced machine learning algorithm to achieve better performance.

3. PROPOSED APPROACH

Human binocular perception is a complex visual process which has not yet been fully understood. In order to design a model that can assess the quality of stereoscopic images, research on human binocular perception is required. The hypothesis of cyclopean image is therefore used with consideration of binocular suppression, and disparity map. Metrics that use this hypothesis, such as the FR stereo IQA model in [17], have achieved good precision.

Back-Propagation (BP) neural network is widely used for regression and classification [27]. Liu *et al.* [28] utilized the idea of AdaBoost algorithms combined with BP neural network. They proposed a NR 2D IQA. It has showed robustness and good performances.

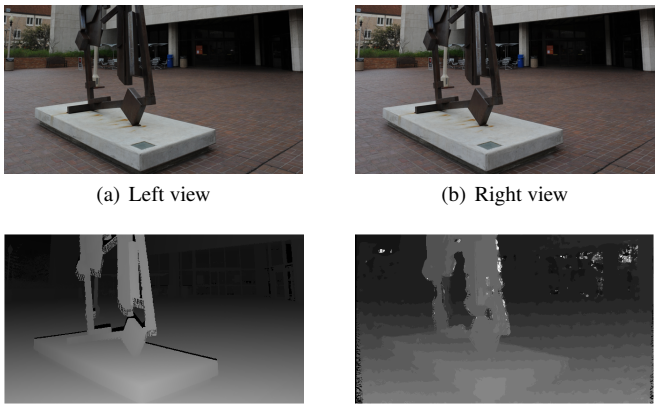
Motivated by these ideas and based on our previous works [5, 29], we develop a new NR quality predictor model for stereoscopic images. In summary, the model involves three steps: first, a cyclopean image is constructed using Gabor filter responses and disparity map. In a second step, gradient characteristics of the cyclopean image and the disparity map are extracted. Finally, to predict a quality score based on feature learning, the AdaBoost algorithm combined to artificial neural network has been used.

3.1. Cyclopean view

The cyclopean image synthesis differs from the usual 2D images for the depth information it contains. The first objective of the proposed stereo IQA algorithm is to estimate the actual cyclopean view formed within the observer's mind while a stereo image is supplied. The HVS processes and combines visual signals from both eyes into a single combined perception [30]. However, the HVS has not been completely understood. Therefore, the cyclopean image that is actually processed in our minds is still unclear.

3.1.1. Disparity map

Nowadays, depth/disparity estimation from stereoscopic images is important in many applications, such as augmented reality, 3D reconstruction and navigation. The disparity information is proven to be a strong effective factor for stereo images and videos quality. Therefore, it is a necessary information for assessing the quality of



(a) Left view (b) Right view

(c) Truth disparity (d) Estimated disparity

Fig. 1: Top: left and right views of the stereo image. Bottom: Estimated disparity versus the ground truth disparity.

the stereo content. Intensive research has been conducted on the design of stereo matching algorithms (disparity estimation). However, there is no agreement, on the type of stereo matching algorithm to be used in stereo IQA, except for those with low complexity. Therefore, a stereo matching model with balanced complexity and performance is deployed.

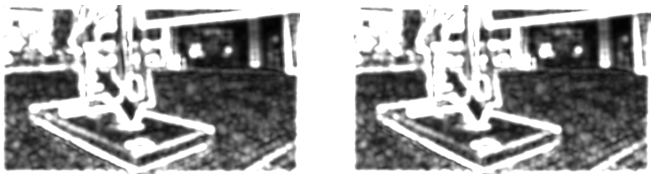
The chosen algorithm is called SSIM-based stereo. It is an improved version of Sum of Absolute Differences SAD stereo matching algorithm [31]. The modification consists in replacing SAD by SSIM in computing disparities. SSIM [32] scores are used to select the best matches. This is done by maximizing the SSIM scores between the current block from left image and right image blocks along the horizontal direction. The maximum number of pixels to be searched for is the maximum disparity. After all, the disparity map values are the difference between the current pixel and the best SSIM location. A 7 by 7 block size has been used, while the maximum disparity distance has been set to 25. Figure 1 shows an estimated disparity versus the ground-truth disparity using the SSIM-based stereo matching algorithm.

3.1.2. Gabor filter responses

Various theories have been proposed to explain binocular rivalry. This visual phenomenon has recently been investigated by many researchers. Binocular rivalry or suppression is a failure of the brain in fusing the left and right views causing fatigue or discomfort to the viewers.

The binocular rivalry as mentioned before is when the two images of a stereo pair present different kinds or degrees of distortion. Therefore, the objective quality of the mostly viewed stereo image cannot be predicted from the average quality of the left and right views. Reasonable explanation for binocular rivalry has been established by Levelt [18]. Levelt *et al*. conducted a series of experiments which clearly demonstrate that binocular suppression or rivalry is strongly governed by low-level sensory factors. Levelt concluded that visual stimuli which have more contours or high contrast, tend to dominate the rivalry. Motivated by this result, the energy of Gabor filter bank responses on the left and right images is used to simulate suppression selection (binocular rivalry) of the cyclopean image when it is computed.

The Gabor filter bank is a band-pass filter. It extracts luminance



(a) Left view (b) Right view

Fig. 2: Gabor filter responses from the left and right views.

and chromatic channels features. The filter is related to the function of primary visual cortex cells in primates [33]. It models the frequency-oriented decomposition in primary visual cortex, and captures energy in both space and frequency in a high localized way [34].

The used Gabor filter is as follows:

$$
\begin{aligned}
&GF(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[(x'/\sigma_x)^2+(y'/\sigma_y)^2]} e^{i(x\zeta_x+y\zeta_y)} \\
&with \\
&x' = (x-m_x).cos(\theta) + (y-m_y).sin(\theta) \\
&y' = -(x-m_x).sin(\theta) + (y-m_y).cos(\theta)
\end{aligned}
\tag{1}
$$

where m_x and m_y define center of the Gabor receptive field (m_x and m_y are the x and y locations of the center with respect to the original coordinate system). σ_x and σ_y are the standard deviations of an elliptical Gaussian envelope along x' and y' directions, ζ_x and ζ_y are spatial frequencies, and θ orients the filter. The design of the Gabor filter bank is based on the work conducted by Chun *et al.* [35].

In visual perception study, a spatial frequency is expressed as the number of cycles per degree of visual angle. In theory, the spatial frequency is that the visual cortex operates not only on the lines and straight edges code but also on a spatial frequency code. To support this theory, a series of experiments have been conducted by P. Issa *et al.* [36]. They studied the effect of spatial frequency on primary visual cortex reaction using cats. The authors concluded that the visual cortex neurons react even more robustly to sine-wave gratings in their receptive fields at specific angles than they do to edges or bars. Therefore, using a band-pass filter over multiple orientations is favorable to extract features which the visual cortex responds to. The choice of the spatial center frequency is inspired by the result of Schor *et al*. [37] who found that the stereoscopic acuity of human vision normally falls off quickly when seeing stimuli dominated by spatial frequencies lower than 2.4 cycles/degree. Based on their findings, this means that using filters having spatial center frequencies in the range from 2.4 to 4 cycles/degree should produce responses to which a human observer would be more sensitive. Therefore, the local energy is estimated by summing Gabor filter magnitude responses over eight orientations at a spatial frequency of 3.67 cycles/degree ($\zeta_x = \zeta_y = 3.67$) . The standard deviations σ_x and σ_y are set to 0.01 ($\sigma_x = \sigma_y = 0.01$). As an example, figure 2 shows the filter outputs on the left and right views.

3.1.3. Cyclopean image construction

The visual signals from the two eyes are added by HVS, a process called binocular summation which enhances vision and increases the ability to detect weak objects [30]. However, current knowledge

Fig. 3: Synthesized cyclopean image by the proposed framework.

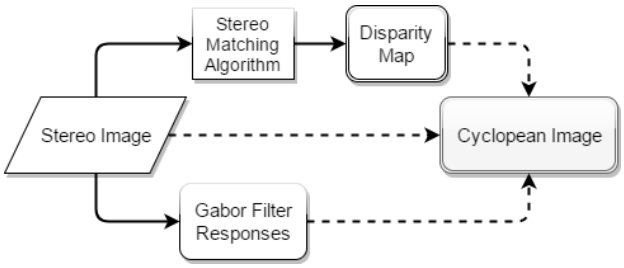


Fig. 4: The flowchart of the formed cyclopean image.

of HVS is very modest to guide the development of a mathematical model that perfectly simulates the process in the human brain. Therefore, a popular choice is to replace the complex simulation by simplified mathematical models. A linear model of a cyclopean image has been firstly proposed in 1968 by Levelt [18] to explain the phenomenon of binocular rivalry.

The model is as follows:

$$C = w_l I_l + w_r I_r \tag{2}$$

where I_l and I_r are respectively the left and right images, both w_l and w_r are the weighting coefficients for the left and right eyes in which $w_l + w_r = 1$.

In order to construct the cyclopean image, Levelt linear model has been used. The energy of Gabor filter bank responses is used to compute the weights, while the SSIM-based stereo matching algorithm is employed to create the disparity map.

The used model is:

$$C(x,y) = w_l(x,y) \times I_l(x,y) + w_r(x+m,y) \times I_r(x+m,y) \tag{3}$$

where the weights w_l and w_r are given by:

$$w_l(x,y) = \frac{GI_l(x,y)}{GI_l(x,y) + GI_r(x+m,y)} \tag{4}$$

$$w_r(x+m,y) = \frac{GI_r(x+m,y)}{GI_l(x,y) + GI_r(x+m,y)} \tag{5}$$

where GI_l and GI_r are the summation of Gabor filter magnitude responses from left and right views respectively, and m is the disparity index that corresponds to pixels from left image I_l to those in right image I_r. The filter of the form (1) is used to compute the magnitude responses over eight orientations for better accuracy. In the equation (1), θ refers to the filter's orientation degree. Table 1 shows the used orientation degrees.

Table 1: Magnitude responses orientation degrees.

Orientation number	1	2	3	4	5	6	7	8
Orientation degree	0	22.5	45	67.5	90	112.5	135	157.5

Results from previous work [5] have shown that the synthesized cyclopean view is reliable for perceptual quality evaluation. Figure 3 shows an example of cyclopean image. In this example, a stereo image without distortion has been used. (Figure 1, left view (a) and right view (b)). Figure 4 summarizes the construction steps of the cyclopean image.

3.2. Feature extraction

In addition to screen height and number of displayed pixels, the viewing conditions, namely: visual angle and viewing distance also influence the stereo image quality for the observer. However, the visual angle and viewing range are not taken into consideration in this study. We precisely simulate the HVS and then focus on local pixel distortions that can occur via 3D image processing applications.

The primary visual cortex receives visual information coming from the eyes. After reaching the visual cortex, the human mind processes that sensory inputs and uses it to realize the scene. Image gradients provide important visual information. They are essential for understanding the scene. Therefore, we believe that such information is important for the human visual system to understand the scene and judge its quality. This theory is supported by numerous FR IQA schemes based on the concept of gradient similarity. In relation to our problem, we use gradient magnitude and orientation as quality-aware features to evaluate the quality of stereoscopic images.

3.2.1. Gradient magnitude and orientation

Three gradient maps are produced from the cyclopean image, and disparity using horizontal and vertical direction derivatives, d_x and d_y respectively. Gaussian distribution function is used as a kernel in a 5 by 5 mask to compute the directional gradient components [$d_x(i,j)$, and $d_y(i,j)$]. The mask weights are samples from 2D Gaussian function which is defined as follows:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{6}$$

where σ controls the amount of smoothing. If σ increases, more samples must be obtained to represent the Gaussian function accurately. The derivatives have been computed using central difference. In our implementation, we used a limited smoothing mask as it tends to extract more edge information which makes the gradients more sensitive to distortions. Thus, σ is fixed to 0.5 ($\sigma = 0.5$).

A Gradient magnitude (GM), relative gradient orientation (RO), and relative gradient magnitude (RM) are computed for each cyclopean image and disparity map. The gradient magnitude is defined as:

$$|\nabla I(i,j)|_{GM} = \sqrt{d_x(i,j)^2 + d_y(i,j)^2} \tag{7}$$

while the gradient orientation is given by:

$$\angle \nabla I(i,j) = \arctan \frac{d_y(i,j)}{d_x(i,j)} \tag{8}$$

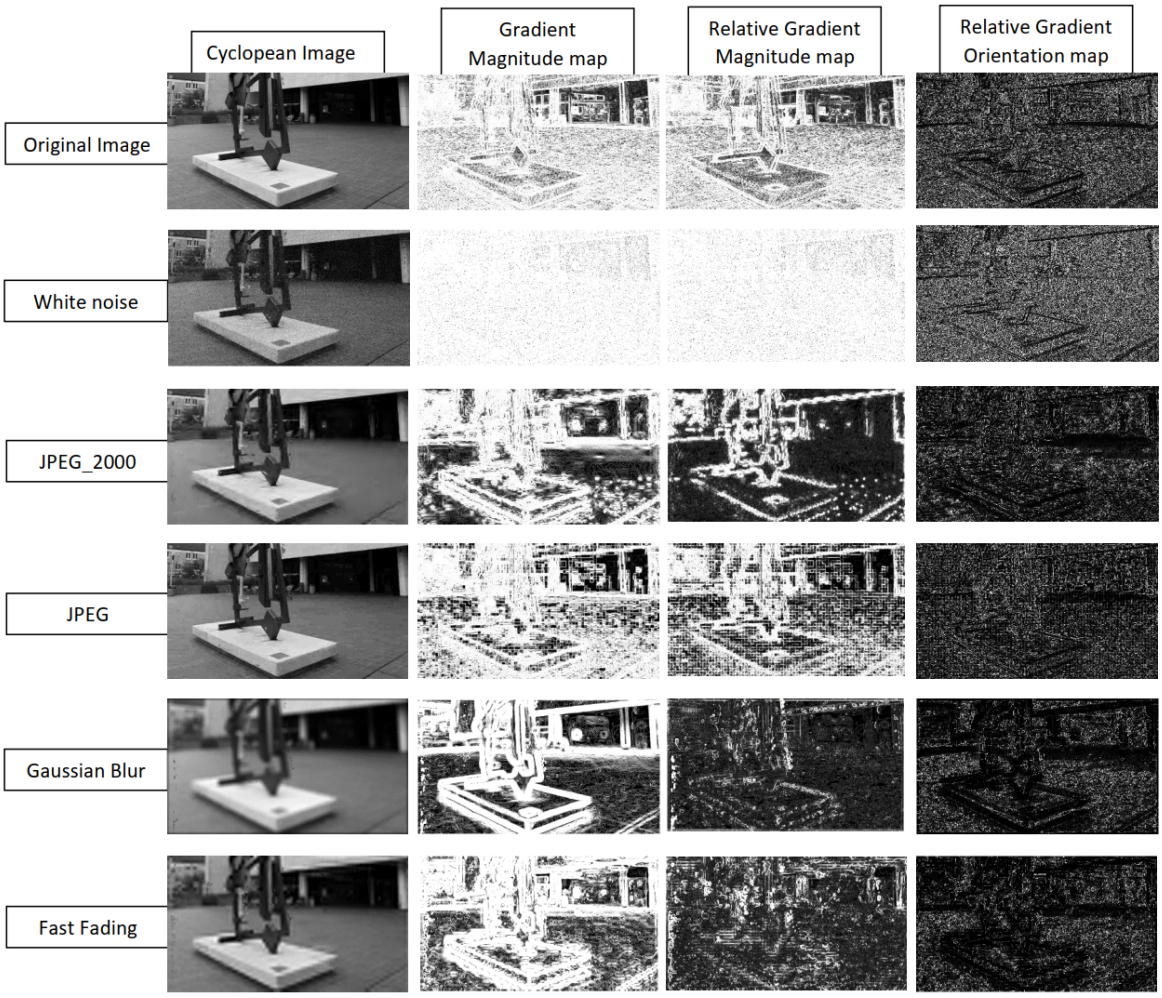


Fig. 5: Examples of the constructed cyclopean image, GM, RM, and RO maps for different type of distortions.

the relative gradient orientation is defined as follows:

$$\angle\nabla I(i,j)_{RO} = \angle\nabla I(i,j) - \angle\nabla I(i,j)_{AV} \tag{9}$$

where the local average orientation is:

$$\angle\nabla I(i,j)_{AV} = \arctan \frac{d_y(i,j)_{AV}}{d_x(i,j)_{AV}} \tag{10}$$

while the average directional derivative over x and y is defined by:

$$I_\gamma(i,j)_{AV} = \frac{1}{MN} \sum_{m,n} I_\gamma(i-m, j-n) \tag{11}$$

where M and N describe the size of the patches, 3×3 square neighborhood has been chosen (M=N=3), γ refers either to the horizontal x or the vertical y direction. Finally the relative gradient magnitude is defined by:

$$|\nabla I(i,j)|_{RM} = \sqrt{(d_x(i,j) - d_x(i,j)_{AV})^2 + (d_y(i,j) - d_y(i,j)_{AV})^2} \tag{12}$$

Standard deviation of each gradient histogram GM, RO, RM is computed as a final features extraction S_{GM}, S_{RO} and S_{RM}, respectively. The standard deviation is known as the square root of variance and defined by:

$$S(h) = \sqrt{\frac{1}{N-1} \sum_{x=1}^{N} (h(x) - \hbar)^2} \tag{13}$$

where $\hbar$ is the sample mean of the histogram $h(x)$ (normalized to unit sum), and N is the number of observations in the sample.

Display resolution is important factor for judging the quality. The subjective evaluation of a given stereo image varies when this factor changes. Therefore for objective evaluation, multiscale method is a convenient way to incorporate image details at different resolutions. Wang *et al* [38] proposed a multiscale quality assessment metric that outperforms the single-scale SSIM [32] model. The authors compared different down-sampling parameters results, and noted that down-sampling with a factor of 0.5 gives the best performance. Consequently, the cyclopean image is down-sampled with a factor of 0.5 (divided by 2), considering the changes in stereo image resolution and visual conditions. For example, distance from

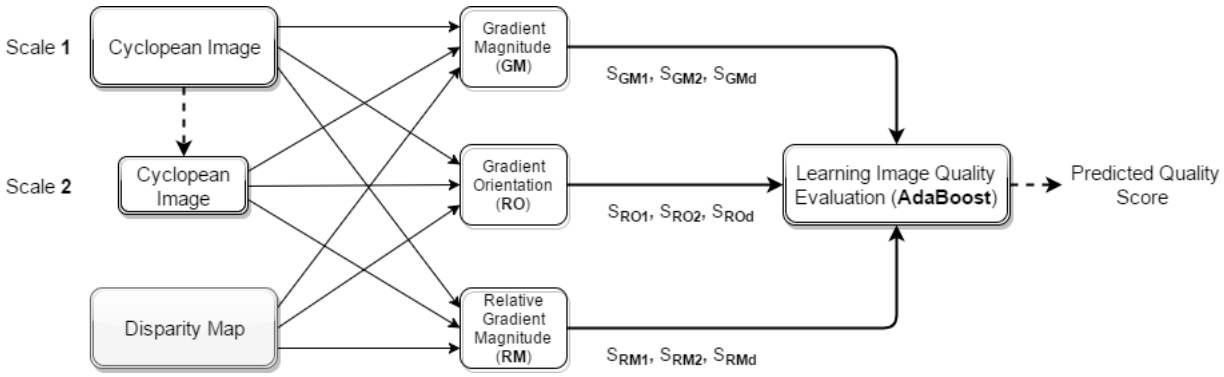


Fig. 6: Flowchart of the proposed measure.

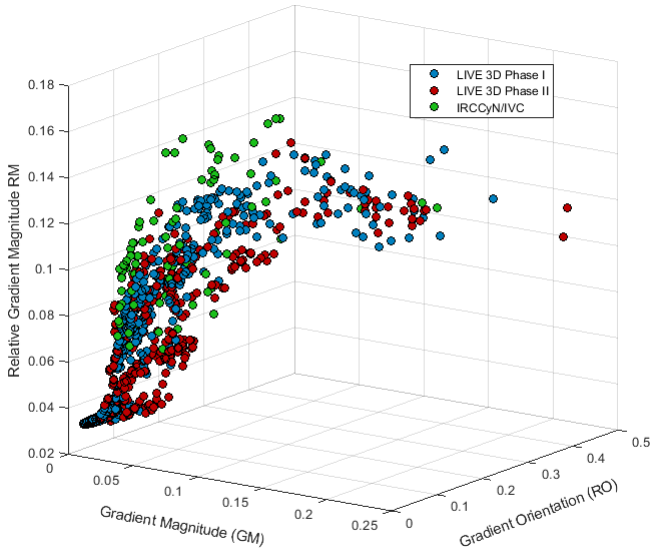


Fig. 7: 3D-plot of the extracted features S_{GM1}, S_{RO1}, and S_{RM1} from the cyclopean image using LIVE 3D phase I, phase II and IVC 3D databases.

the viewer to the screen can change the size of the formed cyclopean view in his brain. The features S_{GM}, S_{RO} and S_{RM} are computed for each scale, yielding 6 features element from the cyclopean image. The final feature vector (F) has nine elements as follows:

$$F = [S_{GM1}, S_{RO1}, S_{RM1}, S_{GM2}, S_{RO2}, S_{RM2}, S_{GMd}, S_{ROd}, S_{RMd}] \quad (14)$$

Figure 5 illustrates the computed maps GR, RM, and RO from the cyclopean image over five well-known distortions. It can be observed that the distortions have affected differently the computed gradient maps GR, RM, and RO.

Figure 6 displays the overview process to measure the quality of stereo images, while Figure 7 exhibits a 3D-plot of the extracted features over three databases. The represented features have been extracted from the cyclopean image in scale 1. The colored dots in the figure represent the extracted indicators in 3-dimensions. This 3D view shows that the features dots follow the same pattern on all databases which contain stereo images of different quality. Consequently, the extracted gradient indicators can be deployed for assessing the quality of stereoscopic images.

3.3. Learning for image quality evaluation

Machine Learning (ML) plays an important role in the development of modern picture quality models. Although a limited number of IQA models have used advanced ML techniques such as AdaBoost [39]. The AdaBoost, short for Adaptive Boosting, is an algorithm that consists in sequentially training a new simple model based on the errors of the previous model. A weight is assigned to each model. In the end, the whole set is combined to become an overall predictor. AdaBoost is one of the most useful ensemble method [40]. It can be used in conjunction with many other types of learning algorithm usually called Weak Learners (WL). The structure of the boosting ensemble generally outperforms a single feature learning model [41]. The boosting procedure tends to discover the examples and data points that are hard to predict and focuses on the next model predicting these examples better, by sequentially building a new simple model based on the errors of the previous model.

3.3.1. AdaBoost neural networks

The use of Back-Propagation neural network is powerful for good prediction. Furthermore, to improve the performance of this neural network regression model, the AdaBoost idea has been implemented. An Artificial Neural Network (ANN) with two hidden layers has been used as WL. However, the AdaBoost neural network can be less susceptible to the over-fitting problem than other learning algorithms. To solve this problem, 15% from training dataset has been dedicated validation for each neural network model.

The overall flow of the AdaBoost BP neural network model that computes the predicted output Q on a test set F is characterized as follows: First, set the quantity L of the Weak Learners (the BP artificial neural network models). Second, train the ith ANN on the sets X_j and Y_j, and estimate the predicted output of the testing set $Y_{i,j}^{pred}$. Afterward, a distribution D_i for the ith ANN is used for computing the evaluation error which is defined as (initial values of

Algorithm 1: Adaptive Boosting (AdaBoost) regression.

Input: L, F

// L: the number of Weak learners (WL), F: stereo image features vector.

Output: Q

// Q: the predicted quality.

Data: dataset for training and testing.

$n \longleftarrow 1; i \longleftarrow 1;$ // Initialization.

$Tr \leftarrow random(Data, 80\%)$

$Te \leftarrow random(Data, 20\%)$ // Divide data randomly for training and testing.

$M \leftarrow size(Te)$ // Get the testing vectors size.

$D_1(1:M) \longleftarrow 1$ // Initialize the first distributions.

for $n = 1 : L$ **do**

  $Ttr \leftarrow random(Tr, 85\%)$

  $Vtr \leftarrow random(Tr, 15\%)$ // Holdout 15% from train set for validation.

  $WL_n \leftarrow random(weights, biases)$

  $Train(WL_n, Ttr)$ // Train and validate the WL.

  $Terr(1:M) \longleftarrow 0$

  $Err \longleftarrow 0$ // Reset the testing and evaluation error for each WL.

  $Terr \leftarrow Test(WL_n, Te)$ // Compute the testing error.

  for $i = 1 : M$ **do**

    if $(Terr(i) > 0.2)$ **then**

      $Err \leftarrow Err + D_n(i)$ // update the distribution $D_{n+1}(i)$ for next WL and compute the evaluation error of the nth WL.

      $D_{n+1}(i) \leftarrow D_n(i) \times (1 + \delta)$

    else

      $D_{n+1}(i) \leftarrow D_n(i)$

  $w_n \leftarrow \frac{1}{e^{Err}}$

  $P_n \leftarrow WL_n(F))$ // Get the prediction of the nth WL.

$Q \longleftarrow \sum_{n=1}^{L} w_n * P_n$ // Compute the final quality score.

D_1 are set to 1):

$$D_{i+1,j} = D_{i,j} \times (1 + \delta . I(Y_j - Y_{i,j}^{pred})) \quad with \begin{cases} i = 1, .., L \\ j = 1, .., M \end{cases} \tag{15}$$

The ith ANN evaluation error Err_i with the corresponding distribution D_i is defined as:

$$Err_i = \sum_{j=1}^{M} \mid D_{i,j} \times I(Y_j - Y_{i,j}^{pred}) \mid \tag{16}$$

where the function I is a binary function in which

$$I(x) = \begin{cases} 1 & \text{if } x > 0.2, \\ 0 & otherwise. \end{cases} \tag{17}$$

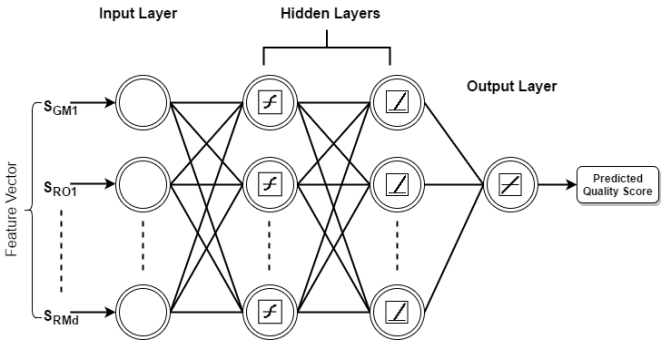


Fig. 8: Structure of the used BP neural network.

Third, assign a weight w_i for the ith ANN using its error Err_i. Finally, the ith ANN predicts the quality P_i for the input F. For each ANN model, the adjusted weights and biases are randomly initialized. Hence, it produces varied L number of WL models with different prediction scores. Error threshold for the binary function I is set to 0.2. M is set to the vector size dedicated for testing. j indexes the jth element in a vector whose range is the integers between 1 and M. For instance, $D_{1,j}$ stands for the jth element in the vector D_1. δ is a constant multiplication factor, both of threshold and δ values are fixed to 0.2.

A convex function is used to convert the error of each ANN into its weight, in order to give the ANN models with a low error a high weight, and models with a high error a small weight. ω_i is the ith ANN weight, given as:

$$\omega_i = \frac{1}{e^{Err_i}} \tag{18}$$

The overall predicted measure is given by the weighted sum of the collection as:

$$Q = \sum_{i=1}^{L} \omega_i \times P_i \tag{19}$$

For the training dataset output, human scores are normalized in the form of DMOS to min-max normalization [0,1]. Hence, the range of the predicted measure values is from 0 to 1. The closer to 0 the better quality of the stereo image is. Algorithm 1 describes the developed AdaBoost regression algorithm.

3.3.2. Network architecture

The AdaBoost neural network has been used to predict the stereo image quality. It has been applied to the obtained features from the disparity and cyclopean image. In the BP neural network, nine inputs cells have been deployed as the size of the final features vector (F) described in equation (14). Elements of the F vector are also mentioned in figure 6 as input elements for the ANN. Two hidden layers have been employed with nine neurons each. The applied transfer functions are tangent sigmoid and ReLU for the first and second hidden layers, respectively as shown in figure 8. A pure linear transfer function $f(x) = x$ has been used for a single node output layer. In hidden layers, a number of tests have been carried out using various activation functions. The tangent sigmoid and ReLU functions have been selected, for their best performance.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed approach has been tested on different databases. The obtained results have been compared to several FR and NR stereo IQA metrics, including six FR and eight NR stereo schemes. The standard performance assessment used in the Video Quality Experts Group (VQEG) has been considered. Objective scores are fitted to the subjective ones using logistic function [42]. This function is based on five parameters (θ_1, θ_2, θ_3, θ_4 and θ_5). The used logistic mapping function for the nonlinear regression is introduced by equation (20).

$$Q_{map} = \theta_1 \left(\frac{1}{2} - \frac{1}{\exp\left(\theta_2 \left(Q - \theta_3\right)\right)} \right) + \theta_4 Q + \theta_5 \tag{20}$$

Where Q and Q_{map} are the objective quality scores before and after the nonlinear mapping, and θ_i (i = 1 to 5) are selected for the most excellent fit.

Three widely-used performance indicators have been chosen to benchmark the proposed metric against the relevant state-of-the-art techniques: Pearson linear correlation coefficient (*LCC*), Spearman's rank-order correlation coefficient (*SROCC*) and root mean squared error (*RMSE*), between the objective and human score (*DMOS*). *LCC* and *RMSE* assess the prediction accuracy while *SROCC* evaluates the prediction notability degree. Higher values for *LCC* and *SROCC* (close to 1) and lower values for *RMSE* (close to 0) indicate superior linear rank-order correlation and better precision with respect to human quality judgments, respectively. For a perfect match between the objective and subjective scores, *LCC* = *SROCC* = 1 and *RMSE* = 0. Cross-validation training and testing provides a more accurate estimate of a model performance. However, several cross-validation techniques have been proposed, such as: LOOCV-Leave one out cross-validation and K-Fold cross-validation. The K-fold technique uses all data points to contribute to an understanding of how well the model performs the task of learning from some data and predicting some new data.

In order to ensure that the proposed approach is robust across content and it is governed by quality-aware indicators, the 5-fold cross validation over the three databases has been used. For every database, the dataset has been divided into 5 folds, where each fold contains a 80%-train set and 20%-test set randomly selected. The overlap between the test and the train set has been avoided to ensure that the reported results do not depend on features derived from known spatial information, which can artificially improve the performance. To demonstrate the generalization of the proposed metric against databases, a cross-database tests have been conducted. For further statistical performance analysis, a T-test scores have been computed over the correlation coefficients *LCC* and *SROCC*. In the different tests, the correlation of the feature vector with subjective human judgment has been studied. Complexity and time consuming of the proposed approach have been computed as well. Finally, influence of the formed cyclopean image and disparity map have been studied.

4.1. Datasets

The different experimentations have been carried out on three databases; including, IRCCyN/IVC 3D, LIVE 3D phase I and LIVE 3D phase II. The IVC 3D Image Quality Database has been established in 2008 [11]. It is the first public-domain database on stereoscopic image quality. Test conditions include JPEG and JPEG2000 compression as well as blur. This dataset contains 96 stereoscopic images and their associated subjective scores. The resolution of these images is 512 $\times$ 512 pixels. 6 different stereoscopic images are used in this database which is composed of 6 reference images and 16 distorted versions of each source generated from 3 different distortion types (JPEG, JP2K, BLUR) symmetrically to the stereopair images.

The popular databases LIVE 3D-I and LIVE 3D-II have been established in 2012 and 2013 respectively [43, 17]. The phase I consists of 365 distorted stereo images with a resolution of 640 $\times$ 360 pixels. There are eighty stereo images for each JPEG, JPEG2000 (JP2K), White Noise (WN), and Fast Fading (FF). The remaining 45 stereo images represent Blur distortion. All distortions are symmetric in nature. Phase II consists of 360 distorted stereoscopic images. This database includes asymmetric and symmetric distorted stereopairs over five types of distortion as the phase I. The two phases constitute the largest and most comprehensive stereoscopic image quality database currently available. The three publicly available datasets have been used to test the performance of the proposed model on several different types of distortion. Table 2 describes the three databases used for the performance evaluation.

Table 2: Summary of the three databases.

Database	$stereopairs\ (sym., asym.)$	$distortions$
LIVE 3D-I	$365(365, 0)$	$JP2K, JPEG, WN, Blur, FF$
LIVE 3D-II	$360(120, 240)$	$JP2K, JPEG, WN, Blur, FF$
IVC 3D	$90(90, 0)$	$JP2K, JPEG, Blur, down/upscaling$

4.2. Feature vector correlation with human score

In this section, the feature vector F correlation with DMOS is evaluated. It is worth recalling that the regression model input is a vector of nine elements, and because of the restriction of human spatial awareness, it is hard to demonstrate the discriminative capacity of the characteristics in a graphical manner, such as a four-dimensional scatter plot. Three plots are used to describe the correlation of the adopted three indicators S_{GM}, S_{RO}, S_{RM} with the human opinion score.

The relationship between stereo image quality and the three features is visually illustrated in the form of three-dimensional plots. The extracted features are used as axes and each stereo image corresponds to a coordinate system scatter point. All the stereoscopic images from the LIVE 3D-I and LIVE 3D-II database are used for this demonstration. As shown in Figure 9, the plots refer to features from cyclopean image scale 1, scale 2, and disparity map respectively from top to bottom. To differentiate the five types of distortion, we use distinct labels and map the DMOS rating of each stereo image to the preset color-map.

The ideal scenario is that the points are well separated with distinct kinds of distortion. It can be seen from Figure 9 that the scatter points of the five distortions are generally distinguished. The used stereo images are distorted increasingly from low to high factor. This can also be observed in the plots, where the adopted features vary smoothly in space with quality correspondence. Although there is some correlation between the extracted features, in particular S_{GM} and S_{RM}, where the coefficient correlation in terms of LCC is equal to 0.751. The deployed features provide good performance, this topic is discussed furthermore in section 4.6.

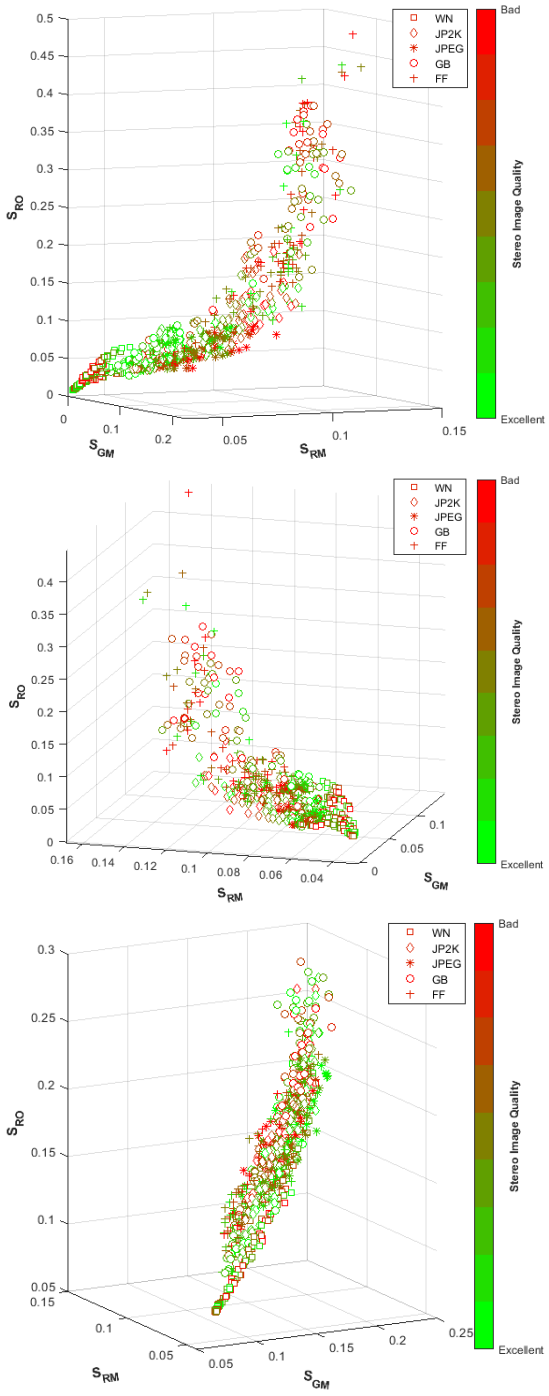


Fig. 9: Illustration of the discriminatory power of the extracted features. Respectively from top to bottom: elements in the axis are from cyclopean image scale 1, scale 2, and disparity. (zoom in to get the markers more discriminative).

4.3. Comparison with other stereo IQA methods

The overall performance of the proposed scheme has shown good efficiency and consistency. The obtained results have been compared with several full-reference and no-reference stereo IQA metrics, including six FR, and eight NR metrics.

For the comparison purpose, two models have been created. The first model called *3D-nnet*, is a normal neural network regression model. It is equivalent to $L = 1$ in the AdaBoost algorithm (Algorithm 1). The second model named *3D-AdaBoost*, is a neural network combined with the AdaBoost technique as previously demonstrated, where 20 neural network models have been employed ($L = 20$). We find that the performance of the proposed measure is improved by using additional neural network models (Weak Learners) with saturation at a certain number and decreasing in the other case. Note that both models have the same network architecture. Also during the training, 15% is taken out from training set for validation. A Box plot in term of *SROCC* of the two models results is displayed in figure 10. Comparing the proposed models indicates that the performance can be improved by the adopted Adaptive Boosting technique. Tables 3,4 and 5 show the results against DMOS

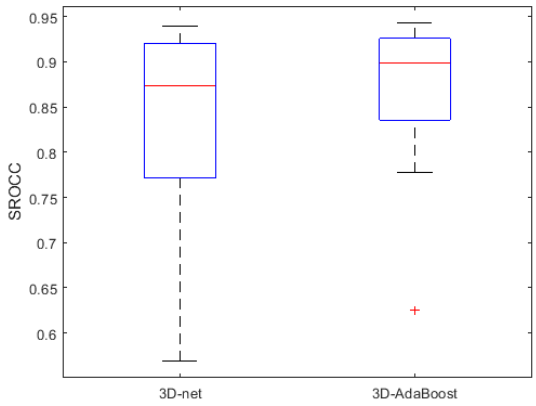


Fig. 10: Comparison Box plots of *SROCC* of the proposed models. The *SROCC* results are split into four groups (quartiles). Each group has 25% of the results. The red line in the rectangle refers to the median value. Upper and lower ends of the rectangle limit the first and third quartiles, respectively. The length of the dashed line means the range of the mild outliers, and the symbol "+" refers to the extreme outlier.

of all stereo IQA algorithms on LIVE 3D-I. Tables 6,7 and 8 show the results against DMOS of all stereo IQA algorithms on LIVE 3D-II. Moreover, the performance on symmetrically and asymmetrically distorted stimuli are shown separately in table 9, while table 10 provides detailed results over the five distortions. The 3D-nnet model has exhibited a good performance. Among all the comparison metrics, the model has obtained the best SROCC score on the two LIVE 3D databases, (SROCC=0.916 on LIVE 3D-I and SROCC=0891 on LIVE 3D-II.) Finally, table 11 shows the results on IVC 3D database. In these tables, the top NR methods results are highlighted in bold.

The proposed model *3D-AdaBoost* has given the best performance among all compared no-reference algorithms, while the full-reference method of Chen [17] yields better performance compared to other FR methods. Figure 11 exhibits the prediction responses against human score DMOS on the three databases. Even though the proposed model is not designed for a specific distortion type. The comparison results on each individual distortion type indicate superiority of the proposed method over the three databases. More specifically, notice that the most of existing stereo IQA methods remain limited in capability and efficiency on asymmetric degra-

dations. These metrics are more appropriate for symmetric distortion, but insufficient for asymmetric one. On the other hand, results in table 10 show that the proposed framework delivers efficient performance over asymmetric/symmetric distortion. The method achieved LCC scores of 0.930 and 0.903 respectively on asymmetric/symmetric degradation. A scatter plots of figures 12 and 13 show the predicted quality on these two types of distortion separately.

It should be noticed that the used neural network model presents better performance than the most commonly used (SVR) Support Vector Regression. The same evaluation process is followed for SVR with 5-fold cross validation. A radial basis function (RBF) kernel has been selected. The other parameters such as the number of support vectors and iterations are adjusted automatically during training for the best fit. Table 12 shows the superiority of the implemented neural network architecture over SVR. The mean scores of each learning method over three databases (LIVE 3D-I, LIVE 3D-II, and IVC 3D) have been calculated (see table 12).

Table 3: SROCC against DMOS on the LIVE 3D phase I dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		0.930	0.910	0.603	0.931	0.699	0.899
You [12]		0.940	0.860	0.439	0.882	0.588	0.878
Gorley [13]	*FR*	0.741	0.015	0.569	0.750	0.366	0.142
Chen [17]		0.948	0.888	0.530	0.925	0.707	0.916
Hewage [14]		0.940	0.856	0.500	0.690	0.545	0.814
Bensalma [7]		0.905	0.817	0.328	0.915	0.915	0.874
DIIVINE [6]		–	–	–	–	–	0.882
Akhter [10]		0.914	0.866	0.675	0.555	0.640	0.383
Chen [19]		0.919	0.863	0.617	0.878	0.652	0.891
Lv [26]	*NR*	–	–	–	–	–	0.897
Appina [25]		0.910	**0.917**	**0.782**	0.865	0.666	0.911
Zhou [23]		0.921	0.856	0.562	**0.897**	0.771	0.901
Fang [22]		0.883	0.880	0.523	0.523	0.650	0.877
3D-nnet		0.938	0.874	0.569	0.866	0.685	0.916
Proposed 3D-AdaBoost		**0.941**	0.899	0.625	0.887	**0.777**	**0.930**

Table 4: LCC against DMOS on the LIVE 3D phase I dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		0.925	0.939	0.640	0.948	0.747	0.902
You [12]		0.941	0.877	0.487	0.919	0.730	0.881
Gorley [13]	*FR*	0.796	0.485	0.312	0.852	0.364	0.451
Chen [17]		0.942	0.912	0.603	0.942	0.776	0.917
Hewage [14]		0.895	0.904	0.530	0.798	0.669	0.830
Bensalma [7]		0.914	0.838	0.838	0.838	0.733	0.887
DIIVINE [6]		–	–	–	–	–	0.893
Akhter [10]		0.904	0.905	0.729	0.617	0.503	0.626
Chen [19]		0.917	0.907	0.695	0.917	0.735	0.895
Lv [26]	*NR*	–	–	–	–	–	0.901
Appina [25]		0.919	**0.938**	**0.806**	0.881	0.758	0.917
Zhou [23]		–	–	–	–	–	0.929
Fang [22]		0.900	0.911	0.547	0.903	0.718	0.880
3D-nnet		0.941	0.919	0.625	0.908	0.777	0.923
Proposed 3D-AdaBoost		**0.941**	0.926	0.668	**0.935**	**0.845**	**0.939**

4.4. Performance using T-test

T-test is one of several types of statistical tests [44]. It questions whether the difference between the groups represents a true difference in the study or if it is likely a meaningless statistical difference, where 1 indicates that the groups are statistically different and 0 indicates that the groups are statistically similar. In order to investigate the statistical performance of the proposed metric, it is compared with the state-of-the-art methods. We conducted a left-tail T-test with confidence at 90% applied over 100 trials for PLCC and SROCC. The results provided in table 13 show the superiority of the proposed method over the existing ones.

Table 5: RMSE against DMOS on the LIVE 3D phase I dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		6.307	4.426	5.022	4.571	8.257	7.061
You [12]		5.621	6.206	5.709	5.679	8.492	7.746
Gorley [13]	*FR*	10.197	11.323	6.211	7.562	11.569	14.635
Chen [17]		5.581	5.320	5.216	4.822	7.837	6.533
Hewage [14]		7.405	5.530	5.543	8.748	9.226	9.139
Bensalma [7]		–	–	–	–	–	7.558
DIIVINE [6]		–	–	–	–	–	7.301
Akhter [10]		7.092	5.483	**4.273**	11.387	9.332	14.827
Chen [19]		6.433	5.402	4.523	5.898	8.322	7.247
Lv [26]	*NR*	–	–	–	–	–	–
Appina [25]		6.664	4.943	4.391	6.938	9.317	6.598
Zhou [23]		–	–	–	–	–	6.010
Fang [22]		–	–	–	–	–	7.191
3D-nnet		5.622	5.083	5.104	6.059	7.819	6.277
Proposed 3D-AdaBoost		**5.593**	**4.867**	4.862	**5.104**	**6.633**	**5.605**

Table 6: SROCC against DMOS on the LIVE 3D phase II dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		0.923	0.751	0.867	0.455	0.773	0.728
You [12]		0.909	0.894	0.795	0.813	0.891	0.786
Gorley [13]	*FR*	0.875	0.110	0.027	0.770	0.601	0.146
Chen [17]		0.940	0.814	0.843	0.908	0.884	0.889
Hewage [14]		0.880	0.598	0.736	0.028	0.684	0.501
Bensalma [7]		0.938	0.803	0.846	0.846	0.846	0.751
DIIVINE [6]		–	–	–	–	–	0.346
Akhter [10]		0.714	0.724	0.649	0.682	0.559	0.543
Chen [19]		0.950	**0.867**	**0.867**	0.900	0.933	0.880
Lv [26]	*NR*	–	–	–	–	–	0.862
Appina [25]		0.932	0.864	0.839	0.846	0.860	0.888
Zhou [23]		0.936	0.647	0.737	0.911	0.798	0.819
Fang [22]		**0.955**	0.714	0.709	0.807	0.872	0.838
3D-nnet		0.939	0.812	0.745	0.900	**0.934**	0.891
Proposed 3D-AdaBoost		0.943	0.842	0.837	**0.913**	0.925	**0.913**

Table 7: LCC against DMOS on the LIVE 3D phase II dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		0.926	0.784	0.853	0.535	0.807	0.784
You [12]		0.912	0.905	0.830	0.784	0.915	0.800
Gorley [13]	*FR*	0.874	0.372	0.322	0.934	0.706	0.515
Chen [17]		0.957	0.834	0.862	0.963	0.901	0.907
Hewage [14]		0.891	0.664	0.734	0.450	0.746	0.558
Bensalma [7]		0.943	0.666	0.857	0.907	0.909	0.769
DIIVINE [6]		–	–	–	–	–	0.442
Akhter [10]		0.772	0.776	0.786	0.795	0.674	0.568
Chen [19]		0.947	**0.899**	**0.901**	0.941	**0.932**	0.895
Lv [26]	*NR*	–	–	–	–	–	0.870
Appina [25]		0.920	0.867	0.829	0.878	0.836	0.845
Zhou [23]		–	–	–	–	–	0.856
Fang [22]		**0.961**	0.740	0.764	0.968	0.867	0.860
3D-nnet		0.948	0.821	0.758	0.960	0.921	0.900
Proposed 3D-AdaBoost		0.953	0.835	0.859	**0.978**	0.925	**0.922**

Table 8: RMSE against DMOS on the LIVE 3D phase II dataset.

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Benoit [11]		4.028	6.096	3.787	11.763	6.894	7.490
You [12]		4.396	4.186	4.086	8.649	4.649	6.772
Gorley [13]	*FR*	5.202	9.113	6.940	4.988	8.155	9.675
Chen [17]		3.368	5.562	3.865	3.747	4.966	4.987
Hewage [14]		10.713	7.343	4.976	12.436	7.667	9.364
Bensalma [7]		–	–	–	–	–	7.203
DIIVINE [6]		–	–	–	–	–	10.012
Akhter [10]		7.416	6.189	4.535	8.450	8.505	9.294
Chen [19]		3.513	**4.298**	**3.342**	4.725	**4.180**	5.102
Lv [26]	*NR*	–	–	–	–	–	–
Appina [25]		4.325	5.087	4.756	6.662	6.519	7.279
Zhou [23]		–	–	–	–	–	6.041
Fang [22]		–	–	–	–	–	5.767
3D-nnet		3.394	5.598	4.780	3.889	4.481	4.905
Proposed 3D-AdaBoost		**3.226**	5.396	3.752	**2.859**	4.352	**4.352**

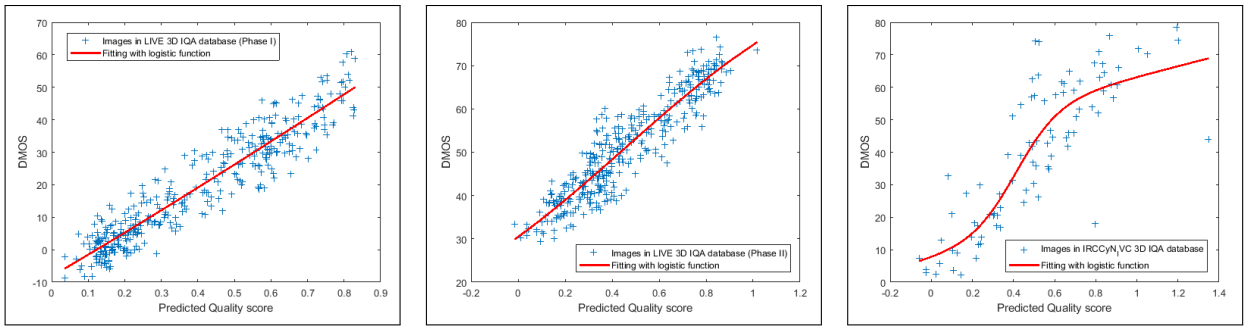


Fig. 11: Scatter plots of subjective scores versus scores from the proposed scheme on the three stereopair IQA databases.

Table 9: SROCC result on Symmetric and Asymmetric distortion from LIVE 3D phase II dataset.

Method	*Type*	*Symmetric*	*Asymmetric*
Benoit [11]		0.860	0.671
You [12]		0.914	0.701
Gorley [13]	*FR*	0.383	0.056
Chen [17]		0.923	0.842
Hewage [14]		0.656	0.496
Bensalma [7]		0.841	0.721
DIIVINE [6]		–	–
Akhter [10]		0.420	0.517
Chen [19]		**0.918**	0.834
Lv [26]	*NR*	–	–
Appina [25]		0.857	0.872
Zhou [23]		–	–
Fang [22]		–	–
3D-nnet		0.861	0.902
Proposed 3D-AdaBoost		0.898	**0.917**

Table 10: Detailed results of SROCC, LCC, and RMSE on symmetric / asymmetric distortion from LIVE 3D-II.

Method	*Indicator*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Proposed 3D-AdaBoost	*SROCC*	0.923	0.829	0.933	0.848	0.889	0.898
Symmetric	*LCC*	0.938	0.922	0.946	0.913	0.903	0.903
	RMSE	3.701	3.709	3.819	3.425	4.876	4.609
Proposed 3D-AdaBoost	*SROCC*	0.897	0.926	0.897	0.921	0.945	0.917
Asymmetric	*LCC*	0.930	0.947	0.917	0.932	0.953	0.930
	RMSE	4.191	4.006	4.747	3.450	3.387	4.216

Table 11: SROCC, LCC, and RMSE against DMOS on the IVC 3D database.

Method	*Type*	*SROCC*	*LCC*	*RMSE*
Benoit [11]		–	–	–
You [12]		–	–	–
Gorley [13]	*FR*	–	–	–
Chen [17]		0.676	0.683	17.100
Hewage [14]		–	–	–
Bensalma [7]		–	–	–
DIIVINE [6]		0.422	0.486	18.259
Akhter [10]		–	–	–
Chen [19]		**0.851**	0.835	12.088
Lv [26]	*NR*	–	–	–
Appina [25]		–	–	–
Zhou [23]		–	–	–
Fang [22]		–	–	–
3D-nnet		0.780	0.779	13.830
Proposed 3D-AdaBoost		0.831	**0.845**	**11.776**

4.5. Cross-database performance

The above tests are useful for assessing robustness and generalization of the proposed metric, since all the results are obtained by

Table 12: Mean of SROCC, LCC, and RMSE results from the three databases using different regressors.

Method	*SROCC*	*LCC*	*RMSE*
SVR	0.8223	0.8406	9.0530
3D-nnet	0.8623	0.8673	8.3373
Proposed 3D-AdaBoost	**0.8913**	**0.9020**	**7.2443**

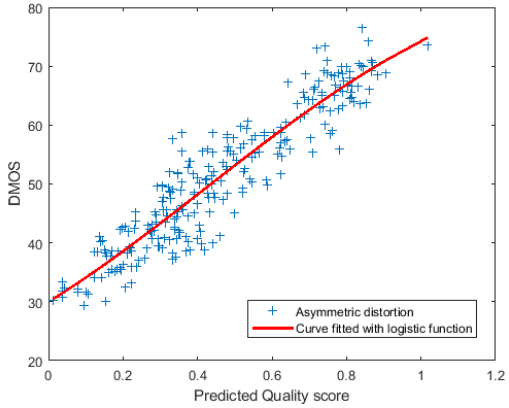


Fig. 12: Scatter plot of asymmetric distortions scores from LIVE 3D phase II IQA database using *3D-AdaBoost* method.

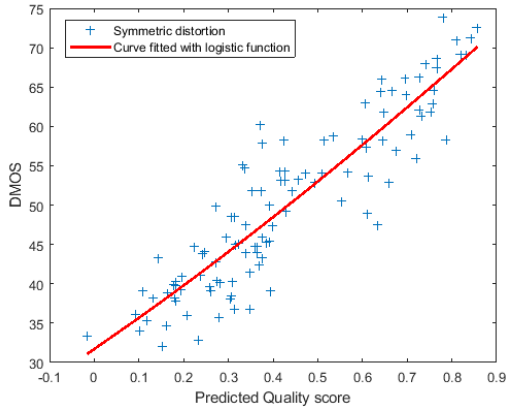


Fig. 13: Scatter plot of symmetric distortions scores from LIVE 3D phase II IQA database using *3D-AdaBoost* method.

Table 13: T-test results with confidence of 90% of the proposed metric against the others using PLCC, SROCC from LIVE I and II

Method		*Akhter*	*Chen*	*Appina*	*Zhou*	*Fang*	*3D-nnet*
LIVE I	LCC	1	1	1	1	1	0
	$SROCC$	1	1	0	1	1	1
LIVE II	LCC	1	1	1	1	1	1
	$SROCC$	1	1	1	1	1	1

training and testing using 5-fold cross validation. We extend cross-database experiments to demonstrate the performance capability of the proposed metric. The LIVE 3D phase I and phase II databases have been selected for these experiments because of the similarity in the number of stereo images. The model is trained on one database and tested on another one.

The Weak Learners (WL) in the *3D-AdaBoost* algorithm 1 are trained, validated, and tested on the LIVE 3D phase I database to obtain a model which will be tested on the LIVE 3D phase II database. Images in the LIVE 3D phase I database have been used for training, validating and testing, and images from the LIVE 3D phase II database are used as a final test set. The obtained results on LIVE 3D-1 are shown in table 15 using LIVE 3D-II for training, whereas table 16 corresponds to the inverse process. Tables 17, and 18 present detailed results over the five distortions. The *SROCC* has been used as a performance index. The best results are highlighted in bold.

It can be noticed that *3D-AdaBoost* trained on the LIVE 3D phase I database achieved lower performance compared to the model trained on phase II. This is due to the lack of asymmetric distortion in the LIVE 3D phase I database. However, it is interesting to observe that the *3D-AdaBoost* method produces good results on the LIVE 3D phase I database. Compared to other methods, although their results are not performed using cross-dataset test, the proposed metric ensures competitive performance on any type of distortion commonly encountered. Scatter plots of figures 14 and 15 show the *3D-AdaBoost* metric responses of cross-dataset test.

The overall experimental results have shown that the proposed method has good consistency among five distortion types with human subjective evaluation. The cross-database test showed the proposed metric reliability for measuring the quality of the stereoscopic image. Among the five distortions, JPEG distortion has the lowest accuracy. We believe this is due to the complexity of the compression distortion. Thus, it should be addressed separately for the stereo image quality assessment.

Table 14: SROCC, LCC, and RMSE results on the LIVE 3D-I dataset. (Trained on LIVE 3D-II).

Method	*Type*	*SROCC*	*LCC*	*RMSE*
DIIVINE [6]		0.882	0.893	7.301
Akhter [10]		0.383	0.626	14.827
Chen [19]		0.891	0.626	7.247
Lv [26]	*NR*	0.897	0.901	–
Appina [25]		**0.911**	0.917	6.598
Zhou [23]		0.901	**0.929**	**6.010**
Fang [22]		0.877	0.880	7.191
3D-nnet		0.880	0.888	7.514
Proposed 3D-AdaBoost		0.887	0.897	7.224

4.6. Influence of cyclopean view and disparity map

In order to demonstrate the efficiency of the proposed approach for measuring the stereo image quality, numerous tests have been conducted that cover the possibilities of feature extraction part. Also a simple feature extraction has been used for comparison. Pixel sum and pixel average have been used. The proposed learning part remains the same as described, using the 5-fold cross validation. The *3D-AdaBoost* model receives different input at each combination, and the mean performance of each combination is calculated over the three databases. The results of the tests are shown in table 18.

Table 15: SROCC, LCC, and RMSE results on the LIVE 3D-II dataset. (Trained on LIVE 3D-I).

Method	*Type*	*SROCC*	*LCC*	*RMSE*
DIIVINE [6]		0.346	0.442	10.012
Akhter [10]		0.543	0.568	9.294
Chen [19]		0.543	**0.895**	**5.102**
Lv [26]	*NR*	0.862	0.870	–
Appina [25]		**0.888**	0.845	7.279
Zhou [23]		0.819	0.856	6.041
Fang [22]		0.838	0.860	5.767
3D-nnet		0.798	0.813	6.561
Proposed 3D-AdaBoost		0.823	0.832	6.253

Table 16: SROCC against DMOS on the LIVE 3D-I dataset. (Trained on LIVE 3D-II).

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
DIIVINE [6]		–	–	–	–	–	0.882
Akhter [10]		0.914	0.866	0.675	0.555	0.640	0.383
Chen [19]		0.919	0.863	0.617	0.878	0.652	0.891
Lv [26]	*NR*	–	–	–	–	–	0.897
Appina [25]		0.910	**0.917**	**0.782**	0.865	0.666	**0.911**
Zhou [23]		0.921	0.856	0.562	**0.897**	**0.771**	0.901
Fang [22]		0.883	0.880	0.523	0.523	0.650	0.877
3D-nnet		0.955	0.873	0.588	0.808	0.527	0.880
Proposed 3D-AdaBoost		**0.956**	0.889	0.556	0.875	0.530	0.892

Table 17: SROCC against DMOS on the LIVE 3D-II dataset. (Trained on LIVE 3D-I).

Method	*Type*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
DIIVINE [6]		–	–	–	–	–	0.346
Akhter [10]		0.714	0.724	0.649	0.682	0.559	0.543
Chen [19]		0.950	**0.867**	**0.867**	0.900	0.933	0.880
Lv [26]	*NR*	–	–	–	–	–	0.862
Appina [25]		0.932	0.864	0.839	0.846	0.860	**0.888**
Zhou [23]		0.936	0.647	0.737	0.911	0.798	0.819
Fang [22]		**0.955**	0.714	0.709	0.807	0.872	0.838
3D-nnet		0.882	0.803	0.772	**0.925**	**0.936**	0.798
Proposed 3D-AdaBoost		0.932	0.826	0.737	0.881	0.924	0.824

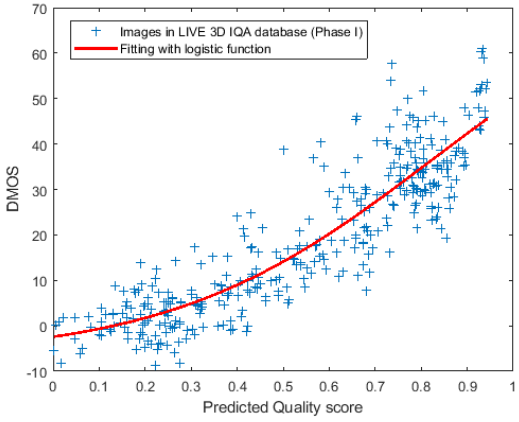


Fig. 14: Scatter plot of cross-dataset scores on LIVE 3D phase I database using *3D-AdaBoost* method.

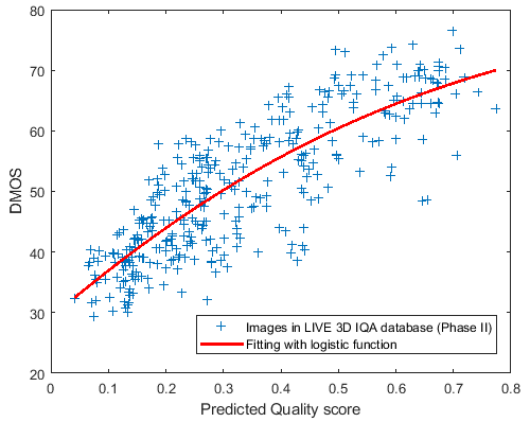


Fig. 15: Scatter plot of cross-dataset scores on LIVE 3D phase II database using *3D-AdaBoost* method.

The pixel sum P_S is defined as follows:

$$P_S = \sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j) \tag{21}$$

where I is the left or right image. The pixel average P_A is defined by:

$$P_A = \frac{1}{m.n}\sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j) \tag{22}$$

From the results it can be observed that the pixel sum and average indicators give a bad performance, because these features do not correlate with the image quality. Meanwhile, the used features give good performance due to their relationship with distortion types and quality degradation as shown previously in Figure 9. It is also noticeable that when using disparity map features, the performance improves which supports the study conducted in [9]. The authors used different measures to illustrated the relationship between the perceptual quality of stereo views and the quality of the disparity map. They concluded that the quality of the depth map is highly correlated with the overall 3D quality.

As discussed earlier, the 2D IQA metrics may not be applied to the stereo IQA problem, since either by averaging the score or the features obtained from left and right image will not consider asymmetrical distortions. The improved 2D IQA metric *DIIVINE* [6] for stereo images provides good performance on the LIVE 3D-I database and low performance on the LIVE 3D-II database. This is because the LIVE 3D-II database mainly contains asymmetric distorted stereo images (see table 2). However, due to the fact that stereo images typically contain redundant information, a feature extraction from the left and right image may result in a redundant features. Therefore, the extracted features (S_{GM}, S_{RO}, and S_{RM}) from left and right images are averaged. Afterward, the *3D-AdaBoost* model has been used to map these features to predict the quality. It is also noticed that the use of 2-scale cyclopean image increases the accuracy of quality prediction. We assume that the space distance between cyclopean scale 1 features and cyclopean scale 2 features is learned while training, helping the model for better prediction. Additionally, the use of pixel sum P_S and pixel average P_A as features decreases the performance as shown in table 18.

Even though the features are somewhat correlated, the model has given good results. Some tests have been carried out to support the use of all gradient extracted features (S_{GM}, S_{RO}, and S_{RM}). The performance deteriorates if one or two features among the three are neglected; as shown in table 19. Therefore, it is important to utilize the three features for better quality assessment accuracy.

Furthermore, an additional experiment has been conducted without using Gabor filter. w_l and w_r have been set to 1 in equation (2). Results given by table 20 indicate that the cyclopean model using Gabor weights is better than the simple cyclopean model, in particular on the LIVE II. Compared to the stereo image model, the simple cyclopean model is also competitive, but the model may not be accurate on the asymmetric distortion situation as discussed earlier. The results of table 20 also support the idea of using cyclopean view rather than using the stereo image for quality assessment problem. The superiority of the cyclopean image on symmetric and asymmetric degradations is also shown in table 21. Notice that the performance of the stereopair image method in table 22 drops significantly on LIVE 3D-II over all distortions. Consequently, for asymmetric distortions, extracting features directly from stereo images is not reliable. Also, the performance of cyclopean image method maintains consistency. In the tables, N refers to the number of input features to the regression model *3D-AdaBoost*. Overall, we can conclude that the adopted cyclopean model and quality indicators (S_{GM}, S_{RO}, and S_{RM}) and the used combination are effective for assessing the quality of stereopair images.

Table 18: Mean of SROCC, LCC, and RMSE results from the three databases using various features and combinations.

The used material	*Features*	*N*	*SROCC*	*LCC*	*RMSE*
Stereopair Image	P_S	2	0.107	0.196	16.182
Stereopair Image	P_A	2	0.255	0.208	16.097
Stereopair Image	P_S, P_A	4	0.244	0.245	15.995
Stereopair Image	S_{GM}, S_{RO}, S_{RM}	3	0.740	0.769	10.386
Stereopair Image, disparity	S_{GM}, S_{RO}, S_{RM}	6	0.858	0.869	8.203
Stereopair Image, disparity	S_{GM}, S_{RO}, S_{RM}, P_S, P_A	12	0.716	0.741	9.803
Cyclopean view (scale 1)	S_{GM}, S_{RO}, S_{RM}	3	0.776	0.801	9.802
Cyclopean view (scale 1)	S_{GM}, S_{RO}, S_{RM}, P_S, P_A	5	0.676	0.705	10.552
Cyclopean view (scale 1, and 2)	S_{GM}, S_{RO}, S_{RM}	6	0.798	0.818	9.444
Cyclopean view (scale 1, and 2)	S_{GM}, S_{RO}, S_{RM}, P_S, P_A	10	0.688	0.715	10.298
Cyclopean view (scale 1, and 2), disparity	S_{GM}, S_{RO}, S_{RM}	9	**0.891**	**0.902**	**7.244**
Cyclopean view (scale 1, and 2), disparity	S_{GM}, S_{RO}, S_{RM}, P_S, P_A	15	0.628	0.635	12.445

Table 19: Mean of SROCC, LCC, and RMSE results from the three databases using different gradient features and combinations.

The used material	*Features*	*N*	*SROCC*	*LCC*	*RMSE*
Cyclopean view (scale 1, and 2), disparity	S_{RO}	3	0.725	0.748	10.698
Cyclopean view (scale 1, and 2), disparity	S_{GM}	3	0.724	0.751	10.722
Cyclopean view (scale 1, and 2), disparity	S_{RM}	3	0.751	0.709	10.995
Cyclopean view (scale 1, and 2), disparity	S_{RO}, S_{RM}	6	0.809	0.817	9.401
Cyclopean view (scale 1, and 2), disparity	S_{GM}, S_{RM}	6	0.844	0.849	8.768
Cyclopean view (scale 1, and 2), disparity	S_{RO}, S_{GM}	6	0.838	0.847	8.823
Cyclopean view (scale 1, and 2), disparity	S_{GM}, S_{RO}, S_{RM}	9	**0.891**	**0.902**	**7.244**

Table 20: Cyclopean view versus Stereopair image method results over the three databases.

The used material	*Features*	*N*	*Indicator*	Live 3D-I	Live 3D-II	IVC 3D
Stereopair Image	S_{GM}, S_{RO}, S_{RM}	3	*SROCC*	0.905	0.725	0.590
			LCC	0.913	0.791	0.602
			RMSE	6.657	6.896	17.606
Cyclopean Image (scale 1)	S_{GM}, S_{RO}, S_{RM}	3	*SROCC*	0.908	0.797	0.622
			LCC	0.920	0.850	0.634
			RMSE	6.417	5.944	17.046
Cyclopean Image Simple (scale 1)	S_{GM}, S_{RO}, S_{RM}	3	*SROCC*	0.904	0.780	0.607
			LCC	0.914	0.828	0.622
			RMSE	6.644	6.323	17.263

Table 21: Cyclopean view versus Stereopair image method results on Symmetric and Asymmetric distortion from LIVE 3D-II dataset.

The used material	*Features*	*N*	*Indicator*	*Symmetric*	*Asymmetric*
Stereopair Image	S_{GM}, S_{RO}, S_{RM}	3	$SROCC$	0.672	0.745
			LCC	0.779	0.802
			$RMSE$	6.746	6.851
Cyclopean Image (scale 1)	S_{GM}, S_{RO}, S_{RM}	3	$SROCC$	0.733	0.822
			LCC	0.840	0.855
			$RMSE$	5.832	5.954
Cyclopean Image Simple (scale 1)	S_{GM}, S_{RO}, S_{RM}	3	$SROCC$	0.734	0.796
			LCC	0.814	0.834
			$RMSE$	6.249	6.337

Table 22: SROCC results of Cyclopean view versus Stereopair image method over LIVE 3D-I and LIVE 3D-II databases.

The used material	*Database*	*WN*	*JP2K*	*JPEG*	*Blur*	*FF*	*All*
Stereopair Image	LIVE 3D-I	0.943	0.867	0.597	0.816	0.615	0.905
	LIVE 3D-II	0.497	0.684	0.606	0.870	0.732	0.725
Cyclopean Image (scale 1)	LIVE 3D-I	0.943	0.869	0.589	0.867	0.680	0.908
	LIVE 3D-II	0.924	0.676	0.678	0.858	0.735	0.797
Cyclopean Image Simple (scale 1)	LIVE 3D-I	0.942	0.872	0.595	0.821	0.678	0.904
	LIVE 3D-II	0.911	0.706	0.674	0.844	0.727	0.780

4.7. Computational Complexity

Computational complexity of the proposed algorithm is discussed in this section. The most computationally expensive stage is the cyclopean image construction, since it involves weights computation of the left and right views by performing a multiscale Gabor filter. The complexity of the proposed measure depends on the size of the testing vectors (M) and the number of the Weak Learners (L). Therefore, the overall complexity of the proposed algorithm is *O(M . L)*. Furthermore, the computation time of the proposed model has been computed using a laptop computer with intel i5-2410M CPU, 2.30 GHz and 8 GB RAM, hence the run time in second is 72.5238 (including training time). There are no details on the complexity of the other NR methods. So, state-of-the-art metrics complexities have not been compared.

The stereoscopic image's pixel resolution may increase or decrease the run time, as well as the hardware computing power. The test has been conducted on the stereo image shown in figure 1, of 640 x 360 pixels resolution. The more neural network models used, the higher the run time is. Clearly, the run time increases with the number of neurons. Note that the run time can be reduced via parallel computing (GPU cards) since the proposed method is based on neural networks.

5. CONCLUSION

In this paper, a new blind stereoscopic IQA metric has been proposed. The model is based on human binocular perception and advanced machine-learning algorithm. Efficient perceptual features have been extracted from the gradient magnitude (GM) map, relative gradient orientation (RO) map and the relative gradient magnitude (RM) map. Experimental results showed that the extracted features are sensitive to the five common distortions. Considering the variations of stereo image resolution and viewing conditions, a multiscale gradient maps of the cyclopean image have been employed. AdaBoost neural network is used to map the stereo image features to quality score. The overall obtained results have indicated that the metric correlates well with subjective scores DMOS over symmetric and asymmetric distortions. The proposed metric performs better in terms of both accuracy and efficiency on the three publicly available stereoscopic IQA databases, LIVE 3D-I, LIVE 3D-II, and IRCCyN/IVC 3D than the state-of-the-art methods.

In feature works, we believe that the use of the extracted features can also be useful for the development of no-reference stereoscopic video quality models. It is also possible to develop the idea of AdaBoost by incorporating other feature-learning algorithms.

6. REFERENCES

[1] Motion Picture Association of America, "2016 theatrical market statistics report," 2016.

[2] Albert M William and Darrell L Bailey, "Stereoscopic visualization of scientific and medical content," in *ACM SIGGRAPH 2006 Educators Program*, New York, NY, USA, 2006, SIGGRAPH '06, ACM.

[3] C. F. Westin, "Extracting brain connectivity from diffusion mri [life sciences]," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 124–152, 2007.

[4] Jacky Baltes, Sancho McCann, and John Anderson, "Humanoid robots: Abarenbou and daodan," *RoboCup-Humanoid League Team Description*, 2006.

[5] O. Messai, F. Hachouf, and Z. Ahmed Seghir, "Blind stereoscopic image quality assessment using cyclopean view and neural network," in *The fifth IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 196–200.

[6] Anush Krishna Moorthy and Alan Conrad Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[7] R. Bensalma and Mohamed-Chaker Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, vol. 24, no. 2, pp. 281–316, 2013.

[8] P. Seuntiens, "Visual experience of 3d tv," *doctor doctoral thesis, Eindhoven University of Technology*, 2006.

[9] Amin Banitalebi-Dehkordi, Mahsa T Pourazad, and Panos Nasiopoulos, "A study on the relationship between depth map quality and the overall 3d video quality of experience," in *2013 3DTV Vision Beyond Depth (3DTV-CON)*. IEEE, 2013, pp. 1–4.

[10] R. Akhter, ZM Parvez Sazzad, Yuukou Horita, and Jacky Baltes, "No-reference stereoscopic image quality assessment," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 75240T–75240T.

[11] A. Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, no. 1, pp. 1–13, 2009.

[12] J. You, Liyuan Xing, Andrew Perkis, and Xu Wang, "Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis," in *Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010.

[13] P. Gorley and Nick Holliman, "Stereoscopic image quality metrics and compression," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 680305–680305.

[14] CTER Hewage, Stewart T Worrall, Safak Dogan, and AM Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-d video," *Electronics letters*, vol. 44, no. 16, pp. 963–965, 2008.

[15] David G Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Ieee, 1999, vol. 2, pp. 1150–1157.

[16] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in computer vision*, pp. 726–740. Elsevier, 1987.

[17] M. Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[18] WJM Levelt, "On binocular rivalry (p. 107)," *The Hague-Paris: Mouton*, 1968.

[19] Ming-Jun Chen, Lawrence K Cormack, and Alan C Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[20] W. Hachicha, Azeddine Beghdadi, and Faouzi Alaya Cheikh, "Stereo image quality assessment using a binocular just noticeable difference model," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 113–117.

[21] Y. Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu, "Binocular just-noticeable-difference model for stereoscopic images," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 19–22, 2011.

[22] Meixin Fang and Wujie Zhou, "Toward an unsupervised blind stereoscopic 3d image quality assessment using joint spatial and frequency representations," *AEU-International Journal of Electronics and Communications*, vol. 94, pp. 303–310, 2018.

[23] Wujie Zhou, Weiwei Qiu, and Ming-Wei Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-d images," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 404–415, 2017.

[24] Che-Chun Su, Lawrence K Cormack, and Alan C Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Transactions on image processing*, vol. 24, no. 5, pp. 1685–1699, 2015.

[25] Balasubramanyam Appina, Sameeulla Khan, and Sumohana S. Channappayya, "No-reference stereoscopic image quality assessment using natural scene statistics," *Signal Processing: Image Communication*, vol. 43, pp. 1 – 14, 2016.

[26] Yaqi Lv, Mei Yu, Gangyi Jiang, Feng Shao, Zongju Peng, and Fen Chen, "No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network," *Signal Processing: Image Communication*, vol. 47, pp. 346–357, 2016.

[27] M. Korytkowski, Leszek Rutkowski, and Rafal Scherer, "On combining backpropagation with boosting," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*. IEEE, 2006, pp. 1274–1277.

[28] L. Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Processing: Image Communication*, vol. 40, pp. 1–15, 2016.

[29] Oussama Messai, Fella Hachouf, and Zianou Ahmed Seghir, "Deep learning and cyclopean view for no-reference stereoscopic image quality assessment," in *2018 International Conference on Signal, Image, Vision and their Applications (SIVA)*. IEEE, 2018, pp. 1–6.

[30] R. Blake, David H Westendorf, and Randall Overton, "What is suppressed during binocular rivalry?," *Perception*, vol. 9, no. 2, pp. 223–231, 1980.

[31] Karsten Mühlmann, Dennis Maier, Jürgen Hesser, and Reinhard Männer, "Calculating dense disparity maps from color stereo images, an efficient implementation," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 79–88, 2002.

[32] Z. Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] John G Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision research*, vol. 20, no. 10, pp. 847–856, 1980.

[34] D. J Field, "Relations between the statistics of natural images and the response properties of cortical cells," *JOSA A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[35] C. Su, Alan C Bovik, and Lawrence K Cormack, "Natural scene statistics of color and range," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 257–260.

[36] Naoum P Issa, Christopher Trepel, and Michael P Stryker, "Spatial frequency maps in cat visual cortex," *Journal of Neuroscience*, vol. 20, no. 22, pp. 8504–8514, 2000.

[37] C. Schor, Ivan Wood, and Jane Ogawa, "Binocular sensory fusion is limited by spatial resolution," *Vision research*, vol. 24, no. 7, pp. 661–665, 1984.

[38] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Ieee, 2003, vol. 2, pp. 1398–1402.

[39] Yoav Freund and Robert E Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[40] Peter L Bartlett and Mikhail Traskin, "Adaboost is consistent," *Journal of Machine Learning Research*, vol. 8, no. Oct, pp. 2347–2368, 2007.

[41] B. P Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor, "Boosted decision trees as an alternative to artificial neural networks for particle identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2, pp. 577–584, 2005.

[42] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[43] A.K. Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[44] ITUT Rec, "P. 1401, methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," *International Telecommunication Union, Geneva, Switzerland*, 2012.

[45] Feng Shao, Weisi Lin, Shanshan Wang, Gangyi Jiang, and Mei Yu, "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 154–165, 2015.

[46] Z. Wang, Alan C Bovik, and Ligang Lu, "Why is image quality assessment so difficult?," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 4, pp. IV–3313.

[47] M. Chen, Alan C Bovik, and Lawrence K Cormack, "Study on distortion conspicuity in stereoscopically viewed 3d images," in *IVMSP Workshop, 2011 IEEE 10th*. IEEE, 2011, pp. 24–29.

[48] N Clayton Silver and William P Dunlap, "Averaging correlation coefficients: should fisher's z transformation be used?," *Journal of applied psychology*, vol. 72, no. 1, pp. 146, 1987.